Probability and Statistics
Russian Papers of the Soviet Period

Selected and translated by Oscar Sheynin
Berlin 2004

Contents

12. Kolmogorov, A.N. Determining the center of scattering and the measure of precision given a restricted number of observations. *Izvestia Akademii Nauk SSSR*, ser. Math., vol. 6, 1942, pp. 3 – 32 …
13. Kolmogorov, A.N. The number of hits after several shots and the general principles of estimating the efficiency of a system of firing. *Trudy* [Steklov] *Matematich. Institut Akademii Nauk SSSR*, No. 12, 1945, pp. 7 – 25
14. Kolmogorov, A.N. The main problems of theoretical statistics. An abstract. *Второе всесоюзное совещание по математической статистике. Ташкент, 1948* (Second All-Union Conference on Mathematical Statistics. Tashkent, 1948). Tashkent, 1948, pp. 216 – 220
15. Kolmogorov, A.N. His views on statistics
15a. Anonymous, Account of the All-Union Conference on Problems of Statistics. Moscow, 1954 (extract). *Vestnik Statistiki*, No. 5, 1954, pp. 39 – 95 (pp. 46 – 47)
15b. Anonymous, On the part of the law of large numbers in statistics. [Account of an aspect of the same Conference] (Extract). *Uchenye Zapiski po Statistike*, vol. 1, 1955, pp. 153 – 165 (pp. 156 – 158)

Foreword

I am presenting a collection of translations of important Russian papers on probability theory and mathematical statiqtics (including applications of these disciplines). With a single exception of Bortkiewicz' article, all these papers belong to the Soviet period of Russian history. Living and working here in Berlin from 1901 to the end of his life, Bortkiewicz published, in 1921, a paper in the Soviet periodical *Vestnik Statistiki* and is known to have communicated with Soviet statisticians (Slutsky) as well as with Chuprov (who had not returned to Russia after 1917) and to have participated in the activities of the Russian scientific institutions in Berlin, see my pertinent paper in *Jahrb. f. Nat. Ökon. u. Statistik*, Bd. 221, 2001, pp. 226 – 236.

One more unusual entry (Anderson's letters to Pearson left in their original German) belongs to a German statistician of Russian extraction who, apparently all his life, justly considered himself Chuprov's student.

In many instances I changed the numeration of the formulas and I subdivided into sections those lengthy papers which were presented as a single whole; in such cases I denoted the sections by numbers in brackets, for example thus: **[2]**. My own comments are in curly brackets.

Almost all the translations provided below were published in microfiche collections by Hänsel-Hohenhausen (Egelsbach; now, Frankfurt/Main) in their series Deutsche Hochschulschriften NNo. 2514 and 2579 (1998), 2656 (1999), 2696 (2000) and 2799 (2004). The copyright to ordinary publication remained with me.

Acknowledgement. Dr. A.L. Dmitriev (Petersburg) sent me photostat copies of several papers published in sources hardly available outside Russia.

Throughout, I am using the abbreviation
M. = Moscow; L. = Leningrad; and (R) = in Russian.

$\int_{-\infty}^{\infty} \int_{0}^{\infty} \int_{-c}^{c} \int_{0}^{c} \int_{c'}^{c''} \int_{0}^{2} \int_{0}^{A_0} \int \quad \int_{0}^{A_0} \int_{0}^{\alpha_0} \int^{B}$

$\int_{B'}^{B''} \int_{\chi'}^{\chi''} \int_{0}^{C_0} \int_{0}^{\gamma_0} \int_{0}^{\gamma} \int_{0}^{\bar{\chi}^*} \int_{0}^{\bar{\chi}_1 *}$

0. An Appeal to the Scientists

of All Countries and to the Entire Civilized World

The political situation in the Soviet Union had invariably and most strongly influenced science. Here, in this book, it is clearly seen in the section devoted to Romanovsky and it can be revealed in the materials concerning Slutsky. For this reason I inserted an appropriate *Appeal* (above) that can serve as an epigraph.

The Congress of Russian Academic Bodies Abroad, in appealing to the scientists of all countries and to the entire civilized world on behalf of more than 400 Russian scholars scattered in 16 states, raises its voice against those conditions of existence and work that the Soviet regime of utter arbitrary rule and violation of all the most elementary human rights laid down for our colleagues in Russia.

Never, under any system either somewhere else or in Russia itself, men of intellectual pursuit in general, or academics in particular, had to endure such strained circumstances and such a morally unbearable situation. Especially disgraceful and intolerable is the total lack of personal immunity that at each turn causes unendurable moral torment and threatens {everyone} with bodily destruction.

The execution, or, rather, the murder of such scientists as Lasarevsky, a specialist in statecraft, and Tikhvinsky, a chemist, cries out to heaven. They were shot, as the Soviet power itself reported, – the first, for compiling projects for reforming the local government and putting in order the money circulation; and the second, for communicating information to the West on the state of the Russian oil industry. And still, these horrible acts are only particular cases {typical} of the brutal political regime denying any and every right and reigning over Soviet Russia.

We would have failed our sacred national and humanitarian duty had we not stated our public protest against that murderous and shameful system to our colleagues and to all the civilized world.

1. S.N. Bernstein. Mathematical problems of Modern Biology

Nauka na Ukraine, vol. 1, 1922, pp. 13 – 20 …

Darwin's ideas are known to have laid the foundation of modern biology and considerably influenced the social sciences as well. To recall briefly the essence of his doctrine of the evolution of creatures: All species are variable; the properties of organisms vary under the influence of the environment and are inherited to some extent by the offspring; in addition, in the struggle for existence, individuals better adapted to life, supplant their less favorably endowed rivals.

The apparent vagueness of these propositions shows that Darwin only formulated the problem of the evolution of creatures and sketched the method for solving it, but he remained a long way from solving it. He himself, better than many of his followers, was aware of this fact, and understood that, for the solution of the posed problem, mathematics along with observations and experiments will play a considerable part. For that matter, in one of his writings he equated mathematics with a sixth sense [1].

The first and very important attempt to present the laws of heredity and the problem of evolution {of species} in a precise mathematical form was due to Darwin's cousin, Galton. The essence of his theory consisted in his law of hereditary regression: children only partly inherit the deviation of their parents from the average type of the {appropriate} race, and, in the mean, the mathematical coefficient of regression (measuring the likeness between father and son in any trait) was roughly equal to 1/4. This means that, if, for instance, the father is 2 *vershok* {1 *vershok* = 4.4*cm*} higher than the mean stature of the race, the son will likely by only 1/2 *vershok* higher.

By carrying out numerous statistical observations, the eminent English biometrician Pearson corroborated, although by introducing small corrections, the Galton law for various physical and even mental {psychological} properties of man. Nevertheless, the law undoubtedly leaves room for exception and in any case demands certain restrictions.

The discovery of Mendel, an Augustinian monk, delivered a heavy blow to the young Biometrical school. His finding, having remained unnoticed for several decades, was discovered in the beginning of our {of the 20th} century and at once determined the direction for further investigations of heredity. Mendel's extremely thorough botanical experiments upon the crossing of pure races had led him to some remarkable laws of heredity which were recently verified by vast tests involving both plants and animals (including man). It occurred that in the first generation the crossing of individuals of different races produces individuals of a new mixed race (hybrids) who sometimes occupy a middle place between the given pure races, and in other cases do not externally differ from one of the parents (whose type is then called dominating). However, the crossing of hybrids with each other results, on the average, in 1/4 of the offspring being of each of the two pure races, and the other half belonging to the mixed race. For example, an epileptic marrying an absolutely healthy woman (with no epileptics having been among her ancestors) begets healthy children; however, the crossing of healthy individuals of such origin produces children 1/4 of whom are, in the mean, epileptics. Epilepsy is transmitted in accord with the Mendelian laws with healthiness dominating over sickness. The Mendelian law thus explains, in particular, the paradoxical phenomenon of the so-called atavism when a disease or some other property passes not directly from parents to children, but jumps over several generations.

The Galton law of regression and the Mendelian law of crossing exclude each other since a hereditary descent of a certain trait apparently ought to follow either the first or the second law (or perhaps none of them). It is therefore necessary to establish in each particular case which of the two laws, or some of their modification, is taking place. However, allowing for the methodologically unavoidable peculiar quantitative nature equally inherent in each of the two laws of heredity, with an essential part played by the notions of probability, probable deviation, etc, the solution of this problem demands an application of mathematical methods of the theory of probability.

Therefore, the application of the mathematical method is equally necessary for Mendelians and for their rivals belonging to the Pearsonian Biometrical school, and, in general, for all biologists wishing to establish precisely the laws of heredity and variability. However, the significance of mathematics is not restricted to the just indicated and essential but nevertheless only auxiliary role.

In biology, as in the sciences dealing with inorganic nature, mathematics not only records facts and checks the agreement of experimental materials with certain laws; it also claims to be a lawgiver, it attempts to become the formal supervisor of the investigations directing all observations and experiments in accord with a single plan. The mathematical direction in biology therefore aims at the main general problem of discovering such a common form of the laws of heredity and variability that would cover, in a single system, both the Mendelian phenomena and the Galton regression, and, in addition, would conform to all the known evolutionary processes (to mutation, for instance) just as theoretical mechanics embraces all types of movement.

In this case, the part similar to the main postulate of mechanics, – to the principle of inertia, – is played here by the law that we may call the Darwin law of stationarity. If the existence of some simple trait does not either enhance or lessen the individual's adaptation to life (including fertility and sexual selection), the rate of individuals possessing it persists (in the stochastic sense) from generation to generation. Thus, no matter what was the physiological nature of the process of the hereditary descent of simple (monogenic) traits, it is formally characterized by its inability to change, all by itself, the percentage of the mass of individuals possessing such a trait.

It is remarkable that the solution of a purely mathematical problem of discovering an elementary form of the law of individual heredity obeying the Darwin law of stationarity leads to the Mendelian law. This fact establishes the equivalence, in principle, of the Darwin law of stationarity and the Mendelian law of crossing which {?}, due to the above, ought to serve as the foundation of the mathematical theory of evolution.

I briefly indicate three main parts of the problems of that theory. The first part studies the processes of heredity irrespective of the influence of selection and environment. In most cases, the traits (for example, the color of an animal's hair) are polygenic, composed of several simple components, and the conditions of its descent are easily derived by elementary mathematical calculations when issuing from the Mendelian main law.

Of special interest is, in principle, the case of a very complicated trait (e.g., stature of man) composed of very many simple traits obeying the Mendelian law. The application of general stochastic theorems shows that such traits ought to comply with the Galton law of regression. The apparent contradiction between the Galton and the Mendelian laws is thus eliminated just as the Newtonian theory of universal gravitation removed the contrariness between the periodic rotation of the planets round the Sun and the fall of heavy bodies surrounding us on the Earth.

The second problem of the theory of evolution, the study of the influence of all types of selection, presents itself as a mathematical development of the same principles. Whereas, in the absence of selection, the distribution of traits persists, the difference in mortality and in fertility between individuals and in sexual selection made by individuals essentially change it and fix one or several types that can be artificially varied by creating appropriate conditions of selection.

Finally, the third problem studies the influence of the environment on the variability of creatures. Life only consists in responses of a creature to its surroundings, its outward appearance is therefore determined by the environment and, in different conditions, individuals originating from identical ova, become very different from each other. In addition, the environment influences the conditions of selection; it thus changes the type both directly and obliquely. As long as such changes are reversible, their study is guided by the principles described above. Irreversible changes (mutations) are however also possible. Their essence is not sufficiently studied for aptly dwelling on this important issue in an essay.

In concluding my note, expanded too widely but still incomplete, I allow myself to express my desire that more favorable conditions were created here {in the Ukraine} for an orderly work of biologists together with mathematicians and directed towards the study of important theoretical and practical issues connected with the problems indicated above.

Note

1. Here is the pertinent passage from Darwin's *Autobiography* (1887). London, 1958, p. 58:

I have deeply regretted that I did not proceed far enough at least to understand something of the great leading principles of mathematics; for men thus endowed seem to have an extra sense.

However, there hardly exists any direct indication for supporting Bernstein's statement about Darwin's understanding the future role of mathematics in some advanced form in biology.

2. S.N. Bernstein. Solution of a Mathematical Problem Connected with the Theory of Heredity (1924).

Собрание сочинений (Coll. Works), vol. 4. N.p., 1964, pp. 80 – 107

Foreword by Translator

This contribution followed the author's popular note (1922) also translated in this book. Already there, he explained his aim, viz., the study of the interrelation between the Galton law of regression and the Mendelian law of crossing and stated that his main axiom was "the Darwin law of stationarity", which, as he added, was as important in heredity as the law of inertia was in mechanics.

Seneta (2001, p. 341) testifies that Bernstein's main contribution, although partly translated (Bernstein 1942), is little known but that it is "a surprisingly advanced for its time … mathematical investigation on population genetics, involving a synthesis of Mendelian inheritance and Galtonian "laws" of inheritance". I would add: translated in 1942 freely and (understandably because of World War II) without the author's knowledge or consent. The translator (Emma Lehner) properly mentioned Bernstein's preliminary notes (1923a; 1923b). Kolmogorov (1938, p. 54) approvingly cited Bernstein's study and Aleksandrov et al (1969, pp. 211 – 212) quoted at length Bernstein's popular note.

Bernstein described his work on 2.5 pages in his treatise, see its fourth and last edition (1946, pp. 63 – 65). Soon, however, the Soviet authorities crushed Mendel's followers (Sheynin 1998, §7). In particular, in 1949 or 1950 a state publishing house abandoned its intention of bringing out a subsequent edition of Bernstein's treatise because the author had "categorically refused" to suppress the few abovementioned pages, see Aleksandrov et al (1969). And the late Professor L.N. Bolshev privately told me much later that the proofs of that subsequent edition had been already prepared, – to no avail!

In the methodological sense, Bernstein wrote his contribution carelessly. Having proved four theorems, he did not number them but he called the last two of them Theorems A and B. The proofs are difficult to follow because the author had not distinctly separated them into successive steps; his notation was imperfect, especially when summations were involved (also see my Note 13). Many times I have shortened his formulas of the type $z_1 = f(a_1; x; y)$, $z_2 = f(a_2; x; y)$, … by writing instead $z_i = f(a_i; x; y)$, $i = 1, 2, .., n$, and quite a few misprints corrupted his text. I have corrected some of them, but others likely remain. Finally, his references were not fully specified.

Fisher's first contribution on the evolutionary theory appeared in 1918 and his next relevant papers followed in 1922 and 1930 (Karlin 1992). The two authors apparently had not known about each other's work.

* * *

Chapter 1

1. Suppose that we have N such classes of individuals that the crossing of any two of them gives birth to individuals belonging to one of these. We shall call the totality of these classes a closed biotype and we leave completely aside the question of whether it is possible to attribute each individual, given only his appearance, to one of them; we only assume, that, when individuals of classes i and k are crossed, the probability that an individual of class l is produced, has a quite definite value $A_{ik}^l = A_{ki}^l$ with

$$A_{ik}^1 + A_{ik}^2 + \ldots + A_{ik}^N = 1.$$

We shall call these probabilities the coefficients of heredity for the given biotype. Then, if the arbitrary probabilities that each individual belongs to one of the N classes are $\alpha_1, \alpha_2, \ldots, \alpha_N$, the corresponding probabilities for the next generation will be determined [1] by the formulas

$$\alpha_1' = \Sigma A_{ik}^1\,\alpha_i\alpha_k,\ \alpha_2' = \Sigma A_{ik}^2\,\alpha_i\alpha_k,\ \ldots,\ \alpha_N' = \Sigma A_{ik}^N\,\alpha_i\alpha_k \quad (1)$$

and in a similar way for the second generation

$$\alpha_1'' = \Sigma A_{ik}^1\,\alpha_i'\alpha_k',\ \alpha_2'' = \Sigma A_{ik}^2\,\alpha_i'\alpha_k',\ \ldots,$$

$$\alpha_N'' = \Sigma A_{ik}^N\alpha_i'\alpha_k',\ \text{etc} \quad (2)$$

where all the summings extend over indices i and k.

By applying the same iterative formulas we can obtain the probability distribution for any following generation. The problem which we formulate for ourselves is this: *What coefficients of heredity should there exist under panmixia for the probability distribution realized in the first generation to persist in all the subsequent generations*? We say that, if these coefficients obey the stipulated condition, the corresponding law of heredity satisfies the *principle of stationarity.*

2. Here [2], I shall not dwell on those fundamental considerations which convinced me in that, when constructing a mathematical theory of evolution, we ought to base it upon laws of heredity obeying the principle of stationarity. I only note that the Mendelian law, which determines the inheritance of most of the precisely studied elementary traits, satisfies this principle (Johannsen 1926, p. 486).

The so-called Mendelian law concerns three classes of individuals, two of them being pure races [3] and the third one, a race of hybrids always born when two individuals belonging to contrary pure races are crossing. Thus,

$$A_{11}^1 = A_{22}^2 = 1,\ A_{11}^2 = A_{22}^1 = 0,\ A_{12}^3 = 1,\ A_{11}^3 = A_{22}^3 = A_{12}^1 = A_{12}^2 = 0.$$

According to the experiments of Mendel and his followers, the other nine coefficients have quite definite numerical values, viz,

$$A_{33}^1 = A_{33}^2 = 1/4,\ A_{33}^3 = 1/2,\ A_{13}^1 = A_{23}^2 = A_{13}^3 = A_{23}^3 = 1/2,\ A_{13}^2 = A_{23}^1 = 0.$$

Formulas (1) therefore become

$$\alpha_1' = [\alpha_1 + (1/2)\alpha_3]^2,\ \alpha_2' = [\alpha_2 + (1/2)\alpha_3]^2,\ \alpha_3' = 2[\alpha_1 + (1/2)\alpha_3]\cdot[\alpha_2 + (1/2)\alpha_3] \quad (3)$$

from which we obtain by simple substitution

$$\alpha_1'' = \{[\alpha_1 + (1/2)\alpha_3]^2 + [\alpha_1 + (1/2)\alpha_3]\cdot[\alpha_2 + (1/2)\alpha_3]\}^2 = \\ = [\alpha_1 + (1/2)\alpha_3]^2(\alpha_1 + \alpha_2 + \alpha_3)^2, \quad (4)$$

which means that $\alpha_1'' = \alpha_1'$ because $\alpha_1 + \alpha_2 + \alpha_3 = 1$.

In the same way we convince ourselves in that $\alpha_2'' = \alpha_2'$ and $\alpha_3'' = \alpha_3'$. Consequently, the Mendelian law indeed obeys the principle of stationarity.

3. The first very important result that we now want to obtain is this:

Theorem. *If three classes of individuals comprise a closed biotype obeying the principle of stationarity with the crossing of individuals from the first two of them always producing* *individuals of the third class, then classes* 1 *and* 2 *are pure races and their crossing obeys the Mendelian law*.

To simplify the writing, we change the notation in formulas (1) by taking into account that we are considering only three different classes. We designate the probabilities that an individual from the parental (filial) generation belongs to classes 1, 2 and 3 by α, β and γ (α_1, β_1 and γ_1) respectively. Formulas (1) will then be written as

$$\begin{aligned}\alpha_1 &= A_{11}\alpha^2 + 2A_{12}\alpha\beta + A_{22}\beta^2 + 2A_{13}\alpha\gamma + 2A_{23}\beta\gamma + A_{33}\gamma^2 = f(\alpha;\beta;\gamma),\\ \beta_1 &= B_{11}\alpha^2 + 2B_{12}\alpha\beta + B_{22}\beta^2 + 2B_{13}\alpha\gamma + 2B_{23}\beta\gamma + B_{33}\gamma^2 = f_1(\alpha;\beta;\gamma),\\ \gamma_1 &= C_{11}\alpha^2 + 2C_{12}\alpha\beta + C_{22}\beta^2 + 2C_{13}\alpha\gamma + 2C_{23}\beta\gamma + C_{33}\gamma^2 = \varphi(\alpha;\beta;\gamma).\end{aligned} \quad (5)$$

In general,

$$A_{ik} + B_{ik} + C_{ik} = 1.$$

Therefore, in accord with the conditions of the Theorem, we conclude that $B_{12} = A_{12} = 0$ since $C_{12} = 1$ because obviously no coefficient is negative.

Our mathematical problem consists in determining the quadratic forms f, f_1, φ with such non-negative coefficients that

$$f + f_1 + \varphi = (\alpha + \beta + \gamma)^2 = 1$$

under the conditions

$$\begin{aligned}&f(\alpha_1;\beta_1;\gamma_1) = f(\alpha;\beta;\gamma) = \alpha_1, f_1(\alpha_1;\beta_1;\gamma_1) = f_1(\alpha;\beta;\gamma) = \beta_1,\\ &\varphi(\alpha_1;\beta_1;\gamma_1) = \varphi(\alpha;\beta;\gamma) = \gamma_1\end{aligned} \quad (6)$$

the last of which follows from the first two of them.

Equations (6) obviously cannot have only a finite number of solutions; α_1, β_1 and γ_1 would then have been functions of (α + β + γ); therefore, we would have

$$\alpha_1 = p(\alpha + \beta + \gamma)^2$$

which is impossible because the coefficient of αβ should be zero. Consequently, equations (6) may be written out in the form

$$\begin{aligned}&\alpha_1 = \alpha(\alpha + \beta + \gamma) + kF(\alpha;\beta;\gamma),\ \beta_1 = \beta(\alpha + \beta + \gamma) + k_1F(\alpha;\beta;\gamma),\\ &\gamma_1 = \gamma(\alpha + \beta + \gamma) - (k + k_1)F(\alpha;\beta;\gamma)\end{aligned} \quad (7)$$

where $F(\alpha; \beta, \gamma)$ is such a homogeneous form that $F(\alpha_1; \beta_1; \gamma_1) = 0$ for any initial values of α, β and γ.

It is easy to see that $F(\alpha; \beta; \gamma)$ should be not a linear, but a quadratic form because there cannot exist a linear relation of the type

$$l\alpha_1 + m\beta_1 + n\gamma_1 = lf(\alpha;\beta;\gamma) + mf_1(\alpha;\beta;\gamma) + n\varphi(\alpha;\beta;\gamma) = 0$$

with $n \neq 0$ between α_1, β_1 and γ_1; indeed, f and f_1 are devoid of the term αβ which is present in φ. And $n = 0$ is also impossible because then $lm < 0$ so that we could have assumed that $l = 1$ and $m = -p$, $p > 0$; the last of the equations (7) would then be

$$\gamma_1 = \gamma(\alpha + \beta + \gamma) + (A\alpha + B\beta + C\gamma)\cdot(\alpha - p\beta)$$

and, since the coefficients of α^2 and β^2 are non-negative, $A \geq 0$ and $B \leq 0$, whereas, according to the condition of the Theorem, $B - Ap = 2$. And so, $F(\alpha; \beta; \gamma)$ is a quadratic form, k and k_1 are numerical coefficients, and without loss of generality we may assume that $k = 1$; then, obviously, $k_1 = 1$ and the coefficient of $\alpha\beta$ in the polynomial $F(\alpha; \beta; \gamma)$ is -1 since neither $f(\alpha; \beta; \gamma)$ nor $f_1(\alpha; \beta; \gamma)$ contain the term $\alpha\beta$.

It is still necessary to determine the coefficients of the polynomial

$$F(\alpha; \beta; \gamma) = a\alpha^2 + b\beta^2 - \alpha\beta + c\alpha\gamma + d\,\beta\gamma + e\gamma^2.$$

First of all, we note that $a = b = 0$. Indeed, a cannot be positive because the coefficient of α^2 in $f(\alpha; \beta; \gamma)$ does not exceed 1; nor can it be negative since then the same coefficient in $f_1(\alpha; \beta; \gamma)$ would be negative. In the same way we convince ourselves in that $b = 0$ as well.

To determine the other coefficients we note, issuing from equations (7), that the equations of stationarity (6) are transformed into a single equation

$$F(\alpha S + F; \beta S + F; \gamma S - 2F) = 0,\ S = \alpha + \beta + \gamma, \qquad (8)$$

which should persist for any values of α, β, γ.

Expanding equation (8) into a Taylor series we find that

$$S^2F + SF(F'_\alpha + F'_\beta - 2F'_\gamma) + F^2F(1; 1; -2) = 0 \qquad (9)$$

or, after cancelling F out of it,

$$F(1; 1; -2)\, F(\alpha; \beta; \gamma) = -S^2 + S(2F'_\gamma - F'_\alpha - F'_\beta). \qquad (10)$$

However, on the strength of the remark above, F cannot be split up into multipliers, therefore $F(1; 1; -2) = 0$, and, after cancelling S out of equation (10), we finally obtain the identity

$$S = 2F'_\gamma - F'_\alpha - F'_\beta \qquad (11)$$

or

$$\alpha + \beta + \gamma = 2(c\alpha + d\,\beta + 2e\gamma) + \beta - c\gamma + \alpha - d\gamma.$$

Therefore

$$c = d = 0,\ e = 1/4,\ F(\alpha; \beta; \gamma) = (1/4)\gamma^2 - \alpha\beta$$

so that

$$\begin{aligned} f(\alpha; \beta; \gamma) &= \alpha(\alpha + \beta + \gamma) + (1/4)\gamma^2 - \alpha\beta = (\alpha + \gamma/2)^2, \\ f_1(\alpha; \beta; \gamma) &= \beta(\alpha + \beta + \gamma) + (1/4)\gamma^2 - \alpha\beta = (\beta + \gamma/2)^2, \\ \varphi(\alpha; \beta; \gamma) &= \gamma(\alpha + \beta + \gamma) + 2\alpha\beta - (1/2)\gamma^2 = 2(\alpha + \gamma/2)(\beta + \gamma/2), \end{aligned} \qquad (12)$$

QED.

4. As we have shown, the Mendelian law is a necessary corollary of the principle of stationarity provided that the crossing of the first two classes always produces individuals of the third class; and we did not even presuppose that the two former represent pure races.

From the theoretical point of view it would be interesting to examine whether there exist other laws of crossing of pure races compatible with the principle of stationarity.

And so, let us suppose now that the coefficients of α^2 in $f(\alpha; \beta; \gamma)$ and of β^2 in $f_1(\alpha; \beta; \gamma)$ are both unity. Repeating the considerations which led us to the just proved theorem, we again arrive at equations (7) where

$$F = -a\alpha\beta + c\alpha\gamma + d\,\beta\gamma + e\gamma^2$$

and we may assume that $k = 1$ and $k_1 = \lambda$. For determining the five coefficients a, c, d, e, λ we have here, instead of (11), the identity

$$S = (1 + \lambda)F'_\gamma - F'_\alpha - \lambda F'_\beta \qquad (13)$$

from which we obtain the values of c, d, e through the two parameters a and λ:

$$d = (-a + 1)/(\lambda + 1),\ c = (-a\lambda + 1)/(\lambda + 1),\ e = (-a\lambda + \lambda + 1)/(\lambda + 1)^2.$$

The most general form of the polynomial F satisfying our condition is therefore

$$F = -a\alpha\beta + \alpha\gamma\,(-a\lambda + 1)/(\lambda + 1) + \beta\gamma\,(-a + 1)/(\lambda + 1) + \gamma^2(-a\lambda + \lambda + 1)/(\lambda + 1)^2$$

so that, assuming that $a\lambda = b$, we may write the right side as

$$-a\alpha\beta + a\alpha\gamma\,(1 - b)/(a + b) + a\beta\gamma(1 - a)/(a + b) + a\gamma^2(a + b - ab)/(a + b)^2$$

and, by means of simple algebraic transformations, we finally determine that

$$f = [\alpha + \gamma a/(a + b)]\{\alpha + (1 - a)\beta + \gamma[1 - ab/(a + b)]\},$$
$$f_1 = [\beta + \gamma b/(a + b)]\{\beta + (1 - b)\alpha + \gamma[1 - ab/(a + b)]\}, \qquad (14)$$
$$\varphi = (a + b)\,[\alpha + \gamma a/(a + b)][\beta + \gamma b/(a + b)].$$

So that the coefficients will not be negative, it is necessary and sufficient to demand in addition that $0 \le a, b \le 1$. In particular, if $a = b = 1$, formulas (14) coincide with (12).

Whether cases of heredity obeying formulas (14) with $a, b < 1$ occur or not, can only be ascertained experimentally. From the theoretical viewpoint, these formulas provide the most general law of heredity for a closed biotype consisting of three classes two of which are pure races. It is easy to see that the only law of heredity for all three classes being pure races is expressed by the formulas

$$f = \alpha(\alpha + \beta + \gamma), f_1 = \beta(\alpha + \beta + \gamma),\ \varphi = \gamma(\alpha + \beta + \gamma) \qquad (15)$$

which follow from (7) if $k = k_1 = 0$.

5. To complete the investigation of all the possible forms of heredity for biotypes consisting of three classes [4] and assuming as before the principle of stationarity, we still have to prove the following proposition.

Theorem. *If each of the classes can be obtained from the crossing of the other ones, then*

$$f = p(\alpha + \beta + \gamma)^2, f_1 = q(\alpha + \beta + \gamma)^2,\ \varphi = r(\alpha + \beta + \gamma)^2. \qquad (16)$$

If, however, only one class is a pure race, then either

$f = (\alpha + \beta)\{[(1 + b)(\alpha + \beta)/2] + (1 - d)\gamma\},$
$f_1 = (\alpha + \beta)\{[(1 - b)(\alpha + \beta)/2] + d\gamma\},\ \varphi = \gamma(\alpha + \beta + \gamma).$ (17)

or

$f = \alpha S + a\alpha(\mu\beta + \gamma),\ \varphi + \mu f_1 = 0.$

Indeed, if equations (6) possess a finite number of solutions, they lead to formulas (16); otherwise, we arrive at formulas (7), and here two cases are possible.

1) F is a quadratic form which cannot be decomposed into multipliers with k and k_1 being numerical coefficients.

2) F is a linear form and k and k_1 are also linear forms.

Suppose at first that F is a quadratic form. If not a single number from among k, k_1 and $(k + k_1)$ is zero, then obviously two of them, for example, k and k_1, can be chosen to be positive, and the form F should then lack terms with α^2 and β^2 so that the form $\varphi(\alpha; \beta; \gamma)$ will have no negative coefficients. This case should therefore be rejected because it returns us to the formulas (14) that correspond to two pure races. And so, we have to assume that one of the numbers k, k_1 and $(k + k_1)$ is zero. We may suppose that $(k + k_1) = 0$, *i.e.*, that the third class is a pure race (the coefficient of γ^2 is unity). Then, the same coefficient in F should be zero, and, for determining the other coefficients by the same method as before, we obtain for $k = 1$

$F = (\alpha + \beta)\{[\alpha(b - 1)/2] + [\beta(b + 1)/2] - d\gamma\} + \beta\gamma,$

and we arrive at (17.1) and (17.2).

We still have to consider the assumption that F is a linear form. Let

$F = \lambda\alpha + \mu\beta + \gamma.$

Then, similar to the above, the condition of stationarity leads to the identity

$S + \lambda k + \mu k_1 - (k + k_1) = 0$

where k and k_1 are linear forms

$k = a\alpha + b\beta + c\gamma,\ k_1 = a_1\alpha + b_1\beta + c_1\gamma.$

Had we been unrestricted with regard to the signs, we could have chosen k arbitrarily, and, supposing that

$k_1 = [S + k(\lambda - 1)]/(1 - \mu),$

we would have obtained solutions for f, f_1 and φ depending on five parameters (λ, μ, a, b, c). However, not a single of these solutions fits in with the first condition of the Theorem. Indeed, since the coefficients of β^2, $\beta\gamma$ and γ^2 in

$f = \alpha S + kF$

are non-negative, $\mu b \geq 0,\ b + \mu c \geq 0,\ c \geq 0.$

And, issuing from the corresponding property of f_1, we find that

$\lambda a_1 \geq 0$, $c_1 \geq 0$, $a_1 + \lambda c_1 \geq 0$.

It follows that, if $\mu, \lambda \geq 0$, the equality of the type

$$\lambda f + \mu f_1 + \varphi = 0$$

is impossible since then all the coefficients would be positive. If, however, $\mu < 0$, then $b = c = 0$ which is incompatible with the assumption that individuals of the first class can be produced when the other classes are crossed. Nevertheless, it is not difficult to conclude that, because the coefficients are non-negative, conditions $b = c = 0$ lead to $\lambda = 0$ and therefore to

$$f = \alpha S + a\alpha(\mu\beta + \gamma),\ \varphi = -\mu f_1 = [\mu/(\mu - 1)]\cdot[S(\beta + \gamma) - a\alpha(\mu\beta + \gamma)] \qquad (17')$$

And so, all possible cases are exhausted and our Theorem is proved.

6. Let us summarize the obtained results. Under the principle of stationarity the laws of heredity for a closed biotype consisting of three classes can be categorized as follows.

1) Two classes represent pure races. Heredity obeys formulas (14) which, specifically, express the Mendelian law (12) if the crossing of pure races always produces a hybrid race.

2) Not a single class is a pure race but each can be produced when the other classes are crossed. Heredity occurs in accord with formulas (16). The distribution of the offspring by classes is constant and independent of the properties of the arbitrarily chosen parents. No correlation between parents and children exists here and the given biotype, in spite of its polymorphism, possesses the essential property characterizing a pure race.

3) All three classes represent pure races. Heredity obeys formulas (15). Arbitrary distributions by classes are passed on without change. Each two classes of the biotype also form a closed biotype.

4) One of the classes represents a pure race. Heredity conforms to formulas (17) or (17'). If the other classes are united, they, taken together, constitute a closed dimorphic biotype whose heredity fits in with the abovementioned Type 2. Together with the class representing a pure race it obeys the law of heredity of Type 3. Since it is reduced to Types 2 and 3, this type of heredity is not interesting in itself. The case (17') is distinguished from (17) in that the latter predetermines a stationary relative distribution of the pure race and the totality of the hybrid races, whereas the former, to the contrary, predetermines the relative distribution of the hybrid classes with respect to each other [5].

In particular, our investigation shows that the equations

$$S\alpha = f(\alpha; \beta; \gamma) \text{ and } S\beta = f_1(\alpha; \beta, \gamma)$$

are always independent if the coefficients in their right sides are positive and $(S^2 - f - f_1)$ also has positive coefficients (not equal to zero).

Chapter 2

7. Passing on to biotypes with a number of classes $N > 3$ we shall solve the problem formulated in the beginning under three main different suppositions. *First case*: Among the biotypes there is a certain number of pure races whose pairwise crossing is known to follow the Mendelian law. It is required to determine the coefficients of heredity when the other classes are crossed.

Second case: Each crossing can reproduce individuals of the entire biotype. *Third case*: The biotype has two pure races, which, when mutually crossed, produce all classes excepting their own. To determine the laws of heredity in these cases as well.

The solution in the first case is not difficult and is provided by the formulas

$$f_{ii} = [\alpha_{ii} + (1/2)\sum_h \alpha_{ih}]^2, \quad (19)$$
$$f_{ik} = 2[\alpha_{ii} + (1/2)\sum_h \alpha_{ik}]\cdot[\alpha_{kk} + (1/2)\sum_h \alpha_{kh}]$$

where the first sum [6] is extended over $h \neq i$ and α_{ii} and α_{ik} are the probabilities that a parent belongs to pure race A_{ii} and A_{ik} respectively, and f_{ii} and f_{ik} are the probabilities that the offspring belongs to the pure race A_{ii} and to the hybrid race [7] A_{ik} respectively.

Indeed, formulas (19) obviously satisfy the principle of stationarity because

$$f_{ii} + (1/2)\sum_l f_{il} = [\alpha_{ii} + (1/2)\sum_l \alpha_{il}]\cdot[\sum_k \alpha_{kk} + (1/2)\sum_{kl} \alpha_{kl}]$$

where the second factor in the right side is unity.

Let us show that the formulas (19) furnish a unique solution. To this end suppose that only pure races are being crossed in the parent generation so that $\alpha_{ik} = 0$ if $i \neq k$ and denote

$$\alpha_{11} = t_1,\ \alpha_{22} = t_2,\ \ldots,\ \alpha_{nn} = t_n.$$

Then, in the next generation,

$$\alpha^1_{ii} = t_i^2,\ \alpha^1_{ik} = 2t_i t_k.$$

Thus, because of the principle of stationarity, we have

$$f_{11}(t_1^2; 2t_1t_2; \ldots; t_n^2) = t_1^2(t_1 + t_2 + \ldots + t_n)^2 \quad (20)$$

and similar equalities for the other functions. Denoting the coefficient of $\alpha_{ik}\,\alpha_{hl}$ in f_{11} by $A_{ik,\,hl}$ we infer that it is zero if less than two numbers from among *i, k*, *h*, and *l* are unity. And, supposing that *h*, *l* and 1 differ one from another, we have

$$A_{11,\,11} = 1,\ A_{11,\,hl} + 2A_{1h,\,1l} = 1,\ A_{11,\,1h} = 1,\ A_{11,\,hh} + 4A_{1h,\,1h} = 1$$

and therefore

$$f_{11}(\alpha_{11}; \alpha_{12}; \ldots; \alpha_{nn}) = [\alpha_{11} + (1/2)\sum_k \alpha_{1k}]^2 +$$
$$\sum_{h,j} A_{11,\,hj}[\alpha_{11}\alpha_{hj} - (1/2)\alpha_{1h}\alpha_{ij}] + \sum_h A_{11,\,hh}[\alpha_{11}\alpha_{hh} - (1/4)\alpha_{1h}^2] \quad (21)$$

and, since $A_{11,\,hh} = 0$ [8], the last term in the right side vanishes.

The equation of stationarity for the class A_{11} will therefore be expressed by the identity

$$(\alpha_{11} + \alpha_{12} + \ldots + \alpha_{nn})^2 f_{11}(\alpha_{11}; \alpha_{12}; \ldots; \alpha_{nn}) =$$
$$[f_{11} + (1/2)\sum_k f_{1k}]^2 + \sum_{hj} A_{11,\,hj}[f_{11}f_{hj} - (1/2)f_{1h}f_{1j}]. \quad (22)$$

Let us equate the coefficients of $\alpha_{11}\alpha_{hj}^3$ in both of its parts. In the left side it will be $A_{11,\,hj}$; in the right side, taking into account that, from among all of its functions, only f_{hj} contains

α_{hj}^2 (with coefficient 1/2), it will be $(1/2)A_{11,\,hj}^2$. Therefore, $A_{11,\,hj} = 0$ and equation (21) becomes (19.1) which we should have established. The other equations are obtained in exactly the same way [9].

Formulas (19) evidently show that the crossing of A_{ik} with A_{il} produces 1/4 of pure individuals A_{ii} and 1/4 of A_{ik}, A_{il} and A_{kl} each; the crossing of A_{ik} with A_{jh} produces 1/4 of A_{ih}, A_{ij}, A_{kh} and A_{kj} each; and, finally, the crossing of A_{ii} with alien hybrids A_{kl}, – 1/2 of the hybrids A_{ik} and A_{il} each. This result completely coincides with Mendel's initial physiological hypothesis but it demands that the hypothesis of the "presence and absence of genes" be revised [10].

8. The solution of the second problem is expressed by the following proposition.

Theorem. *If the crossing of any individuals of a closed biotype consisting of n classes can produce individuals of any class, – i.e., if the coefficients of all the forms* (1) *are not zeros, – then heredity is determined by the formulas*

$$\alpha_1' = \lambda_1(\alpha_1 + \alpha_2 + \ldots + \alpha_n)^2,\ \alpha_2' = \lambda_2(\alpha_1 + \alpha_2 + \ldots + \alpha_n)^2,\ \ldots,\ \Sigma\lambda_i = 1. \qquad (23)$$

This Theorem generalizes the corresponding proposition for $n = 3$ (§5) and we shall apply it now for proving the new statement by the method of mathematical induction. Let $n = 4$ and choose any two classes A_1 and A_2 from among them; the two other ones, A_3 and A_4, will constitute a special totality, which in general will not possess the characteristic property of a class. That is, when its individuals are crossed one with another, or with those of the other classes, the probability of the appearance of individuals of a certain class will not be *constant.* However, we can construct a class $A_3^{(k)}$ from out of this totality in such a way that the ratio of the number of individuals from class A_4 to those of class A_3 will remain constant (and equal to k) in our totality.

And so, suppose that our formulas of heredity are

$$\alpha_i' = f_i(\alpha_1; \alpha_2; \alpha_3; \alpha_4),\ i = 1, 2, 3, 4. \qquad (24)$$

Let $\alpha_4 = k\alpha_3$ and denote

$$\gamma = \alpha_3 + \alpha_4 = \alpha_3(1 + k).$$

Then, restricting α_1, α_2 and γ by an additional condition

$$kf_3[\alpha_1; \alpha_2; \gamma/(1 + k); k\gamma/(1 + k)] - f_4[\alpha_1, \alpha_2; \gamma/(1 + k); k\gamma/(1 + k)] = 0$$

which expresses the equality $k\alpha_3' = \alpha_4'$, we see that the totalities A_3 and A_4 maintain under heredity the property of the class $A_3^{(k)}$.

Thus, supposing that

$$f_i[\alpha_1; \alpha_2; \gamma/(1 + k); k\gamma/(1 + k)] = \varphi_i(\alpha_1; \alpha_2; \gamma),\ i = 1, 2,$$
$$f_3[\alpha_1; \alpha_2; \gamma/(1 + k); k\gamma/(1 + k)] + \qquad (25)$$
$$f_4[\alpha_1; \alpha_2; \gamma/(1 + k); k\gamma/(1 + k)] = \varphi_3(\alpha_1; \alpha_2; \gamma),$$

we express the law of heredity in the transformed biotype by means of the functions φ_1, φ_2 and φ_3. This law satisfies the principle of stationarity if only the initial distribution of individuals by classes obeys the equation

$$kf_3 - f_4 = F_k(\alpha_1; \alpha_2; \gamma) = 0. \qquad (26)$$

On the other hand, for four classes the stationarity condition cannot depend on more than one parameter, because, after representing the equations (24) as

$$\alpha_i' = \alpha_i S + \psi_i(\alpha_1; \alpha_2; \alpha_3, \alpha_4),\ i = 1, 2, 3, 4, \tag{24'}$$

we see that the equations

$$\psi_1 = 0,\ \psi_2 = 0,\ \psi_3 = 0,\ \psi_4 = 0 \tag{27}$$

cannot be equivalent to one equation. Indeed, in this (impossible) case, supposing that $\alpha_4 = 0$ we could have realized for $n = 3$ an infinite set of stationary conditions which contradicts §5.

If it is not satisfied identically for some k, the equation (26) can therefore provide only a finite number of values for α_1', α_2', α_3', α_4' [11]. Consequently, if equation (26) is satisfied, the functions φ_i given by formulas (25) can take only a restricted number of values, and, owing to their continuity, these values are quite definite. We conclude that

$$\varphi_i = \lambda_i(\alpha_1 + \alpha_2 + \gamma)^2 + \mu_i F_k,\ i = 1, 2, 3 \tag{28}$$

if only F_k is not an exact square [12]. And the constants depending on k, λ_1, λ_2, λ_3 are connected by the equality

$$\lambda_1 + \lambda_2 + \lambda_3 = 1$$

whereas μ_1, μ_2 and μ_3 satisfy the condition

$$\mu_1 + \mu_2 + \mu_3 = 0.$$

Substituting the expressions of φ_1, φ_2, φ_3, F_k through f_1, f_2, f_3, f_4 into equations (28) and returning to the initial variables α_1, α_2, α_3, α_4 we obtain, with respect to f_1, f_2, f_3, f_4 and S^2, where $S = \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4$ [13], three homogeneous linear equations whose coefficients depend on k:

$$f_i + \mu_i f_4 - k\mu_i f_3 = \lambda_i S^2,\ i = 1, 2,\ f_4(1 + \mu_3) + f_3(1 - k\mu_3) = \lambda_3 S^2.$$

If $k \geq 0$ these equations are independent and it is therefore always possible to express three of the forms f_i by the fourth one and S^2. Thus, for the sake of definiteness we may assume that

$$f_i = h_i S^2 + m_i f_1, \tag{29}$$

$$\Sigma h_i = 1,\ \Sigma m_i = -1,\ i = 2, 3, 4 \tag{30}$$

where h_1 and m_1 can depend on $k = \alpha_4/\alpha_3$. In any case, it is easy to see [14] that these two magnitudes can only be linear fractional expressions with regard to α_4/α_3.

The equation of stationarity for f_1 provides, however,

$$f_1(f_1, f_2; f_3; f_4) = S^2 f_1(\alpha_1; \alpha_2; \alpha_3, \alpha_4);$$

or, applying equalities (29), we have

$$f_1(f_1;\, h_2S^2 + m_2f_1;\, h_3 S^2 + m_3f_1;\, h_4 S^2 + m_4f_1) = S^2f_1(\alpha_1;\, \alpha_2;\, \alpha_3;\, \alpha_4). \qquad (31)$$

Therefore, expanding the right side of equality (31) into a Taylor series, we have

$$S^4f_1(0;\, h_2;\, h_3,\, h_4) + S^2f_1(\alpha_1;\, \alpha_2;\, \alpha_3;\, \alpha_4)\cdot[h_2\frac{\partial f_1}{\partial \alpha_2}(1;\, m_2;\, m_3;\, m_4) +$$

$$h_3\frac{\partial f_1}{\partial \alpha_3} + h_4\frac{\partial f_1}{\partial \alpha_4}] + f_1^2(\alpha_1,\, \alpha_2;\, \alpha_3,\, \alpha_4)f_1(1;\, m_2;\, m_3;\, m_4) = S^2f_1(\alpha_1;\, \alpha_2,\, \alpha_3,\, \alpha_4). \; (31')$$

Hence we conclude that either $f_1/S^2 = M$, where M can be a function of α_3 and α_4, or the coefficients of S^4, S^2f_1 and f_1^2 are zeros. But the first supposition can only be realized if M is a constant and in that case the Theorem would have been already proved. It remains therefore to consider the second case in which

$$f_1(0;\, h_2,\, h_3,\, h_4) = 0,\, f_1(1;\, m_2;\, m_3;\, m_4) = 0, \qquad (32)$$
$$h_2\frac{\partial f_1}{\partial \alpha_2}(1;\, m_2,\, m_3,\, m_4) + h_3\frac{\partial f_1}{\partial \alpha_3}(1;\, m_2,\, m_3,\, m_4) + h_4\frac{\partial f_1}{\partial \alpha_4}(1,\, m_2,\, m_3,\, m_4) = 1.$$

Supposing now that

$$\psi_1(\alpha_1;\, \alpha_2;\, \alpha_3,\, \alpha_4) = f_1 - \alpha_1 S,$$

we conclude that this function vanishes at all the values of its arguments connected by the equalities

$$[(\alpha_2 - m_2\alpha_1)/h_2] = [(\alpha_3 - m_3\alpha_1)/h_3] = [(\alpha_4 - m_4\alpha_1)/h_4] = p \qquad (33)$$

with any p because

$$\psi_1(0;\, h_2,\, h_3,\, h_4) = 0,\, \psi_1(1;\, m_2;\, m_3;\, m_4) = 0, \qquad (34)$$
$$h_2\frac{\partial \psi_1}{\partial \alpha_2}(1;\, m_2,\, m_3;\, m_4) + h_3\frac{\partial \psi_1}{\partial \alpha_3} + h_4\frac{\partial \psi_1}{\partial \alpha_4} = 0.$$

We also note that the equalities (33) are equivalent to equations

$$h_iS + m_i\alpha_1 - \alpha_i = 0,\; i = 2,\, 3,\, 4 \qquad (35)$$

only two of which are independent because of (30).

For visualizing the obtained result more clearly we can replace the homogeneous coordinates by Cartesian coordinates supposing that for example $\alpha_3 = 1$. Then we may say that the surface of the second order $\psi_1(x;\, y;\, 1;\, z) = 0$ passes through the line of intersection of the surfaces expressed by the equations (35). But, supposing now that

$$\psi_2 = f_2 - \alpha_2 S = h_2 S^2 + m_2f_1 - \alpha_2 S = m_2\psi_1 + S(h_2 S + m_2\alpha_1 - \alpha_2),$$
$$\psi_i = f_i - \alpha_i S = m_i\psi_1 + S(h_i S + m_i\alpha_1 - \alpha_i),\; i = 3,\, 4$$

we conclude that the surfaces $\psi_2 = 0$, $\psi_3 = 0$, $\psi_4 = 0$ also pass through the same line. In addition, the form of the functions ψ_2, ψ_3, ψ_4 shows that these equations cannot admit of any other positive common solutions excepting those given by equations (35). Consequently, noting that, for

$$f_i = \alpha_i S + \psi_i,\ i = 1, 2, 3, 4, \tag{36}$$

all the stationary solutions are determined by the common solution of equations (27), we conclude that all these solutions are determined by formulas (35) with the parameter $k = \alpha_4/\alpha_3$ taking all possible values from 0 to ∞.

There thus exist such positive values α_4/α_3 that the other coordinates α_1/α_3 and α_2/α_3determined by the equations (35) are also positive. Therefore, by continuously varying the parameter we can make at least one coordinate (for example, α_4/α_3) vanish with the other ones being non-negative. Then, with α_1, α_2, α_3 taking the respective positive values and replacing α_4 by zero, we note that (36.4) vanishes which is impossible because all the coefficients there are positive.

Let us now pass on to the general case and show by the same method that if the Theorem is valid for some n it holds for $(n + 1)$. Indeed, if it is valid for n, the equations (36) with $i = 1, 2, \ldots, n$ cannot include dependent equations

$$\psi_1 = 0,\ \psi_2 = 0,\ \ldots,\ \psi_{n-1} = 0$$

when all the coefficients in f_i are positive. Therefore, the similar equations

$$\psi_1 = 0,\ \psi_2 = 0,\ \ldots,\ \psi_n = 0,$$

where f_i are the same as in (36) but with $i = 1, 2, \ldots, (n + 1)$, cannot be connected by more than one dependence; *i.e.*, the stationarity condition for $(n + 1)$ classes cannot depend on more than one parameter.

Consequently, the requirement that

$$kf_n - f_{n+1} = 0,$$

if only it does not hold identically for some k, leads to a restricted number of possible values for $f_1, f_2, \ldots, f_n, f_{n+1}$. Therefore, uniting the n-th and the $(n + 1)$-th classes into one, and assuming that in the initial distribution

$$\alpha_n = \gamma/(1 + k),\ \alpha_{n+1} = k\gamma/(1 + k),$$

the functions

$$\varphi_i = f_i[\alpha_1; \alpha_2; \ldots; \gamma/(1 + k); k\gamma/(1 + k)],\ i = 1, 2, \ldots, n - 1,$$

$$\varphi_n = f_n[\alpha_1; \alpha_2; \ldots; \gamma/(1 + k); k\gamma/(1 + k)] + f_{n+1}[\alpha_1; \alpha_2; \ldots; \gamma/(1 + k); k\gamma/(1 + k)]$$

if only

$$F_k = kf_n[\alpha_1; \alpha_2; \ldots; \gamma/(1 + k); k\gamma/(1 + k)] - f_{n+1}[\alpha_1; \alpha_2; \ldots; \gamma/(1 + k); k\gamma/(1 + k)] = 0 \tag{37}$$

can take only a restricted number of values, and, owing to their continuity, have only one definite system of values [15]. It follows that if{the left side of} equation (37) is not an exact square, then

$$\varphi_i = \lambda_i(\alpha_1 + \ldots + \alpha_{n-1} + \gamma)^2 + \mu_i F_k,\ i = 1, 2, \ldots, n. \tag{38}$$

We conclude that

$$f_i = h_i S^2 + m_i f_1,\ i = 2, 3, \ldots, n + 1 \qquad (39)$$

where $\Sigma h_i = 1$, $\Sigma m_i = -1$.

When compiling the stationary equation for f_1 we shall now find, as we did before, that

$$\psi_1(0; h_2; \ldots; h_{n+1}) = 0,\ \psi_1(1; m_2; \ldots; m_{n+1}) = 0, \qquad (40)$$
$$h_2 \frac{\partial \psi_1}{\partial \alpha_2}(1; m_2; \ldots; m_{n+1}) + \ldots + h_{n+1} \frac{\partial \psi_1}{\partial \alpha_{n+1}}(1; m_2; \ldots; m_{n+1}) = 0$$

so that for all the values of the parameter p

$$\psi_1(\alpha_1; \alpha_2; \ldots; \alpha_n; \alpha_{n+1}) = 0$$

if

$$\alpha_i/\alpha_1 = m_i + h_i p,\ i = 2, 3, \ldots, n + 1.$$

Again, for all these values, the functions

$$\psi_i = m_i\psi_1 + S(h_i S + m_i\alpha_1 - \alpha_i),\ i = 2, 3, \ldots, n + 1 \qquad (41)$$

also vanish.

Consequently, all possible values of the parameter $k = \alpha_{n+1}/\alpha_n$ provide all the stationary values of α_1. Therefore, some values of that parameter correspond also to the totality of the positive solutions, and, when continuously varying k, we could have also obtained such a totality of values that one or some of the α_i's would have vanished with the other ones being positive. This, however, would have contradicted the assumption that all the coefficients in the forms f_i are positive (not zeros).

The Theorem is thus proved except for the case in which the function in (37) is an exact square for any $k \geq 0$. Obviously, the occurring difficulty would be only essential if this property persisted for any combination of the pairwise united classes. This, however, could have only happened if each of the functions f_i represented an exact square when the respective variable $\alpha_i = 0$.

The excluded case therefore demands that all the functions f_i be of the type

$$f_i = \lambda_i P^2 + \alpha_i Q_i,\ i = 1, 2, \ldots, n + 1 \qquad (42)$$

where λ_i are some positive coefficients and P, Q_1, Q_2, …, Q_{n+1} are linear forms. Forming the equation of stationarity for f_1 we will have

$$S^2 f_1 = \lambda_1^{\,2} P^2(f_1; f_2; \ldots; f_{n+1}) + f_1 Q_1(f_1; f_2; \ldots; f_{n+1}),$$

that is

$$f_1[S^2 - Q_1(f_1; \ldots; f_{n+1})] = \lambda_1^{\,2} P^2(f_1; \ldots; f_{n+1}). \qquad (43)$$

Consequently, either

$$f_1 = C_1 P(f_1; \ldots; f_{n+1}),\ C_1 = \text{Const} \tag{44}$$

or f_1 is an exact square. Since the equations of stationarity for the other f_i lead to the same conclusion, we ought to admit that either all the f_i or all but one of them are exact squares, or that owing to the equality (44) there exist at least two functions f_j and f_k differing from each other only by a numerical coefficient. We may reject the last-mentioned case because the previous method of proof is here applicable.

And so, suppose that there exist three functions, f_1, f_2 and f_3, which are exact squares. Then, eliminating $P(f_1; \ldots; f_{n+1})$ from their equations of stationarity, we obtain

$$\lambda_i f_1[S^2 - Q_1(f_1; \ldots; f_{n+1})] = \lambda_1 f_i[S^2 - Q_i(f_1; \ldots; f_{n+1})],\ i = 2, 3$$

and conclude that at least two from among these three functions only differ one from another by a numerical coefficient so that the previous method is again applicable. The Theorem is thus proved in all generality.

9. The proposition just proved for quadratic forms (which correspond to heredity under bisexual reproduction) holds, as it is easy to see, for linear forms (corresponding to unisexual reproduction). Namely, *if*

$$f_i = A_i^1\alpha_1 + \ldots + A_i^n\alpha_n,\ i = 1, 2, \ldots, n$$

are linear forms with positive coefficients satisfying equalities

$$\sum_k A_k^i = 1$$

for any i, then the establishing condition of stationarity is quite determined, and, when the principle of stationarity is maintained, $f_i = \lambda_i S$.

Indeed, supposing that $\varphi_i = f_i - \alpha_i$, we note that under the condition of stationarity $\varphi_i = 0$, and I say that, except for the dependence $\Sigma\varphi_i = 0$, no other restrictions on the forms φ_i can exist. In the contrary case $\Sigma\lambda_i\varphi_i = 0$ which would have meant that

$$\lambda_1(A_2^1 + A_3^1 + \ldots + A_n^1) = \lambda_2 A_2^1 + \lambda_3 A_3^1 + \ldots + \lambda_n A_n^1,$$
$$\lambda_2(A_1^2 + A_3 + \ldots + A_n^2) = \lambda_1 A_1^2 + \ldots + \lambda_n A_n^2, \ldots,$$
$$\lambda_n(A_1^n + \ldots + A_n) = \lambda_1 A_1 + \ldots + \lambda_{n-1} A_n^{n-1}.$$

But since all the coefficients A_k^i are here positive we should conclude that each of the λ_i is some mean of the other similar magnitudes; and, consequently, that all of them are equal one to another and our statement about the impossibility of any other restrictions being imposed on the φ_i's is proved. The condition of stationarity established in the second generation does not therefore depend on the initial values of α_i and $f_i = \lambda_i S$.

By directly going over to the limit as $n = \infty$ both our theorems on the linear and the quadratic forms are obviously extended onto the case of linear and double integrals respectively. We thus obtain the following two propositions.

Theorem A. *The equation*

$$f(y) = \int_0^1 K(x\,;\,y)f(x)\,dx$$

in which $K(x; y)$ *is positive and* $\int_0^1 K(x; y)\,dy = 1$

has only one solution (*up to a constant factor*). *If, however, the equation*

$$\int_0^1 K(x; y)\varphi(x)\,dx = \int_0^1\int_0^1 K(x; x_1)K(x_1; y)\varphi(x)\,dx\,dx_1$$

is satisfied by any positive and integrable function $\varphi(x)$, *then* $K(x, y)$ *is a function of y only.*

Theorem B. *If the equation*

$$\int_0^1\int_0^1 K(x, y; z)\varphi(x)\varphi(y)\,dx\,dy = \int_0^1\int_0^1 K(x; y; z)\varphi_1(x)\varphi_1(y)\,dx\,dy$$

is satisfied by any positive function $\varphi(x)$ *obeying the condition*

$$\int_0^1 \varphi(x)\,dx = 1$$

and

$$\varphi_1(u) = \int_0^1\int_0^1 K(x; y; u)\varphi(x)\varphi(y)\,dx\,dy$$

with a positive function $K(x; y, z)$ *symmetric with respect to x and y and such that*

$$\int_0^1 K(x; y; z)\,dz = 1,$$

then $K(x ; y ; z)$ *is a function of z only.*

Without dwelling in more detail on the case $n = \infty$ or on its connection with the theory of integral equations, we shall consider now the next important case of a finite number of classes.

Chapter 3

10. Suppose that there are in all $N = n + 2$ classes with two of them being pure races. To repeat (cf. Note 3), each of these two produces, under internal crossing, only its own individuals, and, when being mutually crossed, gives rise to individuals of all the other (hybrid) classes. In accord with §6 we would have had Mendelian heredity if the entire totality of the hybrids represented a class. We shall see now that if these hybrids represent several classes, two possibilities should be distinguished from each other:

1) Under internal crossing each of the hybrid classes produces individuals of one of the two pure classes.

2) There exists a hybrid class, which, under the same condition, cannot produce individuals of those two classes.

Denote the functions of reproduction for our N classes by f and f_1 for the pure races and by φ_i, $i = 1, 2, \ldots, n$ for the hybrid races, and the respective probabilities by α, β and γ_i. Then our main assumption means that all the quadratic forms φ_i have terms containing αβ but that they do not include α^2 or β^2. On the contrary, the form f contains α^2 (with coefficient 1) and does not include either αβ or β^2 and f_1 contains β^2 (with coefficient 1) but does not include either αβ or α^2.

It is not difficult to prove, first of all, that in this case f does not at all depend on β, nor does f_1 depend on α; in other words, that crossing with one of the parents belonging to a pure race never produces an individual of the other pure race. Indeed, let us assume that initially $\gamma_i = 0$ for all values of *i;* then, because of the principle of stationarity,

$$(\alpha + \beta)^2 f = f^2 + f\Sigma A_i\,\varphi_i + \Sigma A_{ik}\varphi_i\varphi_k + f_1\Sigma D_i\psi_i,$$

but in this case $f = \alpha^2$, $f_1 = \beta^2$, and $\varphi_1 = 2c_1\alpha\beta$ where $c_1 > 0$. Since β is not included in the left side in a degree higher than the second, $D_i = 0$ for all values of i which confirms the above.

11. Before going on to the proof of the general proposition, we dwell for the sake of greater clearness on the case $N = 4$. The general statement will be its direct generalization demanding some additional essential considerations.

Theorem. *For* $N = 4$ *the formulas of reproduction should have one of the two following forms*: *either*

$$f = [\alpha + (1/2) A_1\gamma_1 + (1/2)A_2\gamma_2]^2, f_1 = [\beta + (1/2)B_1\gamma_1 + (1/2)B_2\gamma_2]^2,$$
$$\varphi_i = 2c_i[\alpha + (1/2)A_1\gamma_1 + (1/2)A_2\gamma_2]\cdot[\beta + (1/2)B_1\gamma_1 + (1/2)B_2\gamma_2], i = 1, 2, \quad (45)$$

where $c_1 + c_2 = 1$, $A_1 + B_1 = A_2 + B_2 = 2$, $A_1c_1 + A_2c_2 = 1$. *Or,*

$$f = (\alpha + \gamma_1)\cdot(\alpha + \gamma_2), f_1 = (\beta + \gamma_1)\cdot(\beta + \gamma_2), \quad (46)$$
$$\varphi_i = (\alpha + \gamma_i)\cdot(\beta + \gamma_i), i = 1, 2.$$

Indeed, let us assume at first that there exists an identical dependence

$$c_2\varphi_1 = c_1\varphi_2 \quad (47)$$

between φ_1 and φ_2. Then, supposing from the very beginning that $c_2\gamma_1 = c_1\gamma_2$, we may unite both hybrid classes in one so as to obtain a biotype of three classes that must obey the Mendelian law. Consequently,

$$f(\alpha; \beta; c_1\gamma; c_2\gamma) = (\alpha + \gamma/2)^2, f_1(\alpha; \beta; c_1\gamma; c_2\gamma) = (\beta + \gamma/2)^2.$$

Therefore, assuming that

$$f = \alpha^2 + \alpha\Sigma A_i\gamma_i + \Sigma A_{ik}\gamma_i\gamma_k, f_1 = \beta^2 + \beta\Sigma B_i\gamma_i + \Sigma B_{ik}\gamma_i\gamma_k,$$

we find that

$$A_1c_1 + A_2c_2 = B_1c_1 + B_2c_2 = c_1 + c_2 = 1,$$
$$\Sigma A_{ik}c_ic_k = \Sigma B_{ik}c_ic_k = (1/4)(c_1 + c_2)^2. \quad (48)$$

But, forming the equation of stationarity for f, we obtain

$$ff_1 = f[(A_1 - 1)\varphi_1 + (A_2 - 1)\varphi] + \Sigma A_{ik}\varphi_i\varphi_k, \quad (49)$$

and, applying equalities (48) and the relation (47), we conclude that

$$ff_1 = (1/4)(\varphi_1 + \varphi_2)^2.$$

It follows that f and f_1 should be exact squares and we immediately arrive at formulas (45).

Let us suppose now that, on the contrary, there is no identical proportionality between the functions φ_1 and φ_2. Then any dependence between the functions of reproduction should contain at least three of them. We have seen, however (§10), that there exists an infinite set of stationary conditions, under which relation (47) holds with $2c_1$ and $2c_2$ being the coefficients of αβ in φ_1 and φ_2 respectively, and satisfying the equation

$$4c_1^2 ff_1 = \varphi_1^2. \qquad (50)$$

Therefore, if there exists a quadratic dependence $F(f; f_1; \varphi_1, \varphi_2) = 0$ between the four arguments (no linear dependence can exist), it should be identically obeyed, when, at the same time, equalities (50) are satisfied and (47) holds. Consequently,

$$F(\alpha; \beta; \gamma_1; \gamma_2) = P(\alpha; \beta; \gamma_1; \gamma_2)\cdot(c_2\gamma_1 - c_1\gamma_2) + k(4c_1^2\alpha\beta - \gamma_1^2)$$

where P is a polynomial of the first degree and k is a constant. Therefore, a second similar restriction together with the first one would have led to a linear dependence which is impossible. We thus conclude that the equations of stationarity for f and f_1

$$ff_1 = f[(A_1 - 1)\varphi_1 + (A_2 - 1)\varphi_2] + \Sigma A_{ik}\varphi_i\varphi_k,$$
$$ff_1 = f[(B_1 - 1)\varphi_1 + (B_2 - 1)\varphi_2] + \Sigma B_{ik}\varphi_i\varphi_k$$

should be equivalent,

$$A_1 = B_1 = A_2 = B_2 = 1, A_{ik} = B_{ik}$$

and the equation of stationarity becomes

$$F = ff_1 - \Sigma A_{ik}\varphi_i\varphi_k = 0. \qquad (51)$$

The forms φ_1 and φ_2 should therefore be

$$\varphi_i = 2c_i(\alpha\beta - \Sigma A_{ik}\gamma_i\gamma_k) + \gamma_i S,\ i = 1, 2. \qquad (52)$$

But $A_{11} = A_{22} = 0$, otherwise our forms will admit negative coefficients. Thus,

$$ff_1 = 2A_{12}\varphi_1\varphi_2 \qquad (53)$$

and we conclude that

$$f = \alpha^2 + \alpha\gamma_1 + \alpha\gamma_2 + 2A_{12}\gamma_1\gamma_2, f_1 = \beta^2 + \beta\gamma_1 + \beta\gamma_2 + 2A_{12}\gamma_1\gamma_2$$

can be decomposed into factors. Therefore, $A_{12} = 1/2$ and

$$f = (\alpha + \gamma_1)\cdot(\alpha + \gamma_2), f_1 = (\beta + \gamma_1)\cdot(\beta + \gamma_2).$$

Noting finally that $c_1, c_2 \le 1/2$ is necessary for the coefficients in φ_1 and φ_2 to be positive, we find that $c_1 = c_2 = 1/2$ and

$$\varphi_1 = (\alpha + \gamma_1)\cdot(\beta + \gamma_1),\ \varphi_2 = (\alpha + \gamma_2)\cdot(\beta + \gamma_2), \text{ QED.}$$

The law of heredity represented by formulas (45) does not fundamentally deviate from the Mendelian law. On the contrary, formulas (46) provide a really peculiar “quadrille” law of heredity when both hybrid classes are pure races. This is the only law (apart from its simple modifications which will follow from the general theorem) admitting a direct appearance of a new pure race when the given pure races are being crossed. It would be interesting to apply it for an experimental investigation of the cases contradicting the Mendelian theory in which the appearance of “constant” hybrids is observed.

I also note the essential difference between the formulas (45) and (46): the former correspond to the case in which each hybrid can reproduce the initial pure races whereas the latter correspond to the contrary case. We go on now to the main proposition.

12. Theorem. *Given, a closed biotype consisting of* $(n + 2)$ *classes two of which are pure races*; *under mutual crossing these two produce individuals belonging to any of the other classes but cannot give rise to individuals of the parent classes. Then, the law of heredity obeying the principle of stationarity must belong to one of the two following types.*

1) *If, under internal crossing, each of the other* (*hybrid*) *classes can produce an individual belonging to one of the abovementioned pure classes, the law of heredity is a generalization of the Mendelian law and is represented by the formulas*

$$f = [\alpha + (1/2)(A_1\gamma_1 + \ldots + A_n\gamma_n)]^2, f_1 = [\beta + (1/2)(B_1\gamma_1 + \ldots + B_n\gamma_n)]^2, \quad (54)$$
$$\varphi_i = 2c_i[\alpha + (1/2)(A_1\gamma_1 + \ldots + A_n\gamma_n)]^2[\beta + (1/2)(B_1\gamma_1 + \ldots + B_n\gamma_n)]^2,$$

$$\Sigma c_i = 1, \Sigma A_i c_i = 1, A_i + B_i = 2.$$

2) *If there exist such hybrid classes which, under the same condition, cannot give rise to individuals of the abovementioned pure races, the law of heredity belongs to the "quadrille" type and is represented by the formulas*

$$f = (\alpha + \gamma_1 + \gamma_2 + \ldots + \gamma_k)\cdot(\alpha + \gamma_{k+1} + \ldots + \gamma_n),$$
$$f_1 = (\beta + \gamma_1 + \gamma_2 + \ldots + \gamma_k)\cdot(\beta + \gamma_{k+1} + \ldots + \gamma_n),$$
$$\varphi_i = c_i(\alpha + \gamma_1 + \ldots + \gamma_k)\cdot(\beta + \gamma_1 + \ldots + \gamma_k), i \le k, \Sigma c_i = 1, \quad (55)$$
$$\varphi_j = d_j(\alpha + \gamma_{k+1} + \ldots + \gamma_n)\cdot(\beta + \gamma_{k+1} + \ldots + \gamma_n), j > k, \Sigma d_j = 1.$$

Keeping to the previous notation, we obtain

$$f = \alpha^2 + \alpha\Sigma A_i\gamma_i + \Sigma A_{ik}\gamma_i\gamma_k, f_i = \beta^2 + \beta\Sigma B_i\gamma_i + \Sigma B_{ik}\gamma_i\gamma_k. \quad (56)$$

Let us first consider the case of $A_i = B_i = 1$ and $A_{ik} = B_{ik}$. The equations of stationarity for the functions f and f_1 will then be identical and have the form

$$F = \alpha\beta - \Sigma A_{ik}\gamma_i\gamma_k = 0. \quad (57)$$

Before proving our proposition, which certainly demands that all the coefficients be non-negative, I indicate, as an overall guidance, the most general solution (without allowing for the signs of the coefficients) under the condition that the stationary distribution is only restricted by one equation. Since this single equation must be (57), the general type of the functions φ_i is

$$\varphi_i = 2c_iF + \gamma_iS, \Sigma c_i = 1.$$

Consequently, the equation of stationarity becomes

$$(\alpha S - F)\cdot(\beta S - F) = \sum_{ik} A_{ik}[2c_iF + \gamma_iS]\cdot[2c_iF + \gamma_kS]$$

or

$$F^2 - (\alpha + \beta)FS + \alpha\beta S^2 = 4F^2\sum_{ik} A_{ik}cc_k + 4SF\sum_{ik} A_{ik}c_i\gamma_k + S^2\sum_{ik} A_{ik}\gamma_i\gamma_k \quad (58)$$

and, after cancelling F out of it,

$$F(1 - 4\sum_{ik} A_{ik}c_ic_k) = S(\alpha + \beta + 4\sum_{ik} A_{ik}c_i\gamma_k). \qquad (59)$$

It is thus necessary and sufficient that

$$4\Sigma A_{ik}c_i = 1,\ \Sigma c_i = 1$$

because these conditions also lead to $4\sum_{ik} A_{ik}c_ic_k = 1$.

The general solution therefore depends on $\{[n(n + 1)/2] + n\}$ parameters connected by $(n + 1)$ equations; that is, actually, on $(n + 2)(n – 1)/2$ independent parameters.

It is, however, easy to see that for $n > 2$ neither of these solutions suits us: in this case the number of independent stationary equations is always greater than unity. And so, we ought to assume, that, in general,

$$\varphi_i = 2cF + \gamma_iS + \alpha S_i + \beta S_i' + \psi_i \qquad (60)$$

where S_i and S_i' are linear functions of $(\gamma_1; \gamma_2; \ldots; \gamma_n)$, ψ_i are quadratic functions of the same variables, and

$$\Sigma S_i = \Sigma S_i' = \Sigma\psi_i = 0. \qquad (61)$$

We shall now determine the functions S_i, S_i' and ψ_i noting that the conditions of stationarity for each φ_i become

$$fS_i(\varphi_1; \varphi_2; \ldots; \varphi_n) + f_1S_i'(\varphi_1; \varphi_2; \ldots; \varphi_n) + \psi_i(\varphi_1; \varphi_2; \ldots; \varphi_n) = 0. \qquad (62)$$

Assume now that

$$S_i = \sum_h A_h^i\gamma_h,\ S_i' = \sum_h B_h^i\gamma_h. \qquad (63)$$

Then, equating to zero the coefficient of α^3 in the identity (62), we obtain

$$S_i(\gamma_1 + S_1; \gamma_2 + S_2; \ldots; \gamma_n + S_n) = 0 \qquad (64)$$

or

$$\sum_h A_h^i S_h + S_i = 0,\ i = 1, 2, \ldots, n.$$

In exactly the same way we could have gotten

$$\sum_h B_h^i S_h' + S_i' = 0 \qquad (65)$$

so that all conclusions which we reach concerning S_i will also hold for S_i'.

Let us compile the table

$$
\begin{array}{lllll}
A_1^{\,1}+1 & A_1^{\,2} & A_1^{\,3} & \dots & A_1^{\,n} \\
A_2^{\,1} & A_2^{\,2}+1 & \dots & \dots & A_2^{\,n} \\
\dots & \dots & \dots & \dots & \dots \\
A_n^{\,1} & A_n^{\,2} & \dots & \dots & A_n^{\,n}+1
\end{array}
\qquad (66)
$$

On the strength of equations (65), for each of the columns

$$\sum_h \lambda_h^{\,i} S_h = 0$$

where $\lambda_h^{\,i}$ is the term in column i and line h (counting from above).

Note that all the coefficients in S_h excepting $A_h^{\,h}$ are non-negative because they are non-negative in φ_n, and that

$$-A_h^{\,h} = A_h^{\,1} + \dots + A_h^{\,h-1} + A_h^{\,h+1} + \dots + A_h.$$

Let $\lambda_r^{\,i}$, $\lambda_s^{\,i}$, $\lambda_t^{\,i}$ be he maximal terms of column i [16]. In general, for any value of p,

$$\sum_h \lambda_k^{\,i} A_p^{\,k} = 0.$$

Therefore, choosing, in particular, $p = r, s, t$, we obtain

$$
\begin{aligned}
&(\lambda_1^{\,i} - \lambda_r^{\,i})A_r^{\,1} + (\lambda_2^{\,i} - \lambda_r^{\,i})A_r^{\,2} + \dots + (\lambda_n^{\,i} - \lambda_r^{\,i})A_r^{\,n} = 0,\\
&(\lambda_1^{\,i} - \lambda_s^{\,i})A_s^{\,1} + (\lambda_2^{\,i} - \lambda_s^{\,i})A_s^{\,2} + \dots + (\lambda_n^{\,i} - \lambda_s^{\,i})A_s^{\,n} = 0,\\
&(\lambda_1^{\,i} - \lambda_t^{\,i})A_t^{\,1} + (\lambda_2^{\,i} - \lambda_t^{\,i})A_t^{\,2} + \dots + (\lambda_n^{\,i} - \lambda_t^{\,i})A_t^{\,n} = 0.
\end{aligned}
\qquad (67)
$$

Noting that, if k differs from r, s, t

$$\lambda_k^{\,i} - \lambda_r^{\,i} = \lambda_r^{\,i} - \lambda_k^{\,i} - \lambda_s^{\,i} = \lambda_k^{\,i} - \lambda_t^{\,i} < 0,$$

we conclude that for these values of k $A_r^{\,k} = A_s^{\,k} = A_t^{\,k} = 0$.

But $\lambda_h^{\,i} = A_h^{\,i}$ for $i \neq h$ and $\lambda_i^{\,i} = A_i^{\,i} + 1$. Therefore, i should be equal to one of the numbers r, s, t. In addition, if the maximal values in the r-th column correspond to the r-th, s-th, t-th lines, then the maximal values of the s-th and the t-th columns will be on the same lines. It follows that

$$
\begin{aligned}
&S_r + \gamma_r = \lambda_r^{\,r}(\gamma_r + \gamma_s + \gamma_t),\ S_s + \gamma_s = \lambda_s^{\,s}(\gamma_r + \gamma_s + \gamma_t),\\
&S_t + \gamma_t = \lambda_t^{\,t}(\gamma_r + \gamma_s + \gamma_t),\ \lambda_r^{\,r} + \lambda_s^{\,s} + \lambda_t^{\,t} = 1.
\end{aligned}
\qquad (68)
$$

In general, all our linear forms S_h break down into several groups so that only the forms of one and the same group depend on the same variables and obey relations of the type of (68). We shall prove now that the number of these groups cannot exceed two.

Indeed, suppose for the sake of definiteness that the first i forms, $S_1, S_2, \dots, S_i$, belong to the same group so that

$$(S_1 + \gamma_1)/\lambda_1 = (S_2 + \gamma_2)/\lambda_2 = \dots = (S_i + \gamma_i)/\lambda_i = (\gamma_1 + \gamma_2 + \dots + \gamma_i). \qquad (69)$$

Then the equations of stationarity for $\varphi_1, \varphi_2, \dots, \varphi_i$ will be

$$f[\lambda_j(\varphi_1 + \varphi_2 + \dots + \varphi_i) - \varphi_i] + f_1 S_j'(\varphi_i; \dots; \varphi_n) + \psi_j(\varphi_1; \dots; \varphi_n) = 0, \qquad (70)$$

where $j = 1, 2, \dots, i$.

Adding up these equalities we see that the term including f vanishes from the left side and the quadratic form, which is the coefficient of α^2 in the sum

$$\psi_1(\varphi_1; \varphi_2; \dots; \varphi_n) + \psi_2(\varphi_1; \varphi_2; \dots; \varphi_n) + \dots + \psi_i(\varphi_1; \varphi_2; \dots; \varphi_n),$$

should therefore be identically equal to zero. Consequently,

$$\psi_1(S_1 + \gamma_1; S_2 + \gamma_2; \dots; S_n + \gamma_n) + \dots + \psi_i(S_1 + \gamma_1; \dots; S_n + \gamma_n) = 0. \qquad (71)$$

However, since the terms not including γ_k in the function $\psi_k(\gamma_1; \gamma_2; \dots; \gamma_n)$ cannot be negative, we ought to conclude that the coefficients of $\gamma_k\gamma_l$ in each function ψ of the group are equal to zero if $k, l > i$. It follows that, for these values of k and l, A_{kl} is all the more zero. The same reasoning might obviously be applied to each group, and, had their number exceeded two, all the A_{kl} would have vanished which is impossible because the equation of stationarity (57) would then be $ff_1 = 0$. Employing the same argument with respect to S_h' we convince ourselves that these forms also break down into two groups possessing the abovementioned properties.

Thus, with respect both to S and S', all the functions φ_i break down into groups constituting not more than four subgroups. We should also note that, when equating to zero the coefficients of $\alpha^3\beta$ in each of the equations (70) we obtain

$$\lambda_1 = c_1/(c_1 + c_2 + \dots + c_i) \text{ etc.}$$

Therefore, supposing that φ_1 and φ_2 belong to one and the same subgroup, we conclude from the respective equations of stationarity that

$$(f + f_1)\cdot(c_2\varphi_1 - c_1\varphi_2) = c_2\psi_1(\varphi_1; \varphi_2; \dots; \varphi_n) - c_1\psi_2(\varphi_1; \varphi_2; \dots; \varphi_n) \qquad (72)$$

and

$$(f + f_1)[(c_2\gamma_1 - c_1\gamma_2)\cdot(\gamma_1 + \gamma_2 + \dots + \gamma_n) + c_2\psi_1 - c_1\psi_2] =$$
$$c_2\psi_1(\varphi_1; \varphi_2; \dots; \varphi_n) - c_1\psi_2(\varphi_1; \varphi_2; \dots; \varphi_n). \qquad (73)$$

Equating the coefficients of α^2 and β^2 in both parts we find that

$$(c_2\gamma_1 - c_1\gamma_2)(\gamma_1 + \gamma_2 + \dots + \gamma_n) + c_2\psi_1 - c_1\psi_2 =$$
$$c_2\psi_1(S_1 + \gamma_1; \dots; S_n + \gamma_n) - c_1\psi_2(S_1 + \gamma_1; \dots; S_n + \gamma_n) =$$
$$c_2\psi_1(S_1' + \gamma_1; \dots; S_n' + \gamma_n) - c_1\psi_2(S_1' + \gamma_1; \dots; S_n' + \gamma_n). \qquad (74)$$

Consequently, if the groups with respect to S and S' do not coincide,

$$(c_2\gamma_1 - c_1\gamma_2)(\gamma_1 + \gamma_2 + \dots + \gamma_n) + c_2\psi_1 - c_1\psi_2 = A(\gamma_1 + \gamma_2 + \dots + \gamma_n)^2 \qquad (75)$$

where A is a numerical coefficient. But since the left side does not include terms with products $\gamma_k\gamma_l$ where neither k nor l belong to the considered subgroup, $A = 0$ and we obtain the very important relation (47). It can also be derived when assuming that the groups with respect to S and S' coincide because then $S_h' = S_h$ so that the quadratic form that serves as the coefficient of $2\alpha\beta$ in the right side of equality (73) and ought to be identically equal to zero, should also be equal to the expression (74).

And so, in any case, the *functions* φ_h *belonging to the same subgroup differ one from another only by numerical coefficients.* It remains to show that there cannot exist more than two such subgroups. To this end we transform our biotype by uniting all the classes of each subgroup. The transformed biotype of each subgroup will then have only one class. It is necessary to check that the assumptions $n = 4$ and $n = 3$ are impossible.

Let at first $n = 4$. Then, in accord with the above,

$$F = \alpha\beta - A_{14}\gamma_1\gamma_4 - A_{23}\gamma_2\gamma_3 \qquad (76)$$

if, for the sake of definiteness, we suppose that S_1 and S_2 belong to the same group, and S_3 and S_4, to another one, and that the same is true for S_1' and S_3' and S_2' and S_4'.

Issuing from the equation of stationarity

$$ff_1 = A_{14}\varphi_1\varphi_4 + A_{23}\varphi_2\varphi_3$$

we obtain, by equating the coefficients of α_2, the equality

$$A_{14}\gamma_1\gamma_4 + A_{23}\gamma_2\gamma_3 = (A_{14}\lambda_1\lambda_4 + A_{23}\lambda_2\lambda_3)(\gamma_1 + \gamma_2)(\gamma_3 + \gamma_4)$$

and arrive at an impossible conclusion that $A_{14} = A_{23} = 0$. In the same way, if $n = 3$,

$$F = \alpha\beta - A_{13}\gamma_1\gamma_3 - A_{22}\gamma_2^2$$

and we get an impossible equality

$$A_{13}\gamma_1\gamma_3 + A_{22}\gamma_2^2 = A_{13}\lambda_1\lambda_3(\gamma_1 + \gamma_2)\gamma_3 + A_{22}(\gamma_1 + \gamma_2)^2.$$

Consequently, $n \le 2$; that is, the number of subgroups, where all the φ_n differ one from another only by numerical factors, is never greater than two if only

$$f = \alpha^2 + \alpha(\gamma_1 + \ldots + \gamma_n) + \Sigma A_{ik}\gamma_i\gamma_k,$$
$$f_1 = \beta^2 + \beta(\gamma_1 + \ldots + \gamma_n) + \Sigma A_{ik}\gamma_i\gamma_k.$$

13. We shall show now that the same conclusion persists also in the general case when f and f_1 can be represented as

$$f = \alpha^2 + \alpha(\gamma_1 + \ldots + \gamma_n) + \alpha S_o + \Sigma A_{ik}\gamma_i\gamma_k,$$
$$f_1 = \beta^2 + \beta(\gamma_1 + \ldots + \gamma_n) + \beta S_o' + \Sigma B_{ik}\gamma_i\gamma_k, \qquad (77)$$

$$S_o = \sum_h A_h^0\gamma_h,\ S_o' = \sum_h B_h^0\gamma_h,\ |A_h^0| \le 1,\ |B_h^0| \le 1.$$

In this case the equations of stationarity for f and f_1 become

$$F = \alpha\beta - \alpha S_o - \Sigma A_{ik}\gamma_i\gamma_k = 0,\ F_1 = \alpha\beta - \beta S_o' - \Sigma B_{ik}\gamma_i\gamma_k = 0. \qquad (78)$$

We may therefore assume that

$$\varphi_i = c_i(F + F_1) + \gamma_i S + \alpha S_i + \beta S_i' + \psi_i \qquad (79)$$

where, as before,

$\Sigma c_i = 1$, $\Sigma S_i = \Sigma S_i' = \Sigma\psi_i = 0$

and the equation of stationarity for φ_i remains in the form (62). Equating to zero the coefficients of α^3 in (62) we obtain now for any i

$$S_i(-c_1S_o + \gamma_1 + S_1; -c_2S_o + \gamma_2 + S_2; \dots; -c_nS_o + \gamma_n + S_n) = 0. \quad (80)$$

It is evident, however, that the equation of stationarity for f leads to

$$S_o(-c_1S_o + \gamma_1 + S_1; \dots; -c_nS_o + \gamma_n + S_n) = 0. \quad (81)$$

Therefore, if we assume that $P_i = S_i - c_iS_o$, then

$$P_i(\gamma_1 + P_1; \gamma_2 + P_2; \dots; \gamma_n + P_n) = 0, \; i = 1, 2, \dots, n. \quad (82)$$

The forms P_i thus have the property

$$\sum_h \lambda_h^i P = 0$$

where λ_h is the term of the i-th column and the h-th line in the table

$$\begin{matrix} A_1^1 - c_1A_1^0 + 1 & A_1^2 - c_2A_1^0 & \dots & A_1^n - c_nA_1^0 \\ A_2^1 - c_1A_2^0 & A_2^2 - c_2A_2^0 + 1 & \dots & A_2^n - c_nA_2^0 \\ A_3^1 - c_1A_3^0 & A_3^2 - c_2A_3^0 & \dots & A_3^n - c_nA_3^0 \\ A_n^1 - c_1A_n^0 & A_n^2 - c_2A_n^0 & \dots & A_n^n - c_nA_n^0 + 1 \end{matrix} \quad (83)$$

In addition,

$$\Sigma A_h^0(P_h + \gamma_h) = 0.$$

Let us now divide the terms of each h-line by $(1 - A_h^0)$ and assume that

$\lambda_r^i/(1 - A_r^0)$ and $\lambda_s^i/(1 - A_s^0)$

represent two (for the sake of definiteness) maximal values which will then be obtained in the i-th column. Then

$$(1 - A_r^0)\sum_h \lambda_h^i P_h + \lambda_r^i \sum_h A_h^0(P_h + \gamma_h) = 0$$

so that, equating the coefficients of γ to zero, we get

$$(1 - A_r^0)\sum_h \lambda_h^i(A_r^h - c_h A_r^0) + \lambda_r^i[\sum_h A_h^0(A_r^h - c_h A_r^0) + A_r^0] = 0 \quad (84)$$

and a similar equality for s.

Noting then that

$$\sum_h (A_r^h - c_h A_r^0) = - A_r^0$$

we transform the equality (84) obtaining

$$(1 - A_r^0)\sum_h (\lambda_h^i - \lambda_r^i)(A_r^h - c_h A_r^0) + \lambda_r^i[\sum_h A_h^0(A_r^h - c_h A_r^0) + (A_r^0)^2] = 0,$$

or, finally,

$$\sum_h [\lambda_h^i(1 - A_r^0) - \lambda_r^i(1 - A_h^0)] (A_r^h - c_h A_r^0) = 0. \quad (85)$$

But, since on the one hand

$$A_r - c_h A_r^0 \geq 0,$$

and the multiplier

$$\lambda_h^i(1 - A_r^0) - \lambda_r^i(1 - A_h^0) = 0$$

when $h = r$ and $h = s$ and is negative at the other values of h, it is necessary that, for the last-mentioned values,

$$A_r^h - c_h A_r^0 = 0, A_s^h - c_h A_s^0 = 0.$$

We conclude therefrom that one of the values r and s should coincide with i and that the maximal values of the r-th and the s-th columns should be situated in the lines with the same numbers. All the forms P_h can be therefore united into groups of the type of

$$(P_1 + \gamma_1)/\lambda_1 = (P_2 + \gamma_2)/\lambda_2 = \ldots = (P_k + \gamma_k)/\lambda_k =$$
$$(1 - A_1^0)\gamma_1 + (1 - A_2^0)\gamma_2 + \ldots + (1 - A_k^0)\gamma_k, \lambda_1 + \lambda_2 + \ldots + \lambda_k = 1. \quad (86)$$

We would have obtained a similar result for $P_h' = S_h' - c_h S_0'$.

Issuing from this main finding, it is easy to determine that the equation of stationarity for φ_1 will become

$$f\{\lambda_1[(1 - A_1^0)\varphi_1 + (1 - A_2^0)\varphi_2 + \ldots + (1 - A_k^0)\varphi_k] - \varphi_1 +$$
$$c_1(A_1^0\varphi_1 + \ldots + A_k^0\varphi_k)\} + f_1\{\lambda_1^1[(1 - B_1^0)\varphi_1 + \ldots + (1 - B_l^0)\varphi_l] -$$
$$\varphi_1 + c_1(B_1^0\varphi_1 + \ldots + B_l^0\varphi_l)\} + \varphi_1(\varphi_1; \varphi_2; \ldots; \varphi_n) = 0. \quad (87)$$

Compiling the stationary equations for φ_i belonging to the same group with respect to S, we find, by equating to zero the coefficients of $\alpha^3\beta$, that

$$\lambda_1/c_1 = \ldots = \lambda_k/c_k = \frac{1 - (A_1^0 c_1 + \ldots + A_k^0 c_k)}{(1 - A_1^0)c_1 + \ldots + (1 - A_k^0)c_k} = 1/(c_1 + \ldots + c_k). \quad (88)$$

Therefore,

$$\text{if} \sum_{i=1}^{k} c_i A_i^0 \neq 0, \text{ then } c_1 + c_2 + \ldots + c_k = 1.$$

Consequently, $k = n$, otherwise all the other $c_i = 0$ which contradicts the condition of the Theorem. We thus have only one group with respect to S except for the case in which the first sum just above vanishes. But then there will be no terms including α in

$$\sum_{i=1}^{k} A_i^0 \varphi_i$$

and the sum $\sum_{i=1}^{k} \psi_i(\varphi_1; \varphi_2, \ldots; \varphi_n)$

will not therefore contain terms with α^2. We conclude that, as before, the products $\gamma_g\gamma_h$ where both g and h are greater than k are absent from any ψ_i belonging to the given group.

Consequently, for these values of g and h, $A_{gh} = B_{gh} = 0$ so that the number of groups with respect to S does not exceed two and we can obtain the same result for S'. Finally, if φ_1 and φ_2 belong to one and the same group with respect to S and S', we find that

$$(f + f_1)(c_2\varphi_1 - c_1\varphi_2) = c_2\varphi_1(\varphi_1; \varphi_2; \ldots; \varphi_n) - c_1\varphi_2(\varphi_1; \varphi_2; \ldots; \varphi_n),$$

and, as before, we convince ourselves that equality (49) holds.

Thus, all the hybrid races again break down in subgroups not exceeding four in number and for whom functions φ_i differ only by numerical factors. It is not difficult to show that, as before, the number of subgroups does not actually exceed two. Hence, the most general instance is also reduced to the case of $n = 2$ considered in §11, and the Theorem is proved.

14. I shall briefly indicate some conclusions following from my investigation. The closed biotype in which each crossing can produce individuals of any class must posses the properties that the proportion of the individuals of different kinds produced after some crossing *does not at all depend on their parents*. In spite of the obvious difference between the parents, the properties of their sexual cells obey one and the same law of randomness. Had a correlation between parents and offspring been nevertheless observed here, its cause should be only sought in the differing influence of the environment and unequal conditions of selection. The considered biotypes, although polymorphic in appearance, do not essentially differ from pure races. I could have proved that the crossing of various biotypes of this kind indeed obeys the same laws as does the crossing of pure races, see formulas (17). The main problem is the crossing of pure races. As is seen from the theorem proved in Chapter 3 [17], if hybrids of different kinds are produced, only two cases are possible:

1) The proportion of the produced hybrids does not depend on the parents [18]; the entire totality of the hybrids here follows the Mendelian law, satisfying, as is seen from formulas (54), the main relation

$$4ff_1 = (\varphi_1 + \varphi_2 + \ldots + \varphi_n)^2.$$

Thus, an usual large-scale statistical investigation, that does not differentiate between the hybrids, would have registered the existence of an elementary Mendelian law and only indicated a more or less greater variance.

2) The hybrids comprise two essentially different groups. Assuming for the sake of simplicity that each group is homogeneous, both these hybrid classes will also represent pure races characterized by producing, under mutual crossing, in turn, the initial pure races. The considered four pure races constitute a peculiar *quadrille*, and I call their law of heredity, essentially differing from the Mendelian, the quadrille law. I have found only a few controversial cases (De Vries) in the pertinent literature suiting the indicated law and it would be necessary to carry out a more thorough check and establish whether the elementary quadrille law, or some generalized form of that law, is applicable here. Finally, the problem, which I solved by means of formulas (19) in the particular case of a simple Mendelian heredity, indicates the nature of the laws of heredity for a complicated biotype consisting of any number of pure races.

Notes

1. Our formulas obviously presuppose an absolute absence of any selection whatsoever. The biotype is reproduced under conditions of panmixia.

2. See my paper (1922).

3. By a pure race we designate a class which, when the crossing is interior, only produces individuals of its own class.

4. The case of two classes is obviously exhausted by the formulas

$$f = \alpha(\alpha + \beta), f_1 = \beta(\alpha + \beta) \text{ and } f = p(\alpha + \beta)^2, f_1 = q(\alpha + \beta)^2.$$

5. An essential part in our derivations, which were meant for biological applications, was played by the restriction imposed on the signs of the coefficients. If we assume that their signs can be arbitrary, the solutions, except for formulas (16), can be of two types. The first one corresponds to a linear function F that depends on five parameters. The second type is characterized by a quadratic function F and is represented by the formulas depending on four parameters P, Q, d and d_1:

$$f = (1/4P)[P\alpha - Q\beta + (d - 1)(d_1 - 1)S][P\alpha - Q\beta + (d + 1)(d_1 + 1)S], \quad (18)$$
$$f_1 = (1/4Q)[P\alpha - Q\beta + (d - 1)(d_1 + 1)S][P\alpha - Q\beta + (d + 1)(d_1 - 1)S],$$

$$\varphi = S^2 - f - f_1.$$

6.{Bernstein only explains the composition of the sum of the terms α_{ih}.}

7. If n is the number of the pure races, then the number of all the classes is $N = n(n + 1)/2$.

8. Because of the assumption that the crossing of races A_{11} and A_{hh} only produces only the race A_{1h}.

9. Incidentally, the law of heredity expressed by equations (19) and representing a simple generalization of the Mendelian law, was applied when studying *Aquilegia* as investigated by Bauer (Johannsen 1926, p. 581).{In this sentence, Bernstein made a grammatical error and "investigated by Bauer" is only my conjecture.}

10. Our theoretical conclusions are fully confirmed by Morgan's (1919) experimental investigations.

11. Otherwise we directly apply the theorem proved for $n = 3$ and obtain

$$\varphi_i = \lambda_i(\alpha_1 + \alpha_2 + \gamma)^2, \; i = 1, 2, 3.$$

Hence we immediately prove the theorem also for $n = 4$.

12. We consider the contrary case below.

13.{In §3, see formula (8), Bernstein made use of the same letter S in another sense.}

14.{It is difficult to understand the end of this sentence. The author wrote: In any case, after equating both parts of (29) to each other …}

15. When $\alpha_1 + \alpha_2 + \ldots + \alpha_{n-1} + \gamma = 1$.

16. For the sake of definiteness we assumed that such terms are three in number, but our reasoning will not change had we chosen another number.

17. The case in which, apart from the hybrids, individuals of the parent classes can also be produced, would have led to an appropriate generalization of the formulas.

18. If the parents themselves are hybrids, a certain part of the offspring, depending on the kind {the sex? – difficult to understand the Russian text}of the parents, belongs to pure races, but the proportion of the different types of the hybrid offspring persists.

References

Aleksandrov, P.S., Akhiezer, N.I., Gnedenko, B.V., Kolmogorov, A.N. (1969), S.N. Bernstein. An obituary. *Uspekhi Matematich. Nauk*, vol. 24, No. 3, pp. 211 – 218. This periodical is being completely translated as *Russian Math. Surveys.*
Bernstein, S.N. (1922), On the application of mathematics to biology. Translated in this collection.
--- (1923a), Démonstration mathématique de la loi d'hérédité de Mendel. *C.r. Acad. Sci. Paris*, t. 177, pp. 528 – 531.
--- (1923b), Principe de stationnarité et généralisations de la loi de Mendel. Ibidem, pp. 581 – 584.
--- (1942), Solution of a mathematical problem connected with the theory of heredity. *Annals Math. Stat.*, vol. 13, pp. 53 – 61.
--- (1946), *Теория вероятностей* (Theory of probability). M. Fourth edition.
Johannsen, Wm (1926), *Elemente der exakten Erblichkeitslehre.* 3. Aufl. Jena.
Karlin, S. (1922), Fisher and evolutionary theory. *Stat. Sci.*, vol. 7, pp. 13 –33.
Kolmogorov, A.N. (1938), Theory of probability and its applications. In *Matematika i Estestvoznanie v SSSR.* M., pp. 51 – 61. (R) Transl.: DHS 2696, 2000, pp. 175 – 186.
Morgan, T.H. (1919), *The physical basis of heredity.* Philadelphia – London.
Seneta, E. (2001), S.N. Bernstein. In *Statisticians of the Centuries.* Editors, C.C. Heyde, E. Seneta. New York, 2001, pp. 339 – 342.
Sheynin, O. (1998), Statistics in the Soviet epoch. *Jahrbücher f. Nat.-ökon. u. Statistik*, Bd. 217, pp. 529 – 549.

3. S.N. Bernstein. An Essay on an Axiomatic Justification of the Theory of Probability

Собрание сочинений (Coll. Works), vol. 4. N.p., 1964, pp. 10 – 60 …

The human mind experiences
less{difficulties}when moving
ahead than when probing itself
(Laplace 1812)

Foreword by Translator

The first to argue that the theory of probability should be axiomatized was Boole (1854, p. 288); Hilbert, in 1901, attributed the theory to physical sciences and formulated the same demand. Bernstein was likely the first to develop an axiomatic approach to probability, and he later described his attempt in each edition of his treatise (1927). Then, in an extremely short annotation to vol. 4 (1960) of his *Собрание Сочинений* (Coll. Works), where his work was reprinted, he stated that his axiomatics was the basis of a "considerable part" of his writings of 1911 – 1946. Slutsky (1922) examined the logical foundation of the theory of probability. Several years later he (1925, p. 27n) remarked that then, in 1922, he had not known Bernstein's work which "deserves a most serious study".

Kolmogorov (1948, p. 69) described Bernstein's work as follows: the essence of his concept consisted "not of the numerical values of the probability of events but of a qualitative comparison of events according to their higher or lower probabilities …" Then, he and Sarmanov (1960, pp. 215 – 216) largely repeated that statement and added that Koopman had "recently" been moving in the same direction. In turn, Ibragimov (2000, p. 85) stated that both Bernstein and Kolmogorov had "adopted the structure of normed Boolean algebras as the basis of probability theory". Ibragimov (pp. 85 and 86) also politely called in question

Bernstein's standpoint regarding infinitely many trials (his §3.2.1) and even his opinion concerning general mathematical constructions such as convergence almost everywhere.

Bernstein's contribution translated below is hardly known outside Russia. Even Hochkirchen (1999) only mentioned it in his list of *Quellen und Fachliteratur* but not at all in his main text.

Bernstein had not systematized his memoir. The numbering of the formulas was not thought out, theorems followed one another without being numbered consecutively; notation was sometimes violated and misprints were left unnoticed. Finally, in §3.2.5 Bernstein introduced function $F(z)$ which appeared earlier as $F(x)$. And, what happens time and time again in the works of many authors, he had not supplied the appropriate page number in his epigraph above. I have not been able to correct sufficiently these shortcomings but at least I methodically numbered the axioms, theorems and corollaries although several propositions not called either theorems or corollaries; again, yet others were named *principles*.

* * *

The calculation of probabilities is based on several axioms and definitions. Usually, however, these main axioms are not stated sufficiently clearly; it remains therefore an open question which assumptions are necessary, and whether they do not contradict one another. The definition itself of mathematical probability implicitly contains a premise (Laplace 1814, p. 4) in essence tantamount to the addition theorem which some authors (Bohlmann ca. 1905, p. 497) assume as an axiom. Consequently, I consider it of some use to explicate here my attempt to justify axiomatically the theory of probability. I shall adhere to a purely mathematical point of view that only demands a rigorous and exhausting statement of independent rules not contradicting each other, on whose foundation all the conclusions of the theory, regarded as an abstract mathematical discipline, ought to be constructed. It is of course our desire for cognizing the external world as precisely as possible that dictates us these rules. However, so as not to disturb the strictly logical exposition, I prefer to touch the issue of the philosophical and practical importance of the principles of probability theory only in a special supplement at the end of this paper.

Chapter 1. Finite Totalities of Propositions

1.1. Preliminary Definitions and Axioms

1.1.1. Equivalent and Non-Equivalent Propositions

Let us consider a finite or infinite totality of symbols A, B, C, etc which I shall call *propositions*. I shall write $M = N$ ($N = M$) and call M and N equivalent after agreeing that, when performing all the operations defined below on our symbols, it is always possible to replace M by N and vice versa. In particular, if $M = N$ and $M = L$, then $N = L$.

Suppose that not all of the given propositions are equivalent, that *there exist two such A and B that* $A \neq B$. *If the number of non-equivalent propositions is finite*, *I shall call their given totality finite*; otherwise, *infinite*. In this chapter, I consider only *finite* totalities.

1.1.2. Axioms Describing the Operation (of Partition) Expressed by the Sign "Or"

1.1. *The constructive principle*: *If* (*in the given totality*) *there exist propositions A and B*, *then proposition* $C = (A$ or $B)$ *also exists*.

1.2. *The commutative principle*: $(A$ or $B) = (B$ or $A)$.

1.3. *The associative principle*: $[A$ or $(B$ or $C)] = [(A$ or $B)$ or $C] = (A$ or B or $C)$.

1.4. *The principle of tautology*: $(A$ or $A) = A$.

By applying the first three principles it is possible to state that, in general, there exists a quite definite proposition $H = (A$ or B or … $E)$. I shall call it a *join* of propositions A, B, …, E. Each of these is called a *particular* case of H.

Corollary 1.1. *If* y *is a particular case of* A, *i.e.*, if (x or y) = A, *then* (A or y) = A. Indeed, given that (x or y) = A, we conclude that

[x or (y or y)] = (A or y),

hence (x or y) = (A or y) = A, QED.

Corollary 1.2. *If* y *is a particular case of* A *and* A *is a particular case of* B, *then* y *is a particular case of* B.

Corollary 1.3. *The necessary and sufficient condition*{s} *for* H *to be a join of propositions* $A_1, A_2, \ldots, A_n$ *is that*

1) *If for some* i, $i = 1, 2, \ldots, n$, *we have* (A_i or y) = A_i, *then* (H or y) = H, and

2) *If for each* i (A_i or M) = M, *then* (H or M) = M.

Indeed, if H = (A_1 or A_2 … or A_n) then (A_1 or y) = A_1 immediately leads to H = (H or y), and in the same way we find that

[(A_1 or M) or (A_2 or M) or … (A_n or M)] = M

follows from (A_1 or M) = M, (A_2 or M) = M etc so that (H or M) = M.

Suppose now that, in addition to the join H, there exists a proposition H_1 possessing the same two properties. Then, since (A_i or A_i) = A_i, we have, for any i, (H_1 or A_i) = H_1, hence (H_1 or H) = H_1. But, since on the other hand (A_i or H) = H for each i, we have, in accord with the second condition, (H_1 or H) = H. Consequently, $H_1 = H$, QED.

Corollary 1.4. If A *is a particular case of* B, *and* B *is a particular case of* A, *then* $A = B$. Indeed, according to the condition, (A or x) = B, (B or y) = A. Consequently, on the strength of Corollary 1.1, (A or B) = B = A, QED.

Corollary 1.5. *Each proposition is equal to the join of all of its particular cases.*

1.1.3. The Existence Theorem for a Certain (a True) Proposition

Theorem 1.1. *In a given totality, there always exists such a proposition* Ω *that, for any proposition* A,

$$(\Omega \text{ or } A) = \Omega. \quad (1)$$

This proposition is called true or certain. Indeed, let us form the join Ω of all the propositions of the totality. Then, in accord with the definition of a join, Ω will satisfy condition (1). The formula of a true proposition means that the correctness either of the true, or of some other proposition, is the same as the correctness of the former.

Corollary 1.6. *All true propositions are equivalent.*

1.1.4. The Existence Theorem for an Impossible (for a False) Proposition

Theorem 1.2. *In a given totality there exists a proposition* O *called false or impossible satisfying the condition that, for each* A,

$$(A \text{ or } O) = A. \quad (2)$$

Thus, to state the existence of a false proposition or of A is tantamount to stating the existence of the latter.

Corollary 1.7. *All false propositions are equivalent*. Indeed, if O and O_1 are two false propositions, then (O or O_1) = O = O_1.

Corollary 1.8. *A true proposition cannot be equivalent to a false proposition*. Indeed, if $\Omega = O$, then, for each A, (A or O) = (A or Ω) = A = Ω = O, which would have meant that all the propositions of a totality were equivalent.

1.1.5. Combining Propositions

If two propositions, A and B, are given, then there always exists a proposition x satisfying the condition{s}

$$(x \text{ or } A) = A,\ (x \text{ or } B) = B. \qquad (3)$$

Indeed, the impossible proposition O in any case obeys (3).

Proposions A and B are called *incompatible* if O is the only proposition satisfying this condition, and *compatible* otherwise. Any proposition x obeying (3) can be called a *particular compatible case* of propositions A and B.

The join H of all the particular compatible cases of A and B, – *that is, of all the propositions x satisfying condition* (3), – *is called the combination of A and B*. It is expressed by the symbol H = (A and B) and is formally determined by the conditions (H or A) = A, (H or B) = B with (x or H) = H if (x or A) = A, (x or B) = B.

Corollary 1.9. *The operation* (*of combining*) *expressed by the symbol* and *is commutative*: (A and B) = (B and A).

Corollary 1.10. *The same operation is associative*: [A and (B and C)] = [(A and B) and C]. Indeed, proposition z satisfying the conditions

$$(z \text{ or } A) = A,\ (z \text{ or } B) = B,\ (z \text{ or } C) = C$$

means that

$$[z \text{ or } (A \text{ and } B)] = (A \text{ and } B),\ (z \text{ or } C) = C.$$

Therefore, the join H of all such propositions will be H = [(A and B) and C] and exactly in the same way we convince ourselves that H = [A and (B and C)], QED.

Corollary 1.11. *If* (A or B) = A, *then* (A and B) = B *and vice versa*. Indeed, if (A or B) = A, then the conditions (z or A) = A, (z or B) = B are equivalent to the second of these and therefore (A and B) = B. Conversely, (A and B) = B means that the equality (z or B) = B always leads to (z or A) = A and, in particular, to (B or A) = A.

Corollary 1.12. (A and O) = O, A and Ω) = A.

Corollary 1.13.*The operation expressed by the symbol* and *satisfies the principle of tautology*: (A and A) = A.

1.1.6. The Restrictive Principle (Or Axiom)

Axiom 1.5. *Each particular case* (A or B) *is a join of some particular cases A* and *B*.

Theorem 1.3. *The First Theorem of Distributivity:*

$$[A \text{ and } (B \text{ or } C)] = [(A \text{ and } B) \text{ or } (A \text{ and } C)].$$

Indeed, from the equalities

$$[(A \text{ and } B) \text{ or } A] = A,\ [(A \text{ and } C) \text{ or } A] = A$$

we conclude that

{[(*A* and *B*) or (*A* and *C*)] or *A*} = *A*.

Exactly in the same way

{[(*A* and *B*) or (*A* and *C*)] or (*B* or *C*)} = (*B* or *C*)

follows from

[(*A* and *B*) or *B*] = *B*, [(*A* and *C*) or *C*] = *C*.

Thus, [(*A* and *B*) or (*A* and *C*)] is a compatible particular case of propositions *A* and (*B* or *C*).
It is still necessary to prove that, conversely, if

(*z* or *A*) = *A*, [*z* or (*B* or *C*)] = (*B* or *C*)

then

{*z* or [(*A* and *B*) or (*A* and *C*)]} = [(*A* and *B*) or (*A* and *C*)].

To this end, we note that on the strength of the restrictive principle *z* = (*x* or *y*)

where *x* and *y* are particular cases of *B* and *C*, respectively. Then, (*x* or *A*) = *A*, (*x* or *B*) = *B*, hence

[*x* or (*A* and *B*)] = (*A* and *B*).

In the same way

[*y* or (*A* and *C*)] = (*A* and *C*).

Consequently,

{(*x* or *y*) or [(*A* and *B*) or (*A* and *C*)} = [(*A* and *B*) or (*A* and *C*)];

that is,

{*z* or [(*A* and *B*) or (*A* and *C*)]} = [(*A* and *B*) or (*A* and *C*)], QED.

Theorem 1.4. *The Second Theorem of Distributivity:*

[*A* or (*B* and *C*)] = [(*A* or *B*) and (*A* or *C*)].

Indeed,

[(*A* or *B*) and (*A* or *C*)] = {[(*A* or *B* and *A*] or [(*A* or *B*) and *C*]} =
{*A* or [(*A* or *B*) and *C*]} = {*A* or [(*A* and *C*) or (*B* and *C*)]} =
[*A* or (*B* and *C*)], QED.

1.1.7. Duality of the Operations of Combining and Joining

The theorems above along with the associative and commutative principles relative to the operations *or* and *and* exhaust the rules of calculation with these symbols.

It is important to indicate that *all the rules concerning the combination of propositions* (having to do with the symbol *and*) *are necessary conclusions from the rules about the joining of the propositions* (involving the symbol *or*). And very remarkable is the duality that exists here: *The rules regarding the symbols "or" and "and" are absolutely identical so that all the formulas persist when the symbols are interchanged if*{only}*the impossible and the true propositions, O and* Ω, *at the same time also replace each other.*

Indeed, suffice it to observe all the above to note that the sole difference between the two sets of rules is that (A and Ω) = A, (A and O) = O, whereas (A or Ω) = Ω and (A or O) = A.

1.1.8. The Principle (Axiom 1.6) of Uniqueness

To complete our system, I introduce one more principle (*the principle of uniqueness* as I shall call it) underpinning the notion of *negation*. It can be stated thus: *If proposition* α *is compatible with all the propositions of a totality* (*excepting O*), *it is true*: α = Ω.

Definition of negation: *The join* $\overline{A}$ *of all propositions incompatible with A is called the negation of A.*

Corollary 1.14. $\overline{\Omega} = O$.

Corollary 1.15. $\overline{O} = \Omega$.

Corollary 1.16. *If* $\overline{x} = O$, *then* $x = \Omega$. Indeed, all the propositions (excepting O) are compatible with x and, consequently, on the strength of the principle of uniqueness, $x = \Omega$.

Corollary 1.17. *If* $\overline{x} = \Omega$ *then* $x = O$. Indeed, since Ω is the join of the propositions incompatible with x, it itself possesses the same property because of the restrictive principle and therefore $x = O$.

Any set of propositions whose join is Ω is called *solely possible*.

Theorem 1.5. *Propositions A and* $\overline{A}$ *are solely possible and incompatible*; *that is*, (A or $\overline{A}$) = Ω, (A and $\overline{A}$) = O.

Indeed, any proposition is either a particular case of $\overline{A}$ or compatible with A; therefore, because of the principle of uniqueness[1], (A or $\overline{A}$) = Ω. On the other hand, since $\overline{A}$ is the join of propositions incompatible with A, (A and $\overline{A}$) = O.

Theorem 1.6. *A negation of proposition* $\overline{A}$ *is equivalent with A*: neg($\overline{A}$) = A.

It is sufficient to prove that $A = A_1$ always follows from

(A or B) = (A_1 or B), (A and B) = (A_1 and B).

Indeed, however,

A_1 = [A_1 and (B or A_1)] = [A_1 and (B or A)] =
[(A_1 and B) or (A_1 and A)] = [(A and B) or (A_1 and A)] =
[A and (B or A_1)] = [A and (B or A)] = A.

Definition. *If* (A or B) = B,
then the join C of all the propositions which are particular cases of B and incompatible with A is called the complement of A to B. Thus, C = (B and $\overline{A}$). Conversely, A is a complement of C to B. Indeed, (A or C) = B, (A and C) = O. Suppose that A_1 is a complement of C to B, then we would also have (A_1 or C) = B, (A_1 and C) = O and $A = A_1$. Therefore, A = (B and $\overline{C}$).

Corollary 1.18. Neg (A and B) = ($\overline{A}$ or $\overline{B}$). Indeed,

{[(A and B) or $\overline{A}$] or $\overline{B}$} = [(Ω and B) or $\overline{B}$] = Ω,

$$[(A \text{ and } B) \text{ and } (\overline{A} \text{ or } \overline{B})] = [A \text{ and } B \text{ and } \overline{A}) \text{ or } (A \text{ and } B \text{ and } \overline{B}) = O.$$

1.1.9. Solving Symbolic Equations

The principles explicated above allow us to solve, or to convince us in the insolubility of relations between propositions connected by symbols *or* and *and*. It is easy to see that any expression including propositions x and $\overline{x}$ is reducible, on the strength of the rules stated above, to $[A$ or $(a$ and $x)$ or $(b$ and $\overline{x})]$.

I call a symbolic equation in one unknown, x, a statement that two expressions, at least one of which depends on this x, are equivalent. Thus, any {such} equation is reducible to

$$[A \text{ or } (a \text{ and } x) \text{ or } (b \text{ and } \overline{x})] = [A' \text{ or } (a' \text{ and } x) \text{ or } (b' \text{ and } \overline{x})]. \qquad (4)$$

This equation is in general equivalent to two different equations which should be satisfied at the same time:

$$\{[A \text{ or } (a \text{ and } x) \text{ or } (b \text{ and } \overline{x})]$$
$$\text{or } [\overline{A'} \text{ and Neg } (a' \text{ and } x) \text{ and } (b' \text{ and } \overline{x})]\} = \Omega \qquad (5)$$

and

$$\{[A \text{ or } (a \text{ and } x) \text{ or } (b \text{ and } \overline{x})]$$
$$\text{or } [\overline{A}' \text{ and Neg } (a' \text{ and } x) \text{ and Neg } (b' \text{ and } \overline{x})]\} = \Omega. \qquad (6)$$

Indeed, equation (5) expresses the fact that the right side of (4) is a particular case of its left side, whereas equation (6) reflects the converse statement. By means of{one of the} theorem{s}of distributivity, equation (5) is transformed into

$$<[A \text{ or } (\overline{A}' \text{ and } \overline{a}' \text{ and } b')] \text{ or } \{[a \text{ or } (\overline{A}' \text{ and } \overline{a}')] \text{ and } x\} \text{ or}$$
$$\{[b \text{ or } (\overline{A}' \text{ and } b') \text{ and } \overline{x}\}> = \Omega \qquad (5\text{bis})$$

so that each of the equations (5) and (6) will be reduced to

$$[B \text{ or } (C \text{ and } x) \text{ or } (D \text{ and } \overline{x})] = \Omega \qquad (7)$$

which means that

$$[(B \text{ or } C \text{ or } D) \text{ and } (B \text{ or } x \text{ or } D) \text{ and } (B \text{ or } C \text{ or } x)] = \Omega$$

and

$$(B \text{ or } C \text{ or } D) = \Omega,\ (B \text{ or } D \text{ or } x) = \Omega,\ (B \text{ or } C \text{ or } \overline{x}) = \Omega. \qquad (8'; 8''; 9)$$

Equality (8′) is a necessary and sufficient condition [2] for the solvability of equation (7). Indeed, the equality (8″) means that

$$[x \text{ or } (\overline{B} \text{ and } \overline{D})] = x. \qquad (10)$$

In the same way the equality (9) leads to

$$[x \text{ or } (B \text{ or } C)] = (B \text{ or } C), \qquad (11)$$

but for the simultaneous realization of (10) and (11) it is necessary and sufficient that

$$[(\overline{B} \text{ and } \overline{D}) \text{ or } (B \text{ or } C)] = (B \text{ or } C)$$

or

$$[\overline{D} \text{ or } (B \text{ or } C)] = (B \text{ or } C). \qquad (12)$$

which is equivalent to (8′).

If condition (12) or its equivalent (8′) holds, then the equations (10) and (11) mean that x is a particular case of (B or C) which includes ($\overline{B}$ and $\overline{D}$); or, otherwise, they mean that

$$x = \{(\overline{B} \text{ and } \overline{D}) \text{ or } [(B \text{ or } C) \text{ and } \delta]\}, \qquad (13)$$

with δ being an arbitrary proposition, is the general solution of equation (7). In particular, condition (8′) is satisfied if (B or D) = Ω; then, the equation (7) becomes

$$[B \text{ or } (C \text{ and } x) \text{ or } \overline{x}] = \Omega$$

whose solution is x = [(B or C) and δ].

Equation (7) will be an identity if and only if (B or D) = Ω, (B or C) = Ω. On the contrary, this equation will admit only one solution only if ($\overline{B}$ and $\overline{D}$) = (B or C), and, therefore, only if $B = O$ and $C = \overline{D}$. We thus obtain

Corollary 1.19. *The equation*

$$[(C \text{ and } x) \text{ or } (\overline{C} \text{ and } \overline{x})] = \Omega$$

has x = C as its only solution.

I shall not dwell further on the application of the above rules of the symbolic calculus, cf. Schröder (1890 – 1895). It is more important for us to pass on to the proof of their independence and the lack of contradictions between them.

1.2. Consistency and Independence of the Axioms

1.2.1. A System of Numbers Corresponding to a Totality of Propositions

I do not aim at justifying arithmetic; on the contrary, for me, an integer and its main properties are simple notions lacking contradictions. For establishing the consistency of the proposed system of definitions and axioms it will therefore be sufficient to construct a system of numbers satisfying all the axioms; and, for proving independence, I shall construct systems of numbers obeying some axioms but violating the other ones.

To this end let us suppose that our symbols A, B, … {now} denote some integers and that the sign of equivalence (=) means equality; then, the join (A or B) is the greatest (from among the numbers considered) common divisor of the numbers A and B. It follows from the properties of the g.c.d. that the *associative* and *commutative* principles as well as the principle of *tautology* are obeyed here. Nothing prevents us from choosing our numbers in such a manner that the common divisor of any two of them is always included there; for example, we can choose 1, 2 and 3. Thus the *constructive* principle will also hold. On the contrary, it will be violated had we chosen the system 2, 3 and 4; for restoring the constructive principle it would have been necessary to adduce here the number 1. The existence of a true proposition, *i.e.*, of the common divisor of all the given numbers, follows, as we saw, from this principle [3]. But the existence of a false proposition imposes a new restriction on our system of numbers because a multiple of all the given numbers should correspond to it. Therefore, the system 1, 2, 3 lacks a number representing a false proposition, and for the axiom of its existence to hold we ought to adduce here either the number 6 or any of its multiples.

1.2.2. Independence of the Restrictive Principle

The least number from among the multiples of the numbers *A* and *B* belonging to a given system of numbers corresponds to the combination (*A* and *B*) of two propositions, *A* and *B.* Since the false proposition, that is a multiple of all the given numbers, exists, this combination always exists in a given system and satisfies, as it was established, the commutative and the associative principles. However, in order to prove the theorems of distributivity I introduced one more axiom, called the restrictive principle: *If* β *is a particular case of* (*A* or *B*), *it must be a join of some particular case of A with some particular case of B.* For our system of numbers the same principle states that *If* β *is a multiple of the g.c.d. of A and B, it is the g.c.d. of some two multiples of A and B.*

The system of numbers

$$p_1^{k_1} p_2^{k_2} ... p_n^{k_n} \qquad (14)$$

where p_i are some primes and k_i are *all* the non-negative integers not exceeding some given numbers c_i ,satisfies this condition. On the contrary, if we choose, for example, a system

$$1, p_1, p_2, \ldots, p_n, p_1 p_2 \ldots p_n, n \geq 3 \qquad (15)$$

obeying all the previous conditions excepting the last one, the restrictive principle will not hold because the g.c.d. of p_1 and p_2 is 1 but p_3 is not the g.c.d. of such numbers as x_1p_1 and x_2p_2 belonging to our system.

1.2.3. The Principle of Uniqueness and Perfect Totalities

It remains finally to consider the principle of uniqueness by whose means we established the concept of negation. This principle expresses the fact that 1 is the only number having, together with any {other} number, a least multiple differing from the general multiple of all the numbers. The necessary and sufficient condition for it to be realized along with all the previous principles consists in choosing, in §1.2.2, all $c_i = 1$. Indeed, the least multiple of the numbers $N = p_1^{\alpha_1} p_2^{\alpha_2} ... p_n^{\alpha_n}$ and $L = p_1^{k_1} p_2^{k_2} ... p_n^{k_n}$ will be a number of the same kind with exponents h_i equal to the greatest of α_i and k_i The principle of uniqueness means that all $\alpha_i = 0$ if

$$\Sigma(c_i - h_i) > 0 \text{ follows from } \Sigma(c_i - k_i) > 0.$$

The principle will therefore hold if all the $c_i = 1$, but will not be satisfied if $c_i > 1$ for at least one *i* [4].

It is necessary to note that *the restrictive principle is also independent of the principle of uniqueness.* This is shown by the example (15) where the latter obviously persists whereas the former, as we proved, is violated.

And so, *the axioms, which we assumed consecutively, are independent and do not contradict each other* because the system of propositions subordinated to them corresponds to a system of integers

$$1, p_1, p_2, \ldots, p_k, p_1 p_2, \ldots, p_1 p_2 p_3, \ldots, p_1 p_2 \ldots p_k$$

devoid of quadratic divisors and representing all the possible products of primes p_1, p_2, …, p_k.

We shall call a totality of propositions satisfying all our axioms perfect, and we shall hence consider only such totalities.

Note. My proof that the introduced axioms are independent, *i.e.*, that they cannot be consecutively derived from the other ones {from those previously formulated}, should hardly cause any objections in principle. On the contrary, the issue of the consistency of the axioms demands explanation. If, for example, we choose the system of numbers

1, 2, 3, 5, 6, 10, 15, 30

and express in words all the relations of divisibility between them, then, as it is possible to check directly, we obtain a series of statements not contradicting one another (we do not arrive at equalities of unequal numbers). Here, the meaning of the words "least multiple" and "greatest divisor", as well as of those general considerations from which our statements followed, are irrelevant for us. It is only important that we have here a definite system of objects whose interrelations satisfy all the axioms. For us, the use of numbers is therefore only a convenient and obvious method for realizing the system of symbols obeying all the axioms. For convincing ourselves in the existence of systems with an arbitrary great number of propositions we only need the concept of counting as a biunique correspondence between elements of two totalities and the principle of mathematical induction.

The independence of axioms 1.2 – 1.4 of §1.1.2 should also be indicated. Without dwelling on this issue which is of no consequence for the sequel, I restrict my remarks to the following. For a *finite* totality the principle of tautology 1.4 occupies a special place because it is necessary that any operation performed a finite number of times on any given symbol should return us to the same symbol. It is therefore always possible to replace the symbol *or* by $(or)^n$, – that is, by a repetition of this operation n times, – so that this principle will hold.

At the same time, this remark allows us to construct easily a system of numbers for which the principle (d) is not satisfied. Indeed, let us choose the numbers

1, 2, – 2, 3, – 3, 6.

Let the operation *or* for positive numbers preserve its previous meaning; otherwise, however, suppose that it leads to their greatest divisor taken with sign (+) if both numbers are negative and with sign (–) if their signs are different. Then, since the number – 1 is lacking in our totality, let us replace it by 1. This leads to the violation of the principle of tautology because (– 2 or – 2) = 2 but all the other principles persist without contradictions. A number of theorems will naturally be lost here, and, in particular, the uniqueness of the negation of any proposition will not follow from the principle of uniqueness.

1.3. The Structure and the Transformation of Finite Perfect Totalities of Propositions

1.3.1. Elementary Propositions

Any proposition of a totality differing from O and having no other particular cases excepting itself and O is called *elementary.*

Corollary 1.20. *Any proposition of a perfect totality has as its particular case at least one elementary proposition.* Indeed, if $A \neq O$ is not an elementary proposition, it has a particular case $B \neq O$ differing from A; if B is not an elementary proposition, it has a particular case C, etc. Since the number of the propositions is finite, we thus ought to arrive finally at an elementary proposition.

Corollary 1.21. *If a perfect totality has two different elementary propositions, they are incompatible.*

Theorem 1.7. *Any proposition excepting O is a join of elementary propositions.* Indeed, if α is an elementary proposition and a particular case of some proposition A, then A = (α or $A_α$) where $A_α$ is the supplement of α to A. If $A_α$ is an elementary proposition, the theorem holds for A; otherwise, $A_α$ has an elementary proposition β and A = (α or β or $A_{α β}$) where $A_{α β}$ is the supplement of β to $A_α$. When continuing in the same manner, we will finally arrive at the last elementary proposition of A so that A = (α or β or … or λ) where α, β, …, λ are elementary propositions.

Corollary 1.22. *If a totality has n elementary propositions, the general number of non-equivalent propositions is* 2^n. Indeed, if A contains at least one elementary proposition not included in B, then $A \neq B$. Consequently, the number of different propositions excluding O is equal to the sum of the appropriate binomial coefficients, or to $2^n - 1$. And, with O now added [5], the total number of propositions will be 2^n.

Theorem 1.8. *There exist perfect totalities with any number n of elementary propositions.* Indeed, if we have the impossible proposition *O and n* incompatible propositions a_1, a_2, …, a_n,, we may consider all their possible joins 2, 3, … at a time as propositions. All the axioms will then hold; in particular, the negation of each of the propositions will be the join of the other given propositions.

Note. A direct introduction of elementary propositions could have simplified the justification of the theory of *finite* perfect totalities; such a manner of exposition should be nevertheless rejected when bearing in mind *infinite* totalities (below).

1.3.2. Decomposition and Connection of Perfect Totalities

When isolating some k incompatible and solely possible propositions B_1, B_2, …, B_k and all their joins, whose number (including O and Ω) is 2^k, from a given perfect totality H, we form a new perfect totality G calling it a part of H. The propositions B_i will be the elementary propositions of G.

Let us choose some other series of incompatible and solely possible propositions C_1, C_2, …, C_l and compose out of them a new totality G'. *The totalities G and G' are called connected if there exists at least one pair of propositions B_i and C_j incompatible one with another:* (B_i and C_j) = O. And, if (B_i and C_j) ≠ O for any values of i and j, these totalities are *unconnected* or *separate.*

If no propositions are included in H, excepting those which are obtained when the propositions of G and G' are combined, this totality is called the *connection* of the two other totalities. In the same way H can be decomposed into 3, 4, etc parts and it will then be called their connection.

Note that a totality H can be decomposed into separate (unconnected) parts if and only if the number n of its elementary propositions is a composite rather than a prime. Indeed, if k elementary propositions B_i from totality G are always compatible with any of the l elementary propositions C_j from G', then (B_i and C_j) will constitute kl elementary propositions of the connection of G and G'.

If, for example, as when throwing a die, we have six elementary propositions A_1, A_2, …, A_6, we can form two separate parts: a totality G with elementary propositions (A_1 or A_2), (A_3 or A_4), (A_5 or A_6) and totality G' with elementary propositions (A_1 or A_3 or A_5) and (A_2 or A_4 or A_6). If we compile a totality G'' consisting of propositions (A_1 or A_2 or A_3) and (A_4 or A_5 or A_6) instead of G', then G and G'' will be connected with their connection being not H but only its part with elementary propositions (A_1 or A_2), A_3, A_4, (A_5 or A_6).

It is always possible to form a perfect totality of propositions H out of two such totalities G and G' with elementary propositions being all the combinations (B_i and C_j) of the elementary propositions B_i from G with the elementary propositions C_j from G'. Some of the propositions (B_i and C_j) can be assumed equivalent to O, then G and G' will be connected; it is only necessary that at least one of the combinations (B_i and C_j) including a definite proposition B_i will not be O, and that the same will also apply to one combination including a definite C_j because

$$[(B_i \text{ and } C_1) \text{ or } (B_i \text{ and } C_2) \text{ or } \dots \text{ or } (B_i \text{ and } C_l)] = B_i.$$

Suppose that, for example, G and G' are composed of three elementary propositions each, B_1, B_2 and B_3 and C_1, C_2 and C_3 respectively, and that among the combinations of these propositions (B_1 and C_1) = O and (B_2 and C_2) = O. Then the unification [*soedinenie*] H of G and G' will consist of the seven others (3·3 – 2) elementary propositions differing from O. Denoting them by

$$A_1 = (B_1 \text{ and } C_2),\ A_2 = (B_1 \text{ and } C_3),\ A_3 = (B_2 \text{ and } C_1),\ A_4 = (B_2 \text{ and } C_3),$$
$$A_5 = (B_3 \text{ and } C_1),\ A_6 = (B_3 \text{ and } C_2),\ A_7 = (B_3 \text{ and } C_3),$$

we see that

$$B_1 = (A_1 \text{ or } A_2),\ B_2 = (A_3 \text{ or } A_4),\ B_3 = (A_5 \text{ or } A_6 \text{ or } A_7)$$

are the elements of G and that

$$C_1 = (A_3 \text{ or } A_5),\ C_2 = (A_1 \text{ or } A_6),\ C_3 = (A_2 \text{ or } A_4 \text{ or } A_7)$$

constitute G.

1.3.3. The Transformation of Perfect Totalities. Realization of a Proposition

Theorem 1.9. *A given perfect totality can be transformed into another perfect totality of propositions by introducing the condition that a definite proposition A not equivalent to O is* Ω: $A = \Omega$. *Such a proposition is called the realization of proposition A* (*or the occurrence of event A*).

Indeed, if $A = \Omega$, $\overline{A}$ as well as all its particular cases are equivalent to O. Therefore, two propositions, B and C, having previously been mutually supplementary to A, become mutual negations and the obtained totality is therefore perfect. This transformation would have only been impossible if $A = O$ because then all propositions become equivalent to one and the same proposition, $O = \Omega$, which contradicts the assumption made in the very beginning. This transformation is obviously irreversible because a totality cannot be deprived of the true proposition.

Theorem 1.10. *Each transformation of a totality of propositions consisting of introducing the condition A = B is just a realization of some proposition C. It is possible if and only if A and B are not mutual negations.*

Indeed, for $A = B$ it is necessary and sufficient (§1.1.9, Corollary 1.19) that

$$C = [(A \text{ and } B) \text{ or } (\overline{A} \text{ and } \overline{B})] = \Omega, \text{ QED.}$$

Remark. If A and B are incompatible, then

$C = (\overline{A} \text{ and } \overline{B}) = \Omega$; *that is* $\overline{A} = \overline{B} = \Omega$ *and therefore* $A = B = O$.

It should be noted that there exists an essential difference between a unification of two totalities and the transformation called realization. A unification of (connected or unconnected) totalities does not introduce any changes into the substance of the *given* propositions. On the contrary, a realization of a proposition changes its substance; namely, it introduces a new condition of equivalence.

For connected totalities their connection, expressed by conditions of the type (B_i and C_k) = O, should not result in a change of the substance of any of the *given* propositions B_i; and for each k the condition above is impossible. However, the establishment of that condition can also be considered as a transformation of the totality obtained as a unification of unconnected totalities. Thus, when totalities are united, one or another connection between them leads to complicated and differing in their substance totalities having the same initial components.

Chapter 2. The Probabilities of the Propositions of Finite Totalities

2.1. Axioms and Main Theorems of the Theory of Probability

2.1.1. Axioms

As we saw, equivalent propositions can be expressed by one and the same symbol or numerical coefficient. We thus obtained a peculiar calculus of propositions that might find application in pure logic. The main new assumption of probability theory is the thesis that *one and the same numerical coefficient called mathematical probability can sometimes be also assigned to non-equivalent propositions.* This coefficient should not change when we connect another totality to the given one [6]. The probabilities of the propositions of a given totality can only vary when the totality is transformed (§1.3) by the realization of some proposition. We will express the statement that the probability of proposition A is equal to that of proposition B (Prob A = Prob B), or that A and B are *equally possible*, by a short formula $A \sim B$. Consequently, $A \sim B$ and $A \sim C$ lead to $B \sim C$.

If $A = B$ *then, all the more,* $A \sim B$. Therefore, in particular, *all true propositions have one and the same probability* (certainty) *and all the impossible propositions also have one and the same probability* (impossibility).

A totality of propositions where a definite mathematical probability is assigned to each of them is called *arithmetized.* If the numerical coefficient which is the mathematical probability of A is not equal to another numerical coefficient, – to the probability of B, – then one of them is larger than the other one and for the sake of brevity we will express this by the inequalities $A > B$ or $B > A$. The following axioms are the only rules which should be obeyed when arithmetizing a finite totality of propositions.

Axiom 2.1 (on the certain proposition). *If* $A \neq \Omega$ *then* $\Omega > A$.

Corollary 2.1. $\Omega > O$.

Axiom 2.2 (on incompatible propositions). a) *If* $A \sim A_1$, $B \sim B_1$, *and, in addition,* (A and B) = (A_1 and B_1) = O, *then* (A or B) $\sim$ (A_1 or B_1).

b) *If, however,* $A \sim A_1$, $B > B_1$, *then* (A or B) > (A_1 or B_1).

Corollary 2.2. *If* $A \neq O$, *then* $A > O$. Indeed, ($\bar{A}$ or A) = Ω, ($\bar{A}$ or O) = $\bar{A}$, but $\Omega > \bar{A}$, therefore $A > O$.

Corollary 2.3. *If A is a particular case of B and* ($\bar{A}$ and B) $\neq O$, *then* $B > A$. Indeed,

B = [A or ($\bar{A}$ and B)], A = (A or O),

and, since ($\bar{A}$ and B) > O, $B > A$.

2.1.2. Independence and Consistency of the Axioms

These axioms obviously do not follow from the previously established preliminary axioms because nothing could gave prevented us from assuming, for example, that, in spite of Axiom 2.1, all propositions are equally possible; or, on the contrary, from stating that, in spite of Axiom 2.2, only one pair of non-equivalent propositions are equally possible so that, given

$A \sim A_1$ and $B = B_1$, we will have (A or B) > or < (A_1 or B).

Let us show that Axiom 2.1 is not a corollary of both parts of Axiom 2.2 either. To this end we choose some totality formed by means of three elementary propositions *a, b* and *c*; suppose that their probabilities are 1, – 1 and – 2 respectively and assume finally that the impossible proposition has probability 0. We shall obtain quite definite values for the probability of each proposition of the totality if, adhering to Axiom 2.2, we assume in particular that Prob (A or B) = Prob A + Prob B when (A or B) = O. It will occur that (a or b) $\sim O$, $c \sim \Omega$, Prob Ω = Prob c = – 2, (a or c) $\sim b$, Prob (a or c) = Prob b = – 1, Prob (b or c) = – 3. It is obvious that neither can the first part of Axiom 2.2 be a corollary of Axiom 2.1 and of the second part of Axiom 2.2 because it is impossible to obtain an equality out of a finite number of inequalities.

Neither is the second part of Axiom 2.2 a corollary of its first part and of Axiom 2.1. Indeed, let us choose some perfect totality consisting of n elementary propositions $A_1, A_2, \ldots, A_n$ and agree to consider them equally possible. Then all their joins taken two at a time, and, in general, all their joins formed out of k elementary propositions, will also be equally possible. This conclusion follows only from Axiom 2.2a. Assuming that Axiom 2.1 is also valid, we ought to add that, if $k \neq l$, a join of k propositions cannot be equipossible with the join of l propositions. Any function $f(k)$ satisfying the condition{s} $f(k) \neq f(l)$ if the integers k and l are not equal and $f(n) > f(k)$ if $k = n - 1, \ldots, 1, 0$ can be the value of the probability of the join of k propositions. Without contradicting our assumptions we may presume, for example, that $f(1) < f(2) < \ldots < f(n - 1) < f(0) < f(n)$. But then Axiom 2.2b will not hold because on its strength we should have obtained

$$(A_1 \text{ or } A_2) < (A_1 \text{ or } O) = A_1$$

since $A_2 < O$ because $f(1) < f(0)$ whereas $f(2) > f(1)$, *i.e.* (A_1 or A_2) > A_1.

On the contrary, if we assume that $f(0) < f(1) < f(2) < \ldots < f(n)$, then both Axioms 2.1, 2.2a, and 2.2b will be fulfilled. We conclude that our new axioms are not only independent one from another, but also consistent with each other. They lead to the following main theorem of the theory of probability.

2.1.3. The Main Theorem

Theorem 2.1. *If propositions* A and B *are joins of some* m (m_1) *propositions chosen out of some* n (n_1) *incompatible, solely and equally possible propositions*, *then* $A \sim B$ *when* $m/n = m_1/n_1$.

Indeed, let $m\ n = m_1/n_1 = \mu/\nu$ where the last-written fraction is irreducible. Then

$$m = k\mu,\ n = k\nu,\ m_1 = k_1\mu,\ n_1 = k_1\nu$$

where k and k_1 are integers. Denote by

$$c_1, c_2, \ldots, c_m, \ldots, c_n$$

incompatible, solely and equally possible propositions the first m of which have A as their join. Supposing also that

$$d_1 = (c_1 \text{ or } c_2 \text{ or } \ldots \text{ or } c_k),\ d_2 = (c_{k+1} \text{ or } \ldots \text{ or } c_{2k}) \text{ etc,}$$

we compile ν incompatible solely and equally possible propositions (Axiom 2.2a) $d_1, d_2, \ldots, d_\nu$ the first μ of which have A as their join. In the same way, denoting by $\gamma_1, \gamma_2, \ldots, \gamma_n$ incompatible, solely and equally possible propositions, m_1 of which have B as their join, we compile ν incompatible, solely and equally possible propositions δ_1, δ_2, …, δ_ν, μ of which have B as their join. But it is obvious that $d_i \sim \delta_k$ because, when assuming for example that $d_1 > \delta_1$, we would have obtained that always $d_i < \delta_i$ and, in accord with Axiom 2.2b, $\Omega > \Omega$ which is impossible. But then, $d_1 \sim \delta_1$, $d_2 \sim \delta_2$ etc, hence

$$(d_1 \text{ or } d_2 \text{ or } \dots \text{ or } d_\mu) \sim (\delta_1 \text{ or } \delta_2 \text{ or } \dots \text{ or } \delta_\mu);$$

that is, $A \sim B$, QED.

2.1.4. Definition of Mathematical Probability

The coefficient, which we called the mathematical probability of A, is thus quite determined by the fraction m/n where n is the number of solely and equally possible incompatible propositions m of which have A as their join. Consequently, this coefficient is a function of m/n which we denote by $\varphi(m/n)$. On the basis of the above, φ should be *increasing* and this necessary condition is at the same time sufficient for satisfying all the assumed axioms if only the function $\varphi(m/n)$ can be fixed once and forever for all the totalities that might be added to the given one. Since such a function can be chosen arbitrarily, we assume its simplest form: $\varphi(m\ n) = m/n$ so that *m/n is called the mathematical probability of A.*

However, in accord with the main axioms we could have just as well chosen m^2/n^2, $m/(n - m)$, etc. The assumption of one or another verbal definition of probability would have obviously influenced the conclusions of probability theory just as little as a change of a unit of measure influences the inferences of geometry or mechanics. Only the form but not the substance of the theorems would have changed; we would have explicated the theory of probability in a new terminology rather than obtained a new theory. The agreement that I am introducing here is therefore of a purely technical nature [7] as contrasted with the case of the main axioms assumed above and characterizing the essence of the notion of probability: their violation would have, on the contrary, utterly changed the substance of probability theory.

Note. Together with Borel (1914, p. 58) we might have called the fraction $m/(n - m)$, – the ratio of the number of the favorable cases to the number of unfavorable cases; or, the expression $(m/n)/[(n - m)/n]$, – the ratio of the probability of a proposition to that of its negation, – the relative probability of the proposition.

Remark. When adding a new totality to the given one, we must, and we can always distribute, in accord with the axioms, the values of the probabilities of the newly introduced propositions in such a manner that the given propositions will still have the same probability {probabilities} as they had in the original totality. Indeed, suppose that the elementary propositions A_1, A_2, …, A_n in the given totality are equally possible; consequently, after choosing the function $\varphi(m/n)$ all the propositions of the totality acquire quite definite values. Add the second totality formed of elementary propositions B_1, B_2, …, B_k and let us agree to consider that, for example, all the combinations (A_i and B_j) in the united totality are also equally possible. If the function φ persists all the propositions of the united totality will obtain definite probabilities and any join of the type (A_1 or A_2 or … or A_m) previously having probability $\varphi(m/n)$ and regarded as a join

$$[(A_1 \text{ and } B_1) \text{ or } (A_1 \text{ and } B_2) \text{ or } \dots \text{ or } (A_m \text{ and } B_k)]$$

must now have probability $\varphi(km/kn)$ equal to its previous value, $\varphi(m/n)$. And all the propositions B_j will also be equally possible.

We may thus agree to consider any incompatible and solely possible propositions

$$A_1, A_2, \dots, A_k \quad (16)$$

included in a given totality as equally possible. After this, those and only those propositions which are joins of (16), or, otherwise, which are included into a totality G as its elementary propositions, obtain definite probabilities. Then another group of solely possible and incompatible propositions $B_1, B_2, \ldots, B_l$ can also be considered as equally possible if the totality G' composed of them is not connected with G, etc.

Indeed, no proposition α excepting Ω is a join of the elementary propositions of G and G' at the same time. If, however, α and β are two incompatible with each other propositions of G, and α' and β' are similar propositions of G', then, on the strength of the definition of probability, the agreement that $\alpha \sim \alpha'$, $\beta \sim \beta'$ will lead to (α or β) ~ (α' or β') and $\alpha \sim \alpha'$, and $\beta > \beta'$ will imply (α or β) > (α' or β') so that our axioms will not be violated.

As to the combinations and joins of compatible propositions, their probabilities are not quite determined and a new agreement is necessary (see below) for determining them. In any case, I have indicated above the possibility of such an agreement.

2.1.5. The Addition Theorem

The Axiom 2.2a can be formulated otherwise: If p is the probability of A, and p_1, the probability of B, the probability of (A or B) for A and B incompatible one with another is a function $f(p; p_1)$. The type of f depends on the choice of the function $\varphi(m/n)$. It is not difficult to derive a general connection between them, but after the above statements it is quite sufficient to restrict our attention to the case of $\varphi(m/n) = m/n$ which leads, as we shall see, to

$$f(p; p_1) = p + p_1. \tag{17}$$

Conversely, if we fix the function f, which on the strength of the axioms should necessarily be increasing, symmetric and satisfying the equation

$$f[p; f(p_1; p_2)] = f[p_1; f(p_1; p_2)],$$

we will obtain the appropriate φ; in particular, from (17) it is possible to derive $\varphi(m/n) = (m/n)H$ where H is an arbitrary positive number.

Theorem 2.2. *If two incompatible propositions A and B have probabilities p and* p_1 *respectively, then the proposition* (A or B) *has probability* ($p + p_1$).

This theorem is usually proved (Markov 1913, pp. 11 and 172) for incompatible joins of solely and equally possible incompatible propositions A and B, *i.e.*, when a direct application of the definition of probability makes it hardly necessary. In actual fact, the theorem is important exactly when it cannot be justified. For the sake of completeness of the proof we only need to cite Axiom 2.2 once more: to refer to its first part if both numbers p and p_1 are rational, and to its second part if they are irrational [8].

Indeed, let us assume at first that p and p_1 are rational so that $p = m/n$ and $p_1 = m_1/n_1$. When adding some totality to the initial one, unconnected with it and containing nn_1 equally possible elementary propositions, proposition A', which is a join of some mn_1 of these, will have the same probability $mn_1/nn_1 = m/n = p$ as A whereas B', which is a join of some other [9] m_1n elementary propositions, will have the same probability $m_1n/n_1n = m_1/n_1 = p_1$ as B. In this case, (A' or B') will be a join of ($m_1n + n_1m$) out of nn_1 elementary propositions and therefore, in accord with the definition of probability,

$$(m_1n + n_1m)/nn_1 = m_1/n_1 + m/n = p_1 + p$$

will be the probability of (A' or B'), and, on the strength of Axiom 2.2a, it will also be the probability of (A or B), QED.

Suppose now that p and p_1 (or only one of them) are (is) irrational. Then p is the limit of rational numbers $\lambda_1 < \lambda_2 < \ldots < \lambda_n < \ldots$ and $\mu_1 > \mu_2 > \ldots > \mu_n > \ldots$, and p_1, the limit of rational numbers $\pi_1 < \pi_2 < \ldots < \pi_n < \ldots$ and $\rho_1 > \rho_2 > \ldots > \rho_n > \ldots$

Denote some proposition having probability λ_n by A_n, and, by B_n, an incompatible proposition [10] having probability π_n. Then, because of Axiom 2.2b, we have (A_n or B_n) < (A or B), that is, $\lambda_n = \pi_n <$ Prob (A or B). Just as above, we denote the propositions having probabilities μ_n and ρ_n by A'_n and B'_n respectively. In accord with the same axiom we will obtain Prob (A or B) $< \mu_n + \rho_n$, and, on the strength of the generally known theorem on limits, Prob (A or B) $= p + p_1$, QED.

2.1.6. A Corollary

Corollary 2.4. It follows that *a necessary and sufficient condition for numbers* $p_1, p_2, \ldots$, *to be, respectively, probabilities of propositions* $A_1, A_2, \ldots$ *of a given finite totality is that the probability of a join of two or several incompatible propositions is equal to the sum of their probabilities*; *that the certain proposition has probability* 1 (and, consequently, that the impossible proposition has probability 0); *and that all the other propositions have probabilities contained between* 0 *and* 1 ($0 < p < 1$).

In particular, it follows that if two totalities, G and G', are unconnected, the probabilities assigned to the propositions of G are not logically connected with the probabilities of the propositions of G'. In other words, the *arithmetization* of one totality does not depend on the arithmetization of the other one. On the contrary, if G and G' are connected, we cannot assign quite arbitrary probabilities to the propositions of G' after establishing the probabilities of the propositions of G.

The probabilities of some propositions of a totality are not always given, but it is then only necessary to leave room for choosing the yet indeterminate probabilities in such a manner that the abovementioned main condition be held [11].

2.2. Combination and Realization of Propositions

2.2.1. Combination of Propositions

The probability of (A and B) obviously cannot at all be a definite function of the probabilities of A and B. Suffice it to note that, if A and B are incompatible, then (A and B) ~ O; on the contrary, if $A = B$, then (A and B) ~ A. The only general statement that we may formulate here is that Prob (A and B) + Prob (A and $\overline{B}$) = Prob A and that, specifically, Prob (A and B) $\leq$ Prob A.

We may always assume that Prob (A and B) $= \lambda p p_1$ where p and p_1 represent the probabilities of A and B respectively and λ is called the *coefficient of compatibility* of A and B In particular, $\lambda = 0$ if these propositions are incompatible.

Suppose that we have an arithmetized totality composed of elementary propositions A_1, $A_2, \ldots, A_n$ whose probabilities $p_1, p_2, \ldots, p_n$ satisfy the condition $p_1 + p_2 + \ldots + p_n = 1$. Adding to it the totality 0, C, $\overline{C}$, and Ω where Prob $C = p$, Prob $\overline{C} = q$, $p + q = 1$, we arithmetize the united totality by assuming that Prob (A_1 and C) $= \lambda_1 p_1 p$, Prob (A_2 and C) $= \lambda_2 p_2 p, \ldots$, Prob ($A_n$ and C) $= \lambda_n p_n p$ where $0 \leq \lambda_i \leq 1/p$ and $\Sigma\lambda_i p_i = 1$. Then

$$\text{Prob } (A_i \text{ and } \overline{C}) = p_i - \lambda_i p_i p = p_i(1 - \lambda_i p)$$

so that, denoting the coefficient of compatibility of A_i and $\overline{C}$ by μ_i, we have

$$\lambda_i p + \mu_i q = 1.$$

For unconnected totalities we would have gotten λ_i, $\mu_i > 0$. The case of independent totalities deserves special attention.

2.2.2. Independent Propositions

Proposition A is called *independent* of B if the coefficient of compatibility of A and B is equal to the same coefficient of A and $\overline{B}$.

Theorem 2.3. *If proposition A is independent of B, then proposition B is independent of A and the coefficient of compatibility of A and B is unity.*

Indeed, if

$$\text{Prob } (A \text{ and } B) = \lambda p_1 p, \text{ Prob } (A \text{ and } \overline{B}) = \lambda p_1 q, \; p + q = 1,$$

then

$$\text{Prob } (A \text{ and } B) + \text{Prob } (A \text{ and } \overline{B}) = \lambda p_1 = p_1$$

so that $\lambda = 1$. But then

$$\text{Prob } (\overline{A} \text{ and } B) = p - p_1 p = q_1 p$$

which means that B is independent of A.

Corollary 2.5. *If propositions A and* A_1 *are incompatible one with another, and both independent of B, then* (A or A_1) *is also independent of B.* In general, if the coefficients of compatibility of A and B, and A_1 and B are both equal to λ, then the same coefficient of (A or A_1) and B is also λ. If all the elementary propositions of totality H are independent of those of totality H_1, then each proposition of H is always independent of each of the propositions of H_1. Such two totalities are called independent of one another. Obviously, only unconnected totalities can be, but nevertheless not always are independent.

Without dwelling on the further development of these considerations, we shall only briefly indicate how the independence of n propositions $A_1, A_2, \ldots, A_n$ having probabilities $p_1, p_2, \ldots, p_n$ respectively is defined. These are called *pairwise* independent if the combination (A_i and A_k) has probability $p_i p_k$; they are called independent by threes if each combination (A_i and A_k and A_l) has probability $p_i p_k p_l$ etc. If the given propositions are independent pairwise, by threes, …, and all together, they are called (absolutely) independent. Consequently, the number of conditions needed for absolute independence of n propositions is equal to the sum of the appropriate binomial coefficients, *i.e.*, to $2^n - n - 1$, and it can be shown that none of them is a corollary of the other ones [12]. For example, if three propositions are pairwise independent it does not follow that they are absolutely independent.

2.2.3. Realization of Propositions

The above shows that the calculation of the probabilities of combinations of propositions does not demand any new assumptions. All the most important chapters of probability theory (the Bernoulli theorem and all of its generalizations known as the law of large numbers) therefore follow exclusively from the axioms introduced by us. However, it is often more convenient for practice to apply another notion rather than the *coefficient of compatibility*, – namely, the probability of a proposition under the condition that another one had happened. It is essentially necessary only for constructing the chapter on the probability of hypotheses. Accordingly, we introduce a new assumption that supplements the above (§1.3.3) definition of transforming a totality called there *the realization of a proposition.*

Axiom 2.3 of realization. *When proposition A of a given totality H is realized, any proposition* α, *having previously been a particular case of A, acquires in the transformed totality a probability that depends only on the probabilities of A and* α *in the given totality H.*

Thus, by definition, the probability α_A of the proposition α after *A* was realized is

$$\alpha_A = f(\text{Prob}\ \alpha; \text{Prob}\ A). \qquad (18)$$

The definition of the realization of proposition *A,* as provided in §1.3.3, determined only the logical structure of the transformed totality; and, since this does not completely determine the probabilities of propositions, our new assumption [13] cannot be a corollary of the previous ones.

Let us now show that, if only the function *f* in formula (18) is [14]

$$\alpha_A = \text{Prob}\ \alpha/\text{Prob}\ A, \qquad (19)$$

the axiom of realization does not contradict those assumed previously. Indeed, since the structure of the propositions remaining in the transformed totality is the same as it was in the initial one,

$$f(\text{Prob}\ \alpha + \text{Prob}\ \beta; \text{Prob}\ A) = f(\text{Prob}\ \alpha; \text{Prob}\ A) + f(\text{Prob}\ \beta; \text{Prob}\ A).$$

However, as is well known [15], for this relation to hold it is necessary that

$$f(\text{Prob}\ \alpha; \text{Prob}\ A) = (\text{Prob}\ \alpha)\ F(\text{Prob}\ A).$$

On the other hand, $A_A = 1$ in accord with the condition; therefore,

$$(\text{Prob}\ A)F(\text{Prob}\ A) = 1$$

and (19) holds. At the same time we see that the function α_A determined by us implies that all the propositions of the transformed totality acquire probabilities satisfying the condition of arithmetization (§2.1.6) if only the probabilities of the propositions in the initial totality obey the same condition. We see therefore that this function does not contradict the main axioms.

2.2.4. The Multiplication Theorem

Theorem 2.4. *The probability of* (*A* and *B*) *is equal to the probability of A multiplied by the probability of B given that A has occurred.*

Indeed, after *A* is realized, the proposition (*A* and *B*) is equivalent to (Ω and *B*) = *B* and

$$\text{Prob}\ (A \text{ and } B) = (A \text{ and } B)_A = B_A = \text{Prob}\ (A \text{ and } B)/\text{Prob}\ A.$$

Consequently, Prob (*A* and *B*) = (Prob *A*) B_A, QED.

Note. The multiplication theorem implies, specifically, that *if* α *is a particular case of A, its probability only depends on the probability of A and on the probability of* α *after A has occurred.* Or, in an equivalent form: *If* α *and β are particular cases of A and B respectively, then their probabilities are equal to each other if A and B are equally probable and the probability of* α *after A has occurred is equal to the probability of* β *after B has occurred.*

This statement can replace the axiom of realization because the multiplication theorem can be derived therefrom similarly to the above. The introduction of the notion of probability of a proposition given that a second one has occurred allows us to offer another definition of independence.

If the probability of B after A has occurred is equal to its initial probability, then B is independent of A.

Corollary 2.6. *If B is independent of A, then A is independent of B and*
Prob (*A* and *B*) = Prob *A*·Prob *B.*

2.2.5. The Bayes Theorem

Theorem 2.5. *The probability of A after B has occurred is*

$$A_B = (\text{Prob } A)\cdot B_A/\text{Prob } B.$$

Indeed,

$$A_B = \text{Prob } (A \text{ and } B)/\text{Prob } B = (\text{Prob } A)\ B_A/\text{Prob } B.$$

The axiom of realization (§2.2.3) is the sole basis of the Bayes theorem and its corollaries whose derivation presents no difficulties in principle.

Chapter 3. Infinite Totalities of Propositions

3.1. Extension of the Preliminary Axioms onto Infinite Totalities

3.1.1. Perfect Totalities

The main condition that we ought to introduce here is that the rules of the symbolic calculus established for finite totalities (§1.1) must persist should we consider the propositions of a finite totality as belonging to some infinite totality. *A totality* (*finite or otherwise*) *of propositions to which all these rules are applicable is called perfect.* However, some of the premises, which were previously corollaries of the other ones, become here, for infinite totalities, new independent axioms.

Indeed, *the existence of a true proposition,* which was the corollary of axioms (a) – (d) of §1.1.2, is such a new assumption. Let us consider the totality of proper fractions $p/2^n$ written down in the binary number system (0.101; 0.011; etc). We will understand the operation *or* as the compilation of a new fraction out of two given ones in such a manner that each of its digits equals the largest from among the two appropriate digits of the given fractions; thus, (0.101 or 0.011) = 0.111. This totality will not contain a number corresponding to the true proposition that should have been represented by the infinite fraction 0.111… = 1. We can, however, add 1 to our fractions so as to realize the axiom of the true proposition. And, by joining 0, we will also obtain the impossible proposition.

The second necessary new assumption (a corollary of the previous premises for finite totalities) is that *the combination of two propositions* (*A* and *B*) *does exist.* Finally, the third and last additional assumption is *the extension of the restrictive principle onto infinite joins of propositions.*

In the example concerning the binary fractions (with 0 and 1 being included) the combination of two propositions existed and was represented by a fraction having as each of its digits the least one from among the two appropriate digits of the given fractions. At the same time, *the restrictive principle* is also applicable here to any pair of propositions so that all the properties of combinations including the theorems of distributivity persist. In addition, *the principle of uniqueness* also holds in our example, but the totality considered will not be perfect. Indeed, in accord with the definition above (§1.1.8), a negation of any proposition *A* is the join of all propositions incompatible with it. But there will be an infinite set of such propositions and we ought to begin by generalizing the definition of *join* as given in §1.1.1.

A join H = (*A* or *B* or *C* or …) *of an infinite set of propositions A, B, C, … is a proposition H satisfying the conditions* [16]

1) *If y is a particular case of proposition A, or B, or C, …, then it is a particular case of H*;

2) *If each of the propositions A, B, C, … is a particular case of M, then H is also its particular case.*

In accord with this definition, the commutative and the associative principles are extended onto infinite joins as also is the principle of tautology. But whether *the restrictive principle* can be extended remains an open question. In the example above, this extension is not realized. Indeed, a join of any infinite totality of various propositions [in the example considered. – Editor of Russian reprint] (for example, the join of 0.01; 0.001; 0.0001; etc) is a true proposition whereas the proposition 0.1 is not a join of the particular cases of these fractions because none of the latter has any other such cases excepting *O* and itself. Owing to the violation of the generalized restrictive principle, Corollary 1.17 does not hold: *If* $\bar{x} = \Omega$ *then* $x = O$ because an infinite join of propositions incompatible with *x* can be compatible with it.

And so, for an infinite totality to become perfect, the last assumption, *the generalization of the restrictive principle*, should be added.

The join α = (*A* or *B* or *C* or …) *of an infinite totality of propositions does not contain any particular cases other than the joins of the particular cases of A, B, C, …*

3.1.2. Combination of Propositions and the Negation

Corollary 1.18 (§1.1.8) establishes the connection between combination and negation of propositions and makes it possible to determine a combination on the strength of the formula

$$(A \text{ and } B) = \text{Neg}\,(\bar{A} \text{ or } \bar{B}). \qquad (20)$$

The additional assumption that a combination of two propositions exists can therefore be replaced by the following statement: *If A is a proposition of a totality there also exists a join of all propositions incompatible* [17] *with A which is indeed called* $\bar{A}$ *or the negation of A.* On the strength of the generalized restrictive principle, *A* and $\bar{A}$ are incompatible; and, because of the principle of uniqueness, (A or $\bar{A}$) = Ω. In addition, *A* = *A*. Indeed, if (α or *A*) = *A* then (α and $\bar{A}$) = *O* so that (α or Neg $\bar{A}$) = Neg $\bar{A}$, hence (*A* or Neg $\bar{A}$) = Neg $\bar{A}$. On the other hand, if (α or *A*) = *A* then (α and $\bar{A}$) = *O* but α is a particular case of Ω = (*A* or $\bar{A}$), and on the strength of the restrictive principle it is a join of particular cases of *A* and $\bar{A}$ and is therefore a particular case of *A*.

I shall now prove that the proposition

$$z = (\text{Neg}[\bar{A} \text{ or } \bar{B}]) \qquad (21)$$

corresponds to the definition of §1.1.5. To this end I begin by noting that ($\bar{C}$ or *D*) = C follows from (*C* and *D*) = *D* because a proposition incompatible with *D*, *i.e.*, belonging to $\bar{D}$, is also incompatible with *C* and therefore also belongs to $\bar{C}$. And so, we have to show, first, that *z* is a particular case of *A* and of *B,* and, second, that each compatible particular case of *A* and *B* is a particular case of *z*. Indeed, $\bar{A}$ is a particular case of ($\bar{A}$ or $\bar{B}$), therefore (21) is, on the strength of what was just proved, a particular case of *A*; by the same reasoning, *z* is a particular case of *B.*

On the other hand, let x be a particular case of A and B, then $\bar{A}$ and $\bar{B}$ are particular cases of $\bar{x}$ so that ($\bar{A}$ or $\bar{B}$) is a particular case of $\bar{x}$. Consequently, x is a particular case of (21), QED.

Issuing from formula (20) we also derive the definition of a combination of an infinite set of propositions

(A and B and C and …) = (Neg[$\bar{A}$ or $\bar{B}$ or $\bar{C}$ or …]). (22)

Associativity and commutativity thus extend onto infinite combinations.

Let us prove that the theorems of distributivity can also be thus extended. At first, we shall show that

[h and (A or B or …)] = [(h and A) or (h and B) or …]. (23)

Indeed,

[A or B or C or …] = [(h and A) or ($\bar{h}$ and A) or (h and B) or …] =
{[(h and A) or (h and B) or …] or [($\bar{h}$ and A) or ($\bar{h}$ and B) or …]},

therefore

[h and (A or B or C or …)] = {h and [(h and A) or (h and B) or …]} or
{h and [($\bar{h}$ and A) or …]} = [(h and A) or (h and B) or …], QED.

Note. It was assumed here that both parts of equality (23) make sense. However, we can convince ourselves that, if (A or B or …) exists, the second part of (23) also exists. Indeed,

z = [h and (A or B or …)]

has as a particular case each of the combinations (h and A), (h and B), etc, therefore

[(h and A) or (h and B) or …]

has meaning and will be equal to z once we show that any proposition M differing from z and having as a particular case{as particular cases} (h and A), (h and B), etc, includes z. Suppose that z is not a particular case of M, then it would have included z_1 = (z and M) with the same particular cases (h and A) etc. so that (z and $\bar{z}_1$) $\neq O$ would have been incompatible with any of the propositions (h and A), (h and B) etc. whereas each particular case of z should be compatible with h and at least with one of the propositions A, B, … etc which contradicts the supposition made.

Assuming that $\bar{h} = h_1$, $\bar{A}_1 = A_1$, $\bar{B} = B_1$ etc. and considering the negations of both parts of (23), we obtain the second theorem of distributivity

[h_1 or (A_1 and B_1 and …)] = [(h_1 or A_1) and (h_1 or B_1) and …]. (23 bis)

3.1.3. The Generalized Constructive Principle

Our assumptions do not at all imply the existence of an infinite join of any chosen propositions. That such a join of propositions of a given totality exists, *i.e.*, that the generalized constructive principle holds, is not necessary for a perfect totality [18]. Once we assume the general principle, the existence of the true proposition and of a combination will in particular follow.

We can provide an example of a perfect totality, in which the generalized constructive principle is satisfied, by supplementing the just considered system of finite binary fractions by the totality of all infinite fractions, provided that the fractions having 1 recurring will not be considered as an equivalent of those finite fractions, to which they must be equal as being the limits of sums of infinite geometric progressions. So as to avoid this contradiction with the generally accepted arithmetical assumptions, it is sufficient to consider our fractions as being written out not in the binary, but in some other, for example decimal system.

We obtain an example of a perfect totality with a violated generalized constructive principle by considering in addition to finite fractions only those infinite fractions whose period is 1. Indeed, as in the previous case, all the necessary conditions for a perfect totality are here fulfilled whereas some joins, as for example the join of all the fractions only having 1 on their even places, do not make sense; any fraction of the type of 0.01111 …; 0.010111 …; 0.0101011 … should have had this join as its particular case whereas the infinite fraction 0.01 (01) … is not included in our totality.

Remark. When considering, in addition to 0.111 …, only finite fractions, we could have stated that this infinite fraction corresponding to the true proposition is a join of any infinite set of propositions. However, to avoid misunderstandings which can occur since we should always have only to do with perfect totalities, we will include the assumption of the restrictive principle in the very notion of join, so that a join not satisfying this principle should be considered senseless. It is necessary to remember, in particular, that the generalized constructive principle postulates the existence of exactly those joins which are characteristic of perfect totalities, – that is, it implies the generalized restrictive principle. Formula (22) shows that the existence of a combination of any infinite set of propositions follows from the generalized constructive principle.

3.1.4. Classification of Infinite Perfect Totalities

In §1.3 I have shown that all finite perfect totalities have one and the same structure: they are formed by means of elementary propositions. On the contrary, the existence of such propositions is not at all necessary for infinite perfect totalities and is therefore never a necessary condition for the applicability of all the rules of logical calculus established above. Consequently, we may separate perfect totalities into four heads.

1. *The first type* for which
   a) Not each proposition is a join of elementary propositions.
   b) The generalized constructive principle does not hold.
2. *The second type* for which
   a) Not each proposition is a join of elementary propositions.
   b) The generalized constructive principle holds.
3. *The third type* for which
   a) Each proposition is a join of elementary propositions.
   b) The generalized constructive principle does not hold.
4. *The fourth type* for which
   a) Each proposition is a join of elementary propositions.
   b) The generalized constructive principle holds.

The examples of perfect totalities considered above belonged to the third and the fourth types for which the principle of the existence of elementary propositions was valid: the fractions containing only one unity corresponded to elementary propositions. Let us now provide examples of totalities of the first two types.

Consider the totality of all the *pure* periodic fractions formed of zeros and unities [19]. Assigning the same meaning to the operation *or* as before, we convince ourselves that the totality is *perfect,* but that the join of an infinite set of differing fractions will either be true (0.111 …) or senseless. The generalized constructive principle is thus violated, and, in addition, no fraction here considered is an elementary proposition; taking a double period and replacing one unity by a zero, we obtain a particular case of the chosen fraction. Consequently, the compiled totality belongs to the first type.

For constructing a totality of the second type, we return to the totality of all the fractions which corresponds to the fourth type, but instead of each fraction x we choose a function $f(x)$ determined by the condition that $f(x) = 0$ if x only has a finite number of unities, and is thus represented by a finite fraction; that $f(x) = 0.111 \ldots$ if x only has a finite number of zeros; and, finally, that $f(x) = x$ for other values of x. Let $f(x)$ represent each proposition and [$f(x)$ or $f(x_1)$] = $f(y)$ where y has in each place the largest digit out of the corresponding digits of $f(x)$ and $f(x_1)$. Then our totality will be perfect and the equality

$$[f(x) \text{ or } f(x_1) \text{ or } \ldots \text{ or } f(x_n) \text{ or } \ldots] = f(y)$$

will always have sense [20], *i.e.*, the generalized constructive principle holds. There will be no elementary propositions here because a fraction having an infinite set of unities can always be decomposed into two similar fractions. We have therefore constructed a totality of the second type.

Note. A unification of totalities of a certain type is a totality of the same type. On the contrary, when realizing some propositions, the type of the totality can change (below).

3.1.5. Totalities of the Second and the Fourth Types. The Cantor Theorem

When decomposing some perfect totality into simple finite totalities $O, A, \bar{A}, \Omega$; $O, B, \bar{B}, \Omega$; etc in all possible ways, the combinations (A and B …) will always make sense for totalities of the second and fourth types. Consequently, they will represent either an impossible or an elementary proposition. Let α, β, γ, etc be all the elementary propositions, and A, some non-elementary proposition. Then, if $\alpha_1, \beta_1, \gamma_1, \ldots$ are all the elementary propositions included in A, and $A_1 = (\alpha_1 \text{ or } \beta_1 \text{ or } \gamma_1 \text{ or } \ldots)$, then $A_2 = (A \text{ and } \bar{A}_1)$ will have no elementary propositions.

Assume that all propositions A_2 are equivalent to O, then our totality will belong to the fourth type. If however all the $A_1 = O$, the totality will be devoid of elementary propositions and will be called a simple totality of the second type. Thus, any proposition of a most general totality of the second type is a join of a proposition of a simple totality of that type and a proposition of a totality of the fourth type.

Denoting the joins of all A_1 and all A_2 by Ω_1 and Ω_2 respectively, we note that $\bar{\Omega}_1 = \Omega_2$. We conclude that, when realizing Ω_2, that is, *when assuming that $\Omega_2 = \Omega$, we transform each totality of the second type into a simple totality of the same type*; *on the contrary, when assuming that $\Omega_1 = \Omega$, we transform our totality into a totality of the fourth type.*

Theorem 2.6. The following Cantor theorem is applicable to the totalities of the second and the fourth types: *The cardinal number of a totality of a second or fourth type is larger than that of the totality of its elementary propositions.*

This follows from the fact that, when compiling all possible joins of elementary propositions, we obtain distinct propositions if only they differ in at least one elementary proposition.

Corollary 3.1. *Totalities of the fourth type are finite or have cardinal number not less than that of continuum.*

Corollary 3.2. *A countable totality of the second type either has no elementary propositions at all or has a finite number of them.*

Now, after proving that perfect totalities of all four types exist, it remains to show that their arithmetization in accord with the principles put forward in Chapt. 2 is possible. Thus, specifically, *when establishing the probabilities of the propositions of an infinite totality, we cannot assign the value* 1 *to the probability of a non-certain proposition A* because, when taking a finite part of our totality ($O, A, \overline{A}, \Omega$), we would have encountered a contradiction. I am emphasizing this obvious point, because, owing to the insufficiently clear statement of the principles of probability theory, many mathematicians apparently reconcile themselves to that contradiction.

3.2. Arithmetization of Infinite Totalities

3.2.1. Arithmetization of Totalities of the First Type

The most important and distinctive specimen of a perfect totality of the first type, to which we may restrict our attention, can be obtained thus. Let us consider a countable totality of finite totalities ($O, A, \overline{A}, \Omega$), ($O, A_1, \overline{A}_1, \Omega$), ($O, A_2, \overline{A}_2, \Omega$) etc unconnected one with another and add them consecutively together. The totality *H* of the propositions here considered is composed of the propositions included in some of the finite totalities thus obtained.

The totality *H* is countable, perfect, and belongs to the first type. In addition to finite joins, it has only those infinite joins (α_1 or α_2 or …) which possess a certain property: only a restricted number of their elements is not a particular case of the previous ones; that is, the totality has only those infinite joins which are directly reducible to finite joins.

An unrestrictedly repeated experiment with throwing a coin provides a concrete example of totality *H.* Any proposition concerning a finite number of throws has a definite sense, but those not included into any finite totality [21] are meaningless.

It is now clear that the probabilities of all the propositions of the totality *H* are defined consecutively on the strength of the agreements and theorems established for finite totalities without encountering contradictions or making any new assumptions. Probabilities of senseless propositions should not even be mentioned; instead, it might often be interesting to calculate the limits of probabilities of some variable and sensible propositions as the number of trials increases unboundedly.

Thus, the probability that heads will occur not less than 10 times in *k* throws tends to its limit, unity, as *k* increases unboundedly. This, however, does not at all mean that we *must* assign a sense to the proposition that this will indeed occur when the number of trials is infinite: it is possible that we will then be unable to establish whether that proposition was realized or not. In the same way, provided that the probability of heads is 1/2, the probability that the ratio of heads to tails will differ from unity as little as desired tends to 1 after a sufficiently large number of throws.

Keeping to the same viewpoint that may be called *finitary* [22], we can also justify without introducing any special assumptions the so-called *geometric probabilities*. We note at once that, from the logical side, the finitary point of view, which only considers totalities of the first type, is quite admissible. However, it becomes somewhat artificial when applied to geometry because {then} the isolation of a special category of propositions making sense is conjecturable and not sufficiently substantiated by intuitive geometric reasoning.

3.2.2. Geometric Probabilities

In geometry, the main problem concerning the calculation of probabilities, to which all the other problems are reducible, consists in determining the probability that a point *M* situated on segment *AB* lies on some of its part *PQ.* To solve this problem, we may act in the following way. Supposing, for the sake of simplifying the writing, that $AB = 1$, and that point *A* coincides with the origin *O,* we choose the abovementioned pattern of a totality of the first type and agree to form a binary fraction with a unity on its first, second, … place corresponding to propositions $A, A_1, \ldots$, and with a zero on places corresponding to $\overline{A}, \overline{A}_1$, …

In this case, all the finite or infinite binary fractions with definite digits being in one or several places will correspond to each proposition of our perfect totality H of the first type. For example, all the fractions beginning with 0.01, *i.e.*, all the numbers x satisfying the condition $0.01 < x < 0.1$, correspond to the proposition ($\overline{A}$ and A_1). The symbols < and > can be replaced by ≤ and ≥ respectively since the propositions $x = a$, where a is a given finite or infinite fraction, ought to be considered as senseless because they correspond to a combination of an infinite set of propositions.

It follows that if P and Q are two points of the segment (0; 1), whose abscissas are expressed by finite binary fractions a and b (a < b), the probability of the inequalities

$$a < x < b, \quad (24)$$

i.e., of x belonging to segment PQ, is obtained by a direct application of the addition and multiplication theorems for the probabilities of propositions A, A_1, A_2, … It is not difficult to see that, for absolutely arbitrary probabilities chosen (they should only correspond to the main assumptions), we come to the expression $F(b) – F(a)$ for the probability of (24). Here, $F(x)$ is an arbitrary non-decreasing function only defined for finite binary values of x and such that

$$F(b) – F(0) \le 1, \text{ and, for } b = 1,\ F(1) – F(0) = 1.$$

In particular, if the propositions A, A_1, … are independent one of another and the probability of each of them is equal to 1/2, we arrive at the usually adopted type of the function, $F(x) = x$.

As stated above, from the finitary point of view the proposition $x = a$ is senseless. Nevertheless, we may consider the *limits* of the probability of the inequalities

$$a < x < a + h \text{ or } a – h < x < a \text{ as } h \to 0.$$

These are, as $h \to 0$,

$$\lim [F(a + h) – F(a)],\ \lim [F(a) – F(a – h)]$$

and they are known to exist; specifically, for a continuous function they are equal to zero.

Similarly, when keeping to the finitary viewpoint, we have no right to mention the proposition $\alpha < x < \beta$ if α and β are not finite binary fractions. Instead, we must consider the limit of the probability of the inequalities

$$a_n < x < b_n \quad (25)$$

where a_n and b_n are finite fractions having α and β respectively as their limits.

If $F(x)$ is a continuous function, the limit of such probabilities will not depend on whether α or β are larger or smaller than a_n and b_n respectively; it will be equal to
$F(\beta) – F(\alpha)$. And in the general case we may, in accord with the usual notation, suppose that

$$F(\beta + 0) = \lim F(b_n),\ b_n > \beta, \text{ and } F(\beta – 0) = \lim F(b_n),\ b_n < \beta.$$

The inequalities (25) thus have probabilities whose limit is $F(\beta \pm 0) – F(\alpha \pm 0)$
depending on whether a_n and b_n tend to their limits, α and β, from the right or from the left.

By applying the well-known theorems of the theory of limits, it becomes possible, in most cases, to deal with the limits of probabilities in the same way as with probabilities themselves. For example, if segments $(\alpha;\ \beta)$ and $(\alpha_1;\ \beta_1)$ have no common part, the probability that either of the inequalities

$$\alpha < x < \beta,\ \alpha_1 < x < \beta_1$$

is valid, provided that they both make sense, is equal to the sum of the probabilities of each of them. If, however, they both, or one of them, ought to be considered senseless, it will be necessary to say, that the limit of the probability that one of the inequalities

$$a_n < x < b_n,\ a_n' < x < b_n'$$

holds, is equal to the sum of the limits of the probabilities of each of them. Here, a_n, b_n, a_n', b_n' have α, β, α_1, β_1 as their limits respectively.

It is now sufficiently clear that if the limit of the probabilities of some of the propositions is zero, it does not mean that the limiting proposition also exists, *i.e.*, that it makes sense under the given formulation; and, if it does make sense, it is impossible. A similar statement concerns the limit of probabilities equal to 1.

Note. Instead of binary fractions it would have been possible to consider, absolutely in the same way, decimal or other fractions. In each case a definite sense is only assigned to propositions establishing that definite digits occupy definite places. Inequalities making sense in one system therefore occur senseless in another and vice versa.

3.2.3. Arithmetization of Totalities of the Second Type

When studying the same countable totality of propositions *H,* that we obtained above by adding together the totalities $(O, A, \overline{A}, \Omega)$, $(O, A_1, \overline{A}_1, \Omega)$, etc, we may agree to regard any infinite combination (A and A_1 and …) as impossible; or, which is the same, to consider any infinite join [23] of the type (A or A_1 or A_2 …) as true. We will thus form *a simple perfect totality of the second type* (devoid of elementary propositions but obeying the generalized constructive principle).

Let us study geometric probabilities from this new viewpoint which is logically as admissible as the finitary approach. By means of the same system of binary fractions we conclude that each definite equality $x = a$ should be considered impossible, *i.e.* as such that could never be precisely realized [24] or established; the signs $\leq$ and $<$ are thus equivalent.

It should be noted that the statement that x is some number contained between 0 and 1 does not at all mean that it can be determined absolutely precisely, and this explains the apparently paradoxical statement that a true proposition is allegedly a join of an uncountable set of impossible propositions. However, we ought to study in more detail the probabilities of geometric propositions belonging to perfect totalities of the second type, and to supplement appropriately the principles of the calculus of probability when generalizing them onto infinite totalities.

3.2.4. The Generalization of the Addition Theorem

Until now, when calculating the probabilities of the propositions of infinite totalities, both of the first and second type, we have made use of the fact that each of the propositions considered belonged also to some finite totality; hence, when applying the principles of probability theory of finite totalities, it was possible to calculate the probabilities sought. Owing to that circumstance, our calculations do not depend on whether we admit that the Axiom 2.2 (§2.1) is generalized onto joins of an infinite set or incompatible propositions; or, which is the same, on whether the addition theorem is generalized onto infinite joins.

Indeed, since that theorem is valid for a finite number of propositions, we can only conclude that, if A is a join of a countable totality of incompatible propositions a_1, a_2, … with probabilities p_1, p_2, … respectively, the probability P of proposition A is higher than, or equal to the sum of the series

$$p_1 + p_2 + \ldots + p_n + \ldots$$

which, consequently, should be convergent.

The issue about the extension or non-extension of the addition theorem cannot at all present itself for totalities of the first type, because there, in accord with the essence of the matter, an infinite join only makes sense if it is reducible to a finite join. The problem differs for totalities of the second type. When introducing the same monotone increasing function $F(z)$ as above, and considering the proposition $\alpha < x < \beta$, we see that here it makes sense for any α and β and that its probability w only obeys two conditions,

$$w \le F(\beta + 0) - F(\alpha - 0),\ w \ge F(\beta - 0) - F(\alpha + 0).$$

If at least one of the points α and β is a point of discontinuity of the function F, the equality sign cannot take place in both conditions. Then, supposing that

$$w > F(\beta - 0) - F(\alpha + 0),$$

the generalized addition theorem will not apply to the proposition that x belongs to segment $(\alpha; \beta)$ considered as the limit of the sum of the segments $(a_1; b_1)$, $(a_1; a_2)$, $(b_1; b_2)$, $(a_2; a_3)$, $(b_2; b_3)$ etc, included into (α, β). If, however,

$$w < F(\beta + 0) - F(\alpha - 0),$$

the addition theorem will not be valid with respect to the negation of that proposition. In both cases it is possible to consider the true proposition $-\infty < x < +\infty$ as an infinite join of such propositions the sum of whose probabilities has a limit smaller than unity.

Note. Since all points cannot be points of discontinuity for a monotone function F, there will always also exist such infinite joins, to whom the generalized addition theorem is applicable. Hence the assumption or the violation of that theorem is equivalent to the extension or non-extension of Axiom 2.2 onto infinite joins. For such joins, the violation of the addition theorem or of Axiom 2.2 implies an infraction [25] of the multiplication theorem for infinite combinations as well as of some properties of expectations. This fact would have presented considerable inconveniences; even when assigning a definite sense to limiting propositions (as for example to such as $x = a$, or "event A will occur at least once when the trials are repeated unboundedly"), we are usually more interested in the probabilities of variable propositions for which the proposition under consideration is a limit, and we therefore desire that the probability of the latter be in turn the limit of those probabilities [26]. This continuity of the dependence between the propositions and their probabilities leads to the need for extending Axiom 2.2, and, along with it, of the addition and the multiplication theorems onto an infinite set of propositions.

And so, the generalization of Axiom 2.2 is the only new assumption being added to the previous ones, and we thus obtain the main general principle of probability theory of infinite or finite totalities.

A necessary and sufficient condition for p_1, p_2, … *to be the probabilities of propositions* A_1, A_2, … *respectively of a given infinite perfect totality, is that the probability of any proposition contained there, and being a join of a finite number or infinite set of its propositions, will be equal to the sum* (*to the limit of the sum*) *of the probabilities of these latter*; *that the true proposition will have probability* 1 (*and, consequently, that the* *impossible proposition, probability* 0); *and that the other propositions will have probabilities contained between* 0 *and* 1 ($0 < p < 1$).

As to the notion of probability of a proposition given that another one has occurred, and the coefficient of compatibility of propositions, I have nothing essentially to add to what was said in §2.5.

3.2.5. Investigation of the function $F(z)$

The assumption that the totality H is a simple totality of the second type; that is, that infinite combinations such as (A and A_1 and …) are impossible because of the generalized addition theorem, means that the limit of the probability of inequalities

$$a - h < x < a + h \text{ as } h \to 0 \qquad (26)$$

is zero; consequently, the function $F(z)$ is *continuous.* Conversely, if it is continuous, the totality of propositions

$$a < x < b \ (0 \le a \le b \le 1),$$

and of all of their possible joins, is a totality of the second type, *i.e.*, a totality devoid of elementary propositions.

A still stronger restriction is usually imposed on $F(z)$: it is supposed to be differentiable, so that

$$F(z) = \int_0^z f(x)\, dx\,. \qquad (27)$$

This restriction is connected with the following property:

Theorem 3.1. *If* ($1/2 + \alpha_n$) *and* ($1/2 - \alpha_n$) *are the probabilities for the n-th digit of a binary fraction to be* 1 *and* 0 *respectively, the necessary and sufficient condition for* (27), *where* $f(x)$ *is bounded, continuous at* $x \ne k/2^n$ *with*

$$\{f[(k/2^n) + 0] - f[(k/2^n) - 0]\} \to 0 \textit{ as n increases,} \qquad (28)$$

is that the series $\Sigma\alpha_n$ *should be absolutely convergent* [27].

Indeed, this convergence is equivalent to the uniform convergence of all the possible products

$$\prod_{i=1}^{\infty} (1 \pm 2\alpha_i). \qquad (29)$$

The same product multiplied by coefficient 1/2 and extending over $[1; n]$ is equal to

$$F[(k + 1)/2^n] - F(k/2^n)$$

and represents the probability of inequalities

$$k/2^n < x < (k + 1)/2^n. \qquad (30)$$

Consequently, the convergence of the products (29) implies the existence of a finite positive limit

$$\lim 2^n\{F[(k + 1)/2^n] - F(k/2^n)\} = f(x) \qquad (31)$$

where x is determined by the inequalities (30) as n increases unboundedly. Since the product (29) is uniformly convergent, the function $f(x)$ is continuous at $x \neq k/2^n$; in addition, the quantity (28) that differs generally from 0 can be made arbitrarily small at sufficiently large values of n (and an odd k). Therefore, $f(x)$ is integrable (in the Riemann sense).

The equality (31) can be represented as

$$F[(k+1)/2^n] - F(k/2^n) = [f(x) + \varepsilon_k]\delta \qquad (31\ \text{bis})$$

where $\delta = 1/2^n$ and ε_k tends uniformly to 0 as n increases. Consequently,

$$F[(k+1)/2^n] = \Sigma[f(x) - \varepsilon_k]\delta,\ k \geq 0.$$

It follows that

$$F(x) = \int_0^x f(w)\,dw, \qquad (32)$$

or $F'(x) = f(x)$ at the points of continuity of $f(x)$, – that is, at $x \neq k/2^n$; at the other points the derivative of $F(x)$ on the right is $f(x+0)$, and its derivative on the left is $f(x-0)$.

Conversely, equality (32) leads to (31) for those values of x for which $f(x)$ is continuous; therefore, all the products (29) converge, QED. A similar theorem can be proved in the same way for other number systems as well, hence the

Corollary 3.3. *A necessary and sufficient condition for the function* $F(x)$ *to have a continuous derivative everywhere* [28] *is that, for two number systems* (*e.g., for the binary and tertiary systems*), *the probability that the n-th digit acquires one of its h possible values is equal to* $[1/h + \alpha_n^{(h)}]$, *and all the products*

$$\prod_{n=1}^{\infty} [1 + h\alpha_n^{(h)}]$$

converge.

Note. The proved theorem [29] shows that all the various laws of distribution of probability determined by the arbitrary function $f(x)$ differ one from another only in the values of the probabilities of the first binary (or, which is the same, decimal) digits, whereas the subsequent digits at least tend to become equally possible. Thus, no definite laws governing the sequences of the digits of the fractions considered, as, for example, the law of periodicity, can be realized whatever is the function $f(x)$. The arithmetization of a geometric totality by means of any continuous function $f(x)$ excludes the possibility of definite equalities $x = a$; conversely, if such equalities can be realized in accord with the nature of a given problem, the possibility of the existence of a continuous function $f(x)$, or even of a continuous $F(x)$, is excluded.

Definition. We call a function $F(x)$ *continuous in the narrow sense if*

$$\Sigma|F(\beta_n) - F(\alpha_n)|$$

always tends to zero together with $\Sigma|\beta_n - \alpha_n|$.

In particular, it is obvious that the existence of a finite derivative [30] is sufficient for $F(x)$ to be continuous in the narrow sense. *A necessary condition for this is that the probabilities of the various digits in all the places of an infinite fraction have a lower limit differing from zero* (*that is, an upper limit differing from* 1).

Indeed, had not this lower limit been different from zero, there would have existed a combination (A_π and A_ρ and …) of an infinite number of digits with probability differing from zero. The sum of the intervals corresponding to the given digits on the π-th, the ρ-th, …, the n-th places, whose total length is equal to $1/2^{n+1}$, and thus tends to 0 together with n, would have differed from zero and the function $F(z)$, contrary to the demand stated, would not have been continuous in the narrow sense.

Without dwelling in more detail on the investigation of the connection between the properties of continuous functions $F(x)$ with the probabilities of the various binary (or decimal) digits of the number x, I pass on to considering discontinuous functions.

3.2.6. Arithmetization of Totalities of the Fourth Type by Discontinuous Functions

We saw that $F(z)$ is an arbitrary monotone function obeying the condition $F(1) - F(0) = 1$. It is known from the theory of functions that, if F has points of discontinuity, *i.e.*, such points a where

$$F(a + 0) - F(a - 0) > 0,$$

their totality is countable.

I showed that the left side of this inequality is the limit of the probability of the inequalities (26). Since the proposition $x = a$ is the combination of all the propositions (26), it will, on the strength of the generalized addition (multiplication) theorem, have probability

$$h_o = F(a + 0) - F(a - 0).$$

Let us isolate all the countable totality of the points of discontinuity, $a_1, a_2, \ldots, a_n, \ldots$ Denote

$$h_n = F(a_n + 0) - F(a_n - 0),$$

and compile such a function $F_1(z)$, for which the same equality will hold and the sum of its variations at all the other points will be zero. Then the function

$$F_2(x) = F(x) - F_1(x)$$

will be continuous [31].

First suppose that

$$F_2(x) = 0, \text{ i.e., that } \Sigma h_n = 1, 1 \le n < \infty.$$

Here, we have only a finite or a countable totality of elementary propositions $x = a_1$, $x = a_2$, … and of all of their possible joins, – that is, a totality *of the fourth type.* Any proposition

$$a < x < b \qquad (33)$$

has probability equal to the sum of the probabilities of all the elementary propositions $x = a_n$ satisfying the inequalities (33). It is self-evident that the symbols $<$ and $\le$ are now equivalent only if they are applied to values differing from {those at} the points of discontinuity.

This case most essentially differs from the instance of a continuous $F(x)$. Here, generally, the consecutive digits are not only not independent, but, after a finite number of them is given, all the infinite set of the other ones is determined with probability approaching certainty. Indeed, the probability of any value is represented by a convergent infinite product whose consecutive multipliers are all rapidly tending to unity. We see thus that *some arithmetization of geometric totalities changes them into totalities of either the second or the fourth type. If both* $F_1(x)$ and $F_2(x)$ *differ from zero, we have a mixed or general case of a* *totality of the second type* (§3.1.5), *which is easily reduced to a totality of the fourth type, and a simple totality of the second type.*

In Chapter 4 I shall return to the considerations that guide us when arithmetizing totalities. Here, it is appropriate to note that difficulties and contradictions appear, because, when establishing a certain arithmetizing function $F(x)$, we keep at the same time to intuitive notions incompatible with it. For example, recognizing that $F(x)$ is continuous, we find it difficult to imagine that the possibility of a definite proposition $x = a$ is incompatible with our assumption; and that, when admitting that possibility, we ought to make a a point of discontinuity of $F(x)$. But it is hardly needed to say that such contradictions between intuitive and logical conclusions are rather usual in mathematics, and that they cannot be resolved by some compromise such as "not any proposition of an infinite totality having probability zero is impossible". Thus, in the theory of functions, we are not embarrassed by the contradiction between our intuitive notion of a curve and the existence of continuous functions lacking a derivative; and it will certainly never occur to anybody to assume, that a continuous function is absolutely arbitrary, and to consider, at the same time, a tangent at some point of the curve depicting that function.

Having any absolutely arbitrary totality of the fourth type of any cardinal number as a totality of all the points of a segment, and of all of their possible joins, we will always preserve, after its arithmetization, only a countable totality of elementary propositions, and we will be obliged to consider the other elementary propositions impossible. Indeed, there cannot be more than one elementary proposition with probability higher than 1/2; or more than two of them having probabilities exceeding 1/3 etc.

The choice of the elementary propositions which should be considered possible, is in many cases an unsolvable problem. Indeed, who, for example, will be able to indicate that countable totality of the points of a segment, which anybody at all had already indicated or chosen, or will indicate or choose (as, for example, 1/2, or $1/\sqrt{2}$, or ln2, etc)? Nevertheless, it is obvious that this totality is countable [32] whereas all the other numbers ought to be considered impossible, because they never were, and never will be actually realized, and, consequently, cannot be realized. This inability is proper; in accord with the demands of experience, practice in most cases compels us to abandon the attempts to arithmetize totalities of the fourth type, and to replace them by those of the second type, naturally without violating the principles of the theory.

Here is the usual reasoning: When having two equal {congruent} finite segments, the probabilities that a definite number is contained in either of them are equal. However, this consideration is not quite rigorous. The less is the length of the intervals, the more considerable is the inaccuracy of that assumption which cannot be absolutely admitted since it would have led us to an arithmetizing function $F(z) = z$, incompatible, as I showed above, with the realization of definite equalities $x = a$. On the contrary, when considering our arithmetization as only *approximate*; when assuming that the probabilities are not equal but differing one from another less than by some very small but not exactly known number ε, we should remember that our arithmetization is *relatively* the less satisfactory the smaller are the segments (so that, in particular, the probability of the equality $x = a$ is not always equal to zero). We have thus solved the paradox consisting in that, for the arithmetizing function $F(z)$ exactly equal to z, all the totality of the never realizable [33] (impossible) numbers with measure 1 would have probability 1 (equal to certainty).

When arithmetizing a totality of the fourth type, the issue of determining the probabilities of the so-called non-measurable totalities of points should also be indicated. For us, this problem does not present difficulties, because, after choosing the arithmetizing function, – that is, after selecting the countable totality of elementary propositions, – any totality of points, whether measurable or not, acquires a probability on the strength of the generalized addition theorem depending on the points which are included in it and are corresponding to the elementary propositions.

As to the totalities of the second type considered above, they include, because of their very structure, only such joins [34], that are reducible to finite or countable joins, so that we do not have to mention non-measurable totalities; consequently, all the propositions of the totalities, both of the fourth and the second type, acquire quite definite probabilities once the arithmetizing function $F(z)$ is chosen.

3.2.7. Arithmetization of Totalities of the Third Type

On the basis of what was said about totalities of the fourth type, we already know that after arithmetization only a countable totality of possible elementary propositions can be left in the considered totality. The difference between arithmetized totalities of the third and the fourth types only consists in that there exist infinite joins making sense in the fourth, but not in the third type; we do not therefore mention here probabilities of such joins. All the other joins will have the same probabilities in both totalities.

Summarizing all that was said about the arithmetization of infinite totalities, we see that, to whichever type they belong, this procedure is entirely determined by the function $F(x)$ [35], on whose choice the very type of the totality also depends because a point of discontinuity of $F(z)$ corresponds to each elementary proposition, and vice versa. If we admit the generalized constructive principle, we obtain, depending on the nature of $F(z)$, totalities of the second and the fourth types. If, however, we hesitate to attach sense to some infinite joins (and combinations), our totalities should be attributed to the first or the third type.

3.2.8. Arithmetizaton of the Totality of Integers

Integers and their finite joins provide an example of a totality of the third type. If we link to them all the possible infinite joins, we obtain a totality of the fourth type with a countable totality of elementary propositions. Its arithmetization is usually achieved on the basis of the assumption that all numbers are equally possible. *This premise however is obviously inadmissible because it would have implied that the probability of each number is zero, i.e., that no number could have been realized, and, in addition, the generalized addition theorem would have been violated because the sum of the probabilities of a countable totality of propositions with probability* 0 *would have been unity.*

The difficulty of selecting a law of probability for the numbers depending, in each concrete case, on the statement of the problem at hand, cannot justify the choice of a law, even if it is simple, contradicting the main principles of the theory of probability. We may consider such a limit of probabilities of some propositions, that corresponds to a gradually increasing restricted totality of numbers, under the assumption that in such totalities the numbers are equally possible, but that limit is not the probability of a definite proposition of our infinite totality.

Another *inadmissible assumption*, connected with the one just mentioned, is made as often as that latter, viz., that the probability for the number N, when divided by a prime number a to provide a remainder α, does not depend on the remainder left after dividing N by a prime number b. Indeed, let $\alpha_o = 0$, $\alpha_1 = 1$ be the two possible remainders after dividing N by two; $\beta_o = 0$, $\beta_1 = 1$ and $\beta_2 = 2$, the remainders after dividing it by three; etc. Then, according to the assumption, the probabilities of all the infinite combinations (α and β and …) are equal, but most of these combinations are impossible, because, when dividing N by a greater number, all the remainders obtained become equal to N, so that, for example, the combination (0, 1, 0, 1, …) of the remainders is impossible. And it would follow that also impossible are those combinations which correspond to integers.

It is also possible to attach another meaning to all the combinations if we connect each of them with the series

$x = \alpha/2 + \beta/(2\cdot3) + \ldots + \lambda/(2\cdot3\cdot\ldots p_n) + \ldots$

where p_n is the n-th prime and $\lambda < p_n$. Then any combinations of the remainders correspond to all the values of x contained between 0 and 1 [36]. The values of x corresponding to integers (according to the condition, this is the only possible case) are characterized by *periodicity* indicated above and are obviously countable, whereas its other values are uncountable. It would therefore be absolutely wrong, when assuming that all the numerical values [37] $\lambda < p_n$ are equally possible, to consider as certain that x belongs to the first, to the countable totality, and that its pertaining to the second one is impossible.

It is thus necessary to admit, that the use of the term *probability* in the theory of numbers (for example, "the probability that a number is a prime is zero") is in most cases unlawful; there, the sense of that term does not correspond to the meaning attached to it in the theory of probability.

4. Supplement. Some General Remarks on the Theory of Probability As Being a Method of Scientific Investigation

4.1. The Possibility of Different Arithmetizations of a Given Totality of Propositions

In the previous{main}chapters, I attempted to establish the formal logical foundation of probability theory as a mathematical discipline. For us, propositions were until now only abstract symbols without any concrete substance having been attached to them. We have only determined definite rules for performing operations on them and on the appropriate numerical coefficients which we called probabilities. We proved that these rules did not contradict one another and allowed under certain conditions to derive by mathematical calculations the probabilities of propositions given the probabilities of some other propositions.

However, only the logical structure of a totality of propositions, which at least for finite totalities is usually understood in each concrete case without any difficulty, does not suffice for arithmetizing totalities; some additional conditions are still needed for calculating all the probabilities by means of the principles of probability theory. Indeed, if we throw a die and restrict our attention on two possible outcomes, on the occurrence and non-occurrence of a six, we have a simple pattern $O, A, \overline{A}, \Omega$. We obtain the same pattern when throwing a coin, and, also, when again considering the throws of a die and regarding the cases of an even (2, 4, 6) or an odd (1, 3, 5) number of points as differing from each other. In the latter experiment we have the same scheme $O, B, \overline{B}, \Omega$ although A is a particular case of B. It is not difficult to conclude now that the same arithmetization (for example, the assumption that all the elementary propositions are equally possible) of all the logically identical totalities would have led to an unavoidable contradiction.

And so, not all the conditions needed for the arithmetization of a totality follow from its formal logical structure; only the real meaning that we attach to probability provides additional information for preliminary agreements which are arbitrary from the mathematical viewpoint. On the other hand, our calculations are practically and philosophically interesting only because the coefficients derived by us correspond to some realities. {Thus, a certain coefficient}(the mathematical probability) should provide the highest possible precision concerning the degree of expectation of some event on the basis of available data; in other words, of the measure of predetermination of the event given some objective information. If we state that the mathematical probabilities of events A and B are equal (*i.e.*, that the events are equally possible), it means that the totality of the available objective data is such that any reasonable person must expect them both to an absolutely the same extent.

4.2. The Origin and the Meaning of the Axioms of the Theory of Probability

For the time being, we leave aside the issue of whether there exists such objective information that any person will agree that they predetermine the events A and B to the same extent, so that they should be equally expected, should be considered equally possible. However, even when denying the availability of such information for each case, each person who attempts to understand to what extent he may count on the occurrence of some event, the following axioms of §2.1 will be compulsory.

1. *We ought to reckon on a certain event more than on an uncertain event.*

2. *If we expect A and* A_1 *to the same extent; if, further, the same is true with respect to B and* B_1; *if A is incompatible with B, and* A_1 *incompatible with* B_1, *– then we should equally expect* (A or B) *and* (A_1 or B_1). *On the contrary, if we expect B rather than* B_1, *then we expect* (A or B) *more than* (A_1 or B_1).

The Axiom of realization 2.3 (§2.2.3) will become just as obvious, if, when stating it, we attach the abovementioned sense to the notion of probability: *If* α *and* β *are particular cases of A and B respectively, we should, when counting on A as much as on B, equally expect the occurrence of* α *and* β; *and, if A occurs, we should expect* α *to the same extent as* β *provided that B occurs.*

Depending on whether our assumptions of the equal probability of the considered events are objective or subjective, the conclusions, following from our objectively compulsory (for a normal state of mind) axioms and theorems, will acquire an objective or a more or less subjective meaning.

We ought to show now that the assumptions about an equal possibility of two phenomena can be as objective as the premise of the equality of any two concrete quantities whatsoever; and to reveal thus the scientific importance of the theory of probability.

4.3. Equipossibility

To this end, let us consider an example. A homogeneous sphere is placed on a cylinder of revolution with horizontal elements in such a manner that its center {of gravity}is on the same vertical line with the point of contact. Had the experiment been realized ideally, the sphere would have been in equilibrium. However, it follows both from mechanics and experience that this equilibrium is unstable. An unyielding to measurement deviation from the conditions of an ideal experiment is sufficient for the sphere to roll to one or to the other side. If the practitioner will realize this experiment with all the possible precision by taking all measures for the deviation of the cylinder to one side not to outweigh its deviation to the other side, the outcome would have remained unknown to him. It is naturally possible that another experimentalist with more precise instruments can foresee the result; but then he should again modify the experiment for arriving at the same pattern of unstable equilibrium, and the new outcome would be just as unknown to him as it was to his predecessor.

When preparing the second experiment identical with the first one as much as possible, our practitioner will have the same grounds for expecting a similar result. *Had the experiment been stable with its outcome not being influenced by such differences in its arrangement, which are not allowed for, we might have foreseen that the results in both cases would be the same. However, owing to the mechanical instability of the realized layout, we restrict our statement by concluding that a definite outcome of the second experiment* (*the movement of the sphere to the right*) *has the same probability as in the first one.*

In general, *if the difference between the causes leading to the occurrence of events A and B is so negligible as not to be detected or measured, these events are recognized as equally probable.*

Once this definition is admitted, it also directly leads to the previously assumed axioms. However, our axiomatic construction of the theory of probability is not connected with accepting or disregarding it. An absolute equality of probabilities naturally represents only a mathematical abstraction, just as the equality{congruence}of segments does; for establishing that the fall of a given die on any of its faces has one and the same probability, we may only use those objective, but not absolutely precise methods of measurement, which are usually applied in geometry and physics.

Just as it happens when applying any mathematical theory, when precise equalities have to be replaced by approximate equalities, it is therefore essential to study how will the theorems of the theory of probability change if the probabilities mentioned there acquire slight arbitrary variations. Extremely important in this respect is, for example the Poisson theorem, without which the Bernoulli theorem would have been devoid of practical interest.

4.4. Probability and Certainty

On the strength of the considerations above, mathematical probability represents a numerical coefficient, a measure of expectation of the occurrence of an event given some concrete data. It thus characterizes the objective connection between the observed information and the expected event. In particular, the dependence between the given data and a future event can be such that the data imply its certainty; that they are its cause, and its probability is 1. It should be borne in mind that certainty, just as probability, is always theoretical; it is always possible that an incomplete correspondence between reality and our theoretical pattern disrupts or modifies the expected act of the cause.

By definition, unconditionally certain can only be a result of an agreement or a logical conclusion, whereas any prevision of a future fact is always based on induction, – that is, in the final analysis, on a direct or oblique assumption that an event invariably occurring under given conditions will happen again under similar circumstances. By applying the principles of the theory of probability, it can be shown that such a prevision has probability very close to unity, *i.e.*, to certainty. The other statements of the theory, having the same degree of probability, should therefore be considered as practically certain, bearing however in mind that the mistake, caused by the incomplete correspondence between the preliminary assumptions and reality, has no less chances of undermining the correctness of any proposition than the fact that its probability does not entirely coincide with certainty.

The study of many experiments, each of which is represented by some unstable pattern of the type indicated above, where the conditions to be allowed for compel us to assign definite probabilities to their outcomes, leads us, on the basis of calculations belonging to probability theory, to statements known as the law of large numbers, which have approximately the same high probability as our inductive inferences. When employing that law, as when applying inductive laws of nature, we ought to consider the possibility that the concrete conditions of our experiment do not quite correspond to the theoretical layout. A definite result of the experiment in both cases has therefore only a high probability rather than an absolute certainty. A mistake, *i.e.*, a non-occurrence of our prevision, is not impossible, it is only highly improbable. But it is the feature of the law of large numbers that an occurrence of an unbelievable fact is not an unconditional indicator of the incorrectness of our theoretical assumptions: the law tolerates exceptions. A detailed study, of how should we regard an admitted hypothesis if the previsions based on it are often wrong, is beyond the boundaries of this paper. The theory of the probability of hypotheses, to which this issue belongs, is entirely founded on the axiom of realization [38]. Restricting our attention to general considerations, we may only note that a prior estimate of the probability of some arrangements is usually very arbitrary, so that only those conclusions, made in the context of that chapter of probability theory, are of special interest which are more or less independent of such estimates.

In itself, an occurrence of an unbelievable fact does not refute an hypothesis and only represents new information that can change its probability; there exists no pattern under which all the occurring phenomena have considerable probabilities [39]. Our only demand on the accepted hypothesis is that a greater part of the occurred facts should have possessed a high degree of probability with only a comparatively few of them having been unlikely. The vagueness of this remark corresponds to the essence of the matter: the impossibility of allowing for the entire unbounded totality of causes influencing an isolated concrete phenomenon excludes infallibility in previsions. Instead of the certain, representing a theoretical abstraction, we have to be satisfied with the probable (the practically certain), and we only ought to try that this substitute will result in mistakes as rarely as possible.

It is clear now, that the application of the theory of probability contains a portion of subjectivity, but only a portion, to a certain degree inherent in any method of cognition that interprets facts and connects them by definite abstract interrelations. These relations, which in our theory are characterized by a coefficient, – by the mathematical probability, – are able to interpret reality more or less precisely; and the conclusions implied by the application of probability theory must then possess an appropriate degree of accuracy. Indeed, the few axioms underpinning this mathematical theory represent a necessary feature of the concept of probability as a measure of expectation, independent of the objective meaning of the relevant data.

4.5. Infinite Totalities

When considering some experiment admitting of a finite number of outcomes and stating that result *A* is possible, we mean that, bearing in mind all the experiments corresponding to the same theoretical pattern, we believe that in some of them *A* really takes place. Had we the possibility of glancing over all the previous and the future experiments covered by that layout, and concluding that *A* had{and will}never happen(ed), we would have been compelled to say that, once the arrangement of these experiments is correct, *A* is impossible. The usual inductive inferences are reached in the same way: if *A* had not occurred in a large number of experiments, we conclude that it is impossible.

A similar remark is also applicable to infinite totalities. If a totality of logically possible incompatible outcomes is uncountable, as for example the number of points of the segment [0; 1], *i.e.*, the totality of the values of x satisfying the inequalities $0 < x < 1$, it can occur that actually possible here is only a countable totality of outcomes, and any arithmetization of the{initial} totality, in accord with the theoretical principles established in Chapt. 3, should consider all the actually (or mentally) never happening outcomes as impossible. The totality of the realizable results is unknown to us. And we have still less prior grounds for believing, in accord with what was said about the objective indications of equipossibility, that it is equally probable that the number mentally chosen by someone is 1/2, or that it is a result of calculation impossible by means of contemporary analysis [40].

It is therefore necessary to distinguish between arbitrary undefinable numbers and those definable by some means. It should be noted however, that, only if these means are indicated, we obtain a *definite* totality of definable numbers (for example, of algebraic numbers), and we may only state that there should exist numbers which will never be defined. But the very boundary between these two categories of numbers cannot be precisely pointed out.

Choosing an arbitrary number written down as an infinite decimal fraction, and asking ourselves what is, for example, the probability that 0 will not occur there at all, we ought to answer it depending on the category to which that number belongs. Assume that the probability for each digit to be in each place is [41] 1/10 because we may assume that there is no objectively revealable cause for one digit to have preference over another one in any possible case. Under these conditions the probability that 0 does not occur will be

$$\lim (9/10)^n = 0 \text{ as } n \to \infty.$$

However, our assumption obviously concerns only absolutely arbitrary *undefinable* numbers, whose composition does not obey any law, so that only some finite number of digits can be indicated in each number. Such numbers cannot ever be completely defined because infinitely many digits only depending on chance is still always left, and no experiment can establish that 0 will not yet occur. On the contrary, its occurrence is compatible with any observed result, *i.e.*, it is certain (in accordance with the principle of uniqueness). The matter is different if we believe that the composition of our fraction obeys some law. If we precisely indicate that law, – for example, if we choose proper rational fractions whose denominators have no multipliers excepting twos and fives, – then we should first of all investigate whether there exists a direct causal connection between the law and the occurrence of the zero; here, arithmetic teaches us that the non-occurrence of 0 is impossible. However, had we chosen fractions of the type

$$(10^n - 2)/(10^n - 1),$$

then, on the contrary, the occurrence of 0 would have been impossible. If no direct causal connection is seen, we should nevertheless remember that our law connects the sequence of digits in a certain way so that their total independence and equiprobability cannot be admitted. The less definite is the law, the more difficult it is to indicate a priori the exact value of the probability of each digit being in a definite place; in such cases, it is more correct to calculate the probabilities a posteriori, and, although it ought to be thought that mostly the value of that probability will be very close to 1/10, it is really possible that for a diverse totality of numbers a supernormal variance testifying to the lack of constant probability will be revealed in some cases [42].

The study of totalities (always countable) of some categories of definable numbers is of little practical interest. On the contrary, experimental science usually has to do with infinite totalities of the second, and partly of the first type lacking elementary propositions, – that is, with totalities of undefined numbers because no experiment can precisely establish non-integral numbers. An experiment can only determine a few decimal places of an unknown number not admitting experimental determination. Accordingly, we should choose a continuous arithmetizing function; bearing in mind the considerations of §3.2.5, it is almost always possible to assume that

$$F(z) = \int f(z)dz$$

where $f(z)$ is some non-negative continuous function whose value{s}is{are} determined a priori by the conditions of the experiment, or a posteriori by the results of its numerous repetitions.

We also encounter infinite totalities when applying the law of large numbers to some experiment repeated without restriction. In most cases the number of the experiments is supposed to be finite although very large. The law therefore provides its inherent practical, but not logical certainty, and, as noted above, admits exceptions. However, had we created a pattern realizing the limiting case of an infinite number of repetitions for interpreting some phenomenon, we could have arrived at conclusions possessing logical certainty. For example, admitting the possibility of a gradual speeding-up of the experiment with throwing a coin, or of another one where the probability of the event is 1/2, so that the first experiment lasts 1 *min,* the second one, 1/2 *min,* the third one, 1/4 *min,* etc, then the total number of experiments lasting 2 *min* would have been infinite.

Assuming that there exists some stable mechanical device consecutively recording the ratio of the number of the occurrences of the event to the number of the experiments (although the recording of the result of each separate experiment then becomes impossible), we will notice that until the end of the second minute this ratio is somewhat varying. However, after that the pointer of the dial of our device will occupy quite a definite position corresponding as much as it is possible for our device to the number 1/2. This conclusion is theoretically certain and its non-occurrence can take place only because of an incomplete accordance between the actual conditions and our theoretical pattern. Thus, when forming a definite infinite binary fraction, for example, 8/15 = 0.10001 …, where the limit of the relative number of the units is 1/4, we must state that its composition is incompatible with the assumption that the occurrence of 1 and 0 on each place is equally probable.

And it is generally impossible to indicate a method of forming an infinite binary fraction, where the sequence of unities and zeros would have obeyed an infinite number of conditions implied by the laws of large numbers. Infinite series composed absolutely arbitrarily, randomly (so that each number is arbitrary in itself) essentially differ from series compiled in accord with a definite mathematical law no matter how arbitrary it is. A confusion of these two notions, caused by the fact that such differentiation does not exist between random and regular finite series, is one of the main sources of paradoxes to which the theory of probability of infinite totalities is leading.

Notes

1. (§1.1.8). Conversely, the principle of uniqueness follows if we suppose that (A or $\overline{A}$) = Ω; that is, when assuming that a proposition and its negation are solely possible. Indeed, if α is compatible with any proposition (excepting O), then $\overline{\alpha} = O$, hence (α or O) = Ω and $\alpha = \Omega$.

2. (§1.1.9). When applying it to the given equation (4), we find that conditions

$$\{(A \text{ or } a \text{ or } b) \text{ or } [\overline{A}' \text{ and } (\overline{a}' \text{ or } \overline{b}')]\} = \Omega,$$
$$\{(A' \text{ or } a' \text{ or } b') \text{ or } [\overline{A} \text{ and } (\overline{a} \text{ or } \overline{b})]\} = \Omega$$

are necessary and sufficient for its solvability.

3. (§1.2.1). It could have been proved that, also conversely, the assumption that a true proposition exists, leads to the constructive principle. Thus, for a finite totality {of propositions}, this principle and the axiom of the existence of a true proposition are equivalent. {Not axiom but Theorem 1.1.}

4. (§1.2.3). For example, the principle of uniqueness is not valid for the system 1, 2, 3, 4, 6, 12. The proposition corresponding to the number 2 would have been compatible with all the propositions and its negation would therefore be only the false proposition. This last-mentioned would therefore possess the most important property of a true proposition without however being true.

5. (§1.3.1). So as not to exclude the false proposition, it is possible to agree that it is a join of zero elementary propositions, – that it does not contain any of them.

6. (§2.1.1). The theory of probability considers only perfect totalities of propositions.

7. (§2.1.4). Had we called the fraction $m/(n - m)$ probability, the ratio of the number of the occurrences of an event to the total number of trials should have been replaced, in the Bernoulli theorem for example, by the ratio of the former to the number of the non-occurrences. The statement of the addition theorem would have also been appropriately changed: the probability of (A or B) would have been equal not to the sum of the probabilities, $(p + p_1)$, but to

$$(p + p_1 + 2pp_1)/(1 - pp_1).$$

8. (§2.1.5).{Below, the author specifies his incomplete distinction.}

9. (§2.1.5) If $p + p_1 > 1$, we would have to choose $\overline{A}'$ instead of B' and would have thus convinced ourselves in the inadmissibility of such an assumption.

10. (§2.1.5). Such two propositions could not be chosen only if $\lambda_n + \lambda'_n > 1$. Replacing then proposition B_n by $\overline{A}_n$, we would have determined, when applying Axiom 2.2, that the proposition (*A* or *B*) has probability higher than unity, – that is, higher than Ω has, which contradicts Axiom 2.1. Consequently, in this case *p* and p_1 cannot be the probabilities of incompatible propositions.

11. (§2.1.6). Note that Axioms 2.1 and 2.2b taken together are equivalent to the following single axiom: *Inequality A > B means that there exists* (or can be added) *such a proposition* B_1 *being a particular case of A that* $B_1 \sim B$.

12. (§2.2.2). Issuing from the notion of realization of one or several propositions, Markov (1913, p. 19) provides another definition:

We call several events $E_1, E_2, \ldots, E_n$ *independent one of another if the probability of none of them depends on the existence or non-existence of the other ones, so that no indication that some of these events exist or do not exist changes the probabilities of the other events.*

It is not difficult to satisfy ourselves that the two definitions are equivalent, but it ought to be noted that some conditions in the latter necessarily follow from the other ones. This is obvious because the number of these conditions is here $n(2^{n-1} - 1)$, or $[(n - 2)2^{n-1} + 1]$ greater than in the former definition. These redundant conditions are therefore corollaries of the other ones. For $n = 2$ the independence of *B* of *A* is a corollary of the independence of *A* of *B*. Note that in many cases (for example, in the {Bienaymé –} Chebyshev inequality), it is essential to break up the notion of independence, and the *pairwise* dependence or independence plays an especially important part.

13. (§2.2.3). A similar axiom only concerning totalities with equally possible elementary propositions is found in Markov (1913, p. 8). Let us explain our axiom by an example. If any permutation of a complete deck of playing cards taken two at a time has the same probability 1/(52·51), then, in accord with the addition theorem, the probability that the first or the second drawn card is the knave of hearts is 51/(52·51) = 1/52; upon discovering that the first card was the queen of hearts, all the permutations containing that queen remain equally possible only because of the axiom of realization, and the probability for the second card to be the knave of hearts becomes equal to 1/51. Had we only assumed that at each single drawing the possibility of the occurrence of each card was one and the same, this axiom would have been insufficient for recognizing that all the permutations taken two at a time were equally possible. This fact becomes natural once we note that it is easy to indicate an experiment where these permutations are not equally possible.

14. (§2.2.3). {Literal translation.}

15. (§2.2.3). At first, issuing from the functional equation $f(x + y) = f(x) + f(y)$, we obtain, for any integer *n*,

$$f(n x) = n f(x). \qquad (*)$$

Then, assuming that $nx = my$, where *m* is also an integer, we get $nf(x) = mf(y)$, hence $f(nx/m) = n/mf(x)$. Since $f(x)$ is finite ($|f(x)| \le 1$ for $0 \le x \le 1$), we infer from (*) that it tends to zero with *x* so that it is continuous and the equality $f(tx) = tf(x)$, proven for any rational *t,* is then valid for any *t*. Consequently,
$f(t) = tf(1)$.

16. (§3.1.1). It is obvious that, once join *H* exists, it is unique. Indeed, if H_1 also satisfies the first condition, then (*A* or H_1) = H_1, (*B* or H_1) = H_1, etc, so that (*H* or H_1) = H_1. But since H_1 also obeys the second condition, (*H* or H_1) = *H* and $H = H_1$.

17. (§3.1.2). The previous definition of *incompatibility* of two propositions can persist.

18. (§3.1.3). The generalized constructive principle (along with the restrictive principle) is realized in the pattern of §1.2 if only we extend this latter onto an infinite set of primes and of all of their products devoid of quadratic factors. Unity will still correspond to the true proposition, and 0, which we shall define as a multiple of all integers, to the impossible proposition. However, a compiled system where also the combination of two propositions (the least multiple) always exists, will not be perfect because there the principle of uniqueness is violated: any pair of propositions (excepting *O*) is here compatible, therefore *O* will be the negation of any proposition. Nevertheless, the theorems of distributivity persist.

19. (§3.1.4). We may even suppose that they are written down in the binary number system. They will then be rational numbers of the type $a/(2^n - 1)$.

20. (§3.1.4). In each place *y* has the largest digit out of the corresponding digits of $f(x)$, $f(x_1), \ldots, f(x_n), \ldots$ so that, if $x, x_1, \ldots$ are finite fractions, $f(x) = f(x_1) = \ldots = 0$ and $f(y) = 0$.

21. (§3.2.1). For example, "The number of heads, when the experiment is repeated indefinitely, is equal to the number of tails", or "There will be no less than 10 heads".

22. (§3.2.1). {The author's term was *finitistic.*}

23. (§3.2.3). As to all the infinite joins of the type (A_α or A_β or …), they can represent new propositions; otherwise, they will all be true as being compatible with any proposition of the totality.

24 (§3.2.3). In §1.3 I explained that an impossible proposition is characterized by the fact that it cannot become true or certain, – that it cannot be realized.

25. (§3.2.4). The generalization of the multiplication theorem is a corollary of the generalized addition theorem. Indeed,

$$\text{Prob}\,(A \text{ and } B \text{ and } \ldots \text{ and } L \text{ and } \ldots) = 1 - \text{Prob}\,(\overline{A} \text{ or } \overline{B} \text{ or } \ldots \text{ or } \overline{L} \text{ or } \ldots)$$
$$= 1 - \text{limProb}\,(\overline{A} \text{ or } \overline{B} \text{ or } \ldots \text{ or } \overline{L} \text{ or } \ldots) = \text{limProb}\,(A \text{ and } B \text{ and } \ldots \text{ and } L \text{ and } \ldots).$$

26. (§3.2.4). If, for example, event *A* has probability 1/2 at the first trial, 1/4 at the second one, 1/8 at the third one, etc, then, irrespective of these values of the probabilities, we might have stated that the occurrence of the event at least once is certain because this is compatible with any result of a finite number of trials. However, the limit of the probability that *A* will occur in *k* trials will be, as *k* increases unboundedly,

$$1/2 + (1/2)\,(1/4) + (1/2)\,(3/4)\,(1/8) + (1/2)\,(3/4)\,(7/8)\,(1/16) + \ldots < 3/4 < 1.$$

Here, the violation of the addition theorem leads to obscuring the fact that the realization of the proposition *A* in some finite experiment becomes ever less probable with time.

27. (§3.2.5). If the subsequent digits are not independent of the previous ones, this condition should be replaced by a uniform convergence of all the different products $\Pi(1 + 2\alpha_n)$.

28. (§3.2.5). Note that for the existence of a finite derivative of $F(x)$ at a given point *x* it is sufficient that, uniformly with respect to λ,

$$(1/\lambda) \lim\{[F(x + \lambda h) - F(x)]/[F(x + h) - F(x)]\} = 1 \text{ as } h \to 0.$$

Indeed, for an arbitrarily small ε it is possible to choose such a small α that

$$[F(x + \lambda h) - F(x)]/[F(x + h) - F(x)] = \lambda(1 + \varepsilon')$$

where $|\varepsilon'| < \varepsilon$ as soon as $|\lambda h| < \alpha$, $|h| < \alpha$. By choosing some definite value of *h* in this manner, we will obtain

$F(x + h) - F(x) = Mh$, therefore $F(x + \lambda h) - F(x) = M\lambda h(1 + \varepsilon')$,

hence

$$[F(x + \lambda h) - F(x)]/\lambda h = M(1 + \varepsilon').$$

However, since ε can be chosen no matter how small, the left side here differs from M, which does not depend on λ if $\lambda h \to 0$, as little as desired, QED. In most cases, when applying the theory of probability, this condition is obviously fulfilled. It can be shown that the condition stated above is also necessary for the existence of a finite derivative (differing from 0).

29. (§3.2.5).{Apparently, Theorem 3.1.}

30. (§3.2.5). For continuity in the narrow sense, it is sufficient that the Lipschitz condition be satisfied for the whole interval excepting a restricted number of points where the function can simply be continuous.

31. (§3.2.6). Let $\psi(x) = 0$ at $x < 0$ and $= 1$ at $x > 0$. Then $F_1(x)$ can be represented as an absolutely convergent series

$$F_1(x) = \Sigma h_n\psi(x - a_n),\ 1 \le n \le \infty.$$

32. (§3.2.6). It would have been finite for a world restricted in time.

33. (§3.2.6). The totality of the realizable numbers is countable, so that, even being everywhere dense, it has measure 0.

34. (§3.2.6) It is of course possible to choose an arbitrary totality of points S and to define a proposition A as a combination of all the propositions, *i.e.*, of the sums of the segments which include those points. Then A will correspond to the outer measure of totality S which always exists. But the inner measure of that totality, if S is non-measurable, can lead to another proposition $B \ne A$. These propositions will always have definite probabilities whereas the totality S does not represent a proposition.

35. (§3.2.7). In more complicated cases, by one or several such functions of several variables.

36. (§3.2.8). In particular,

$$1 = 1/2 + 2/(2\cdot3) + \dots + (p_n - 1)/(2\cdot3 \dots p_n) + \dots$$

37. (§3.2.8). On the strength of the above, this assumption leading to the arithmetizing function $F(z) = z$, always excludes the possibility of any precise equality $x = a$, and, in particular, of any integer.

38. (§4.4). {Note that the author had not mentioned the then originating mathematical statistics.}

39. (§4.4). We may consider, for example, as practically certain that a first-class chess player paying full attention to his game will beat a novice who had just learned the rules of chess. Nevertheless, it is not absolutely impossible, that all the moves made by this beginner by chance, satisfy the demands of the art of the game and led him to victory. A combination of isolated unlikely facts of this kind can indeed happen. And such a result (especially repeated twice or thrice) would have placed us in a very difficult situation concerning the expected outcome of the next game. Can we be sure that our novice had indeed nothing to do with chess, as all those knowing him are stating; can we deny the possibility of such an unprecedented and unknown until now talents, that were revealed so brilliantly at the very first game? But, although we cannot completely answer these questions, the games that took place represent a specimen of the wittiest chess maneuvers whose study will reveal the profound expediency of the separate moves. Therefore, no matter how we would be inclined to uphold our prior confidence in that the novice could not have played deliberately, we would still be compelled to admit that the connection between his moves was advisable and appropriate.

A similar remark is applicable to the hypothesis on the regularity of the phenomena of nature. As much as we would like to believe in miracles, we will inevitably have to admit the regularities in order to explain the data available. However, it is impossible to dissuade anyone from believing in the existence of wonders taking place beyond the domain of precise observation; and the laws, until now considered as indisputable, will perhaps turn out to be freaks of chance.

40. (§4.5). For example, before the theory of logarithms was discovered, that it had been ln2.

41. (§4.5). That is, $F(z) = z$.

42. (§4.5). {The author apparently had in mind the Lexian theory of dispersion. See the description of the appropriate work of Markov and Chuprov in Sheynin (1996, §14).}

References

Bernstein, S.N. (1927), *Теория вероятностей* (Theory of Probability). Fourth ed.: M., 1946.

Bohlmann, G. (ca. 1905), Technique de l'assurances sur la vie. *Enc. Sci. Math.*, t. 1, vol. 4. German text: Lebensversicherungsmathematik. *Enc. Math. Wiss.*, Bd. 1, Tl. 2, pp. 852 – 917.

Boole, G. (1854), On the conditions by which the solution of questions in the theory of probabilities are limited. *Studies in Logic and Probability.* London, 1952, pp. 280 – 288.

Borel, E. (1914), *Le hasard.* Paris. Several subsequent editions.

Hochkirchen, T. (1999), *Die Axiomatisierung der Wahrscheinlichkeits-rechnung.* Göttingen.

Ibragimov, I.A. (2000). Translation: On Bernstein's work in probability. *Transl. Amer. Math. Soc.*, ser. 2, vol. 205, 2002, pp. 83 – 104.

Kolmogorov, A.N. (1948, in Russian), E.E. Slutsky. An obituary. *Math. Scientist*, vol. 27, 2002, pp. 67 – 74.

Kolmogorov, A.N., Sarmanov, O.V. (1960), Bernstein's works on the theory of probability. *Teoria Veroiatnostei i Ee Primenenia*, vol. 5, pp. 215 – 221. That periodical is being translated as *Theory of Probability and Its Applications.*

Laplace, P.S. (1812), *Théorie analytique des probabilités. Oeuvr. Compl.*, t. 7. Paris, 1886.

Laplace, P.S. (1814), *Essai philosophique sur les probabilités.* English translation: *Philosophical Essay on Probabilities.* New York, 1995.

Markov, A.A. (1912), *Wahrscheinlichkeitsrechnung.* Leipzig – Berlin. This is a translation of the Russian edition of 1908. Other Russian editions: 1900, 1913 (to which Bernstein referred) and 1924.

Schröder, E. (1890 – 1895), *Vorlesungen über der Algebra der Logik*, Bde 1 – 3. Leipzig.

Sheynin, O. (1996), *Chuprov.* Göttingen.

Slutsky, E.E. (1922), On the logical foundation of the calculus of probability. Translated in this collection.

--- (1925), On the law of large numbers. *Vestnik Statistiki*, No. 7/9, pp. 1 – 55. (R)

4. S.N. Bernstein. On the Fisherian "Confidence" Probabilities

Bernstein, S.N. *Собрание сочинений* (Coll. Works), vol. 4. N.p., 1964, pp. 386 – 393

1. This paper aims at making public and at somewhat extending my main remarks formulated after the reports of Romanovsky and Kolmogorov at the conference on mathematical statistics in November 1940 [1]. So as better to ascertain the principles of the matter, I shall try to write as elementary as possible and will consider a case in which the "confidence" probability is based on one observation. However, all my conclusions are applicable to more complicated cases as well.

Suppose that one observation of a random variable x is made providing $x = x_1$ and that x obeys a continuous law of distribution depending on one parameter a, so that

$$P(t_0 < x - a < t_1) = \int_{t_0}^{t_1} f(t)dt, f(t) = (1/\sqrt{2\pi}) \exp(-t^2/2) \text{ (say)}. \quad (1)$$

According to the classical theory, after securing the observation x_1 it only makes sense to say that the probability that an unknown parameter obeys the inequalities

$$t_0 < x - a < t_1 \quad (2)$$

if, even before the observation, a could have been considered as a stochastic variable. In particular, if $p(a)$ is the prior density of a, the prior density of x is

$$P(x) = \int_{-\infty}^{\infty} p(a) f(x - a)\, da. \quad (3)$$

Therefore, on the strength of the Bayes theorem, the probability of the inequalities (2) is equal to

$$\Phi(x_1; t_0; t_1) = \frac{\int_{x_1-t_1}^{x_1-t_0} p(a) f(x_1 - a) da}{\int_{-\infty}^{\infty} p(a) f(x_1 - a) da} = \frac{1}{P(x_1)} \int_{t_0}^{t_1} p(x_1 - t) f(t)dt. \quad (4)$$

Thus, this is the probability of inequalities (2) for any interval $(x_1 - t_1; x_1 - t_0)$ considered by us. Regrettably, our information about the function $p(a)$ is usually very incomplete, and, consequently, formula (4) only provides an approximate value of Φ and the precision of this approximation depends on the measure of the precision of our knowledge about the function $p(a)$.

2. This inconvenience, that lies at the heart of the matter, became the reason why the British statisticians led by Fisher decided to abandon the Bayes formula and to introduce some new notion, or more precisely, some new term, *confidence*. They consider some pair of values, t_0; t_1 such that

$$\int_{t_0}^{t_1} f(t)dt = 1 - \alpha(t_0; t_1) \quad \text{(i)}$$

is very close to unity: differs from unity (for example, $\alpha(t_0; t_1) = 0.05$); these values therefore possess the property according to which the probability of (1) differs from unity by a given small variable $\alpha(t_0; t_1)$; and, after observation provided $x = x_1$, the interval $(x_1 - t_1; x_1 - t_0)$ is called the *confidence* region of magnitude a corresponding to *confidence* $1 - \alpha(t_0; t_1)$.

It would have been possible to agree with the introduction of the new term, *confidence*, if only new contents, differing from, and even fundamentally contradicting the previous ones adopted when defining it, were not read into the word. Indeed, Fisher and his followers believe that, once x took value x_1, the magnitude (i) is the confidence probability of a being situated in the interval $(x_1 - t_1; x_1 - t_0)$. However, since t_0 and t_1 can in essence take any values, the *confidence* probability satisfies all the axioms characterizing the classical concept of probability, and all the theorems of probability theory are applicable to it. It follows that for some choice of the function $p(a)$ in (4), the *confidence* probability must coincide with $\Phi(x_1; t_0; t_1)$ so that we would have obtained

$$\int_{t_0}^{t_1} f(t)dt = \frac{1}{P(x_1)} \int_{t_0}^{t_1} p(x_1 - t) f(t)dt \qquad (4bis)$$

for any values of x_1, t_0 and t_1. Then, for any values of t_1 and x_1; we would have arrived at

$$f(t_1) = [1/P(x_1)]\, p(x_1 - t_1) f(t_1)$$

so that $p(a) = P(x)$ should be constant over all the real axis which is impossible (a uniform distribution over all this axis is impossible). In addition, it is not difficult to see that the equality (4) leads to the same contradiction if we only assume its validity for one x_1 for a given t_0 and any $t_1 > t_0$.

It is not necessary to prove that, if the confidence region $(x_1 - t_1; x_1 - t_0)$ becomes a part of a region where a certainly cannot be situated, then $\Phi(x_1; t_0; t_1) = 0$. For example, if it is known that $|a| < 3$, then no sensible statistician will apply confidence probability

$$\int_{-2}^{2} f(t)dt$$

for the interval $-2 < 5 - a < 2$ if by chance the observation yields $x_1 = 5$.

3. The equality (4) can be *approximately* correct under some more or less definite assumptions regarding the prior probability $p(a)$ and meaning, in essence, that the *confidence* region is a part of a sufficiently large region where $p(a)$ is more or less constant. More exactly, the following proposition takes place:

A limit theorem. *If, for any positive magnitudes* $\varepsilon < 1$ *and L, it is possible to indicate such a* n_0 *that, for all the integer* $n > n_0$,

$$[p_n(x'')/p_n(x')] < 1 + \varepsilon$$

when $|x'| \le L$, $|x''| \le L$, *and, in general, if*

$[p_n(a)/p_n(x')] < c$ *for any a and* $c \ge 1 + \varepsilon$ *is a constant not depending on n*,

then, as $n \to \infty$,

$$\Phi(x_1; t_0; t_1) = \frac{\int_{t_0}^{t_1} p_n(x_1 - t) f(t)\, dt}{\int_{-\infty}^{\infty} p_n(x_1 - t) f(t)\, dt} \int_{t_0}^{t_1} f(t)dt$$

for any x_1, t_0 *and* t_1. The convergence to the limit is uniform if the confidence interval $(x_1 - t_1; x_1 - t_0)$ is situated within an arbitrary finite region.

Indeed, for any observed x_1 we can, given an arbitrary small $\varepsilon > 0$, choose such a large number L, that

$$\int_{x_1 - L}^{x_1 + L} f(t)dt > 1 - \varepsilon$$

and, in addition, $L > |x_1 - t_0|$, $L > |x_1 - t_1|$. Then, when determining n_0 in accord with the conditions of the theorem, we have, for $n > n_0$,

$$p_n(x') \int_{t_0}^{t_1} f(t)dt < \int_{t_0}^{t_1} p(x_1 - t) f(t)dt < (1 + \varepsilon)\, p_n(x') \int_{t_0}^{t_1} f(t)dt,$$

$$p_n(x')\int_{x_1-L}^{x_1+L} f(t)dt < \int_{x_1-L}^{x_1+L} p(x_1-t)f(t)dt < (1+\varepsilon)\,p_n(x')\int_{x_1-L}^{x_1+L} f(t)dt$$

where $p_n(x')$ ist he least value of $p_n(x)$ when $|x| \le L$. Consequently,

$$p_n(x')(1-\varepsilon) < \int_{-\infty}^{\infty} p(x_1-t)f(t)dt < p_n(x')(1-\varepsilon^2+\varepsilon c) < p_n(x')(1+\varepsilon c),$$

hence

$$\frac{1}{1+\varepsilon c}\int_{t_0}^{t_1} f(t)dt < \Phi(x_1;\, t_0;\, t_1) < \frac{1+\varepsilon}{1-\varepsilon}\int_{t_0}^{t_1} f(t)dt, \qquad (5)$$

QED.

4. It is doubtless that, in general, we have no special reasons to hope that the conditions of the theorem are fulfilled in practice. Therefore, the theory of probability definitely declares that *a conclusion based on one observation is, generally, unreliable*; however, if we assume that $a = \mathrm{E}x$, then the limit theorem will be applicable for determining $a\sqrt{n}$ after n observations $x_1, x_2, \ldots, x_n$. Indeed, it follows from the theorem that, *after observing* $[(x_1+x_2+\ldots+x_n)/\sqrt{n}] = X_1$ *and having a sufficiently large n, the probability of the inequalities*

$$t_0 < X_1 - a\sqrt{n} < t_1 \qquad (6)$$

will differ arbitrarily little from the their confidence probability [2], *i.e., from*

$$\int_{t_0}^{t_1} f_n(t)dt,$$

under only one assumption that $p(a)$ *is continuous in the vicinity of* $a = X_1$ *and* $[p(a)/p(X_1)] < \infty$ *for all* $a \neq X_1$.

Thus, the theory of probability applies *confidence* understood as a limiting probability without making use of this term and offers absolutely precise indications about when it is admissible. In particular, since the law of large numbers in its wide sense is the only foundation for studying stochastic phenomena, the fundamental superiority of the *confidence* determination of a from inequalities (6) when issuing from n observations (instead of one) consists not so much in that a $\sqrt{n}$ smaller region corresponds to the same *confidence* as in that this region is more reliable *for any* X_1 since, *as n increases*, *the confidence probability tends to the actual posterior probability.*

5. The application of the Fisherian confidence probability to a definite region $(x_1 - t_1;\, x_1 - t_0)$ is practically admissible from the classical viewpoint also when we may be sure that the true probability of inequalities (2) cannot be considerably lower than the *confidence* probability; that is, when

$$\Phi(x_1;\, t_0;\, t_1) \ge \frac{1-\alpha(t_0;t_1)}{1+\delta\alpha(t_0;t_1)} \qquad (7)$$

where δ is not large.

For this to be valid, *it is sufficient that* $p(b) \le (1+\delta)p(a)$ *if a is any point within the confidence region*, $x_1 - t_1; < a < x_1 - t_0$, *whereas b is any point outside it*: $b < x_1 - t_1$ *or* $b > x_1 - t_0$. Indeed, in this case

$$\frac{\Phi(x_1;t_0;t_1)}{1-\Phi(x_1;t_0;t_1)} = \frac{\int_{t_0}^{t_1} p(x_1-t)f(t)dt}{\int_{-\infty}^{t_0} p(x_1-t)f(t)dt + \int_{t_1}^{\infty} p(x_1-t)f(t)dt} =$$

$$[p(\xi)/p(\eta)]\frac{\int_{t_0}^{t_1} f(t)dt}{\int_{-\infty}^{t_0} f(t)dt + \int_{t_1}^{\infty} f(t)dt}$$

where ξ is some point within the interval $(x_1 - t_1; x_1 - t_0)$ and η is some point outside it. Therefore, in accord with the adopted condition, $p(\eta) \le (1 + \delta)p(\xi)$ and

$$(1+\delta)\,\frac{\Phi(x_1;t_0;t_1)}{1-\Phi(x_1;t_0;t_1)} \ge \frac{\int_{t_0}^{t_1} f(t)dt}{1-\int_{t_0}^{t_1} f(t)dt} - \frac{1-\alpha(t_0;t_1)}{\alpha(t_0;t_1)},$$

hence (7).For example, if $\alpha(t_0; t_1) = 1/20$, $\delta = 1$, then

$$\Phi(x_1; t_0; t_1) > 1 - 2/21 > 1 - 1/10.$$

6. Excepting the indicated or similar cases, the application of confidence probability for estimating the probability of inequalities (2) after determining a *definite* value $x = x_1$ can sometimes lead to blunders. However, it is obviously true that when $\alpha(t_0; t_1) = 0$ we will have, for any assumptions about $p(a)$,

$$\Phi(x_1; t_0; t_1) = \int_{t_0}^{t_1} f(t)dt = 1$$

so that we should believe that the smaller is $\alpha(t_0; t_1)$ the less probable is a considerable difference between $\Phi(x_1; t_0; t_1)$ and the integral above. The exact meaning of this statement becomes evident from the following remark: When integrating the equality

$$P(x_1)\Phi(x_1; t_0; t_1) = \int_{t_0}^{t_1} p(x_1 - t) f(t)dt$$

with respect to x_1 from $-\infty$ to $+\infty$ we have

$$\int_{-\infty}^{\infty} P(x_1)\Phi(x_1; t_0; t_1)\, dx_1 = \int_{-\infty}^{\infty}\int_{t_0}^{t_1} p(x_1 - t) f(t)dt\, dx_1 = \int_{t_0}^{t_1} f(t)dt \qquad (8)$$

or, in other words, the confidence probability is the expected posterior probability of the inequalities (2).

It is quite natural to desire, when knowing *absolutely nothing* about $p(a)$, and, consequently, about $P(x)$, to determine the unknown magnitude $\Phi(x_1; t_0; t_1)$ through its known expectation. Taking into account that the variance

$$\mathrm{E}[\Phi(x_1; t_0; t_1) - \int_{t_0}^{t_1} f(t)dt]^2 = \mathrm{E}\Phi^2(x_1; t_0; t_1) - [\int_{t_0}^{t_1} f(t)dt]^2 <$$

$$[1 - \alpha(t_0; t_1)] - [1 - \alpha(t_0; t_1)]^2 < \alpha(t_0; t_1)\,[1 - \alpha(t_0; t_1)] < \alpha(t_0; t_1),$$

we may conclude, on the strength of the Chebyshev – Markov lemma, that the probability of the inequality

$$|\Phi(x_1; t_0; t_1) - \int_{t_0}^{t_1} f(t)dt| > z\sqrt{\alpha(t_0;t_1)} \qquad (9)$$

is lower than $(1/z^2)$ for any $z > 1$. For example, if $(1/z^2) = \sqrt[3]{\alpha(t_0;t_1)}$, we find that

$$P[|\Phi(x_1; t_0; t_1) - \int_{t_0}^{t_1} f(t)dt| \le \sqrt[3]{\alpha(t_0;t_1)}\ ] \ge 1 - \sqrt[3]{\alpha(t_0;t_1)},$$

i.e., that this probability is arbitrarily close to 1 if $\alpha(t_0; t_1)$ is sufficiently small.

The statistician may therefore, in each particular instance, only apply confidence 0.999 when it is practically admissible to neglect probability 0.1 (not 0.001). But, in general, when knowing nothing about $p(a)$, *confidence* only acquires real sense in accord with its initial definition when applied to *a large number of independent observations* $x_1, x_2, \ldots, x_n$. Indeed, on the strength of the law of large numbers, the frequency of cases in which a will be situated in the appropriate confidence regions, just as the arithmetic mean of $\Phi(x_1; t_0; t_1)$, will be close [3] to confidence. However, if a remained invariable, then, as we saw above, its determination through $\bar{x}$, the arithmetic mean of x_i, will not only be $\sqrt{n}$ times more precise than the same *confidence* probability; it will already be very close to the individual probability of a being situated in a *given* interval $(\bar{x} - t_1\sqrt{n};\ \bar{x} - t_0\sqrt{n})$: for a sufficiently large n the regulating action of the law of large numbers already leads to practically identical values for the posterior probability at any assumptions about the prior probability $p(a)$, – excepting suppositions intentionally invented contrary to any common sense.

7. To prevent any misunderstandings that can arise in practice when putting too much trust in *confidence* probability, let us consider the following example. A large number of boxes A_1, $A_2, \ldots, A_i, \ldots$ contain {several} objects {each} whose {total} values $x_1, x_2, \ldots, x_i, \ldots$ obey the Gaussian law $(1/\sqrt{2\pi})\exp[-(x - a_i)^2/2]$ with variance 1 and location parameter a_i. Because of technical or economic considerations, the value x_{1i} of only one object is checked in each box. Assuming that $t_1 - t_0 = 2$, we obtain for each a_i a confidence region $(x_{1i} - 2; x_{1i} + 2)$ corresponding to confidence probability

$$(1/\sqrt{2\pi})\int_{-2}^{2} \exp(-t^2/2)\,dt \approx 0.95.$$

The salesman fastens confidence tags on each box saying that $|a_i - x_{1i}| < 2$ believing, as Fisher does, that it is unnecessary to warn the buyer that the inscription is only based on one observation, he only honestly declares that the guarantee is not true for 5% of the boxes.

If the buyer intends to buy any box, then its contents will not justify the tag in only one case out of 20. However, if he wants to obtain a box, or several boxes, with a definite value $a_i = a$ to within ± 2, I would advise him to go to another store where the goods are sorted out more cautiously. Indeed, if the buyer required, for example, high-valued objects corresponding to the mean $a_i = a$ which are comparatively rare (for large values of a the prior probability $p(a)$ is low), it is highly probable that, from among the 20 boxes with the appropriate tags bought by him, not one, but ten boxed will be worthless.

If the technical difficulty or the high cost of sampling are such that only one object can be inspected in each box, it would be necessary to know the (prior) density $p(a)$. Then the tags ensuring that the errors occur not more often than in 5% of the boxes will correspond to intervals $(x_i \pm t_i)$ of unequal lengths, but they will guarantee that whichever *definite* box the buyer chooses, the probability of error will actually be equal to 0.05 with t_i being determined from the equality

$$\Phi(x_{1i}; -t_i; t_i) = \frac{\int_{-t_i}^{t_i} p(x_{1i} - t)\exp(-t^2/2)dt}{\int_{-\infty}^{\infty} p(x_{1i} - t)\exp(-t^2/2)dt} = 0.95.$$

I shall not dwell on the problem of practically determining $p(a)$ given the distribution $P(x_{1i})$ of the x_{1i}'s over all the stock. Theoretically, it is solved by means of the equality

$$\exp(-t^2/2)\theta_a(t) = \theta_{x_1}(t) \qquad (10)$$

and

$$\theta_{x_1}(t) = \int_{-\infty}^{\infty} P(x)e^{itx}dx,\ \theta_a(t) = \int_{-\infty}^{\infty} p(a)e^{ita}da$$

are the characteristic functions of x_{1i} and a_i respectively.

It is also advantageous to bear in mind equality (10) in order to warn the too zealous advocates of Fisher against a rash conclusion that, since for a very large number of experiments the frequency of cases in which $|x_{1i} - a_i| < t$ tends to the same value

$$(1/\sqrt{2\pi})\int_{-t}^{t} \exp(-t^2/2)dt$$

independently of whether we consider a_i or x_{1i} as given magnitudes [4], the limiting frequencies or probabilities $p(a_i)$ and $P(x_{1i})$ should coincide [5], or at least be symmetrically connected.

Summing up the above, we see that the abandoning of the Bayes formula by Fisher leads to the confusion of the statistical frequency with mathematical probability [6] and to the identification of the case in which confidence is very close to the actual posterior probability with the instance in which it is only some mean of various posterior probabilities. In the latter case, the application of confidence probabilities can therefore lead to the same mistakes that are connected with considering general means for involved populations, as for example the application of a mean coefficient of mortality or literacy calculated for a given nation boasting a hundred million inhabitants to its various groups.

Notes

1. It is perhaps not amiss to indicate that, while criticizing here the concept of *confidence*, I would not at all desire to belittle the importance of that part of the investigations done by Fisher and other British statisticians which is connected with the problem of constructing functions $F(x_0; x_1; \ldots; x_p; a_1; \ldots; a_k)$ of $(p + 1)$ independent variables $x_0, x_1, \ldots, x_p$ $(0 < p \leq k)$ obeying one and the same law $P(x; a_1; a_2; \ldots; a_k)$ whose laws do not depend on $a_1, a_2, \ldots, a_p$ at any given values of a_i $(i = 1, 2, \ldots, k)$. {This is hardly understandable.}My objections only concern the *confidence* interpretation of the results.

2. See my *Теория вероятностей* (Theory of Probability). M. – L., 1946, Suppl. 4. It is known that, if $E(x - a)^2$ exists, then, for any given function $f(t) = f_1(t)$,

$$\lim \int_{t_0}^{t_1} f_n(t)dt = (1/\sigma\sqrt{2\pi})\int_{t_0}^{t_1} \exp(-t^2/2\sigma^2)\, dt \text{ as } n \to \infty.$$

3. In accord with what was said above, this statement would have been obviously wrong if the condition of independence of the x_i's be violated, and only those x_i's which are situated, for example, within some given interval would have been considered.

4. I hope that the previous pages have explained that this statement is mistaken.

5. The distribution of $\bar{\xi}$, the arithmetic mean of $x_{1i}, x_{2i}, \ldots, x_{ni}$, when having a sufficiently large n, will really tend to $p(a_i)$. Indeed, denoting the characteristic function of $\bar{\xi}$ by $\theta_n(t)$, we obtain, instead of (10), the equality $\theta_n(t) = \exp(-t^2/2n)\,\theta_a(t)$ so that $\theta_n(t) \to \theta_a(t)$ as $n \to \infty$.

6.{About 25 years later, Chuprov criticized Pearson and his school for confusing theoretical and empirical magnitudes, see my book (1990, in Russian), *Chuprov*. Göttingen, 1996, §15.3.}

5. L.N. Bolshev. Commentary on Bernstein's Paper on the Fisherian "Confidence" Probabilities

In: Bernstein, S.N. *Собрание сочинений* (Coll. Works), vol. 4. N.p., 1964, pp. 566 – 569 …

Bernstein's paper is the first one published in the Soviet mathematical literature that critically analyzed the theory of fiducial inferences developed by one of the founders of the modern mathematical statistics, Fisher (1935). The formal (the computational) side of the Fisherian theory usually leads to the same results as does the now widely spread theory of confidence intervals offered by the American mathematician Neyman (1938). The fundamental difference between the concepts of the two scholars lies in the interpretation of the obtained results. According to Neyman, the estimated parameter a is treated as an *unknown constant*. A confidence interval $[a_1(\xi); a_2(\xi)]$ is constructed beforehand for its experimental (prior) determination with its ends being functions of a random variable ξ subject to observation and obeying a distribution with unknown parameter a. If the probability of the simultaneous existence of two inequalities,

$$a_1(\xi) < a < a_2(\xi) \qquad (1)$$

does not depend on a, it is called the *confidence probability* or *confidence coefficient* [1].

In other words, the confidence probability is the *prior probability* of "covering" the unknown true value [2] of the parameter by the confidence interval. This interpretation of the confidence probability naturally persists when a is a random variable; indeed, that probability is calculated before the experiment and does not therefore depend on the prior distribution of the parameter a.

According to Fisher, the confidence coefficient is interpreted as the *posterior probability* of the simultaneous existence of the two inequalities

$$a_1(x) < a < a_2(x) \qquad (2)$$

after it becomes experimentally known that the random variable ξ had taken the value x. This interpretation fails for an unknown constant a: once x becomes known, the boundaries $a_1(x)$ and $a_2(x)$ are not random anymore; the *confidence probability* can be either zero or unity. In attempting to surmount this difficulty in interpreting the confidence coefficient as posterior probability, Fisher additionally assumes that each observation $\xi = x$ provides the so-called *fiducial* distribution of a depending [3], in the general case, on the parameter x. This distribution is such that the *fiducial* probability of the event (2) usually coincides with the prior confidence probability of (1) [4].

From the viewpoint of the classical theory of probability, the artificial introduction of *fiducial* distributions ought to be considered as an attempt at excluding the influence of the unknown prior distribution of a on the posterior probability

$$P[a_1(x) < a < a_2(x) \,/\, \xi = x]. \qquad (3)$$

Bernstein showed that the Fisherian interpretation of *confidentiality* as posterior probability contradicts the foundations of probability theory, and, specifically, is at variance with the Bayes theorem. This proposition proves that the confidence probability is actually the expectation of the posterior probability (3), see Bernstein's formula (8).

The acceptance of the Fisherian concept will consequently imply the replacement of the posterior probability by its mean value; for a small number of observations this can lead to large mistakes. Thus, as stressed by Kolmogorov & Sarmanov (1962, p. 200, Note 4), it is proved, that, having

one or a small number of observations, it is not possible, generally speaking, to preclude the role, indicated by the Bayes theorem, of prior information about the value of a.

Modern textbooks on mathematical statistics interpret the meaning and the practical use of the confidence intervals in accord with Neyman whose concept advantageously differs from the Fisherian notion by being logically irreproachable. More details are to be found in Kolmogorov (1942) and Neyman (1961).

Remark 1. The term *доверительная вероятность* (confidence probability) as applied by Bernstein is equivalent to the English Fisherian expression *fiducial probability*. Attempting to stress the fundamental distinction between his concept and Fisher's notion, Neyman proposed a new term, *confidence probability*, which is again translated into Russian in the same way. It would have therefore been more natural and precise to say, confidence probability according to Fisher, or according to Neyman, respectively. More often, however, the attribution to Neyman is left out because it is only in that sense that the theory of confidence intervals is explicated in modern Soviet statistical literature. But, when desiring to stress that the Fisherian notion is meant, it is said *fiducial probability* rather than *confidence probability in accord with Fisher*.

Remark 2. The concluding section of Bernstein's paper is devoted to an example showing that a wrong interpretation of the confidence coefficient as posterior probability can lead to considerably mistaken results. The problem of sorting formulated there can be solved rather simply by the Bayes formula provided that the prior distribution of the expectations of a_i is known. It is still an open question whether this problem can be solved by means of the theory of confidence intervals (according to Neyman); it is unknown whether there exists a method of sorting ensuring (independently of the prior distribution of a_i) the consumer, with a sufficiently low risk, the receipt of not less than 95% boxes satisfying his demand that

$$|a_i - a| < 2. \tag{4}$$

In this connection, it ought to be stressed once more that Bernstein demonstrates not the deficiency of the method of confidence intervals, but only the absurdity of the wrong interpretation of confidence probability. It is easy to convince ourselves that, in the example offered, a reasonable interpretation and application of this method do not lead to nonsense; they only testify that, for the procedure of sorting introduced in this example, the problem is unsolvable without prior information on the distribution of a_i. Indeed, if the customer intends to select several boxes for which

$$|x_{1i} - a| < h \tag{5}$$

($h > 0$ and a, determining the method of sorting indicated by the author, are constants given beforehand), the distribution functions of the material values x_{1i} of the chosen boxes will be represented by the formula

$$P(x_{1i} < x) = F(x; a_i) = C(a_i) \int_{a-h}^{x} \exp [-(u - a_i)^2/2]du$$

where $a - h < x < a + h$ and $[1/C(a_i)]$ is equal to the same integral extended over $a - h \le x \le a + h$. These functions determine the conditional distribution of x_{1i} given (5).

Consider random variables $F(x_{1i}; a_i)$. It is easy to see that, if a_i varies from $-\infty$ to $+\infty$, F decreases monotonically from 1 to 0. The difference $(1 - F)$ as a function of a_i represents a distribution function [5].

Let A_i and B_i be random variables determined as solutions of the equations

$$F(x_{1i}; A_i) = 0.975, F(x_{1i}; B_i) = 0.025.$$

It is not difficult to convince ourselves that, for any fixed a_i, the probabilities of the events $(A_i \ge a_i)$ and $(B_i \le a_i)$ are equal to 0.025. It follows that for any a_i

$$P(A_i < a_i < B_i / a_i) = 0.95. \quad (6)$$

This means that (6) persists for any prior distribution of a_i.

In other words, $(A_i; B_i)$ is a confidence interval for a_i with (prior) confidence probability 0.95. The length of this interval varies and essentially depends on x_{1i}; if x_{1i} tends to $(a - h)$ or $(a + h)$, then $(B_i - A_i) \to \infty$. Therefore, it is impossible to make any conclusions about whether, for the selected boxes, the inequalities (4) hold or do not hold given any $h > 0$.

Notes

1. In the more general case this coefficient is defined as the exact lower bound of the probabilities of inequalities (1) over all the admissible values of a.

2. {This is a rare example of a statistician applying a term usually restricted to the theory of errors.}

3. According to Fisher (1935), until the experiment, a is an unknown constant; its *fiducial* distribution, in his own words, *is only revealed by sample observations*. His approach can apparently be justified, if at all, only beyond the conventional theory of probability. {I have not found the quoted phrase in exactly the same wording.}

4. The difference between the *fiducial* and the confidence probabilities is only revealed when the fiducial distribution of the parameter a represents a convolution of some other *fiducial* distributions (Neyman 1941).

5. According to Fisher, this distribution should be called *fiducial*.

References

Fisher, R.A. (1935), The fiducial argument in statistical inference. *Annals Eugenics*, vol. 6, pp. 391 – 398.

Kolmogorov, A.N. (1942), Determining the center of scattering, etc.

Kolmogorov, A.N., Sarmanov, O.V. (1960, in Russian), The work of S.N. Bernstein on the theory of probability. *Theory of Probability and Its Applications*, vol. 5, 1962, pp. 197 – 203.

Neyman, J. (1938), L'estimation statistique traitée comme un problème classique de probabilité. *Sel. Early Stat. Papers*. Cambridge, 1967, pp. 332 – 353.

--- (1941), Fiducial argument and the theory of confidence intervals. *Biometrika*, vol. 32, pp. 128 – 150.

--- (1961), Silver jubilee of my dispute with Fisher. *J. Oper. Res. Soc. Japan,* vol. 3, pp. 145 – 154.

Seneta, E. (2001), Bernstein. In Heyde, C.C., Seneta, E., Editors. *Statisticians of the Centuries*. New York, pp. 339 – 342.
{Additional items: two articles in *Statistical Sci.*, vol. 7, No. 3, 1992, by T. Seidenfeld and S.L. Zabell.}

6. E.E. Slutsky. On the Logical Foundation of the Calculus of Probability[1]

Report read at the Section of theoretical statistics
of the Third All-Union Statistical Congress, November 1922.
Избранные труды (Sel. Works). Moscow, 1960, pp. 18 – 24 …

Foreword by Translator

Evgeny Evgenievich Slutsky (1880 – 1948) was an eminent mathematician and statistician, see the translation of Kolmogorov (1948) with my Foreword. Kolmogorov (p. 69) stated that Slutsky "was the first to draw a correct picture of the purely mathematical essence of probability theory" and cited the paper here translated ("the present paper", as I shall call it) and a later contribution (Slutsky 1925b). Kolmogorov (1933) referred to both these articles but did not mention the former in the text itself; curiously enough, that inconsistency persisted even in the second Russian translation of Kolmogorov's classic published during his lifetime (Kolmogorov 1974, pp. 54 and 66).

The present paper first appeared in 1922, in *Vestnik Statistiki*, Kniga (Book) 12, No. 9 – 12, pp. 13 – 21. Then Slutsky published it in a somewhat modified form (see his Notes 1 and 2) in 1925, in *Sbornik Statei Pamiati Nikolaia Alekseevicha Kablukova* [Collected papers in Memory of Nikolai Alekseevich Kablukov], vol. 1. Zentralnoe Statisticheskoe Upravlenie, Moscow, pp. 254 – 262, and, finally, it was reprinted in the author's *Selected Works.* Several years after 1922 Slutsky (1925a, p. 27n) remarked that back then he had not known Bernstein's work (1917) which "deserves a most serious study".

Bernstein, S.N. (1917), An essay on an axiomatic justification of the theory of probability. Translated in this book.
Kolmogorov, A.N. (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung.* Springer, Berlin.
Kolmogorov, A.N. (1948). Obituary: Evgeny Evgenievich Slutsky. Translation (2002): *Math. Scientist* **27,** 67 – 74.
Kolmogorov, A.N. (1974). Second Russian translation of Kolmogorov (1933). Nauka, Moscow.
Slutsky, E.E. (1925a), On the law of large numbers. *Vestnik Statistiki*, No. 7/9, pp. 1 – 55. (R)
--- (1925b). Über stochastische Asymptoten und Grenzwerte. *Metron* 5, No. 3, pp. 3 – 89.

* * *

The calculus of probability is usually explicated as a purely mathematical discipline, and it is really such with respect to its main substance when considered irrespective of applications. However, the pure mathematical nature of one element that enters the calculus from its very beginning is very questionable: any detailed interpretation of that element involves our thoughts in a domain of ideas and problems foreign to pure mathematics. Of course, I bear in mind none other than the notion of probability itself. As an illustration, let us consider, for example, the classical course of Academician Markov.

In the Introduction to its second edition (1908), he declares that he will treat the calculus of probability as a branch of mathematics, and each attentive reader knows how strictly he regards his promises. Markov shows this strictness at once, in the extremely typical of him comment on the second line of the very first page. There, he elaborates on the word *we*:

The word we *is generally used in mathematics and does not impart any special subjectivity to the calculus of probability*.

Let us however compare this pronouncement with Markov's definition of equipossibility that he offers on p. 2:

We call two events equally possible if there are no grounds for expecting one of them rather than the other one.

He adduces a note saying that, according to his point of view,

various concepts […] *are defined not so much by words, each of which in turn demands a definition, as by our attitude to them, which is ascertained gradually.*

It is doubtless, however, that, in the given context, this remark should only be considered as a logically hardly rightful way out for the author's feeling of some dissatisfaction with his own definition. That subjective element, whose shadow he diligently attempted to drive out by his remark on the first page, appears here once more so as to occupy the central position in the structure of the main notion, that must serve as the foundation for all the subsequent deliberations.

In my opinion, there is a means for countering this difficulty; there exists an absolutely drastic measure that many will however be bound to consider as cutting rather than untangling the Gordian knot. The legend tells us, that, nevertheless, such an attitude proved sufficient for conquering almost the whole world. I shall sketch the idea of my solution.

First of all, it is necessary to introduce the main notions defined in the spirit of strict formalism by following the classical example of Hilbert's *Grundlagen* (1899). Such notions as *event*, *trial*, *the solely possible events*, (*in*)*compatible events*, etc, ought to be established in this way, *i.e.* with the removal of all the concepts concerning natural sciences (time, cause, etc).Let us call the complex of the solely possible and incompatible events *A, B, …, H* an *alternative*, and the relation between them, *disjunction.* Then, instead of introducing the notion of equipossibility, we shall proceed as follows.

We shall consider such relations, which take place if some number is associated with each of these solely possible and incompatible *events,* under the condition that, if any of them (for example, *A*) is in turn decomposed into an alternative (either α, or β, or γ, …, or η), then the sum of those numbers, that occur to be associated with α, β, γ, …, η, will be equal to the number associated with *A*.

The association just described should be understood as the existence of some one-valued but not one-to-one relation *R* between the *events* included in the alternative and the numbers. In addition, *R* is the same for all the *events* and possesses the abovementioned formal property, but in essence it remains absolutely arbitrary in the entire domain of the calculus considered as a purely mathematical discipline. It can even happen that, in the context of one issue, each term of an alternative is connected with some number by relation *R* , whereas each term of another alternative is in turn connected with some number by relation *R′* not identical with *R* ; the relation *R″* will take place for a third alternative, and so on. If, in addition, a formal connection between the relations *R,, R′, R″, …* is given, purely mathematical complications, which the classical calculus of probability had never studied in a general setting, will arise. Leaving them aside, I return however to the simplest type.

Suppose that an alternative can be decomposed into the solely possible and incompatible *events* with which some fundamental relation *R* connects numbers equal one to another. I shall call such elementary *events* isovalent , and I shall introduce the notion of valency of an event as a proper fraction whose numerator is equal to the number of the elementary *events* corresponding to the given *event,* and whose denominator is the number of all the solely possible elementary and incompatible *events* included in the given alternative.

It is absolutely obvious that this foundation formally quite coincides with the classical foundation; hence, all the former's purely mathematical corollaries will formally be the same. The word *probability* will everywhere be substituted by *valency*; the formulation of all the theorems will, *mutatis mutandis,* persist{with necessary alterations}; all the proofs will remain valid. The only change consists in that the very substance of the calculus will not now have any direct bearing on probability.

For example, the addition theorem will be formulated thus: If *A* and *B* are events incompatible one with another, the valency of the event "either *A* or *B*" is equal to the sum of their valencies. The multiplication theorem will be: For compatible events *A* and *B*, the valency of the event "both *A* and *B*" is equal to the valency of one of them multiplied by the conditional valency of the other one; etc.

The purport of any theorem obviously remains purely formal until we, when somehow applying it, associate some material sense with the fundamental relation *R*; that is, until we fix the meaning of those numbers, that in the given case are attached to the terms of the alternative. Knowing the sense in which such and such *events* are isovalent, we will be logically justified, on the grounds of our calculus, to state that some other definite *events* will also be isovalent or have such and such valency, again in the same sense. It will now be naturally inconsistent to call our science calculus of probability; the term *disjunctive calculus* will apparently do [2].

This science will be as formal and as free of all the non-mathematical difficulties as the theory of groups. There, we are known to be dealing with some things, but it remains indefinite with which exactly. Then, we have to do there with some relation that can pairwise conjugate any two things one with another so that the result of this operation will be some third thing from the same totality. Under these conditions, the theory of groups develops an involved set of theorems, mathematically very elegant and really important for various applications. Within the bounds of the theory itself, the material substance of that set remains indefinite which leads to formal purity and variety of applications, and is indeed one of the theory's most powerful points. If the group consists of natural numbers, and the main operation providing a third thing out of the two given ones is addition, we obtain one possible interpretation; if the main operation is multiplication, we arrive at another interpretation; then, when compiling a group out of all possible permutations of several numbers, we get a still differing interpretation; and if, instead, we consider all possible rotations of some regular polyhedron, we have a yet new interpretation, etc.

In our case, we also have something similar. The formal notion of valency can have more than one single sense, and the meaning of the theorems known to us long since in their classical form is in essence also many-valued. Their nature remains however hidden and is only dwelt with during disputes, to a considerable extent fruitless, on the notion of probability. I shall attempt to sketch several possible interpretations of the calculus of alternatives.

First of all, we certainly have its classical form. We come to it by replacing isovalency by equipossibility and substituting probability for valency. This change may be considered from a purely formal, and from a material point of view. When keeping to the former, which is the only interesting one for a mathematician, we introduce the concepts purely conventionally. Suppose that the *possibility* of an event can be higher than, or equal to the *possibility* of another event. Presume also that two events, each decomposable into the same number of other solely and equally possible incompatible events, are themselves necessarily equally possible. Then, irrespective of either a more definite meaning of, or of the conditions for equipossibility, we are able to introduce, in the usual way, the notion of *probability* in its purely mathematical aspect. Or, otherwise, when keeping closer to the reasoning above, we may say: Suppose that *possibility* can be expressed numerically and that the *possibility* of an event is equal to the sum of the *possibilities* of those solely possible and incompatible events into which it is decomposable. Then, etc, etc.

This deliberation is tantamount to the following. We have a finished formal mathematical calculus complete with its notions and axioms. When applying it, we suppose that those axioms, that underpin the formal disjunctive calculus, are valid for some chosen concept, – for example, as in our case, for *possibility.* Thus, we presume that *possibilities* can be expressed by numbers; that all the terms of a given alternative are connected with these numbers by a one-to-one correspondence; that these latter obey those formal relations which we introduced for the numbers connected with the former by valency. From a purely mathematical viewpoint, this is apparently quite sufficient for passing on from the calculus of alternatives to the calculus of probability.

It is obvious however that all this only covers one aspect of the matter, and that here also exists another, material, so to say, side lying entirely beyond the bounds of purely mathematical ideas and interests. Indeed, for settling the issue of whether all the abovementioned notions and axioms categorically rather than conditionally suit the concepts *possibility* and *probability*, we ought to know what exactly do we mean by these notions. It is clear that this problem is of an absolutely special type requiring not a mathematical, but an essentially different phenomenological and philosophical approach. I think that for my formulation of the issues, the line of demarcation appears with sufficient clearness as though all by itself.

Let us now go somewhat farther in another direction. I have remarked that the calculus of alternatives admits not a single interpretation, but rather a number of them, and this formal generality is indeed one of its most important logical advantages over the classical calculus of probability. So as to justify this idea, I ought to indicate at least one more of its differing interpretations. Let us have a series of trials where each of the events *A, B, …, H* is repeated several times. The numbers of these repetitions, *i.e.*, the actual absolute frequencies of the events, are uniquely connected with these because each event has one certain frequency. This relation is not biunique, because, inversely, two or more events can have one and the same frequency. Then, if some event is decomposable into several solely possible and incompatible kinds, the sum of their frequencies is equal to the frequency of the given event. Frequency thus satisfies those conditions under which I introduced the concept of valency into the calculus of alternatives.

We may therefore replace valency by relative frequency and thus obtain a number of theorems with respect to the latter without repeating all the deliberations or calculations, but jokingly, so to say, by a single substitution of the new term into the previous formal statements. Thus, we will have the addition and the multiplication theorems for frequencies, absolutely analogous to the known propositions of the calculus of probability. How far-reaching are such similarities? Obviously, they go as far as the general foundation of definitions and axioms do. Had there been no other independent entities except these notions and axioms in the disjunctive calculus, or, respectively, the calculus of probability, then the calculus of frequencies would have formally covered the entire contents of both the two former calculuses. This, however, is not so. The issue of repeated trials and the concept of frequency enter the calculus of probability at one of its early stages; in addition, we should naturally find out whether, when assuming our second interpretation of the calculus of alternatives, the formal conditions that correspond to that stage can be, and are actually satisfied.

More interesting is a third interpretation. It goes much farther, covers a large and perhaps even the entire domain of our calculus provided only that we can agree with a purely empirical understanding of *probability.* Suppose that it makes sense to consider any number of trials under some constant conditions. Presume also that there exists a law on whose strength the relative number of the occurrences of any of the alternatively possible events must tend to some limit as the number of trials increases. This limiting relative frequency{These … frequencies}apparently satisfies{satisfy}those conditions, under which I introduced the notions of isovalency and valency of events. Hence, as far as the general foundation of the axioms reaches, all the theorems of the calculus of valency will possess this, as well as the classical interpretation. That the analogy goes very far is unquestionable. To say nothing about the almost trivial addition and multiplication theorems, it also covers the doctrine of repetition of events including such of its propositions as the Jakob Bernoulli, and the{De Moivre –}Laplace theorems. Small wonder that sometimes all civic rights are granted to this interpretation. Thus, we find it as a special favorite in the British school. What has it to do with the classical interpretation? Does it entirely cover the latter? And, if not, where do they diverge? Only in the understanding of the sense of the theorems, or perhaps in the extent of mathematical similarity? Until now, there are no definitive answers to any of these questions.

A rigorous revision of all the fundamentals of the calculus of probability, a creation of a rigorous axiomatics and a reduction of the entire structure of this discipline to a more or less visible mathematical form, are necessary. This however is only possible on the basis of a complete formalization of the calculus with the exclusion from it of all not purely mathematical issues. Neither probability, nor the potential limiting frequency possess such a formal nature. The calculus of probability should be converted into a disjunctive calculus as indicated above, and only then will it enter the system of mathematical sciences as its branch and become definitive, a quality which it is still lacking, and enjoy equal logical rights with the other branches.

My solution is however something more than a simple methodical device for disentangling the issues of the logic of the calculus of probability. So as to convince ourselves in this fact, suffice it to imagine that nature of logical purity which our calculus will obtain as a result of the indicated conversion. This nature is something objective, as are all the borders separating the sciences one from another. We reveal them, but we do not create them. Indeed, it needs only to compare with each other even those few theorems whose statements in terms of *probabilities, frequencies* and potential *limiting frequencies* is unquestionable.

Let us only imagine three such absolutely parallel series of definitions, axioms and theorems explicated independently, and, consequently, roughly speaking, separately from each other in three different treatises devoted to three supposedly separate calculuses respectively. In each case we will have independent series of ideas, definitions and proofs. We ask ourselves, whether the similarity between them is objective or subjective. The answer is self-evident. The general pattern and the course of reasoning are the same. Once we perceive this, we also observe that the likeness exists irrespective of our subjective arbitrariness.

It may be objected, that the formalization of the calculus of probability postulated here avoids exactly the most essential and the most interesting for theoretical statistics issues. This however is no objection. The essence of *probabilities,* the relations between *probability* and *limiting frequency*, and between the calculus of probability and the real course of things, – all these problems are important and interesting, but they are of another logical system, and, moreover, such, whose proper statement is impossible without solving simpler and logically more primitive problems. Their definitive and complete solution, as dictated by the entire development of mathematical thought, lies exactly in the direction whose defence is the subject of my study. We only have to dart a look on the issues, concerning the essence of the notion of probability and of its relation to reality, for understanding with full clearness their utter distinction from the formal mathematical problems comprising the subject of the disjunctive calculus and its axiomatics and logic. Thus, only a logical and phenomenological analysis absolutely not of a formal mathematical nature can indicate that probability is a category unto itself, completely independent of the notion of *limiting frequency.*

Now I allow myself a remark as a hint of a solely preliminary nature. Suppose that we have a number of frequencies which must surely approach some limit as the number of repetitions{of trials}increases unboundedly. It does not however follow at all that in some initial part of the trials the event could not have been repeated with a frequency sharply different from its limiting value. Suppose for example that a sharper deals the cards unfairly; that he cheats relatively less as the game goes on; and that in the limit, as the number of rounds increases unboundedly, each card will appear with frequency 1/52 as it should have happened under fair circumstances [3]. Even without knowing anything about the law governing the composition of the series of trials, we would nevertheless be sure to discover, after observing the actual behavior of the frequencies, that, with probability extremely close to certainty, the probability of the event during the first series of the trials diverges from its limiting value not less than by such-and-such amount. True, the notion of limiting frequency can also be applied to the proportion of right and wrong judgements, but neither here is the issue definitively decided: just as in the case above, we may ask the{same}questions about the probability of judgement, about the frequency and the probability of that proportion [4].

The same is true with respect to the possibility of applying the calculus of probability to empirical experience. Not the latter guides us when we establish the calculus' theorems, but, on the contrary, they, and only they, provide us with a prior compulsory clue for regulating it. From the calculus of probability we borrow the type of that law, which, following N.A. Umov [5], we might have called the law of chaos, of complete disorder. There exist domains of phenomena where the chain of causes and effects on the one hand, and the arrangement of idiographic information [6] on the other hand, ensure, in conformity with natural laws, the regularity of such a sequence: if the occurrence (non-occurrence) of some event is denoted by *A* (by *B*), then, in the limit, as the number of trials increases unboundedly, *A* ought to appear with the same relative frequency both in the entire series and after any combination of the events; equally often after *A*, and after *B*; after *AA, AB,* or *BB*; after *AAB, AAA, ABA,* etc, etc.

That such domains actually exist is shown by experience, but only when the idea of probability guides it and provides the very pattern of the law of chaos and the tests for establishing its action in one or another field, and for appraising the judgement which establishes it. Hence, in this respect the notion of probability also becomes indispensably necessary and logically primary. Is it even possible to justify the natural philosophical premises of the law of chaos without applying the notion of probability? I think that this is questionable.

Now, I have however went out of the boundaries of my main subject although this was apparently not quite useless for its elucidation. These concluding remarks will perhaps amplify the purely logical arguments by a vivid feeling, caused not by a logically formal consideration, but by direct vision and comprehension of the essence of things and issues.

Notes

1. After my text had appeared in *Vestnik Statistiki*, I improved some formulations making them more intelligible and introduced a few editorial corrections, but I did not change anything in essence.

2. After my report was published, Professor Bortkiewicz, in a letter{to me}, kindly suggested this term.

3.{I have omitted some details in this passage because Slutsky had not explained the essence of the game.}
4. {Some explanation is lacking.}
5.{Russian physicist (1846 – 1915). Slutsky provided no reference.}
6.{Idiography, the science of single facts, of history. This notion goes back to the philosophers Windelband and Rickert. Also see Sheynin (1996, p. 98).}

References

Hilbert, D. (1899), *Grundlagen der Geometrie.* Teubner, Leipzig – Berlin, 1930.
Markov, A.A. (1900), *Исчисление вероятностей* (Calculus of Probability). Subsequent editions: 1908, 1913 and 1924. German translation of the second edition: *Wahrscheinlichkeitsrechnung.* Teubner, Leipzig – Berlin, 1912.
Sheynin, O. (1990, in Russian), *A.A.Chuprov: Life, Work, Correspondence.* Göttingen, 1996.

6a. E.E. Slutsky. [Earlier] Autobiograhy

My grandfather on my father's side, Makary Mikhailovich Slutsky, served in Kiev in the Judicial Department. He began his career already before the Judicial Reform {of 1864} and stood out against the civil service estate of those times because of his exceptional honesty. He died in poverty, but he had been nevertheless able to secure higher education for my father, Evgeny Makarievich, who graduated in 1877 from the Natural-Scientific Department of the Physical and Mathematical Faculty at Kiev University.

From the side of my mother, Yulia Leopoldovna, I descend from Leopold Bondi, a physician of French extraction who, together with others {?}, moved to Russia under circumstances unknown to me. A part of his numerous descendants from two marriages established themselves as Russians. Thus, his son Mikhail, who joined the Russian Navy, was the father of the well-known Pushkin scholar S.M. Bondi. However, some of his children regarded themselves as Poles, and became Polish citizens after Poland was established as an independent state.

Soon after my birth my mother adopted Orthodoxy and, under the influence of my father, became an ardent Russian patriot in the best sense of that word and the Polish chauvinism of our relatives always served as a certain obstacle to more close relations. For about 30 years now, I have no information about these, absolutely alien {to me} representatives of our kin. After the death of my grandmother all the contacts between me and my relatives {in Russia} with them have been absolutely broken off [1].

I was born in 1880 in the village Novoe, former Mologsky District, Yaroslavl Province, where my father was a teacher and tutor-guide in the local teacher's seminary. In 1886, not willing to cover up for his Director, who had been embezzling public funds, he lost his job. For three years we were living in poverty in Kiev after which my father became the head of a Jewish school in Zhitomir. There, he had been working until his resignation in 1899, again caused by a clash with his superiors.

But then, in 1899, I had just graduated from a classical gymnasium with a gold medal and entered the Mathematical department of the Physical and Mathematical Faculty at Kiev University. I earned my livelihood by private tutoring. In January 1901, I participated in a {student} gathering demanding the return to the University of two of our expelled comrades, and we refused to obey our superiors' order to break up. In accordance with the then current by-laws of General Vannovsky [2], I, among 184 students, was forcibly drafted into the Army. Student unrest broke out in Moscow and Petersburg and in the same year the government was compelled to return us to the University.

However, already in the beginning of 1902 I was expelled once more because of {my participation in} a demonstration against the Minister Zenger and this time prohibited from entering any higher academic institution of the Russian Empire. My maternal grandmother whom I mentioned above had helped me to go and study abroad. From 1902 to 1905 I studied at the Machine-Building Department at Munich Polytechnical School. I had not graduated from there. When, in the fall of 1905, owing to the revolutionary movement in Russia, it became possible for me to enroll in a university in Russia, I entered the Law Faculty at Kiev University.

Munich was a turning point in my development. Circumstances imposed the machine-building speciality on me; it oppressed me, and, as time went on, I liked it ever less. I was forced to analyze my situation and I discovered that my visual memory was very weak. Therefore, as I understood, I could not become a good mechanical engineer. And, by the same reason, I very badly memorized people by sight and mistook one person for another one even if having met them several times so that I was unable to be a political figure either. A further analysis of my abilities confirmed this conclusion. I studied mathematics very well and everything came to me without great efforts. I was able to rely on the results of my work but I was slow to obtain them. A politician, a public speaker, however, needs not only the power of thought but quick and sharp reasoning as well. I diagnosed my successes and failures and thus basically determined the course of my life which I decided to devote exclusively to scientific work.

I became already interested in economics during my first student years in Kiev. In Munich, it deepened and consolidated. I seriously studied Ricardo, then Marx and Lenin's *Development of Capitalism in Russia*, and other authors. Upon entering the Law Faculty, I already had plans for working on the application of mathematics to economics.

I only graduated from the University in 1911, at the age of 31. The year 1905 – 1906 {the revolutionary period} was lost since we, the students, barely studied and boycotted the examinations, and one more year was lost as well: I was expelled for that time period because of a boyish escapade. At graduation, I earned a gold medal for a composition on the subject *Theory of Marginal Utility* [3]. However, having a reputation as a *Red Student*, I was not left at the University and {only} in 1916/1917 successfully held my examinations for becoming Master of Political Economy & Statistics at Moscow University.

In 1911 occurred an event that determined my scientific fate. When beginning to prepare myself for the Master examinations, I had been diligently studying the theory of probability. Then, having met Professor (now, academician) A.V. Leontovich and obtaining from him his just appeared book on the Pearsonian methods, I became very much interested in them. Since his book did not contain any proofs and only explained the use of the formulas, I turned to the original memoirs and was carried away by this work. In a year, – that is, in 1912, – my book *Теория корреляции* (Theory of Correlation) had appeared. It was the first Russian aid to studying the theories of the British statistical school and it received really positive appraisal.

Owing to this book, the Kiev Commercial Institute invited me to join their staff. I worked there from January 1913 and until moving to Moscow in the beginning of 1926 as an instructor, then Docent, and, from 1920, as an Ordinary Professor. At first I took courses in mathematical statistics. Then I abandoned them and turned to economics which I considered my main speciality, and in which I had been diligently working for many years preparing contributions that remained unfinished. Because, when the capitalist economics {in the Soviet Union} had been falling to the ground, and the outlines of a planned socialist economic regime began to take shape, the foundation for those problems that interested me as an economist and mathematician disappeared. The study of the economic processes under socialism, and especially of those taking place during the transitional period, demanded knowledge of another kind and other habits of reasoning, other methods as compared with those with which I had armed myself.

As a result, the issues of mathematical statistics began to interest me, and it seemed to me that, once I return to this field and focus all my power there, I would to a larger extent benefit my mother country and the cause of the socialist transformation of social relations. After accomplishing a few works which resulted from my groping for my own sphere of research, I concentrated on generalizing the stochastic methods to the statistical treatment of observations not being mutually independent in the sense of the theory of probability.

It seemed to me, that, along with theoretical investigations, I ought to study some concrete problems so as to check my methods and to find problems for theoretical work in a number of research institutes. For me, the methodical approach to problems and the attempts to prevent deviations from the formulated goal always were in the forefront. In applications, I consider as most fruitful my contributions, although not numerous, in the field of geophysics.

I have written this in December 1938, when compiling my biography on the occasion of my first entering the Steklov Mathematical Institute at the Academy of Sciences of the Soviet Union. I described in sufficient detail the story of my life and internal development up to the beginning of my work at Moscow State University and later events are sufficiently well outlined in my completed form. I shall only add, that, while working at the University, my main activity had been not teaching but work at the Mathematical Research Institute there. When the Government resolved that that institution should concentrate on pedagogic work ({monitoring} postgraduate studies) with research being mainly focussed at the Steklov Institute, my transfer to the latter became a natural consequence of that reorganization.

Notes

1. {It had been extremely dangerous to maintain ties with foreigners, and even with relatives living abroad, hence this lengthy explanation. A related point is that Slutsky passed over in silence his work at the Conjuncture Institute, an institution totally compromised by the savage persecution of its staff.}

2. {Vannovsky as well as Bogolepov mentioned in the same connection by Chetverikov in his essay on Slutsky (also translated here) are entered in the third edition of *Большая Советская Энциклопедия*, vols 4 and 3 respectively, whose English edition is called *Great Soviet Encyclopedia.* It is not easy, nor is it important, to specify which of them was actually responsible for expelling the students.}

3. {This unpublished composition is kept at the Vernadsky Library, Ukrainian Academy of Sciences.}

6b. E.E. Slutsky. [Later] Autobiography

I was born on 7(19) April 1880 in the village Novoe of the former Mologsky District, Yaroslavl Province, to a family of an instructor of a teacher's seminary. After graduating in 1899 from a classical gymnasium in Zhitomir with a gold medal, I entered the Mathematical Department of the Physical and Mathematical Faculty at Kiev University. I was several times expelled for participating in the student movement and therefore only graduated in 1911, from the Law Faculty. Was awarded a gold medal for my composition on political economy, but, owing to my reputation of a *Red Student*, I was not left at the University for preparing myself for professorship. I passed my examinations in 1917 at Moscow University and became Master of Political Economy and Statistics.

I wrote my student composition for which I was awarded a gold medal from the viewpoint of a mathematician studying political economy and I continued working in this direction for many years. However, my intended {summary?} work remained unfinished since I lost interest in its essence (mathematical justification of economics) after the very subject of study (an economic system based on private property and competition) disappeared in our country with the revolution. My main findings were published in three contributions ([6; 21; 24] in the appended list {not available}). The first of these was only noticed 20 years later and it generated a series of Anglo-American works adjoining and furthering its results.

I became interested in mathematical statistics, and, more precisely, in its then new direction headed by Karl Pearson, in 1911, at the same time as in economics. The result of my studies was my book *Теория корреляции* (Theory of Correlation), 1912, the first systematic explication of the new theories in our country. It was greatly honored: Chuprov published a commendable review of it and academician Markov entered it in a very short bibliography to his *Исчисление вероятностей* (Calculus of Probability). The period during which I had been mostly engaged in political economy had lasted to ca. 1921 – 1922 and only after that I definitively passed on to mathematical statistics and theory of probability.

The first work [8] of this new period in which I was able to say something new was devoted to stochastic limits and asymptotes (1925). Issuing from it, I arrived at the notion of a random process which was later destined to play a large role. I obtained new results, which, as I thought, could have been applied for studying many phenomena in nature. Other contributions [22; 31; 32; 37], apart from those published in the *C.r. Acad. Sci. Paris* (for example, on the law of the sine limit), covering the years 1926 – 1934 also belong to this cycle. One of these [22] [1] includes a certain concept of a physical process generating random processes and recently served as a point of departure for the Scandinavian {Norwegian} mathematician Frisch and for Kolmogorov. Another one [37], in which I developed a vast mathematical apparatus for statistically studying empirical random processes, is waiting to be continued. Indeed, great mathematical difficulties are connected with such investigations. They demand calculations on a large scale which can only be accomplished by means of mechanical aids the time for whose creation is apparently not yet ripe.

However, an attempt should have been made, and it had embraced the next period of my work approximately covering the years 1930 – 1935 and thus partly overlapping the previous period. At that time, I had been working in various research institutions connected with meteorology and, in general, with geophysics, although I had already begun such work when being employed at the Central Statistical Directorate.

I consider this period as a definitive loss in the following sense. I aimed at developing and checking methods of studying random empirical processes among geophysical phenomena. This problem demanded several years of work during which the tools for the investigation, so to say, could have been created and examined by issuing from concrete studies. It is natural that many of the necessary months-long preparatory attempts could not have been practically useful by themselves. Understandably, in research institutes oriented towards practice the general conditions for such work became unfavorable. The projects were often suppressed after much work had been done but long before their conclusion. Only a small part of the accomplishment during those years ripened for publication. I have no heart for grumbling since the great goal of industrializing our country should have affected scientific work by demanding concrete findings necessary at once. However, I was apparently unable to show that my expected results would be sufficiently important in a rather near future. The aim that I formulated was thus postponed until some later years.

The next period of my work coincides {began} with my entering the research collective of the Mathematical Institute at Moscow State University and then {and was continued}, when mathematical research was reorganized, with my transfer to the Steklov Mathematical Institute under the Academy of Sciences of the Soviet Union. In the new surroundings, my plans, that consumed the previous years and were sketchily reported above, could have certainly met with full understanding. However, their realization demanded means exceeding any practical possibilities. I had therefore moved to purely mathematical investigations of random processes [43; 44]; very soon, however, an absolutely new for me problem of compiling tables of mathematical functions, necessary for the theory of probability when being applied in statistics, wholly absorbed my attention and activity.

Such tables do exist; in England, their compilation accompanied the entire life of Karl Pearson who during three decades published a number of monumental productions. Fisher's tables showed what can be attained on a lesser scale by far less work. Nevertheless, a number of problems in this field remained unsolved. The preparation of Soviet mathematical-statistical tables became topical and all other problems had to be sacrificed. The year 1940 – 1941 was successful. I was able to find a new solution of the problem of tabulating the incomplete gamma-function providing a more complete and, in principle, the definitive type of its tables. The use of American technology made it possible to accomplish the calculations during that time almost completely but the war made it impossible to carry them through.

I described all the most important events. Teaching had not played an essential part in my scientific life. I had been working for a long time, at first as a beginning instructor, then as professor at a higher academic institution having a purely practical economic bias, at the Kiev Commercial Institute, which under Soviet power was transformed into the Kiev Institute for National Economy. I had been teaching there from 1912 to 1926. The listeners' knowledge of mathematics was insufficient which demanded the preparation of elementary courses. I do not consider myself an especially bad teacher, but I had been more motivated while working as professor of theoretical economy since my scientific constructions conformed to the needs of my listeners. During a later period of my life the scientific degree of Doctor of Sciences, Physics & Mathematics, was conferred on me as an acknowledgment of the totality of my contributions and I was entrusted with the chair of theory of probability and mathematical statistics at Moscow State University. However, soon afterwards I convinced myself that that stage of life came to me too late, that I shall not experience the good fortune of having pupils. My transfer to the Steklov Mathematical Institute also created external conditions favorable for my total concentration on research, on the main business of my scientific life.

A chain of events, which followed the war tempest, took me to Uzbekistan. But it is too soon to write the pertinent chapter of my biography. I shall only say that I am really happy to have the possibility of continuing my work which is expected to last much more than a year and on which much efforts was already expended, – of continuing it also under absolutely new conditions on the hospitable land of Uzbekistan.

Note

1. Its new version [42] was prepared on the request of *Econometrica*.

7. N.S.Chetverikov. The Life and Scientific Work of Slutsky

In author's *Статистические исследования*
(Statistical Investigations. Coll. Papers). Moscow, 1975, pp. 261 – 281

Foreword by Translator

Evgeny Evgenievich Slutsky (1880 – 1948) was an outstanding economist, statistician and mathematician. Kolmogorov (1948, p. 69) stated that, in 1922 – 1925, he was "the first to draw a correct picture of the purely mathematical essence of probability theory" and that (p. 70) "the modern theory of stationary processes … originated from Slutsky's works" of 1927 – 1937 "coupled" with Khinchin's finding of 1934. Earlier Kolmogorov (1933) referred to Slutsky's papers [9] and [13] but did not mention the former in the text itself; curiously enough, the same happened even in the second Russian translation of 1974 of his classic.

Nevertheless, Slutsky is not sufficiently known in the West. In 1995,Von Plato, when studying *modern probability*, certainly mentioned him, but did not describe at all his achievements; see however Seneta (2001).

My contributions (1993; 1996) contain archival materials concerning Slutsky; my article (1999a; b) is devoted to him and also includes such materials and the latter is fuller in this respect.

The essay below complements other pertinent sources, notably Kolmogorov (1948). Regrettably, however, two negative circumstances should be mentioned. First, Chetverikov quoted/referred to unpublished sources without saying anything about their whereabouts. Second, Chetverikov's mastery of mathematics was not sufficient, – he himself said so before adducing a long passage from Smirnov (1948), – and I had to omit some of his descriptions.

As compared with the initial version of this essay, the second one lacks a few sentences; I have inserted them in square brackets. Then, being able to see the texts of Slutsky's autobiographies, I note that Chetverikov quoted them somewhat freely (although without at all corrupting the meaning of the pertinent passages).

A special point concerns terminology. Slutsky's term "pseudo-periodic function" also applied by Smirnov, see above, and retained in the English translation of Slutsky's paper [17], is now understood in another sense, see *Enc. of mathematics*, vols 1 – 10, 1988 – 1994. Chetverikov, moreover, applied a similar term, quasi-periodic function, in the same context. It is now understood differently and, in addition, does not coincide with "pseudo-periodic function" (Ibidem). Note that Seneta (2001) applies the adjective *spurious* rather than *pseudo*. Unlike Chetverikov and Kolmogorov, he also mentions Slutsky's discovery [13] that, if a sequence of random variables $\{\xi_i\}$ tends in probability to a random variable ξ, then $f(\xi_i)$, where f is a continuous function, tends in probability to $f(\xi)$.

* * *

[The sources for this paper were Slutsky's biography written by his wife (manuscript); Kolmogorov (1948) and Smirnov (1948); Slutsky's autobiographies the first of which he presented when joining the Steklov Mathematical Institute in 1939, and the second one which he compiled for submitting it to the Uzbek Academy of Sciences on 3 December 1942; Slutsky's note [27]; his letters to his wife and to me; and my personal recollections.]

An historical perspective and a long temporal distance are needed for narrating the life and the work of such a profound researcher as Evgeny Evgenievich Slutsky (7(19) March 1880 – 10 March 1948). Time, however, is measured by events rather than years; in this case, first and foremost, by the development of scientific ideas.

Only a little more than ten years have passed since E.E. had died, but the seeds of new ideas sown by him have germinated and even ripened for the first harvest, – I bear in mind the rapid development of the theory of random functions.{To repeat,}however, a comprehensive estimation of his total rich and diverse heritage will only become possible in the future.

The description of Slutsky's life presents many difficulties occasioned both by complications and contradictions of his lifetime and the complexity of his spiritual make-up: a mathematician, sociologist, painter and poet were combined in his person. In essence, his life may be divided into three stages: the periods of seeking his own way; of passion for economic issues; and the most fruitful stage of investigations in the theory of probability and theoretical statistics. The fourth period, when he went away into pure mathematics [1], had just begun and was cut short by his death.

E.E. grew up in the family of a teacher and educator of the Novinsk teachers' seminary (former Yaroslav province). His father was unable to get along with the Director who had not been averse to embezzlement of state property, and, after passing through prolonged ordeals, his family settled in Zhitomir. There E.E. had learned in a gymnasium which later on he was unable to recall without repugnance. His natural endowments enabled him to graduate with a gold medal. His exceptional mathematical abilities and the peculiar features of his thinking had been revealed already in school. Having been very quick to grasp the main idea of analytic geometry, he successfully mastered its elements all by himself and without any textbooks.

After graduating in 1899, he entered the physical and mathematical faculty of Kiev University. There, he was carried away by the political wave of the student movement, and already in 1901, for participating in an unauthorized gathering (*skhodka*), he was expelled, together with 183 students, and drafted under compulsion into the Army on the order of Bogolepov, the Minister of People's Education. Because of vigorous public protests coupled with disturbances at all higher academic institutions, that order was soon disaffirmed. Nevertheless, already next year, for participating in a demonstration against the Minister Senger, E.E. was again thrown out of the University, and this time banned from entering any other Russian higher institution.

Only fragmentary information about Slutsky's active political work at that time, including the performance abroad of tasks ordered by a revolutionary group, is extant, but even so it testifies to the resolve and oblivion of self with which he followed his calling as understood at the moment. Owing to financial support rendered by his grandmother, E.E. became able to enter the machine-building faculty of the Polytechnical High School in Munich. Being cut off from practical political activities, he turned to sociology and was naturally enthralled by its main field, political economy{economics} He had begun by studying the works of Ricardo, then Marx' *Kapital* and Lenin's *Development of Capitalism in Russia*, and turned to the classics of theoretical economy. Although technical sciences provided some possibilities for his inclination to mathematics to reveal itself, he felt a distaste for them. He mostly took advantage of the years of forced life abroad for deep studies of economic problems. At the end of 1904 E.E. organized in Munich a group for studying political economy and successfully supervised its activities.

After the revolutionary events of 1905{in Russia}he became able to return to his homeland. He abandoned technical sciences and again entered Kiev University, this time choosing the law faculty whose curriculum included political economy. His plans contemplating long years of studying theoretical economy with a mathematical bias have ripened.

Slutsky's mathematical mentality attracted him to the development of those economic theories where the application of mathematics promised tempting prospects. However, now also his scientific activities and learning went on with interruptions. The years 1905 and 1906 were almost completely lost {because of revolutionary events?}and in March 1908 he was expelled from the University for a year. As E.E. himself admitted, that disciplinary punishment followed after a "boyish escapade" resulting from his "impetuous disposition". Nevertheless, in 1911, being already 31 years old, he graduated from the law faculty with a gold medal awarded for his diploma thesis "Theory of marginal utility", a critical investigation in the field of political economy [2], but his firmly established reputation of being a "red student" prevented his further studies at the university.

These were the external events that took place during Slutsky's first stage of life. They should be supplemented by one more development, by his marriage, in November 1906, to Yulia Nikolaevna Volodkevich.

Before going on to his second stage, let us try to discuss what were the inner motives, the vital issues, the inclinations that had been driving E.E. at that time. Those years may be called the period when he had been searching his conscience. An indefatigable thinker was being born; a person who criticized everything coming from without, who avidly grabbed all the novelties on which he could test his own ripening thoughts. He looked for his own real path that would completely answer his natural possibilities and inclinations. [He withdrew from practical revolutionary activities because he had soon understood that the path of a revolutionary was alien for him: in dangerous situations he was unable to orient himself quickly enough, he had no visual memory and he lacked many more of what was necessary for a member of an underground organization.]

E.E. was attracted by creative scientific work and he examined himself in various directions, – in technology and economics, in logic and the theory of statistics. In any of these domains, however, he only became aware of his real power when becoming able to submit his subject of study to quantitative analysis and mathematical thought. In one of his letters he wrote:

The point is not that I dream of becoming a Marx or a Kant, of opening up a new epoch in science, etc. I want to be myself, to develop my natural abilities and achieve as much as <u>*is*</u> *possible for me. Am I not entitled to that?*

He aimed at finding his place in science that would be in keeping with his natural gifts. In 1904, he wrote:

A man must certainly be working {only} *in that field to which his individuality drives him. … He must be living only there, where he is able to manifest it more widely, more completely, and to create, i.e., to work independently and with loving care.*

The word "independently" was not chosen randomly; it illuminated his creative life-work. When taking up any issue, he always began by thinking out the initial concepts and propositions. He always went on in his own, special manner, and the ideas of other authors only interested him insofar as they could serve for criticisms. This originality of thought deepened as the years went by and gradually led Slutsky to those boundaries after which not only new ways of solving{known}problems are opening up, but new, never before contemplated issues leading the mind to yet unexplored spaces, are discovered.

The most remarkable feature of Slutsky's scientific work was the selfless passion with which he seeked the truth and which he himself, in a letter to his wife, compared with that of a hunter:

You are telling me of being afraid for my work, afraid of the abundance of my fantasy … Is it possible to work without risk? And is it worthwhile to undertake easy tasks possible for anyone? I am pleased with my work {published}*in Metron exactly because it contains fantasy. For two hundred years people have been beating about the bush and no-one ever noticed a simple fact, whereas I found there an entire field open for investigation … It is impossible to avoid wrong tracks. Discovery is akin to hunting. Sometimes you feel that the game is somewhere here; you poke about, look out, cast one thing aside, take up another thing, and finally you find something …*

My present work [17] [of 1925 – 1926 on pseudo-periodic waves created by the composition of purely random oscillations – N.C.] *is, however, absolutely reliable. I am not pursuing a chimera, this is absolutely clear now, but it does not lessen the excitement of the hunt. In any case, I found the game, found it after the hunt was over… I am afraid that the purely mathematical difficulties are so great as to be unsurmountable for me. But neither is this, after all, so bad.*

After graduating from the University, Slutsky plunged into the work of numerous scientific societies, and, at the same time, being compelled to earn money and wishing to pass on his views, knowledge, and achievements, into teaching. It seemed that he had left little time and strength for scientific work, but his creative initiative overcame every obstacle, and even during that difficult and troublesome period E.E. was able to publish his first, but nevertheless important investigations. Already the list of the scientific societies whose member he was, shows how diverse were his interests and how wide was the foundation then laid for future investigations. In 1909, still being a student, he was corresponding member of the Society of Economists at Kiev Commercial Institute; in 1911, he became full member, in 1911 – 1913, he was its secretary, and, in 1913 – 1915, member of its council. In 1912 E.E. was elected full member of the{Kiev?}Mathematical Society; later on he joined the Sociological Society at the Institute for Sociological Investigations in Kiev, and in 1915 became full member of the A.I. Chuprov [3] Society for Development of Social Sciences at Moscow University.

Owing to his *disreputable political reputation*, Slutsky's pedagogical work at once encountered many obstacles. In 1911 he was not allowed to sit for his Master's examinations at Kiev University [4] and in 1912 he was not approved as teacher. The same year his father-in-law, N.N. Volodkevich, an outstanding educationalist of his time, took him on as teacher of political economy and jurisprudence at the{commercial}school established and headed by himself, but the Ministry for Commerce and Industry did not approve him as a staff worker. Only Slutsky's trip to Petersburg and his personal ties made it possible for him to remain in that school and to be approved, in 1915, in his position. Yulia Nikolaevna taught natural sciences at the same school. The apartment of the young married couple was attached to the school building and it was there that his life became then "mostly tied to the desk and illuminated by the fire of creative life" (from his biography written by his wife).

Slutsky first became acquainted with theoretical statistics in 1911 – 1912 having been prompted by Leontovich's book (1909 – 1911). It is impossible to say that that source, whose author later became an eminent physiologist and neurohistologist, member of the Ukrainian Academy of Sciences, was distinguished by clearness or correct exposition of the compiled material. Nevertheless, it was there that the Russian reader had first been able to learn in some detail the stochastic ideas of Pearson and his collaborators, and there also a list of the pertinent literature was adduced. That was enough for arousing Slutsky's interest, and we can only be surprised at how quickly he was able to acquaint himself with the already then very complicated constructions of the English statisticians-biologists by reading the primary sources; at how deeply he penetrated the logical principles of correlation theory; and at how, by using his critical feelings, he singled out the most essential and, in addition, noticed the vulnerable spots in the Pearsonian notions.

It is almost a miracle that only a year later there appeared Slutsky's own book [1] devoted to the same issues, explicated with such clearness, such an understanding of both the mathematical and logical sides, that even today it is impossible to name a better Russian aid for becoming acquainted with the principles of the constructions of the British school of mathematical statistics. And less than in two years the *Journal of the Royal Statistical Society* carried Slutsky's paper [5] on the goodness of fit of the lines of regression criticizing the pertinent constructions of the English statisticians. A short review published in 1913 [3] shows how deeply E.E. was able even then to grasp such issues as Markov chains{a later term}and how ardently he defended Markov's scientific achievements against the mockery of ignoramuses.

During those years, economic issues had nevertheless remained in the forefront. Even as a student, E.E. decided not to restrict his attention there to purely theoretical constructions and contemplated a paper on the eight-hour working day. He buried himself in factory reports, established connections with mills, studied manufacturing and working conditions. Issuing from the collected data, he distributed the reported severe injuries in accord with the hours of the day and established their dependence on the degree of the workers' tiredness. Earnestly examining economic literature, he connected his studies with compilation of popular articles on political economy as well as with his extensive teaching activities which he carried out up to his move to Moscow in 1926.

E.E. remained in Volodkevich's school until 1918 although school teaching was difficult for him. In the spring of 1915 he became instructor at the Kiev Commercial Institute. There, he read courses in sampling investigations and mathematical statistics, and, after the interruption caused by the World War, both there and at the Ukrainian Cooperative Institute, the history of economic and socialist doctrines. In 1917 Slutsky began teaching the discipline most congenial to him at the time, – theoretical economy. After the October revolution he taught in many newly established academic institutions: an elementary course in political economy at the Cooperative Courses for the disabled; introduction to the logic of social sciences at People's University. The Commercial Institute remained, however, his main pedagogical place of work, and there he also read courses in theoretical economy and political economy (theory of value and distribution).

The listing above, even taken in itself, clearly shows that in those years Slutsky concentrated on issues of theoretical economy and, more specifically, on those that admitted a mathematical approach. After his diploma thesis and a small essay on Petty [4], E.E. published an investigation about "equilibrium in consumption" [6] that only much later elicited response and due appreciation in the Western literature [5]. Less important research appeared in Kiev and Moscow respectively [10; 11]. However, two considerable contributions to economics [14; 15] were still connected with Kiev.

By 1922 Slutsky had already abandoned theoretical economics and afterwards devoted all his efforts to statistics. He himself, in his autobiography, explained his decision in the following way:

When the capitalist economics had been falling to the ground, and the outlines of a planned socialist economic regime began to take shape, the foundation for those problems, that interested me as an economist and mathematician, disappeared [6].

This is a very typical admission: the decisive significance for E.E., when choosing a field for scientific work, was the possibility of applying his mathematical talent. His inclination was caused not as though by an artisan's joy of skilfully using his tools, – no, he was irrepressibly attracted to abstract thinking, be it mathematics, logic, theory of knowledge or his poetic creativity.

[2] We already know that E.E. began his investigations in the theory of statistics in 1911 – 1914, his first contributions having been the book on correlation theory [1] and a paper on the lines of regression [5]. In the beginning of September 1912, in Petersburg, where E.E. had come to plead for being approved as teacher, he became acquainted, and fruitfully discussed scientific and pedagogic issues with A.A. Chuprov, who highly appraised his book [7]. During 1915 – 1916 Slutsky's name regularly appeared in the *Statistichesky Vestnik*, a periodical issued by the statistical section of the A.I. Chuprov Society at Moscow University. There, he published thorough reviews [7] or short notes [8] directed against wrong interpretation of the methods of mathematical statistics.

In 1922, after an interval of many years, Slutsky returned to the theory of statistics. He examined the logical foundation of the theory of probability and the essence of the law of large numbers from the viewpoint of the theory of knowledge. In November 1922, at the section on theoretical statistics of the Third All-Russian Statistical Conference, he read a report of great scientific importance. It touched on the main epistemological issue of the theory of probability, was published [9] and then reprinted with insignificant changes. In 1925 he issued another important paper [12] introducing the new notions of stochastic limit and stochastic asymptote, applied them for providing a new interpretation of the Poisson law of large numbers and touched on the logical aspect of that issue by critically considering the

Cournot lemma as formulated by Chuprov [8]. Also in 1925, he published a fundamental contribution [13] where he defined and investigated the abovementioned notions, applied them for deriving necessary conditions for the law of large numbers, which he, in addition, generalized onto the multidimensional case. Later on this work became the basis of the theory of stochastic functions.

By 1926, Slutsky's life in Kiev became very complicated. He did not master Ukrainian, and a compulsory demand of the time, that all the lectures be read in that language, made his teaching at Kiev higher academic institutions impossible. After hesitating for a long time, and being invited by the Central Statistical Directorate, he decided to move to Moscow. However, soon upon his arrival there, he was attracted by some scientific investigations (the study of cycles in the economy of capitalist countries) made at the Conjuncture Institute of the Ministry of Finance. E.E. became an active participant of this research, and, as usual, surrendered himself to it with all his passion. Here also, a great creative success lay ahead for him. In March of that year he wrote to his wife:

> *I am head over heels in the new work, am carried away by it. I am almost definitively sure about being lucky to arrive at a rather considerable finding, to discover the secret of how are wavy oscillations originating by a source that, as it seems, had not been until now even suspected. Waves, known in physics, are engendered by forces of elasticity and rotatory movements, but this does not yet explain those wavy movements that are observed in social phenomena. I obtained waves by issuing from random oscillations independent one from another and having no periodicities when combining them in some definite way.*

The study of pseudo-periodic waves originating in series, whose terms are correlatively connected with each other, led Slutsky to a new important subject, to the errors of the coefficients of correlation between series of that type. In both his investigations, he applied the "method of models", of artificially reproducing series similar to those actually observed but formed in accord with some plan and therefore possessing a definite origin.

The five years from 1924 to 1928, in spite of all the troubles, anxieties and prolonged housing inconveniences caused by his move to Moscow, became a most fruitful period in Slutsky's life. During that time, he achieved three considerable aims: he developed the theory of stochastic limit (and asymptote); discovered pseudo-periodic waves; and investigated the errors of the coefficient of correlation between series consisting of terms connected with each other.

In 1928, E.E. participated at the Congress of Mathematicians in Bologna. The trip provided great moral satisfaction and was a grand reward deserved by sleepless nights and creative enthusiasm. His report on stochastic asymptotes and limits attracted everyone. A considerable debate flared up at the Congress between E.E. and the eminent Italian mathematician Cantelli concerning the priority to the strong law of large numbers. Slutsky [16] had stated that it was due to Borel but Cantelli considered himself its author. Castelnuovo, the famous theoretician of probability, and other Italian mathematicians rallied together with Cantelli against Slutsky, declared that Borel's book, to which E.E. had referred to, lacked anything of the sort attributed to him by the Russian mathematician, and demanded an immediate explanation from him. E.E. had to repulse numerous attacks launched by the Italians and to prove his case.

The point was that Slutsky, having been restricted by the narrow boundaries of a paper published in the *C.r. Acad. Sci. Paris*, had not expressed himself quite precisely. He indicated that Borel was the first to consider the problem and that Cantelli, Khinchin, Steinhaus and he himself studied it later on. However, he should have singled out Cantelli and stressed his scientific merit. Borel was indeed the first to consider the strong law, but he did it only in passing and connected it with another issue in which he was interested much more. Apparently for this reason Borel had not noticed the entire meaning and importance of that law, whereas Cantelli was the first to grasp all that and developed the issue, and his was the main merit of establishing the strong law of large numbers. E.E. was nevertheless able to win. Understandably, he did not at all wish to make use of his victory for offending Cantelli. He appreciated the Italian mathematician; here is a passage from his letter to his wife (Bologna, 6 September 1928) [9]:

[He is] *not a bad man at all, very knowledgeable, wonderfully acquainted with Chebyshev, trying to learn everything possible about the Russian school* (*only one thing I cannot forgive, that he does not esteem Chuprov*). *In truth, he has brought fame to the Russian name in Italy, because he doesn't steal but honestly says: that is from there, that is Russian, and that is Russian … Clearly one must let him keep his pride*.

After a prolonged discussion of the aroused discord with Cantelli himself, and a thorough check of the primary sources, E.E. submitted an explanation to the Congress, agreed beforehand with Cantelli. The explanation confirmed his rightness but at the same time had not hurted Cantelli's self-respect. After it was read out, Cantelli, in a short speech, largely concurred with E.E. This episode vividly characterizes Slutsky, – his thorough examination of the problems under investigation, an attentive and deep study of other authors, and a cordial and tactful attitude to fellow-scientists. He was therefore able not only to win his debate with Cantelli, but to convince his opponent as well.

In 1930, the Conjuncture Institute ceased to exist, the Central Statistical Directorate was fundamentally reorganized, and Slutsky passed over to institutions connected with geophysics and meteorology where he hoped to apply his discoveries in the field of pseudo-periodic waves. However, he did not find conditions conducive to the necessary several years of theoretical investigations at the Central Institute for Experimental Hydrology and Meteorology. [These lines smack of considerable sadness but they do not at all mean that Slutsky surrendered.] In an essay [27] he listed his accomplished and intended works important for geophysics. He also explicated his related findings touching on the problem of periodicity, and indicated his investigation of periodograms, partly prepared for publication [26] Slutsky then listed his notes in the *C.r. Acad. Sci. Paris* [16; 18; 20 – 23] where he developed his notions as published in his previous main work of 1925 [13].

To the beginning of the 1930s belong Slutsky's investigations on the probable errors of means, mean square deviations and coefficients of correlation calculated for interconnected stationary series. He linked those magnitudes with the coefficients of the expansion of an empirical series into a sum of (Fourier) series of trigonometrical functions and thus opened up the way of applying those probable errors in practice.

Slutsky himself summarized his latest works in his report at the First All-Union Congress of Mathematicians in 1929 but only published (in a supplemented way) seven year later [30]. Owing to the great difficulties of calculation demanded by direct investigations of the interconnected series, Slutsky developed methods able to serve as an ersatz of sorts and called by a generic name "statistical experiment". Specifically, when we desire to check the existence of a connection between two such series, we intentionally compare them in such a way which prevents a real connection; after repeating such certainly random comparisons many times, we determine how often parallelisms have appeared in the sequences of the terms of both series. They, the parallelisms, create an external similarity of connection not worse than the coincidences observed by a comparison of the initial series. Slutsky developed many versions of that method and applied it to many real geophysical investigations of wavy oscillating series.

E.E. did not belong to those statisticians-mathematicians for whom pure mathematics overshadowed the essence of studied phenomena. He thought that the subject of a methodological work should be determined by its material substance.

It seemed to me that, along with theoretical investigations, I ought to study some concrete problems so as to check my *methods and to find the problems for theoretical work,*

he wrote in his autobiography submitted in 1939. Bearing in mind such aims, he studied the series of harvests in Russia over 115 years (compiled by V.G. Mikhailovsky), those of the cost of wheat over 369 years (Beveridge), series of economic cycles (Mitchell), etc. Passing on from economic to geophysical series, Slutsky then examined the periodicity of sunspots checking it against data on polar aurora for about two thousand years (Fritz [10]) and studied the peculiar vast information *stored* as annual rings of the giant sequoia of Arizona (mean data for eleven trees covering about two thousand years) [11].

And yet fate directed Slutsky to the domain of pure mathematics. In 1934 he passed on to the Mathematical Institute of Moscow University and in 1935 abandoned geophysics. In 1939 he established himself at the Steklov Mathematical Institute of the Soviet Academy of Sciences. At the same time, having been awarded by Moscow University the academic status of Doctor of Mathematical and Physical Sciences *honoris causa* on the strength of his writings, and entrusted by the chair of mathematical statistics there, Slutsky apparently resumed the long ago forsaken teaching. [However, because of the situation that took shape at the University in those years,] teaching demanded more strength than he{still}possessed at that time. As he himself wrote,

Having been entrusted with the chair of the theory of probability{!}*and mathematical statistics at Moscow University, I have convinced myself soon afterwards, that that stage of life came too late, and I shall not experience the good fortune of having pupils.*

It seemed that, having consolidated his position at the Mathematical Institute, E.E. will be able to extend there his work on the theory of statistics. But his plans were too extensive, they demanded the establishment of a large laboratory, and, therefore, large expenses. That proved impossible, and Slutsky had to concentrate on investigations in the theory of stochastic processes and to plunge ever deeper into pure mathematics.

At the end of October 1941, when Moscow was partly evacuated, Slutsky moved with his family to Tashkent. A part of his{unpublished}works was lost. And still he [considered the year 1940/1941 as *lucky* and]wrote about that period:

I was able to find a new solution of the problem of tabulating the incomplete Γ*-function providing a more complete, and, in principle, the definitive type of its table. The use of American technology allowed to accomplish the calculations almost completely in one year. But the war made it impossible to carry them through.*

The work had indeed dragged on, and even after his return to Moscow three more years were required for their completion. I cannot fail to mention the selfless help rendered by N.V. Levi, a woman employee of the Mathematical Institute, who accomplished that task when Slutsky had already begun feeling himself ill. He developed lung cancer, and it was a long time before the disease was diagnosed although E.E. himself never got to know its nature. He continued to work on the Introduction to the tables where he was explaining the method of their compilation, but it was N.V. Smirnov who wrote the definitive text. On 9 March 1948 Slutsky was still outlining the last strokes of the Introduction, but next day he passed away.

[3] Already in Kiev Slutsky had been deeply interested in the cognitive and logical side of the problems that he studied, especially concerning his investigations in mathematical statistics. His first independent essential writings were indeed devoted to these general issues. Later on, he essentially cooled down for them; he either solved them to a required by him degree, or his great success in more concrete investigations overshadowed philosophical problems. In any case, in the middle of the 1940s, E.E. even with some irritation refused to discuss purely logical concepts although he had been unable to disregard the then topical criticism levelled by Fisher against the problem of calculating the probabilities of hypotheses (of the Bayes theorem).

First of all Slutsky took it upon himself to ascertain the relations of the theory of probability to statistical methodology. To this aim, he singled out the formal mathematical essence of the theory itself by expelling from it all that, introduced by the philosophical interpretation of the concept of *probability*. So as to achieve complete clearness, he proposed to abandon the habitual terms and to make use of new ones: *disjunctive calculus, valency* (*assigned* to events), etc. To assign, as he stated, meant to establish some relation *R* between an event and its valency in accord with only one rule: if event *A* breaks down into a number of alternatives, the sum of all of their valencies must be equal to the valency of *A*. The valency of the joint event *AB*, that is, of the occurrence of the events *A* and *B*, was determined just as formally. These relations between valencies were included in the axiomatics of the *disjunctive calculus*, sufficient for developing it as a mathematical discipline. Its applications depended on the contents which we might introduce into the term *valency* and which can be probability, frequency, or, as a special notion, *limiting frequency*. To what extent will these *interpreted calculuses* coincide and cover each other, depends on the contents of their axiomatics, which, under differing interpretations, can be distinct one from another. However, these distinctions cannot concern the purely mathematical discipline, the *disjunctive calculus*, because its axiomatics is constructed independently of the interpretation of the subject of *valency* [9].

When explaining his understanding of the logic of the law of large numbers, Slutsky issued from those considerations, and he also made use of the notions of stochastic asymptote and stochastic limit.{Chetverikov describes here Slutsky's paper [13]: I advise readers to look up that contribution itself.}

Slutsky also criticized the purely logical Chuprov – Cournot{Cournot – Chuprov}construction that aimed at connecting probabilities with frequencies of the occurrence of phenomena in the real world, at throwing a "bridge" between them. He thought that the essence of the so-called Cournot lemma consisted in attaching to the law of large numbers the importance of a law of nature without any qualifying remarks about the probability of possible, although extremely rare exceptions. The notion of probability cannot be removed from the Cournot lemma, so, as he concluded, the logical value of the "bridge" itself is lost [12] [12].

Having been especially prompted by the need to work with time series and issuing from the concept of stochastic limit (asymptote), E.E. also constructed a theory of random functions.{A description of Slutsky's pertinent findings follows.}

An important discovery made by Slutsky in the mid-1920s consisted in that he connected wavy oscillations with random oscillations and showed how the latter can engender the former […] Wavy oscillations are extremely common (for example, in series occurring in economics and meteorology), whereas unconnected randomly oscillating series are met with not so often. A practically important problem is, therefore, to derive the errors of the various general characteristics, – of the mean, the standard deviation, the correlation coefficient, – for connected series [13].

E.E. devoted much effort to the solution of that problem. His formulas are bulky, see for example the expression for the error of the correlation coefficient [24, p. 75]. Simpler formulas for particular cases are in [27]. Later Slutsky examined the possibility of applying the χ^2 test and its distribution to connected series as well as of determining the required magnitudes through the Fourier coefficients [25; 26].

By issuing from his theory of connected series, and allowing for the course of random processes, Slutsky was able to provide a methodology of forecasting them, including sufficiently long-term forecasting, with given boundaries of error [29].

We ought to dwell especially on his *method of models* (of *statistical experimentation*) for discovering connections between phenomena. His idea was as follows. When studying many problems not yet completely solved by theory, it is possible to arrange a statistical "experiment" and thus to decide whether the statistical correspondence between phenomena is random or not. For example, when selecting a number of best and worst harvests in Russia from among the series collected by Mikhailovsky for 115 years, we can compare them with the series of maximums and minimums of the number of sunspots for more than 300 years. If such comparisons are{if the correspondence is} only possible after shifting one of the series with respect to the other one, then, obviously, the coincidences will be random. However, since the sum of the squares of the discrepancies **14** is minimal when those series are compared without such shifting, we may be sufficiently convinced in that the coincidences are not random [28].

I am unable to appraise Slutsky's purely mathematical studies and am therefore quoting most eminent Soviet mathematicians. Smirnov (1948, pp. 418 – 419), after mentioning Slutsky's investigation [13], wrote:

The next stage in the same direction was his works on the theory of continuous stochastic processes or random functions. One of Slutsky's very important and effective findings here was the proof that any random stochastically continuous function on a segment is stochastically equivalent to a measurable function of an order not higher than the second Baire class. He also derived simple sufficient conditions for a stochastic equivalence of a random function and a continuous function on a segment, conditions for the differentiability of the latter, etc. These works undoubtedly occupy an honorable place among the investigations connected with the development of one of the most topical issues of the contemporary theory of probability, that{issue or theory?}*owes its origin to Slutsky's scientific initiative.*

The next cycle of Slutsky's works (*1926 – 1927*) *was devoted to the examination of random stationary series, and they served as a point of departure for numerous and fruitful investigations in this important field. Issuing from a simplest model of a series obtained by a multiple moving summation of an unconnected series, he got a class of stationary series having pseudo-periodic properties imitating, over intervals of any large length, series obtained by superposing periodic functions. His finding was a sensation of sorts; it demanded a critical revision of the various attempts of statistical justification of periodic regularities in geophysics, meteorology, etc. It occurred that the hypothesis of superposition of a finite number of regularly periodic oscillations was statistically undistinguishable from that of a random function with a very large zone of connectedness.*

His remarkable work on stationary processes with a discrete spectrum was a still deeper penetration into the structure of random functions. In this case, the correlation function will be almost periodical. Slutsky's main result consisted here in that a random function was also almost periodic, belonged to a certain type and was almost everywhere determined by its Fourier series.

These surprisingly new and fearlessly intended investigations, far from exhausting a very difficult and profound problem, nevertheless represent a prominent finding of our science. With respect to methodology and style, they closely adjoin the probability-theoretic concepts of the Moscow school (*Kolmogorov, Khinchin*), *that, historically speaking, originated on a* *different foundation. The difficult to achieve combination of acuteness and wide theoretical reasoning with a quite clearly perceived concrete direction of the final results, of the final aim of the investigation, is Slutsky's typical feature.*

Proving that Slutsky's works were close to those of the Moscow school, Kolmogorov (1948, p. 70) stated:

In 1934, Khinchin showed that a generalized technique of harmonic analysis was applicable to the most general stationary processes considered in Slutsky's work [...] *The modern theory of stationary processes, which most fully explains the essence of continuous physical spectra, has indeed originated from Slutsky's works, coupled with this result of Khinchin.*

After E.E.'s interest in applications had shifted from economics to geophysics, it was quite natural for him to pass from considering connected series of random variables to random functions of continuous time. The peculiar relations, that exist between the different kinds of continuity, differentiability and integrability of such functions, make up a large area of the modern theory of probability whose construction is basically due to Slutsky [19; 20; 26; 30 – 33] [15]. *Among the difficult results obtained, which are also interesting from the purely mathematical viewpoint, two theorems should be especially noted. According to these, a 'stochastically continuous' random function can be realized in the space of measurable functions* [31; 33]; *and a stationary random function with a discrete spectrum is almost periodic in the Besikovitch sense with probability* 1 [32].

Kolmogorov then mentions the subtle mastery of Slutsky's work on the tables of incomplete Γ- and B-functions that led him to the formulation of general problems. The issue consisted in developing a method of their interpolation, simpler than those usually applied, but ensuring the calculation of the values of these functions for intermediate values of their arguments with a stipulated precision. For E.E., this, apparently purely "technical", problem became a subject of an independent scientific investigation on which he had been so enthusiastically working in his last years. He was able, as I indicated above, to discover a new solution of calculating the incomplete Γ-function, but that successful finish coincided with his tragic death.

Notes

1.{Chetverikov thus separated the theory of probability from pure mathematics.}

2. {Still extant at the Vernadsky Library, Ukrainian Academy of Sciences, Fond 1, No. 44850 (Chipman 2004, p. 355.}

3.{A.I. Chuprov, father of the better known A.A. Chuprov.}

4. He only held them in 1918, after the revolution, at Moscow University.

5. Slutsky made the following marginal note on a reprint of Schults (1935): "This is a supplement to my work that began influencing{economists}only 20 years after having been published".

6.{This explanation would have only been sufficient if written before 1926. Below, Chetverikov described Slutsky's work in theoretical economics during 1926 – 1930 at the Conjuncture Institute and then implicitly noted that in 1930 the situation in Soviet statistics had drastically worsened. I (2004) stated that Slutsky had abandoned economics largely because of the last-mentioned fact. On the fate of the Conjuncture Institute see also Sheynin (1996, pp. 29 – 30). Kondratiev, its Director, who was elbowed out of science, persecuted, and shot in 1938 (Ibidem), had studied cycles in the development of capitalist economies. In at least one of his papers, he (1926) had acknowledged the assistance of Chetverikov and

Slutsky, a fact that Chetverikov naturally had to pass over in silence. Three papers devoted to Kondratiev are in *Ekonomika i Matematich. Metody*, vol. 28, No. 2, 1992.}

7.{I (1996, p. 44) reprinted Chuprov's review originally published in a newspaper. I also made public Slutsky's relevant letters to Markov and Chuprov and Slutsky's scientific character compiled by Chuprov (pp. 44 – 50). Slutsky's correspondence with Chuprov discussed, among other issues, the former's encounter with Pearson. Three letters from Slutsky to Pearson dated 1912 are now available (Sheynin 2004, pp. 227 – 235.}

Chuprov was six years older than Slutsky, had much more teaching experience, and was the generally accepted head of the{Russsian}statistical school. In the{Petersburg}Polytechnical Institute, he laid the foundation of teaching the theory of statistics.

8.{Chetverikov repeated the mistake made by Chuprov (1909, pp. 166 – 168). The latter stated that Cournot had provided a "canonical" proof of the law of large numbers. In actual fact, Cournot did not even formulate that law (and did not use that term), and his "Lemma" (a term only used by Chuprov himself) had simply indicated (after Dalembert!) that rare events did not happen (Cournot 1843, §43). Chuprov, however, interpreted that statement as "did not happen often". Chetverikov was translator of Cournot (Moscow, 1970). Note that Slutsky [12, p. 33] followed Chuprov.}

9.{The translation of the passage below is due to Seneta (1992, p. 30) who published the letter (as well as another relevant one from Slutsky to his wife) in full. In 1970 Chetverikov had given me copies of these letters and about 1990 I sent them to Seneta. Seneta acknowledged my help in obtaining "important materials" but, being concerned that I could have problems with the Soviet authorities, did not elaborate. I (1993) explained all that and provided additional material concerning Chuprov, Slutsky and Chetverikov.}

10.{Hermann Fritz (1830 – 1893), see the appropriate volume of Poggendorff's *Handwörterbuch.*}

11. Slutsky's large work on those annual rings including all the pertinent calculations got lost during his evacuation from Moscow.

12.{Chuprov and Slutsky formulated the "Cournot lemma" not as Cournot himself did, see Note 7.}

13. These errors are usually many times greater than the respective errors in unconnected series.

14.{A loose but understandable description.}

15.{I changed the numbering, here and below, to conform to that in the present paper.}

References

E.E. Slutsky

Kolmogorov (1948) adduced a complete list of Slutsky's contributions compiled by Chetverikov. For this reason, I only provide references to those writings, which the latter mentioned in his text.

Abbreviation: *GIIA* = *Giorn. dell'Istituto Italiano degli Attuari*

1. *Теория корреляции и элементы учения о кривых распределения* (Theory of Correlation and the Elements of the Theory of Distribution Curves). Kiev, 1912.
2. Essence and form of cooperatives. *Справочный календарь земледельца* (Farmers' Reference Calendar) for 1913. Kiev, 1912, pp. 1 – 15.
3. Review of Markov (1913). Newspaper *Киевская мысль* (Kiev Reflexions), 30 March 1930, p. 5.
4. *Сэр Вильям Петти* (Sir William Petty). Kiev, 1914.

5. On the criterion of goodness of fit of the regression lines and on the best method of fitting them to the data. *J. Roy. Stat. Soc.*, vol. 77, 1914, pp. 78 – 84.
6. Sulla teoria del bilancio del consumatore. *Giorn. del Economisti,* vol. 51, 1915, pp. 1 – 26. Translation: On the theory of the budget of the customer. In: *Readings in Price Theory*. Editors, G.J. Stigler, K.E. Boulding. Homewood, Ill., 1952, pp. 27 – 56.
7. Statistics and mathematics, this being a review of Kaufman (1912). *Statistich. Vestnik* No. 3/4, 1915 – 1916, pp. 1 – 17. (R)
8. On an error in the application of a correlation formula. Ibidem, pp. 18 – 19. (R)
9. On the logical foundations of the theory of probability. Read 1922. Translated. in this collection.
10. Calculation of state revenue from the issue of paper money. An appendix to an article of another author. *Mestnoe Khoziastvo* (Kiev) No. 2, Nov. 1923, pp. 39 – 62. (R)
11. Mathematical notes to the theory of the issue of paper money. *Ekonomich. Bull. Koniunkturn. Inst.* No. 11/12, 1923, pp. 53 – 60. (R)
12. On the law of large numbers. *Vestnik Statistiki* No. 7/9, 1925, pp. 1 – 55. (R)
13. Über stochastische Asymptoten und Grenzwerte. *Metron*, t. 5, No. 3, 1925, pp. 3 – 89.
14. Ein Beitrag zur formal-praxeologischen Grundlegung der Ökonomik. *Acad. Oukrainienne Sci.*, Annales Cl. Sci. Soc.-Écon., t. 4, 1926, pp. 3 – 12. Also in Ukrainian in same year. Translation: An enquiry into the formal praxeological foundations of economics. *Structural Change and Econ. Dynamics*, vol. 15, 2004, pp. 371 – 380.
15. Zur Kritik des Böhm-Bawerkschen Wertbegriffs und seiner Lehre von der Messbarkeit des Wertes. *Schmollers Jahrb. Gesetzgeb., Verwalt., Volkswirt. im Dtsch. Reich*, Bd. 51, No. 4, 1927, pp. 37 – 52. Translation: A critique of Böhm-Bawerk's concept of value and his theory of the measurability of value. Ibidem, pp. 357 – 369.
16. Sur un théorème limite rélatif aux séries des quantités éventuelles. *C.r. Acad. Sci. Paris,* t. 185, 1927, pp. 169 – 171.
17. Revised translation of Russian paper of 1927: Summation of random causes as the source of cyclic processes. *Econometrica*, vol. 5, 1937, pp. 105 – 146.
18. Sur un critérium de la convergence stochastique des ensembles de valeurs éventuelles. Ibidem, t. 187, 1928, pp. 370 – 372.
19. Sur les fonctions éventuelles compactes. *Atti del Congresso Intern. dei Matematici* (Bologna, 1928), t. 6. Bologna, 1932, pp. 111 – 115.
20. Sur les fonctions éventuelles continues, intégrables et dérivables dans les sens stochastiques. *C.r. Acad. Sci. Paris*, t. 187, 1928, pp. 878 – 880.
21. Quelques propositions sur les limites stochastiques éventuelles. Ibidem, t. 189, 1929, pp. 384 – 386.
22. Sur l'erreur quadratique moyenne du coefficient de corrélation dans le cas des suites des épreuves non indépendantes. Ibidem, pp. 612 – 614.
23. Sur l'extension de la théorie de périodogrammes aux suites de quantités dépendantes. Ibidem, pp. 722 – 725.
24. On the standard error of the correlation coefficient in the case of homogeneous coherent chance series. *Trudy Koniunkturn. Inst.*, vol. 2, 1929, pp. 64 – 101. (R) English summary on pp. 150 – 154.
25. On the distribution of the error of the correlation coefficient in homogeneous coherent chance series. *Zhurnal Geofiziki*, vol. 2, 1932, pp. 66 – 98. (R)
26. Alcune applicazioni dei coefficienti di Fourier all'analisi funzioni aleatorie stazionarie. *GIIA*, t. 5, No. 4, 1934, pp. 3 – 50.
27. *Десять лет работы для статистической геофизики* (1924 – 1933). (Ten Years of Research in Statistical Geophysics …). Published by Central Inst. Experimental Hydrology and Meteorology as a mimeographed manuscript. Moscow, 1934. (R)

28. Statistical experiment as a method of research. Critical essays on the Earth – Sun problem. *Zhurnal Geofiziki*, vol. 5, 1935, pp. 18 – 38. (R)
29. On extrapolation in connection with the problem of forecasting. Ibidem, pp. 263 – 279. (R)
30. On connected random functions of one independent variable. *Труды первого всесоюзного съезда математиков* (Proc. First All-Union Congress Mathematicians), Kharkov, 1929. M. – L., 1936, pp. 347 – 357. (R)
31. Qualche proposizione relativa alla teoria delle funzione aleatorie. *GIIA*, t. 8, No. 2, 1937, pp. 3 – 19.
32. Sur les fonctions aléatoires presque périodiques et sur la décomposition des fonctions aléatoires stationnaires en composantes. In *Colloque consacré à la théorie des probabilités*, 5ième pt. (Actualités scientifiques et industrielles No. 738). Eds, S. Bernstein, E. Slutsky, H. Steinhaus. Paris, 1938, pp. 33 – 55.
33. Some propositions on the theory of random functions. *Труды Среднеазиатск. Гос. Унив.* (Proc. Sredneasiatsk. State Univ.), ser. Math., No. 31. Tashkent, 1939, pp. 3 – 15. (R) Several issues collected, with separate paging, in *Сборник, посвященный 30-летию деятельности В.И. Романовского* (Collection Honoring 30 Years of Romanovsky's Work).
34. *Избранные труды. Теория вероятностей, математическая статистика* (Sel. Works in Theory of Probability and Math. Statistics). M., 1960. Includes, among others, reprints or translations of [5; 9; 18 – 24, 26; 30; 32; 33] and a biography compiled by B.V. Gnedenko.

Other Authors

Chipman, J.S. (2004), Slutsky's praxeology and his critique of Böhm-Bawerk. *Structural Change and Econ. Dynamics*, vol. 15, pp. 345 – 356.
Chuprov, A.A. (1909), *Очерки по теории статистики* (Essays on the Theory of Statistics). References in text from third (last) edition: M., 1959.
Cournot, A.A. (1843), *Exposition de la théorie des chances et des probabilités.* Paris, 1984.
Kaufman, A.A. (1916), *Теория и методы статистики* (Theory and methods of statistics). M. First published in 1911. Several later editions and German translation of 1913.
Kolmogorov, A.N. (1933), *Grundbegriffe der Wahrscheinlichkeitsrechnung.* Berlin.
--- (1948, in Russian), Obituary. E.E. Slutsky. *Math. Scientist*, vol. 27, 2002, pp. 67 – 74.
Kondratiev, N.D. (1926), The problem of forecasting. *Ekonomika i Matematich. Metody*, vol. 24, 1988, pp. 245 – 268. (R)
Leontovich, A. (1909 – 1911), *Элементарное пособие к применению методов Гаусса и Пирсона* (Elementary Textbook on the Use of the Gauss and Pearson Methods), pts 1 – 3.Kiev. (R)
Markov, A.A. (1913), Essai d'une recherche statistique sur le texte du roman «Eugène Onegin». *Izvestia Imp. Akad. Nauk*, ser. 6, t. 7, No. 3, pp. 153 – 162. Additional French title. (R)
--- (1924), *Исчисление вероятностей* (Calculus of Probability). M. Previous editions: 1900, 1908, 1913. German translation: 1912.
Schults, H. (1935), Interrelation of demand, price and income. *J. Polit. Econ.*, vol. 43.
Seneta, E. (1992), On the history of the strong law of large numbers and Boole's inequality. *Hist. Math.*, vol. 19, pp. 24 – 39.
--- (2001), Slutsky. In Heyde, C.C., Seneta, E., Editors, *Statisticians of the Centuries.* New York, pp. 343 – 345.
Sheynin, O. (1993), Chuprov, Slutsky and Chetverikov: some comments. *Hist. Math.*, vol. 20, pp. 247 – 254.
--- (1990, in Russian), *Chuprov.* Göttingen, 1996.

--- (1999, in Russian), Slutsky: commemorating the 50th anniversary of his death. IMI, vol. 38, pp. 128 – 137. Translated with additional material in author's *Russian papers on the History of probability and Statistics*. Berlin, 2004, pp. 222 – 240.
Smirnov, N.V. (1948), Slutsky. *Izvestia AN*, ser. Math., vol. 12, pp. 417 – 420. (R)

8. Romanovsky. V.I. His reviews of R.A. Fisher; official attitude towards him; his obituary

Vsevolod Ivanovich Romanovsky (1879 – 1954) was an outstanding mathematician and statistician. He also originated statistical studies in Tashkent and might certainly be also remembered as an educationalist. Bogoliubov & Matvievskaia (1997) described his life and work but have not dwelt sufficiently on his ties with Western scientists (or on the ensuing criticism doubtless arranged from above), and this is what I am dealing with by translating his reviews of Fisher's books, pertinent materials of a Soviet statistical conference of 1948[1] (fragments of one report and the resolution) and a publisher's preface to the Russian translation of Fisher's *Statistical Methods*… My manuscript "Romanovsky's correspondence with Pearson and Fisher" intended for the forthcoming volume commemorating A.P. Youshkevich, will make public eight of his letters to Pearson (1924 – 1925) and 23 letters exchanged in 1929 – 1938 between him and Fisher [2].

Some of Romanovsky's writings (1924; 1934) not translated below also bear on my subject. In the first of them, he (pt. 1, p. 12) called Pearson "our celebrated contemporary", and, in the second part, discussed the latter's work. Lenin had severely criticized Pearson's ideology (Sheynin 1998, p. 530), and, beginning with ca. 1926, Soviet statisticians had been rejecting Pearson's work out of hand [3]. In the second writing, Romanovsky discussed the work of Fisher. He argued that it should be developed and propagandized (p. 83), that Fisher's *Statistical Methods* …was almost exclusively "prescriptive", which "distressed" those readers, who wished to study the described issues deeper, but that that book ought to be translated into Russian (p. 84). He also noted that Fisher's methods were unknown to Russian statisticians and put forward recommendations concerning the teaching of statistics. In particular, Romanovsky (p. 86) advised the introduction of optional studies of statistical applications to genetics; cf. the Resolution of the Soviet conference below!

Notes

1. The criticism levelled there against Romanovsky was comparatively mild, no doubt because the participants simply obeyed ideological orders. Nevertheless, it led to a further attack against him, see my introductory remarks to the translation of Sarymsakov (1955) below. A highly ranked geodesist noted mistakes in Romanovsky's treatise on the theory of errors and supported his remarks by stupid ideological accusations. He even alleged that Romanovsky's expression "probability … is described by the law …" was unacceptable because Marx had declared that the world needed change rather than description.

2. In one of his letters to Fisher written in Paris in 1929, Romanovsky called the Soviet political police "the most dreadful and mighty organization".

3. Even Fisher became suspect. Here is an editorial note to Romanovsky's paper (1927, p. 224): "The editorial staff does not share either the main suppositions of Fisher, who belongs to the Anglo – American empiricists' school, or Romanovsky's attitude to Fisher's constructions …"

8a.V.I. Romanovsky.
Review of R.A. Fisher "Statistical Methods for Research Workers". London, 1934

Sozialistich. Rekonstruksia i Nauka, No. 9, 1935, pp. 123 – 127

One of the most typical features in the history of mathematical statistics during the last 20 – 25 years is the rapid development of the theory of small samples, rich in important achievements. Its main problem is to enable to make as justified as possible inferences about phenomena or properties, for whose complete and adequate description very, if not infinitely many observations are needed, by issuing from a small number of statistical experiments or observations. Such problems are most often encountered in natural sciences and their applications, – in biology, agronomy, selection, etc. There, it is frequently required to obtain as reliable as possible conclusions about a phenomenon that can only be studied statistically, and, therefore, from the viewpoint of classical statistics, that ought to be observed many times. However, the real situation, for example when investigating the action of various fertilizers or the properties of several varieties of a cultivated plant, does not allow to make many observations.

The development of the theory of small samples is mainly due to a most prominent scientist, the English mathematician and statistician Ronald Aylmer Fisher, Professor of eugenics at London University. For a long time (for 15 years), until becoming, in 1933, the Galton chairperson, he was head of the statistical section of the oldest English agronomical station at Rothamsted. It was there that his practical work led to theoretical and practical achievements in mathematical statistics, which made him known the world over and ensured his election to the Royal Society, and which are almost completely reflected in the book under review. From among 82 of Fisher's contributions published until now, 40 are devoted to theoretical research in mathematical statistics and 42 deal with its applications to the theory of Mendelian inheritance, agronomy, methodology of field experimentation, soil science, meteorology, theory of evolution, etc.

The first edition of Fisher's book appeared in 1925 and the fifth in 1934 which testifies to its exceptional success among researchers in biology, agronomy, etc. The book was indeed mainly intended for them since it cannot at all be a theoretical guide to statistics, to say nothing about beginners or those inadequately versed in that discipline. The reader will not find a coherent theoretical exposition, proofs or derivations of the theorems applied there. On the other hand, he will see many very interesting and topical problems, mostly biological; exact indications about, and precise interpretation of their solution; and auxiliary tables whose compilation is not explained but whose usage and possible applications are described in detail.

The author quite deliberately eliminated a consistent theory whose foundation consists in logical and mathematical ideas, which are too difficult for an ordinary natural-scientific researcher and demand outstanding mathematical training. Their avoidance is therefore quite natural since the book aims at helping biologists, agronomists and others. According to Fisher's opinion, the practical application of the general theorems of mathematical statistics is a special skill, distinct from that required by their mathematical justification. It is useful for many specialists for whom mathematical substantiations are not necessary.

The most typical feature of Fisher's book is the novelty of the explicated methods, their deep practicality which meets the real demand made on statistical investigations in natural sciences. Classical statistics, the pre-Pearsonian, and to a large extent even the Pearsonian statistics, is characterized by the development of methods only suitable for statistical series or totalities comprising a large amount of data. The newest statistics, however, is being developed, as mentioned above, in the direction of deriving methods applicable for any, even for a very small number of observations. The theory of small samples is the theoretical foundation, the vital nerve of Fisher's book. This will become especially clear when I shall go over to a systematic review of its contents.

The book begins with an introductory chapter providing a general definition of statistics as a branch of applied mathematics [1] where that science is made use of for treating observations.

Here also, the author considers the main problems of statistics: the study of populations (totalities, collectives) and variability as well as of reduction of data. Population is understood not only as a collection of living individuals or of some objects, but, in general, as a totality of any unities yielding to statistical methods of study.

The theory of errors, for instance, is engaged in studying totalities of measurements of definite magnitudes. The study of populations naturally leads to the examination of variability which in turn involves the notions on the distribution of frequencies in finite and infinite totalities, and on correlation and covariation. Reduction of data consists in the description of involved totalities consisting of large amounts of data by means of a small number of numerical characteristics. These must as far as possible exhaust all that which is interesting or important for a given researcher in the totality.

When considering the study of populations, Fisher introduces the notion of a hypothetically infinite totality of values, which are possible under the same general conditions along with those actually observed, and such with respect to which the observed data constitute a random sample; the concept of a statistic, – of an empirical magnitude formed from the experimental data and serving for estimating the parameters of an infinite totality.

Then he formulates the three main problems originating when reducing the data: specification (the choice of a special form for the law of distribution of an infinite totality); estimation (the choice of statistics for estimating the parameters of infinite totalities) and the problem of distributions (derivation of precise laws of the distribution of the possible values of the statistics in samples similar to that under consideration). The third problem also includes the issue of determining tests of goodness of fit between empirical and theoretical distributions. Then follows a classification of statistics (consistent and inconsistent, effective and sufficient) which is one of Fisher's most original and interesting statistical ideas. […]

All the book is a successive consideration of problems that can be solved by means of three distributions: the χ^2 discovered by Pearson in 1900 and essentially supplemented by Fisher; the *t*-distribution first established by Student (Gosset) in 1908; and the *z*-distribution introduced by Fisher in 1924. […]

The end of Chapter 2 is devoted to cumulants, these being special statistics introduced by Fisher instead of the moments and possessing some advantages as compared with them. […] At the end of{Chapter 4}Fisher considers the important issue of expanding χ^2 in components and applies his findings to studying the discrepancies between experiment and theory in Mendelian heredity. […] The second part{of Chapter 5}deals with the coefficients of linear and curvilinear regression in the case of two or more variables and with estimation of them, and of the discrepancies between them, which, again, is done by means of the *t*-distribution. […]

The second table{in Chapter 6}enables to determine the appropriate value of the correlation coefficient *r* given

$$Z = (1/2)\log \text{nat}[(1 + r)/(1 - r)].$$

The function *Z* is a transformation of *r* discovered by Fisher and remarkable in that its distribution is very close to normal with variance approximately equal to $1/\sqrt{n-3}$ for any values of *r* and the number of experiments *n.* The coefficient *r* is derived if these constitute a random sample from a normal population. This function enables a more precise estimate of *r* or of the discrepancies between its various values than that achieved by the usual formula for the mean square error of *r* which for small samples is not reliable at all. […]

The analysis of variances … developed by Fisher has most various and important applications. […]

Drawing on examples from genetics, Fisher shows{in Chapter 9}how to apply his method of maximum likelihood for estimating the sum{the quantity}of information inherent in the available data as well as its part made use of by various methods of their treatment.
[…] The book is a remarkable phenomenon in spite of the existence, at present, of many writings devoted to mathematical statistics among which excellent contributions can be indicated, especially in English (for example, those of Yule, Tippett, Bowley, Kelley). Fisher indeed describes modern methods of mathematical statistics, efficacious and deeply practical on the one hand and based on rigorous stochastic theory on the other hand. There is no other such source except for Tippett (1931) that is close to the problems of investigators in natural sciences, thoroughly and comprehensively providing for all his requirements and completely and deeply describing all the details of statistical methodology in biology and at the same time being just as profoundly substantiated by theory.

The exposition is not everywhere clear, in some places it demands efforts for grasping the author's ideas, but it is always consistent, comprehensive, rich in subtle and original remarks and is fresh as a primary source. Indeed, its subject-matter is almost completely Fisher's own creation, checked in practice by him or his students; and what is borrowed, is deeply thought out and recast.

It is in the highest measure desirable to publish Fisher's book in Russian, the more so since it is already translated and issued as a manuscript in a small number of copies by the Lenin All-Union Agricultural Academy [2].

8b. V.I. Romanovsky.
Review of R.A.Fisher "The Design of Experiments". Edinburgh, 1935
Sozialistich. Nauka i Tekhnika, No. 7, 1936, pp. 123 – 125

The newest methods of mathematical statistics based on the theory of small samples are called upon for playing an extremely important role in modern scientific methodology. In their context, many problems of the classical inductive logic are being solved anew, more precisely and deeply.

The new book of the English statistician R.A. Fisher, known the world over, that appeared in 1935, is indeed devoted to throwing light on the issues of inductive logic from the viewpoint of statistical methodology. It is a development of one of the chapters of his generally known contribution, *Statistical Methods for Research Workers*, and its title is typical. Since it discusses fundamental problems of the methodology of scientific investigation and the methods of their best and most proper solution, it is in some aspects even more important and more deserving of attention than that previous book. Its importance is enhanced by its considerations being based on long years of practical research and corroborated by various and numerous applications.

The book begins by indicating the two-pronged criticism usually levelled against experimental conclusions: It is alleged that either the experiment was mistakenly interpreted, thought or poorly carried out. The interpretation of trials or observations is, in essence, a statistical problem, but it does not exceed the bounds of inductive logic. The design of experiments is an issue of their logical structure and is most closely connected with their interpretation. The nature of both problems is the same and consists in revealing the logical conditions whose fulfillment can extend our knowledge by experimentation. It is very important that inductive inferences can be inaccurate but that the very nature of the inaccuracy and its degree can be determined absolutely precisely.

From Bayes and Laplace onwards, scientists have attempted to subordinate inductive inferences to mathematical estimation by issuing from that just mentioned proposition. Because of three circumstances, Fisher, however, entirely denies the theory of posterior probability constructed to this aim. First, it leads to obvious mathematical contradictions;

second, it is not based on assumptions, that everyone would have felt absolutely clear and necessary; third, it is hardly applied in practice. On the contrary, Fisher puts forward his own method of estimating inferences, a direct method based on the logical structure of appropriately designed experiments. In his theory of experimentation, the main part is played by the concept of null hypothesis, by randomization of trials and their replication.

Each time, when we ought to formulate and estimate certain inferences from some experiments, we construct one or another hypothesis on the nature or essence of the studied phenomena. Fisher understands a *null hypothesis* as a proposition that some definite feature of those phenomena does not exist. For example […] *Tests of significance* can serve for checking null hypotheses by calculating, under some general assumption, their probability. If it occurs that that probability is lower than some definite boundary determining practically impossible phenomena, we may conclude with a very low risk of being mistaken, that the facts contradict the null hypothesis, and, consequently, refute it. If, however, its probability is not very low, the facts do not contradict it, the hypothesis might be admitted, although not as the truth (facts cannot absolutely admit any hypothesis, but they can refute it), but as a probable basis for further inferences and investigations. Fisher called those probabilities, which serve, in accord with his theory, for refuting or admitting a null hypothesis, *fiducial or confidence probabilities* [3], and introduced them instead of posterior probabilities which are based on theories connected with Bayes' name.

The notion of probability plays the main part in estimating null hypotheses. An experiment should therefore be designed in such a way that that notion could be rightfully applied to it. Consequently, the experiment should be randomized and then repeated a sufficient number of times. *Randomization* ought to be understood as achieving such a pattern that subordinates to randomness those circumstances, which can lead to constant error, hardly or not at all yielding to estimation in repeated trials. In the mean, the action of that error will then vanish and the studied causes revealed all the more clearly. Replication aims at ensuring a more reliable examination of causes and makes possible their quantitative rather than only qualitative comparison.

In accord with the law of large numbers, statistical inferences are the more reliable the larger is the number of the pertinent trials. Owing to various circumstances, a large number of experiments is often impossible to achieve, but that condition might be slackened either by applying the theory of small samples or by carrying out complex experiments. The former enables the estimation of the results of a small number of trials; the latter possibility provides, even when the trials are repeated a small number of times, a sufficiently large number of cases that may be made use of for calculating the mean errors required for estimating various factors or phenomena.

In each of these directions, we are much indebted to Fisher. The benefits accrued to statistical analysis from the theory of small samples may now be considered more or less known. Lesser known is Fisher's new concept of complex design of experiments, of examining all at once the influence of a number of factors rather than of one factor at a time, on the studied phenomenon. Contrary to the generally held opinion that the latter pattern, when one factor is varied and the other ones are kept constant, is optimal, Fisher shows that in many cases agronomical experiments can be carried out in such a way that a number of interacting and varying factors are examined at the same time. The experiment thus becomes more productive, acquires an extended basis for inductive inferences and makes higher precision possible without enlarging the number of observations [4].

The principles of complex experimentation may certainly be applied not only to agronomy but in any field where statistical methods of investigation are used and where the most effective and at the same time the simplest methods are desirable. They doubtless have a great future and will incalculably benefit science.

The above is a very short essay on the main notions to which Fisher's new book is devoted. I shall now briefly discuss its separate chapters. The first two of them deal with the principles of experimentation; with general notions on the tests of significance; on null hypotheses; and on randomization illustrated by a psychophysical experiment.

The third chapter describes and examines Darwin's experiment on the growth of plants under cross-fertilization and spontaneous pollination. There also Fisher discusses the Student *t*-test; mistaken applications of statistical data and their manipulation; connection between substantiation of inferences and randomization; and the possibility of applying statistical tests more widely than stipulated by the theorems on which they are based.

Chapters four and five discuss the methods of randomized blocks and Latin squares where the Fisherian general ideas on experiments and their interpretation are applied. These methods are especially fruitful in agronomical investigations, and Fisher explicates them in detail, thoroughly examines them from the viewpoint of his general principles and illustrates them by interesting examples.

The next four chapters are devoted to the development of the principles of complex experimentation. In addition to the main issues of the new method, Fisher considers some possible forms of its further development: confounding several factors when their interaction does not essentially influence the studied phenomenon; partial confounding; and the examination of concomitant changes for increasing the precision of the experiment.

In the last two chapters Fisher returns to the general principles. He introduces the concept of *amount of information* that serves for measuring the effectiveness of various statistical tests constructed for examining one and the same null hypothesis; generalizes the last-mentioned notion by showing how to estimate entire classes of such hypotheses; considers the extention of the *t*-and χ^2-tests and the problem of estimation, – of the empirical derivation of the unknown parameters of the general population for which the experiment is a random sample. The end of the last chapter is given over to the calculation of the amount of information with regard to various statistical characteristics and to the application of that issue to the determination of the linkage of hereditary factors.

Fisher's book deserves greatest attention. Here and there, it contains many interesting remarks and new ideas. It is written quite intelligently and only makes use of elementary mathematics. Nevertheless, being concise and rich in issues, it demands concentration and mental effort. […]

Many methods described by Fisher have already been applied in fields other than agronomy and biology, viz., in textile industry, horticulture, animal husbandry, chemical industry, etc. A Russian translation of the book is very desirable [5]. It would be a very valuable supplement to the Soviet literature on statistical methodology.

8c.V.I. Romanovsky.
Review of Fisher, R.A., Yates, F. "Statistical Tables for Biological, Agricultural and Medical Research", 6th edition. New York, 1938

Sozialistich. Nauka i Tekhnika, No. 2/3, 1939, pp. 106

In 1938, Fisher, the leader of modern mathematical statistics in England, and Yates, head of the statistical section of the Rothamsted experimental station, issued new statistical tables […] The tables had grown out of Fisher's diverse practice of many years; for 15 years he had been head of the abovementioned section at Rothamsted, known the world over for its agricultural investigations. A few years ago, he turned over his job to his student Yates who successfully replaces him and is well known due to his statistical research.

Some tables are included in Fisher's *Statistical methods* … the seventh edition of which appeared in 1938 and which enjoys a wide and sound reputation as a first-rate statistical aid for experimentalists. The tables make it possible to solve many topical problems occurring in experimentation and connected with the case of scarce data. The more important of these problems are comparison and estimation of means and of variances, methodology of the analysis of variances, estimation and comparison of dependences between statistical magnitudes, comparison of experimental data with theoretical values and estimation of the discrepancies between them.

A large number of other tables are here added to the 34 due to Fisher. The new ones are intended for the solution of further practical problems not considered in Fisher's book and for facilitating yet more the practical methods expounded there. To the second category belong Tables 15 and 16 of Latin squares; of patterns of experiments by the method of randomized blocks for various numbers of replication and different versions of experimenting, of blocks and of the number of experiments in a block (Tables 17 – 19); tables of orthogonal polynomials (Table 23) for facilitating the laborious and time-consuming computation of curvilinear regressions; tables of common and natural logarithms (25 and 26) compiled in a new manner, compact but sufficient for usage; […] tables of random numbers compiled thoroughly and anew, not as extensive as the generally known tables of Tippett (1927) or the similar tables by the Soviet author Kadyrov (1936), but quite sufficient for the solution of most practical problems. All these tables are really necessary and valuable.

We also find a number of other tables (for example, 9 – 14, 20 – 22) intended for the solution of new important and interesting statistical problems, such as estimating the mortality of animals and plants made use of in experiments; estimating the frequencies of the arrangement of phenomena or objects in accord with certain groups or categories; for making statistical inferences when the order of some experimental materials is known rather than the pertinent exact values, etc.

The tables are preceded by an Introduction that describes them, illustrates their use by examples and indicates various practical problems whose solution is made easier by them. The tables are not bulky, and, together with the Introduction, occupy 90 pages. We, in the Soviet Union, in spite of the grand scale of our experimental work, regrettably do not have such tables. They are urgently needed but we should not simply reissue the Fisher & Yates *Tables*. That would have been the easiest way out, but their tables can be partly shortened, partly supplemented by some other important tables, whereas the Introduction, not everywhere clear, should be replaced by another one, more systematic and clear, and supplemented by new examples.

8d. T.A. Sarymsakov. Statistical Methods and Issues in Geophysics (Fragments)

Второе всесоюзное совещание по математической статистике. Ташкент, 1948
(Second All-Union Conference on Mathematical Statistics. Tashkent, 1948). Tashkent, 1948, pp. 221 – 239 …

[…] We ought to comprehend, interpret and appraise the theoretical and empirical regularities revealed by the theory of probability and mathematical statistics, first and foremost, by issuing from the physical nature and essence of the studied phenomenon. Indeed, when applied to some domain of knowledge, these two disciplines mostly play, as a rule, a subordinate part. To base oneself, when applying them, always (or only) on deductive mathematical considerations is in most cases dangerous because the inferences thus obtained can be incompatible with the physical or natural side of the studied phenomenon, and even fruitless. As an example, I may indicate the fruitless applications of the theory of probability and mathematical statistics to the theory of heredity [6]… […]

As far as mathematical statistics, which is more practical{than the theory of probability}in its subject, is concerned, it was mostly developed{in the Soviet Union}by Romanovsky and several of his students, by Slutsky, who devoted very many writings to issues of statistical geophysics, and others. However, we ought to indicate a fundamental shortcoming occurring in a number of works in mathematical statistics carried out in Tashkent. When choosing their subject and the problems to be solved, Romanovsky strictly followed the Anglo-American direction. At the same time we must point out that a considerable part of investigations in mathematical statistics that was performed in Tashkent was concerned with concrete applications, and made wide use of those methods for their accomplishment which were developed by the Chebyshev school, – although certainly adapted for the solution of statistical problems. […]

…the specific nature of the socialist production in our country, as distinct from production based on private property in the capitalist nations, in some of which statistical research is widely enough used, is not yet sufficiently taken account of in{our}investigations of the applications of mathematical statistics, especially to industry. A number of our works both in this direction and in other fields mostly flow into the common channel of foreign investigations.

These circumstances demand a serious ideological and methodological reconstruction in mathematical statistics. The issue already raised by Kolmogorov (1938, p. 176) ten years ago, now becomes especially topical. He wrote then: "The development and reappraisal of the general concepts of mathematical statistics, which are now studied abroad under great influence of the idealistic philosophy, remain a matter for the future". However, recent investigations of methodological issues of mathematical statistics pay considerably more attention to formal rigorous determination of its scope rather than to the extremely important, in the sense of ideology and subject, topic on the essence of its problems and on how should concrete phenomena be statistically examined.

8e. Resolution of the Second All-Union Conference on Mathematical Statistics 27 September – 4 October 1948, Tashkent

Второе всесоюзное совещание по математической статистике. Ташкент, 1948 (Second All-Union Conference on Mathematical Statistics. Tashkent, 1948). Tashkent, 1948, pp. 331 – 317

The Five-Year Plan of reconstruction and further development of the national economy of the Soviet Union raises before Soviet science fundamentally new problems formulated by Comrade Stalin in his slogan addressed to Soviet scientists, – "to overtake and surpass the achievements of the science abroad". The Great Patriotic War{1941 – 1945}put forward for the statisticians the topical issues concerning the theory of the precision of machinery, rejection of defective articles and inspection of the quality of mass products, etc. After the war, the part played by statistics in a number of branches of the national economy increased still more. The role of statistics in the development of the main directions of natural sciences is also great.

Some statisticians took up idealistic positions, supported the Weismann direction in biology and developed abstract patterns of formal genetics cut off from reality. This, however, does not at all discredit statistics itself as being a most important tool of investigation in biology and other sciences. The Conference resolutely condemns the speech of V.S. Nemchinov, made at the session of the Lenin All-Union Agricultural Academy, for his attempt statistically to "justify" the reactionary Weismann theories [7]. Objectively, Academician Nemchinov adhered to the Machian Anglo-American school which endows statistics with the role of arbiter situated over the other sciences, a role for which it is not suited.

The latest decisions of the Central Committee of the All-Union Communist Party (Bolsheviks) concerning ideological issues raised the problem of rooting out the survivals of capitalism from people's minds, which, among the Soviet intellectuals, are expressed by servility and kow-towing to foreign ideas, by lack of a proper struggle for the priority of the

Russian, and especially of the Soviet science. Together with the past discussions on issues of philosophy, literature, music, and, finally, biology, these decisions directly indicate that it is necessary to revise the contents of statistics from the viewpoint of the struggle against bourgeois ideology as well as for attaining closer proximity between theoretical investigations and the problems of socialist practice.

Among statisticians, the passion for the theories of foreign, and especially English and American scientists, is still great. Along with these theories, often uncritically grasped, a Weltanschauung alien to Soviet scientists, and in particular the Machian concepts of the Anglo-American statistical school of Pearson, Fisher and others, had sometimes been introduced. Even during this Conference attempts had been made to force through the Machian Weltanschauung disguising it by loud revolutionary phrases (Brodovitsky, Zakharov).

The Conference accepts with satisfaction the statement of a most eminent Soviet statistician, the Full Member of the Uzbek Academy of Sciences, Professor Romanovsky, who confessed to having made ideological mistakes in some of his early works. *The Conference considers it necessary to list the following essential shortcomings in the work of Soviet statisticians.*

1. The existence of a *gap between theory and practice* resulting in that some serious theoretical findings were not carried out onto practical application.

2. The lack of prominent monographs generalizing numerous theoretical achievements of Soviet statistics and harmoniously explicating the concepts of Soviet statistics [8].

3. The methods of bourgeois statistics were not always critically interpreted; sometimes they had been propagandized and applied.

4. The teaching of the theory of probability and mathematical statistics, in spite of their ever increasing significance for studying most important issues in natural sciences, technology and economics, is either altogether lacking in the appropriate academic institutions or carried out insufficiently and sometimes on a low level. In particular, utterly insufficient attention is given to the training of specialists in mathematical statistics in the universities and the teaching of the elements of statistics is not at all introduced in technological academic institutions.

5. The publication of a special statistical periodical has yet not begun [9] which greatly hampers the intercourse and the exchange of experience between theoreticians and practitioners.

6. The existing educational literature and monographs on statistics and probability theory are meager, their quality is sometimes unsatisfactory and they are insufficiently connected with concrete applications.

In mathematical statistics, the Conference considers research in the following directions as most topical.

1. A construction of a consistent system of mathematical statistics embracing all of its newest ramifications and based on the principles of the Marxist dialectical method.

2. A further development of the theory of estimation of parameters and of checking hypotheses. In particular

a) The development of such a theory for series of dependent observations.

b) The development of methods for an unfixed number of observations (of the type of sequential analysis).

3. The development of statistical methods of inspection of manufacture and of finished articles; and in particular of methods not assuming a stationary condition of manufacturing.

4. Construction of a rational methodology of designing and treating field experiments and agricultural forecasts.

5. The further development of statistical methods of investigation in geophysics; in particular, in synoptic meteorology.

6. Stochastic justification of the various types of distributions occurring in natural sciences and technology by issuing from the essence of the studied phenomena [10], and the development of methods of sampling without the assumption of normality.
7. The compilation of necessary tables and the use of calculating machines as much as possible.
For propagating statistical knowledge the Conference believes that the following is necessary
1. The introduction of courses in mathematical statistics and theory of probability in the largest technical institutes and in the physical, biological, geographical and geological faculties of the universities, for selectionists in the agricultural institutes and especially for postgraduate students of all the abovementioned disciplines. The curricula and the scope of these courses should be closely coordinated with the main disciplines […]
2. A wide expansion of training of highly qualified specialists in the theory of probability and statistics; restoration of the pertinent training in the universities in Moscow, Leningrad and Tashkent as well as of postgraduate training in the all-union and republican academies of sciences and in the abovementioned universities.
3. Implementation of desirable additional training of specialists whose work is connected with statistics by organizing special courses and by individual attachment to scientific institutions and chairs.
4. The compilation, as soon as possible, of textbooks and educational aids on general courses in mathematical statistics and the theory of probability for mathematical faculties of the universities as well as on special courses for technical, geological and other institutions and faculties; the completion of a number of special monographs.
5. We charge Kolmogorov, Romanovsky and Gnedenko with preparing a plan of compilation of textbooks and monographs in mathematical statistics, theory of probability and their applications [11].
6. The publication of a journal of mathematical statistics for reflecting theoretical and methodological achievements of Soviet statisticians, applications of statistics to various fields of technology and natural sciences and methodological issues of teaching [12].
7. The establishment of an All-Union Statistical Society under the Soviet Academy of Sciences for uniting the work of theoreticians and practitioners in all fields of national economy so as to foster the development and propagation of statistical knowledge.
8. The organization, with the help of the All-Union Society for Propagating Political and Scientific Knowledge, of seminars and lectures on the application of statistical methods for a wider circle of listeners.
9. A regular convocation, once in three years, of all-union conferences on mathematical statistics.
10. A convocation, in 1949, of an all-union conference on statistical methods of inspection and analysis of manufacturing.
11. The publication of the transactions of this Conference.
12. The pertinent editorial committee will include Romanovsky, Kolmogorov, Sarymsakov and Eidelnant.

8f. The Publisher's Preface to the Russian Translation of R.A. Fisher "Statistical Methods for Research Workers". Moscow, 1958

The work of Fisher strongly influenced the development of mathematical statistics. Gosstatizdat{the State Publishers of Statistical Literature}therefore issues{a translation of}his book […]. The author is a prominent English theoretician of mathematical statistics. For many years he had been head of the statistical section of the Rothamsted experimental station in England. This book, as he indicates in the Introduction, was created as a result of his collaboration with experimentalists in biology.

It should be borne in mind that, being a theoretician of modern bourgeois statistics, bourgeois narrow-mindedness and formal viewpoints are in his nature. According to his concepts, quantitative analysis is an universal and absolute statistical means of cognition. In actual fact, he completely ignores the qualitative aspect of phenomena [13]. Suffice it to indicate his statement that social teachings can only be raised up to the level of real science to that degree to which they apply statistical tools and arrive at their inferences by issuing from statistical arguments. Here, he bears in mind mathematical statistics which he considers as an universal science.

A duality, usual for bourgeois scientists, is typical of Fisher. On the one hand, submitting to objective reality, they make valuable observations and conclusions; on the other hand, being under the influence of an idealistic Weltanschauung, they color their findings in an appropriate way. When considering the methods that Fisher describes, it is impossible to deny their logicality; when, however, passing on to concrete examples illustrating them, we meet with some non-scientific propositions concerning sociological issues. Thus, it results from one of his examples, that if one of the two monozygotic twins became a criminal, the second twin will follow suit almost surely. It follows that criminality was already present in that impregnated egg, from which those twins were later developed. It is clear that such a concept of a "biological" origin of criminality absolutely ignores the social conditions of the life of men from which bourgeois sociologists and economists disengage themselves [14].

Fisher's exposition is irregular: in some places he describes an issue in all details including calculations, elsewhere he only sketches a problem. His discussion of statistical methods cannot be considered simple and quite popular; an acquaintance with his book demands certain effort. Although the book is intended for researchers in biology and agronomy, it will also be somewhat interesting for statisticians working in economics. Making use of the correct propositions of mathematical statistics, Soviet readers will cast away all the conclusions and considerations unacceptable to real statistical science.

Notes

1.{On the relation between statistics and mathematics see Sheynin (1999, pp. 707 – 708).}

2.{An ordinary edition of the translation only appeared in 1958, and was accompanied by negative commentaries, see Item 6 of this section.}

3.{In accord with present-day terminology, I translated Romanovsky's *doveritelnye* as *confidence* (probabilities). These, however, were introduced by Neyman in 1937.}

4. I advise readers to take notice of Yates (1935). Yates replaced Fisher at Rothamsted when the latter had passed on to London University. Yates minutely examined complex experiments in the issue of the *J. Roy. Stat. Soc.* devoted to applications of statistics to agronomy and industry.

5.{No such translation ever appeared.}

6.{The last two sentences were considered extremely important and their essence was repeated in the Resolution, see Item 5 of this section. From 1935 onward, genetics in the Soviet Union came under fierce attacks and was rooted out in 1948 as being contrary to dialectical materialism; it only reappeared in 1964. The contrasting of mathematics and the "physical or natural" side of phenomena, – of quantitative and qualitative, – was typical for Soviet statisticians, see Sheynin (1998, pp. 540 – 541 and elsewhere) and may be explained in a similar way: quantitative considerations had not been allowed to interfere with Marxist (or Leninist, or Stalinist) dogmas.}

7.{Nemchinov had to abandon his post of Director of the Timiriazev Agricultural Academy, to leave his chair of statistics there (Lifshitz 1967, p. 19), and to confess publicly his guilt (Sheynin 1998, p. 545).}

8.{Soviet statistics may obviously be understood as a discipline obeying ideological dogmas, cf. Note 6 above. Below, the Resolution stated that statistics should be based on the Marxist dialectical method.}

9.{The most influential Soviet statistical periodical, *Вестник статистики*, was suppressed in 1930 and did not reappear until 1948.}

10.{Cf. Note 6.}

11.{In 1950, Gnedenko published his generally known *Курс теории вероятностей* (Course in the Theory of Probability; several later editions and translations). He "followed the path suggested by Kolmogorov" (p. 47 of the edition of 1954).}

12.{The periodical *Теория вероятностей и ее применения* (Theory of probability and Its Applications) is only being published since 1955. No Statistical Society (see below) was ever established.}

13.{Cf. Note 6.}

14.{In such cases, similarity of the main conditions (of the conditions of life of the twins) is always presupposed.}

References

Bogoliubov, A.N., Matvievskaia, G.P. (1997), *Романовский* (Romanovsky). M.

Chuprov, A.A. (1918 – 1919), Zur Theorie der Stabilität statistischer Reihen. *Skand. Aktuarietidskr.*, t. 1, pp. 199 – 256; t. 2, pp. 80 – 133.

Kadyrov, M. (1936), *Таблицы случайных чисел* (Tables of Random Numbers). Tashkent.

Kolmogorov, A.N. (1938), The theory of probability and its applications. In *Математика и естествознание в СССР* (Math. and Natural Sciences in the USSR). M., 1938, pp. 51 – 61. Translation:

Lipshitz, F.D. (1967), Nemchinov as a statistician. In Nemchinov, V.S. (1967), *Избранные произведения* (Sel. Works), vol. 2. M., pp. 5 – 22. (R)

Romanovsky, V.I. (1923), Review of Chuprov (1918 – 1919). *Vestnik Statistiki*, No. 1/3, pp. 255 – 260. Translated in Chuprov, A. (2004), *Statistical Papers and Memorial Publications*. Berlin, pp. 168 -174.

--- (1924), Theory of probability and statistics. Some newest works of Western scientists. *Vestnik Statistiki*, No. 4/6, pp. 1 – 38; No. 7/9, pp. 5 – 34. (R)

--- (1927), Theory of statistical constants. On some works of R.A. Fisher. Ibidem, No. 1, pp. 224 – 266. (R)

--- (1934), On the newest methods of mathematical statistics applied in field experimentation. *Sozialistich. Nauka i Tekhnika* (Tashkent), No. 3/4, pp. 75 – 86. (R)

Sarymsakov, T.A. (1955). Obituary of Romanovsky. Translated in this book.

Sheynin, O. (1998), Statistics in the Soviet epoch. *Jahrbücher f. Nat.-Ökon. u. Statistics*, Bd. 217, pp. 529 – 549.

--- (1999), Statistics: definitions of. *Enc. Stat. Sciences*, Update volume 3. New York, pp. 704 – 711.

Smirnov, N.V. (1948), Mathematical statistics. In *Математика в СССР за 30 лет* (Math. in the USSR during 30 years). M. – L., 1948, pp. 728 – 738. Translation:

Tippett, L.H.C. (1927), Random sampling numbers. *Tracts for Computers*, No. 15.

--- (1931), *Methods of Statistics.* London.

Yates, F. (1935), Complex experiments. *J. Roy. Stat. Soc. Suppl.*, vol. 2, pp. 181 – 247.

8g. T.A. Sarymsakov. Vsevolod Ivanovich Romanovsky. An Obituary

Uspekhi Matematich. Nauk, vol. 10, No. 1 (63), pp. 79 – 88

Romanovsky, the eminent mathematician of our country, Deputy of the Supreme Soviet of the Uzbek Soviet Socialist Republic, Stalin Prize winner, Ordinary Member of the Uzbek Academy of Sciences, Professor at the Lenin Sredneaziatsky {Central Asian} State University {SAGU}, passed away on October 6, 1954.

He was born on Dec. 5, 1879, in Almaty and received his secondary education at the Tashkent non-classical school {Realschule} graduating in 1900. In 1906 he graduated from Petersburg University and was left there to prepare himself for professorship. After passing his Master examinations in 1908, Romanovsky returned to Tashkent and became teacher of mathematics and physics at the non-classical school. From 1911 to 1917 he was reader {Docent} and then Professor at Warsaw University. In 1912, after he defended his dissertation *On partial differential equations*, the degree of Master of Mathematics was conferred upon him. In 1916 Romanovsky completed his doctor's thesis but its defence under war conditions proved impossible. The degree of Doctor of Physical and Mathematical Sciences was conferred upon him in 1935 without his presenting a dissertation.

From the day that the SAGU was founded and until he died, Romanovsky never broke off his connections with it remaining Professor of the physical and mathematical faculty. For 34 years he presided over the chairs of general mathematics and of theory of probability and mathematical statistics; for a number of years he was also Dean of his faculty.

Romanovsky was Ordinary Member of the Uzbek Academy of Sciences from the moment of its establishment in 1943, member of its presidium and chairman of the branch of physical and mathematical sciences. His teaching activities at SAGU left a considerable mark. Owing to the lack of qualified instructors in the field of mathematics, he had to read quite diverse mathematical courses, especially during the initial period of the University's existence. Romanovsky managed this duty with a great success presenting his courses on a high scientific level.

Romanovsky undoubtedly deserves great praise for organizing and developing the higher mathematical education in the Central Asiatic republics {of the Soviet Union} and especially in Uzbekistan. He performed a considerable and noble work of training and coaching scientific personnel from among the people of local nationalities.

Modernity of the substance of the courses read; aspiration for coordinating the studied problems with the current scientific and practical needs of our socialist state, and, finally, the ability to expound intelligibly involved theoretical problems, – these were the main features of V.I. as a teacher. Add to all this his simplicity of manner and his love for students, and you will understand that he could not have failed to attract attention to himself and to his subject. Indeed, more than sixty of his former students are now working in academic institutions and research establishments of our country.

Romanovsky always combined teaching activities with research, considerable both in scale and importance. He published more than 160 writings on various fields of mathematics with their overwhelming majority belonging to the theory of probability and mathematical statistics. He busied himself with other branches of mathematics, mostly with differential and integral equations and some problems in algebra and number theory, either in the first period of his scientific work (contributions on the first two topics) or in connection with studying some issues from probability theory and mathematical statistics.

The totality of Romanovsky's publications in probability and statistics (embracing almost all sections of mathematical statistics) unquestionably represents a considerable contribution to their development in our country. Accordingly, he became an eminent authority on these branches of the mathematical science not only at home, but also far beyond the boundaries of our country.

Among Romanovsky's most fundamental and important studies in probability is his work on Markov chains (which he began in 1928) and their generalizations (correlation chains and polycyclic chains) and on generalizing the central limit theorem onto the multidimensional case. He was the first to study exhaustively by algebraic methods the limiting (as $n \to \infty$) behavior of the transition probabilities describing the change of state during n steps for homogeneous Markov chains with a finite number of states [96].

In the same paper and in his later work [112; 121; 126; 132; 142] Romanovsky was engaged in proving a number of other limit theorems for the same kind of Markov chains. This research also became the starting point for many other studies of Markov chains and their various generalizations by algebraic methods. In [63], applying the method of characteristic functions, he extended the central limit theorem onto sums of independent random vectors.

In statistics, Romanovsky's work cover an extremely wide range of problems. It is hardly possible to point out any large section of this discipline, whether modern or classical, in whose development he did not actively and authoritatively participate. Especially great are Romanovsky's merits in widely popularizing mathematical-statistical methods in our country as well as in heightening the mathematical level of statistical thought. Here, his course [144] published in 1924 and 1939 and his books [50; 105; 136; 137] played a very large part.

I shall now briefly describe some of his important studies in mathematical statistics. Depending on the form of the theoretical law of distribution and on the organization of observations, there appear various methods of approximately estimating the different characteristics of the parent population. The most prominent research in our country in this sphere was done by Romanovsky.

A large cycle of his writings [38; 39; 45; 46; 48; 54; 60; 62] concerned with the theory of sampling was generally recognized. With regard to their substance, they adjoin the studies of the British school of statistics, but they advantageously differ from the latter by rigor of their methodological principles. In addition, when choosing methods for solving his problems, Romanovsky exclusively used those developed by the Chebyshev school; to be sure, he perfected and adopted these methods for achieving new goals. That Romanovsky followed here Chebyshev can partly be explained by his belonging to the latter's school and having attended the course in probability theory read by the celebrated Markov. Keeping in his studies to that mathematical rigor that distinguished his teacher, Markov, he used the theory as the main tool for logically irreproachably justifying mathematical statistics. Such a substantiation was indeed lacking in the constructions of British statisticians whose works served Romanovsky as a starting point for choosing his problems. The rigorous theoretical underpinning of mathematical statistics is one of his merits that promoted its development in our country.

Romanovsky was the first to offer an analytical derivation of the laws of distribution of the well-known criteria, of the Student – Fisher t and z, of empirical coefficients of regression and other characteristics [105]. He also provided a more general theory of the Pearson chi-squared test [65] and studied problems connected with checking whether two independent samples belonged to one and the same normal population.

From among Romanovsky's work on probability and statistics deserving serious attention I also mention [113; 120; 122; and 115]. In the first of these, he shows that the θ criterion, that he himself introduced in 1928 [60], is much easier to apply to all the problems where the Fisherian z test based on the tables of that scholar is made use of; that it leads to the same qualitative solutions; and that it often solves these problems more precisely than the latter. In addition, the construction of the former is simpler.

The second writing [120] is very interesting methodologically. There, Romanovsky attempts to review systematically the main statistical concepts and problems. Given the variety and detachment of those latter, and the availability of a diverse set of methods applied by statistics, his endeavor was absolutely necessary.

The third paper provided an elementary and simple solution to a topical statistical problem connected with objectively estimating unknown characteristics of parent populations by means of observation. In the last-mentioned work he calculated transition and other kinds of probabilities for Markov chains and offered their statistical estimates given the appropriate observations.

The classical theory of periodograms enables to analyze a number of random variables under the assumption that several periodic oscillations and additional perturbations independent from one trial to another one are superposed. Romanovsky devoted a series of important studies [78; 79; 81] to the circumstances occurring when admitting dependence between random perturbations.

Systematically and intensively carrying out his scientific work for half a century, Romanovsky, especially from the 1930s onwards, had been paying more attention than he did before to problems directly connected with practical needs, e.g., in [86; 87; 98; 146]. And during the last period of his life he was much engaged in the important problem of contemporary statistics, – in the statistical estimation of the quality of production [134; 145; 153; 151; and other ones].

From among contributions not connected with probability or statistics I mention [68; 72; 104; 92]. The first three of these, although originating from problems connected with Markov chains, provided findings of independent mathematical interest.

In concluding this brief review, it is also necessary to indicate that, not restricting his efforts to publishing scientific contributions, Romanovsky unceasingly counselled, verbally and in writing, most diverse productive establishments and scientific institutions and answered questions arriving from all quarters of our country. Until his last days he combined scientific studies with active social work. He was permanent chairman of Tashkent Mathematical Society and an active member of the Society for Dissemination of Political and Scientific Knowledge. The people and the government of Uzbekistan estimated his merits at their true worth. He was three times elected Deputy of the Republic's Supreme Soviet and decorated with three orders of Lenin and an Order of the Red Banner of Labor; he became Honored Science Worker of Uzbekistan, and he wined a Stalin prize (1948).

The sum totals of all his scientific work and teaching activities was the mathematical school that he created in Tashkent.

Bibliography

{A comprehensive bibliography of Romanovsky's contributions is in Bogoliubov, A.N., Matvievskaia, G.P. (1997), *Romanovsky*. M. (R), against which I checked and specified the entries below. With the exception of contributions mentioned by Sarymsakov in his main text, I omitted those not belonging to probability or statistics. Nevertheless, I preserved his numeration; for example, the first entry below has number 6. Contributions 52, 57, 80, 92 and 104 are reprinted/translated in Romanovsky's *Избранные труды* (Sel. Works), vol. 2. Tashkent, 1964.}

Abbreviation

AN = Akademia Nauk (of Soviet Union if not stated otherwise)
C.r. = *C.r. Acad. Sci. Paris*
GIIA = *G. Ist. Ital. Attuari*
IMM = Inst. Math. & Mech.
SAGU = Srendeaziatsk. Gosudarstven. Univ.
SSR = Soviet Socialist Republic
VS = *Vestnik Statistiki*

6. Jakob Bernoulli's theorem and the law of large numbers. *Protokoly Obshchestvo Estestvoispytatelei Warsaw Univ.*, No. 4, 1912, pp. 39 – 63. (R)
13. Some problems in the calculus of probability. Ibidem, 1915. (R)
14. Generalization of the Pearson curves. Ibidem, 1916. (R)
17. *Элементарный курс математической статистики* (Elementary Course in Mathematical Statistics). Tashkent, 1921. Mimeogr. edition.
18. On linear correlation of two dependent magnitudes. VS No. 9 – 12, 1922, pp. 23 – 28. (R)
19. On the probability of connected indications and on their application in statistics. Ibidem, pp. 34 – 41. (R)
20. On the correlation ratio. Ibidem, pp. 29 – 33. (R)
21. *Элементы теории корреляции* (Elements of the Theory of Correlation). Tashkent, 1923.
23. Review of A.A. Chuprov, Zur Theorie der Stabilität statistischer Reihen, 1918 – 1919. VS No. 1 – 3, 1923, pp. 255 – 260. Translated in Chuprov, A. (2004), *Statistical Papers and Memorial Publications*. Berlin.
24. Note on the moments of a binomial about its mean. *Biometrika*, vol. 15, 1923, pp. 410 – 412.
26. *Элементарное изложение теории средних* (Elementary Exposition of the Theory of Means). Zentraln. Statistich. Upravlenie Turkestan, 1924. {The republic of Turkestan ceased to exist at the end of 1924.}
27. On the generalization of the Markov inequality. *Bull. SAGU* No. 8, 1925, pp. 107 – 111. (R)
28. Generalization of some types of the frequency curves of Prof. K. Pearson. *Biometrika*, vol. 16, 1924, pp. 106 – 117.
29. Second edition of [17]. M., 1924.
30. Theory of probability and statistics. VS, 1924, No. 4 – 6, pp. 1 – 38 and No. 7 – 9, pp. 5 – 34. (R)
32. The contemporary state of the theory of correlation. VS No. 1 – 3, 1925, pp. 1 – 46. (R)
33. A new proof of the Poisson theorem. *Bull. SAGU* No. 9, 1925, pp. 95 – 101. (R)
34. On the distribution of sums of normally distributed variables. Ibidem, pp. 89 – 94. (R)
35. On statistical tests of a given individual belonging to some closed species. Trudy Turkestan Nauchn. Obshchestvo. SAGU No. 2, 1925, pp. 173 – 184. (R)
36. On the moments of hypergeometric series. *Biometrika*, vol. 17, 1925, pp. 57 – 60.
37. Généralisations d'un inégalité de Markoff. *C.r.*, t. 180, 1925, pp. 1468 – 1470.
38. Sur certaines espérances mathématiques et sur l'erreur moyenne du coefficient de corrélation. Ibidem, pp. 1897 – 1899.
39. Sur la distribution des écarts quadratiques moyens dans les observations sur les quantités à distribution normale. Ibidem, pp. 1320 – 1323.
40. Sur une méthode d'interpolation de Tschébycheff. Ibidem, pp. 595.
41. A short proof of the Pearson formula for the moments of the hypergeometric series. *Bull. SAGU*, No. 12, 1926, pp. 127 – 129. (R)
42. On a problem due to Fisher. Ibidem, No. 15, 1927, pp. 313 – 317. (R)
43. On the central moments of two normally distributed random variables. Ibidem, pp. 307 – 312. (R)
44. A new derivation of the Chebyshev method of parabolic interpolation. *Vestnik Irrigatsii*, No. 1, 1926, pp. 29 – 51. (R)
45. On the distribution of the arithmetic mean in series of independent trials. *Izvestia AN*, ser. 6, vol. 20, 1926, pp. 1087 – 1106. (R)

46. On the distribution of the regression coefficient in samples from a normal population. Ibidem, pp. 643 – 648.
47. On certain expansions of the incomplete beta function. MS, vol. 33, 1926, pp. 201 – 229. (R)
48. On the moments of the standard deviations and correlation coefficient in samples from a normal population. *Metron*, t. 5, 1926, pp. 3 – 46.
49. Theory of statistical constants. VS, No. 1, 1927, pp. 224 – 266. (R)
50. This is the 2nd edition of [21]. Tashkent, 1927.
51. Note on orthogonalizing series of functions and interpolation. *Biometrika*, vol. 19, 1927, pp. 93 – 99.
52. On criteria of the belonging of two random samples to one and the same normal population. *Труды Всесоюзного съезда математиков 1927г.* (Proceedings All-Union Conference of Mathematicians 1927). M., 1928, pp. 273 – 274. (R)
53. A stochastic justification of the theory of association. Ibidem, p. 276. (R)
54. On statistical tests of belonging to a group. *Doklady AN*, ser. A, No. 16/17, 1928, pp. 347 – 352. (R)
55. Additions inserted in Kaufman; A.A. *Теория и методы статистики* (Theory and Methods of Statistics), 5th edition. M., 1928.
56. Sur une théorème limite du calcul des probabilités. MS, vol. 36, 1929, pp. 36 – 64.
57. Sur la comparaison des moyennes des variables aléatoires dépendantes. *Atti Congr. Intern. Matematici Bologna, 1928*. Bologna, 1932, t. 6, pp. 101 – 102.
58. Sur le calcul des moments des moyennes. Ibidem, pp. 103 – 105.
59. Sur la généralisation des courbes de Pearson. Ibidem, pp. 107 – 110.
60. On the criteria that two given samples belong to the same normal population. *Metron*, t. 7, 1928, pp. pp. 3 – 46.
61. Extension of the Laplace theorem on the limit of probability onto the case of two dependent events. *Trudy SAGU*, ser. Math., No. 1, 1929, pp. 1 – 11. (R)
62. On the moments of means of functions of one and more random variables. *Metron*, t. 8, 1929, pp. 251 – 290.
63. Sur une extension du théorème de Liapunoff sur la limite de probabilité. *Izvestia AN*, ser. 7, phys. & math. sci., 1929, pp. 209 – 225.
64. Sur les chaînes de Markoff. *Doklady AN*, ser. A, No. 9, 1929, pp. 203 – 208.
65. Sur la loi de probabilité de frequences assujetties aux conditions linéares et le critérium χ^2 de Pearson. Ibidem, No. 4, pp. 63 – 105.
66. Sur les probabilités à posteriori. *C.r.*, t. 189, 1929, pp. 515 – 517.
68. Sur une classe d'équations integrales linéares. Ibidem, t. 191, 1930, pp. 552 – 557.
69. Sur les chaînes discrètes de Markoff. Ibidem, pp. 450 – 452.
70. Sur les chaînes biconnexes continues de Markoff. Ibidem, t. 190, 1930, pp. 695 – 697.
71. On stochastic estimates of the results of field experiments. *Vestnik Irrigatsii*, No. 6, 1928, pp. 39 – 55; NNo. 1 and 3, 1929, pp. 41 – 45 and 13- 26. (R)
72. Sur les zeros de matrices stochastiques. *C.r.*, t. 192, 1931, pp. 266 – 269.
73. Généralisations d'un théorème de Slutsky. Ibidem, pp. 718 – 721.
74. Sulle regressioni multiple. GIIA, t. 2, No. 2, 1931, pp. 161 – 171.
75. Sulle probabilità a posteriori. Ibidem, No. 4, pp. 493 – 511.
76. Due nuovi criteri di controllo sull'andamento casuale di una successione di valori. Ibidem, t. 3, No. 2, 1932, pp. 210 – 222.
78. Sur la loi sinusoidale limite. *Rend. Circ. Mat. Palermo,* t. 56, 1932, pp. 82 – 111.
79. Same title. Congr. Intern. Math. Zürich 1932.
80. On a property of mean ranges in samples from normal population and on some integrals of Prof. T. Hojo. *Biometrika*, vol. 25, 1933, pp. 195 – 197.

81. Sur une généralisation de la loi sinusoidale limite. *Rend. Circ. Mat. Palermo*, t. 57, 1933, pp. 130 – 136.
83. New criteria of randomness of a series of trials. *Trudy SAGU*, ser. Math., No. 9, 1933, pp. 18 – 31. (R)
84. On posterior probabilities. Ibidem, No. 8, pp. 18 – 32. (R)
85. Sui due problemi di distribuzione casuale. GIIA, No. 2 – 3, 1935, pp. 196 – 218.
86. On the newest methods of mathematical statistics applied in field experiments. *Sozialistich. Nauka i Tekhnika* No. 3 -4, 1934, pp. 75 – 86. (R)
87. Mathematical statistics and industry. Ibidem, No. 9, pp. 1 – 14. (R)
89. On the Chebyshev inequality for the bivariate case. *Trudy SAGU*, ser. Math., No. 11, 1934, pp. 1 – 6. (R)
91. Sur une formula de A.R. Crathorne. *C.r.*, t. 200, 1935, pp. 105 – 107.
92. Integration of involutory systems of any class with partial derivatives of any order. *Bull. SAGU*, No. 20 – 1, 1935. (R)
93. Sui momenti della distribuzione ipergeometrica. GIIA, t. 6, No. 2, 1935.
94. Review of Fisher, R.A. *Statistical Methods for Research Workers*. *Sozialistich. Rekonstruktsia i Nauka*, No. 9, 1935, pp. 123 – 127. (R)
95. On the application of mathematical statistics and theory of probability in the industries of the Soviet Union. *J. Amer. Stat. Assoc.*, vol. 30, 1935, pp. 709 – 710.
96. Recherches sur les chaînes de Markoff. *Acta Math*., t. 66, 1936, pp. 147 – 251.
97. The meaning and the power of statistical criteria. *Sozialistich. Rekonstruktsia i Nauka*, No. 3, 1936, pp. 84 – 89. (R)
98. On a method of industrial control. *Problemy Ucheta i Statistiki*, vol. 2, No. 2, 1936, pp. 3 – 13. (R)
99. Note on the method of moments. *Biometrika*, vol. 28, 1936, pp. 188 – 190.
100. Review of Fisher, R.A. *Design of Experiments*. *Sozialistich. Nauka i Tekhnika*, No. 7, 1936, pp. 123 – 125. (R)
104. On a new method of solving some linear finite equations with two independent variables. MS, vol. 3(45), No. 1, 1938, pp. 144 – 165. (R)
105. *Математическая статистика* (Mathematical Statistics): M. – L., 1938.
106. Analytical inequalities and statistical tests. *Izvestia AN*, ser. Math., No. 4, 1938, pp. 457 – 474. (R)
107. Quelques problèmes nouveaux de la théorie des chaînes de Markoff. *Conf. Intern. Sci. Math. Genevé 1938*. Paris, 1938, No. 4, pp. 43 – 66.
108. Mathematical statistics. *Bolshaia Sovetskaia Enziklopedia* (Great Sov. Enc.), vol. 38, 1938, pp. 406 – 410.
109. Least squares, method of. Ibidem, vol. 41, 1939, pp. 53 – 56.
111. Mathematical research in Uzbekistan during 15 years. In *Наука в Узбекистане за 15 лет* (Science in Uzbekistan during 15 years). Tashkent, 1939, pp. 49 – 52.
112. *О цепных корреляциях* (On Chain Correlations). Tashkent, 1939.
113. On analysis of variance. *Trudy sektora matematiki Komiteta Nauk Uzbek Republic*, No. 1, 1939, pp. 3 – 12. (R)
114. *Элементарный курс математической статистики* (Elementary Course in Mathematical Statistics). M. – L., 1939.
115. *Статистические задачи связанные с цепями Маркова* (Statistical Problems Connected with Markov Chains). Tashkent, 1940.
116. On inductive conclusions in statistics. *Trudy Uzbek Filial AN*, ser. 4, math., No. 3, 1942, pp. 3 – 22. (R)
117. Chain connections and tests of randomness. *Izvestia Uzbek Filial AN*, No. 7, 1940, pp. 38 – 50. (R)

119. On the Markov method of establishing limit theorems for chain processes. Ibidem, No. 3, 1941, pp. 3 – 8. (R)
120. The main concepts and problems of mathematical statistics. *Trudy Uzbek Filial AN*, ser. Math., No. 2, 1941, pp. 3 – 23. (R)
121. Bicyclic chains. Ibidem, No. 1, 1941, pp. 3 – 41. (R)
122. Same title as [116]. Ibidem, No. 3, 1942. (R)
123. On the main aims of the theory of errors. Ibidem, No. 4, 1943. (R)
124. On a case of chain correlation. *Bull. SAGU*, No. 23, 1945, pp. 17 – 18. (R)
125. On the limiting distribution of sample characteristics. Ibidem (R)
126. The main property of polycyclic chains. Scientific Conference SAGU. Tashkent, 1945, pp. 20 – 21. (R)
127. On probabilities of the repetition of cycles in polycyclic chains. *Trudy SAGU*, No. 7, 1946. (R)
129. On Markov – Bruns chains. Ibidem, No. 23, 1945, pp. 23 – 25. (R)
130. On some theorems concerning the method of least squares. *Doklady AN*, vol. 51, No. 4, 1946, pp. 259 – 262. (R)
131. On the Pearson curves. *Trudy IMM AN Uzbek SSR*, No. 1, 1946, pp. 3 – 13. (R)
132. On limiting distributions for stochastic processes with discrete time. *Trudy SAGU*, No. 4, 1945. (R)
133. On specifying the location of a target by straddling and recording the shell bursts. *Trudy IMM AN Uzbek SSR*, No. 2, 1947, pp. 3 – 10. (R)
134. Estimating the quality of production by progressive small samples. Ibidem, pp. 11-15. (R)
135. On some problems of long-term regulation of drainage. Ibidem, pp. 126 – 27. (R)
136. *Применение математической статистики в опытном деле* (Application of Mathematical Statistics in Experimentation).M. – L., 1947.
137. *Основные задачи теории ошибок* (Main Goals of the Theory of Errors). M. – L., 1947.
139. On estimating the successfulness of forecasting. *Doklady AN Uzbek SSR*, No. 8, 1948, pp. 3 – 7. (R)
140. Observations of unequal weight. Ibidem, No. 1, 1949, pp. 3 – 6. (R)
142. *Дискретные цепи Маркова* (Discrete Markov Chains). M., 1949.
143. On ordered samples from one and the same continuous population. *Trudy IMM AN Uzbek SSR*, No. 7, 1949, pp. 5 – 17. (R)
144. On the limiting distribution of relative frequencies in samples from a continuous distribution. *Юбилейный сборник, посвященный 25.летию Узбекской ССР* (Jubilee Coll. Articles Devoted to the 25th Anniversary of the Uzbek SSR). Tashkent, 1949, pp. 31 – 38. (R)
144a. On an implicit Markov lemma and on similar lemmas. *Doklady AN Uzbek SSR*, No. 4, 1949, pp. 3 – 5. (R)
145. On statistically checking the course of production by ordered samples. *Vestnik Mashinostroenia*, 1950. (R)
146. On the distribution of a target. *Artilleriisk. Zhurnal*, No. 12, 1950, pp. 20 – 24. (R)
147. On the best critical region for checking simple alternative hypotheses. *Doklady AN Uzbek SSR*, No. 5, 1950, pp. 3 – 6. (R)
148. On a method of double-level checking. In *Взаимозаменяемость и контроль в машиностроении* (Interchangeability and Checking in Machine-Building). M., 1950, pp. 167 – 181. (R)
149. Comparing consecutive samples from one and the same normal population by means and variances. *Doklady AN Uzbek SSR*, No. 3, 1951, pp. 3 – 5. (R)

150. On revealing unknown periods of periodic phenomena. *TrudyTashkent Geofizich. Observatoria*, No. 5 – 6, 1951, pp. 3 – 8. (R)

151. On statistical methods of checking manufacturing and the quality of the production. *Труды второго всесоюзного совещания по математической статистике* (Proceedings Second All-Union Conference on Mathematical Statistics). Tashkent, 1948, pp. 24 – 25. (R)

152. On straddling by observing short falls and shots over the target. *Trudy Akad. Artilleriisk. Nauk*, 1951. (R)

153. On statistical checking the production. Ibidem. (R)

154. Comparing hypergeometric, Bernoulli and Poisson probabilities. *Doklady AN Uzbek SSR*, No. 10, 1952, pp. 3 – 7. (R)

155. On duality theorems for the hypergeometric distribution. Ibidem. (R).

156. Same title. *Trudy IMM AN Uzbek SSR*, No. 11, 1953, pp. 22 – 28. (R)

158. Quartile criteria of belonging. Ibidem, pp. 3 – 11. (R)

159. On estimating whether a random sample belongs to a continuous distribution. Ibidem, pp. 12 – 15. (R)

9.V.I. Bortkevich (L. von Bortkiewicz).Accidents

Энциклопедич. словарь Брокгауз и Ефрон
(Brockhaus & Efron Enc. Dict.), halfvol. 40, 1897, pp. 925 – 930

In the most general sense, an accident is understood as any unforeseen event causing harm to life or property. Such an understanding of accidents seems to be too general for the relevant events to be usefully considered under a single head. Special attention, owing to their importance for both the public health and national economy, is due to accidents whose victims are people engaged as manual laborers in industrial enterprises (*accidents du travail, Betriebsunfälle*). By demanding legislative intervention, this group of accidents has lately become an independent object of study. There even exists a special *International Congress on Accidents.* It took place three times (Paris, 1889; Bern, 1891; Milan, 1894), and, beginning with this year, it is called *Congrès International des Accidents du Travail et des Assurances Sociales.*

Whether the study of accidents is an aim in itself, or caused by practical considerations (for instance, in connection with the appropriate insurance), the method of examination is always mostly *statistical.* This is quite natural since there hardly exists any other domain of facts where the action of the so-called *random causes* is felt just as clearly. A proper collection and organization of statistics of accidents of the indicated type (of the so-called *occupational accidents*) and its correct application is impossible without specifying beforehand the concept of occupational accident.

It may be defined as a bodily injury unwillingly and suddenly caused to a person working in a certain industrial{in a productive}enterprise by some external process (for example, by falling down from a high place) or conditions (e.g., by hightened air temperature) during work. The suddenness and external influence mainly serve as indications for separating accidents and diseases from each other. Events pathologically quite identical one with another become, or not become accidents depending on the type and the method of the action of their causes. Lumbago, for example, Márestaing says, should be considered an accident when proven that it was caused suddenly by a single abrupt effort during work. The same author indicates a number of other examples (rupture of a varicose vessel or a muscle, cases of frost-bitten limbs etc) whose attribution to an accident often becomes questionable. This is especially so when various causes were acting at the same time, some of them purely external, the other ones intrinsic, as a certain predisposition.

No lesser difficulties than those encountered by medical examination are met with when discerning those economic indicators that characterize the notion of occupational accident. Usually it is comparatively easy to decide whether the victim of the accident was engaged in the industrial enterprise; however, his relation with it (more precisely, with its owner) could well be understood in different ways. And it is much more often doubtful whether he had indeed been *working* at the moment of the accident. It is sometimes hardly possible to ascertain when does a certain man begin, and end his occupational activities. Especially difficult are accidents taking place en route, for instance to, or from the enterprise. An accident occurring during a break can also cause doubt. No general solution exists for such cases. It may be stated about each of them that they constitute a *quaestio facti.*

The definition above provides a guiding principle for solving such problems, but even only in principle the concept of occupational accident seems to be questionable. Some authors follow the indications stated above, others understand that concept in a narrower sense demanding that there exist a causal relation between a given accident and the victim's kind of work. According to the latter interpretation, an occupational accident may be recognized as such only when the danger, that brought it about, is exclusively, or in the highest measure (as compared with the conditions of ordinary life) peculiar to the kind of work of the victim. From that viewpoint, an employee, sent by the owner on an errand, even concerning business, and run over by a horse-car, is not a victim of an occupational accident. The *theory of occupational risk* based on that point of view considerably narrows the domain of the concept of occupational accident, and, therefore, of the sphere of the ensuing legal consequences (payment of recompense to victim and his family on the strength of contract or law).

The issue of whether the theory is true is connected with the law current in force and cannot therefore be solved independently. German laws concerning insurance against accidents provide no definite answer to that question and the practice of the *Reichsversicherungsamt*, the supreme legal instance on that kind of insurance, reveals numerous examples of conflicting decisions: first, they recognize the theory of occupational risk, then, they reject it as discordant with the will of the lawgiver.

Somewhat complete statistical data on accidents exist in those states which introduced pertinent compulsory insurance of considerable groups of the population. To these belong Germany and Austria. Special investigations of accidents were also made in other countries, for instance in Switzerland and France, and the compiled information deserves attention all by itself [1]. However, the probability of leaving accidents out is apparently lower when insurance is compulsory. Scientifically and practically important are, in addition, not the absolute, but rather the relative number of accidents, – the ratio of their absolute number to the number of people exposed to the danger of accidents. Consequently, we ought to know how many people are engaged in each branch of industry. And it is very desirable and almost necessary that the information on those exposed to the danger, and on the actual victims issues from the same source; otherwise, disparity in the counting of the same individuals or of those belonging to one and the same category will easily occur. This condition can best be fulfilled for statistics connected with insurance, and, because of the circumstances described above, we ought to restrict our attention to the results recorded in German and Austrian statistics.

In 1894, in the German industrial associations established for insuring against accidents (*gewerbliche Berufsgenossenschaften*) having 5.2 *mln* insured members, there were more than 190 thousand accidents, and, in the similar *landwirthschaftliche Berufsgenossenschaften*, 12.3 *mln* and about 70 thousand, respectively. To these, about 23 thousand industrial accidents in German enterprises managed by the state, the provinces and the communities (0.7 *mln* insured) should be added.

In accord with the gravity of their consequences, accidents are separated into several categories:

1. Mild accidents causing incapacity for work not more than for 13 weeks (in Austria, not more than for 4 weeks).
2. Those causing a longer but still temporary incapacity.
3. Accidents resulting in complete or partial permanent disability.
4. Fatal accidents.

I adduce the *relative numbers* of accidents of these categories; note that the number of accidents coincides with the number of victims. In this table [2], attention is turned first of all on the gradual increase in the relative numbers of accidents, both in their recorded totals and in those of the two last-mentioned categories. Such an increase is observed not only for all the insured taken as a single whole, but also for the separate branches of industry, in Germany as well as in Austria. The opponents of the system of compulsory insurance are apt to interpret this fact, not foreseen beforehand, as an argument favoring their viewpoint and attribute the increase to the action of the insurance. Even assuming, however, that confidence in a partial recompense for the harm caused by an accident can sometimes relax a worker's vigilance and prudence (mostly when handling machines and tools dangerous for life and health), we conclude, as von Mayr absolutely correctly remarked at the Milan Congress, that that premise speaks against insurance as such rather than against the compulsion.

Actually, the connection between insurance and the increase in the number of accidents cannot be considered as proven by statistical data. It is mush more probable that the increase is purely fictitious, that it may be explained by the improvement, over the years, of the system of recording accidents, and by the population's ever better understanding the institution of insurance and of the ensuing rights to demand recompense for the consequences of accidents.

Another possible cause of the relative increase of those recompensed in Germany may be seen in a transition from the narrow interpretation of an occupational accident to its wide definition according to which the presence of a special occupational danger connected with the victim's kind of work is not demanded anymore. In general, given the vagueness and relativity of the concept of occupational accident, not each distinction between statistical numbers concerning different countries or periods corresponds to the same distinction in real life.

In this respect, interesting are the facts reported by Greulich at the Milan Congress. As a result of a law passed in 1877 and imposing on manufacturers the duty to report each more or less serious accident, a continuous increase in their numbers was being noted in the Zürich canton during 1878 – 1883. Then, a directive demanding a stricter observance of the law was issued and the number of accidents in 1885 increased at once by 50%. The law of 1887 introduced a new, more favorable for the victims, procedure at investigations of civil responsibility, and proclaimed free legal hearings of cases dealing with recompense for the consequences of accidents, – and this fact was again reflected in the considerable increase in the pertinent statistical figures. And so, it is not proven, and neither is it likely that the introduction of compulsory insurance in Germany and Austria led to an actual increase in the number of accidents. And there are still less grounds for recognizing a causal relationship between the two facts.

The categories of occupational accidents listed above are vague and relative to the same extent and perhaps even more so. Thus, rather often sure objective indications of whether a victim is disabled forever or only temporary are lacking. Not less difficult is the exact distinction between complete and partial disability. Consequently, it might be thought that the variation of the numbers in columns 5 and 7 in the first part of Table 1 was occasioned by a gradual change of the views held by the authorities responsible for granting pensions to the victims.

Separate spheres of work considerably differ in the degree of danger for those involved. The relative numbers given above provide an idea about that difference between manufacturing industry and agriculture. German statistics furnishes relative numbers of accidents for each industrial G*enossenschaft* separately. Each of these associations covers industrial enterprises of a certain branch. Some branches are united in a single Imperial *Genossenschaft*, others spread over certain industrial regions and are divided between several associations. Here are some figures for 1894 giving an idea about the variation of the number of relative (per 1000 insured) accidents [3]. Note that those insured in a given *Genossenschaft* are often, according to their occupation, very heterogeneous. Were it possible to separate the totality of those insured and engaged in a definite branch of industry into several groups in accord with their kind of work, the relative number of accidents (*i.e.*, the coefficients of risk) for them would have considerably differed from each other. This is evident, for example, for employees of a railway and it should especially be especially borne in mind when comparing two nations, or different parts of one and the same country. In such cases the observed statistical variations are possibly caused by the different composition of the pertinent working populations that can also exist when the technology is the same because of the distinct economic conditions and forms of production and sale. Issuing from such considerations, Jottrand [4] put forward the demand that statistics of accidents collect workers into groups not covering one or another branch of industry, but by their kind of work (miners, foundry workers, […], – in all, 33 categories). Austrian statistics meets that demand to a certain extent; it occurs, for instance, that among builders the coefficient of risk for roofers is considerably higher than for carpenters. Indeed, in 1893 the pertinent figures were […]

The investigation of the *causes* of accidents is of special interest. A cause is here understood as either those technical structures or tools which led to the accident while working on/with them, and those external processes which directly caused the bodily harm, or certain personal attitudes of the victims and their employers to the conditions of occupational danger causing the accident. In the following tables [5] 100 grave accidents that occurred in Germany are distributed in accord with their main causes [6]. Lines 1 – 3 in Table 4 taken together show accidents that ought to be blamed on the employers and managers whereas lines 4 – 8 cover the cases having occurred through the workers' fault. Line 9 illustrates accidents of which both the employers and the workers were guilty. When being described in such a manner, the indications of the table become very valuable both for the lawgiver and those responsible for preventing accidents by adopting appropriate general measures.

Notes

1. The general statistics of accidents is not trustworthy. For example, the French official statistics showed 1959 accidents having occurred during 1885 – 1887 in the coalmines whereas a special investigation covering the same time period and not even extending over all the mines gave 48,344 cases, *i.e*. 24 times more.

2.{Table 1 provides the yearly numbers of accidents per thousand insured for German industrial (1887 – 1895) and agricultural *Genossenschaften* and for Austria (1891 – 1894). Its columns 5 and 7 of its first part, mentioned below, concern Germany and show a steady decrease of accidents resulting in a complete permanent disability and a steady increase of those leading to temporary disability.}

3.{Table 2 lists 10 *Genossenschaften* covering various industrial branches and provides the number of grave accidents and the total number of them in each of the 10 cases. Here are a few figures from this table. The Nordrhein-Westfalen blast-furnace and rolling industry: 145.3 accidents (the highest figure), 9.9 of them grave. The carrier's trade (Fuhrgewerbe?), 13.8 grave accidents (the highest number) with the total number of them being only 44.1.}

4. He thoroughly remarks:

A large number of accidents in breweries (14% in 1887 according to German data) is occasioned by carting, 18%, by loading, unloading and carrying barrels. The coefficient of risk will consequently be quite different for breweries selling beer on the spot and those having a vast region of customers; for those carting beer and breweries sending it by railway.

5. {Table 3 provides separate data on industrial (1887) and agricultural (1891) enterprises in connection with 16 different structures, procedures etc. For example, Item 3, machinery in the proper sense except for lifting gear in industrial enterprises: 17.55% of all the accidents, 3.21% of all the accidents, 3.21 % of that per cent proving fatal; Item 13, water-borne transportation, 0.99% and 74.05% respectively. Table 4 shows figures for the same two kinds of enterprises and years and lists 13 items connected with the second main cause. Thus, lack of safety structures in agricultural enterprises (Item 2), 10,64% and 11.35% respectively; thoughtlessness and intoxication (Item 6), 1.98% and 1.51% respectively.}

6. However interesting are these data in themselves, they should be interpreted very cautiously. In logging, Jottrand indicates,

459 accidents were caused by machines, 273 of them by circular saws and only 17 by band-saws. Are we therefore justified in recommending the latter as comparatively less dangerous instead of the former? Not at all since it is likely that the difference between the numbers was simply caused by the circular saw being much more in use than the band-saw.

In general, it would be delusive to regard the numbers in the first column as indicators of the appropriate risks.

References

On the legal aspect of occupational accidents:

1. *Handbuch der Unfallversicherung*, 2nd edition. Leipzig, 1896, pp. 28 – 78.
2. Rosin, *Das Recht der Arbeiterversicherung*, Bd. 1, 1893, pp. 273 – 290.
3. Piloty, R. *Das Reichs-Unfallversicherungsrecht*, Bd. 1, 1891, pp. 184 – 213.
4. Menzel, Adolf, *Die Arbeiterversicherung nach österreichischem Rechte.* Leipzig, 1893, pp. 279 – 295.

On the methods and results of the statistics of occupational accidents:

1. *Amtliche Nachrichten des Reichs-Versicherungsamt.* Berlin, from 1885. Special statistics of industrial *Genossenschaften* for 1887 is in its Sechster Jahrgang (1890, p. 199ff). Similar statistics of agricultural *Genossenschaften* for 1891 is in its Neunter Jahrgang (1893, p. 231ff).
2. *Statistisches Jahrbuch für das Deutsche Reich*, beginning with its 9th Jahrgang (1888) contains brief information on the same subject.
3. *Die Gebärung und die Ergebnisse der Unfallstatistik der Arbeiterunfallversicherungs-Anstalten.* Wien. Data for Austria.
4. *Congrès International des Accidents du Travail. Exposition universelle de 1889.* Paris, 1889 and 1890. *Congrès* …, 2nd session, Bern, 1891. *Congrès* …, 3rd session, Milan.
5. Bödiker, T. *Die Arbeiterversicherung in den europäischen Staaten*, 1895.
6. Yarotskij, V.G. *Страхование рабочих в связи с ответственностью предпринимателей* (Insurance of Workers in Connection with the Responsibility of the Owners). Psb, 1895.

10. Anderson, O. Letters to Karl Pearson

Unpublished; kept at University College London, Pearson Papers, NNo. 442 and 627/2

Foreword by Translator

Oskar (Nikolaevich) Anderson (1887 – 1960) was Chuprov's student and the last representative of the Continental direction of statistics. Little known information about him is in Sheynin (1996, §7.8). There also, in §15.6, the reader will see that on June 9, 1925, Anderson had sent Pearson a manuscript and that Pearson at once agreed to publish it in *Biometrika*. Note that the letters below were written later.

I am grateful to University College London for allowing me to publish the following letters which are kept by them (Pearson Papers NNo. 442 and 627/2). Both letters devoted to the variate difference method are from Anderson who studied it from 1911 (when he wrote his dissertation on the coefficient of correlation and its application to secular series) onward and, together with Student (Gosset), was its co-creator. In particular, two from among his papers that appeared in *Biometrika* (1923; 1926 – 1927) treated the same issue. In the second of these, Anderson (p. 299/45) briefly outlined the history and the essence of that method.

His letters below precede the second *Biometrika* paper; the second letter makes it clear that Pearson had tentatively agreed to grant Anderson 15 – 20 pages for publishing his not yet completed manuscript. Actually, however, the article of 1926 – 1927 occupied some 60 pages which testifies that Pearson highly appreciated it. Then, Anderson desired to see his manuscript translated from German to English, but that did not happen, perhaps owing to its great length.

I believe that the publication of the Anderson letters, in spite of the existing *Biometrika* article, will not be amiss. His *Ausgewählte Schriften*, Bde 1 – 2, were published in Tübingen in 1963 and his latest biography written by Heinrich & Rosemarie Strecker is in *Statisticians of the Centuries* (2001), Editors C.C. Heyde, E. Seneta. New York, pp. 377 – 381.

* * *

Letter No. 1

Professeur Oskar Anderson, Varna, Ecole Supérieure de Commerce (Bulgarie)
Varna, den 27. November 1925

Sehr geehrter Herr Professor!

In Ihrer Publikation (Pearson & Elderton 1923, p. 308) ist es zu lesen:

We think it safe to say that there really does exist a substantial <u>*negative*</u> *correlation between deaths of the same group in the first and second years of life. It is not as great as we found it in the previous paper using hypotheses, which, we admit, ought to have been tested; but it is quite adequate to indicate that natural selection is really at work.*

Ich glaube beweisen zu können, daß Sie und Frl. E.M. Elderton hier der W. Person'schen{Persons'schen}Kritik gegenüber *unnützerweise nachgegeben haben* und daß Ihre ursprünglichen Koeffizienten, welche Sie in der Arbeit Pearson & Elderton (1915) veröffentlicht haben, denjenigen der eingangs erwähnten Schrift *jedenfalls in nichts nachstehen und ihnen zum mindesten ganz ebenbürtig an die Seite gestellt werden können.*

Wie Sie sich erinnern werden, haben Sie in der letzteren zur ursprünglichen Methode Ihres Biometr. Laboratoriums zurückgegriffen (1923, p. 284):

To fit high order parabolae to both variates and then take the differences between the ordinates of these parabolae and the observed data for x and y.

Und zwar benutzten Sie dabei “the Rhodes’ and Sheppard’s systems of smoothing”. Worauf beruht denn eigentlich dieses so populäre Verfahren, welches der Variate-Difference-Methode entgegengestellt wird? Doch offenbar darauf, daß man voraussetzt, daß gerade der “ausgeglichene“ (smoothed) Wert der gegebenen evolutorischen Variablen deren “glatte“ Komponente genau wiedergiebt [1].

Es sei $u_1, u_2, u_3, ..., u_N$ eine evolutorische Reihe, deren jedes Glied aus einem “glatten“ Element G und einem restlichen “zufälligen“ (random) Element x bestehen möge, so daß

$$u_i = G_i + x_i.$$

Soviel ich übersehen kann, wird bei allen Ausgleichungsmethoden (Periodogramm und kleinste Quadrate mit eingerechnet) u_i', d. h. der “ausgeglichene“ Wert von u_i, durch eine mehr oder weniger komplizierte *lineare* Funktion einer Anzahl seiner Nebenglieder dargestellt (oder auch durch eine Funktion von deren Gesamtzahl):

$$u_i' = F(u_{i-j}; u_{i-j+1}; \ldots; u_i; u_{i+1}; \ldots; u_{i+r}),$$

oder kurz $u_i' = F(u_{j,\,i,\,r})$.

(Es werden ja bekanntlich bei Ausgleichungsrechnungen sogar alle nichtlineare Relationen zwischen den beobachteten und den unbekannten Größen durch bestimmte Kunstgriffe in angenäherte lineare verwandelt.)

Ist $F(u_{j,\,i,\,r})$ linear, so kann man offenbar setzen:

$$u_i' = F(u_{j,\,i,\,r}) = F(G_{j,\,i,\,r}) + F(x_{j,\,i,\,r})$$

und daher

$$u_i - u_i' = G_i - F(G_{j,\,i,\,r}) + x_i - F(x_{j,\,i,\,r}).$$

Folglich führt die *übliche Hypothese* $u_i - u_i' = x_i$ zwangsläufig zur Annahme

$$G_i - F(G_{j,\,i,\,r}) - F(x_{j,\,i,\,r}) = 0, \text{ oder } G_i = F(G_{j,\,i,\,r}) + F(x_{j,\,i,\,r}).$$

Ich kann nun aber absolut nicht einsehen, warum das “glatte” Element G unter anderem auch durch eine lineare Funktion seiner *zufälligen* “*Beobachtungsfehler*” ausgedrückt werden sollte. Man wollte ja gerade durch die Ausgleichung sich von deren Einfluss befreien! Ich glaube vielmehr, daß man ohne Bedenken nur

$$G_i - F(G_{j,\,i,\,r}) = 0$$

setzen kann, und daß folglich *das restliche Glied* $u_i - u_i'$ nicht durch x_i sondern durch $[x_i - F(x_{j,\,i,\,r})]$ *darzustellen ist.* Schliesslich kann man ja auch die mathematische Erwartung nur dieses Ausdruckes gleich Null setzen, keinesfalls aber diejenige von x_i an und für sich.

Man darf also durchaus nicht die mathematische Erwartung von $(u_i - u_i')\,(u_j - u_j')$ gleich derjenigen von $x_i\, x_j$ setzen, oder den Korrelationskoeffizienten zwischen zwei *zufälligen* Komponenten x_i und y_i zweier evolutorischer Reihen U und S als gleich dem Koeffizienten zwischen $(u_i - u_i')$ und $(s_i - s_i')$ betrachten.

Hier sind, im Allgemeinen, ganz analoge Korrektionen anzubringen, wie Sie, Herr Professor, diese beim Falle der Variate-Diff. Methode auf S. 309 Ihrer [Pearson & Elderton (1923)] ganz richtig angedeutet haben. Ein konkretes Beispiel. Auf S. 295 – 296 von Whittaker & Robinson (1924) finden wir die Sheppard'schen Ausgleichungsformeln für Parabeln bis zur 5. Ordnung und für Anzahl der bei der Ausgleichung benutzten Glieder von 3 bis 21 (d.h. bis $n = 10$).

Nehmen wir an: Ausgleichungsparabel 2ter Ordnung; $n = 2$. Dann ist

$$u_o - u_o' = u_o - (1/35)[17u_o + 12(u_1 + u_{-1}) - 3(u_2 + u_{-2})] =$$
$$(1/35)[18u_o - 12(u_1 + u_{-1}) + 3(u_2 + u_{-2})] =$$
$$(3/35)[u_{-2} - 4u_{-1} + 6u_o - 4u_1 + u_2] = (3/35)\delta^4 u_o$$

wenn man δ^4 für zentrale 4te Differenz setzt.

Es ist also, in mir geläufigeren Bezeichnungen,

$$u_i - u_i' = (3/35)\Delta^4 u_{i-2}.$$

Wird, folglich, angenommen dass in $(u_i - u_i')$ der letzte Rest einer G-Komponente verschwunden ist, so läuft das auf die Annahme

$$u_i - u_i' = (3/35)\Delta^4 x_{i-2}$$

hinaus; *man erhält also hier, im Grunde genommen, eine Formel der Variate-Diff. Methode wieder*!

Greifen wir das selbe Problem von einer anderen Seite an. Die geläufigen Ausgleichungsmethoden werden gewöhnlich aus bestimmten Hypothesen über die Beschaffenheit der *glatten* Komponente G abgeleitet, und die *zufällige* x-Komponente wird eben als *Beobachtungsfehler* betrachtet. Wenn aber unser Interesse auf der x-Komponente konzentriert wird, so dürfte gerade das entgegengesetzte Verfahren am Platze sein. Von diesem Standpunkt ausgehend, wollen wir versuchen, *die Variate-Difference Methode als ein neues Ausgleichungsverfahren auszubauen.*

Es sei wieder

$$u_1 = G_1 + x_1,\ u_2 = G_2 + x_2,\ ...,\ u_N = G_N + x_N\,;$$

und Δ^{2k} bedeutet die $(2k)$-te endliche Differenz, so daß

$$\Delta^{2k}x_i = x_i + 2kx_{i+1} + C_{2k}^{\ 2}\, x_{i+2} + ... + C_{2k}^{\ k}\, x_{i+k} + ...$$

Nehmen wir nun an, daß bei dieser $2k$-ten Differenz die evolutorische Komponente G endgiltig verschwunden ist. Dann ist

$$u_i - u_i' = z\Delta^{2k}u_{i-k} = z\Delta^{2k}x_{i-k}.$$

Hier bedeutet u_i' den nach unserem Ausgleichungsverfahren zu bestimmenden *genauen* Wert von G_i, und z – einen vorläufig unbestimmten Multiplikator, dessen *vorteilhaftester* Wert noch gefunden werden muß. Da nun

$$u_i' = G_i + x_i - z\Delta^{2k}x_{i-k}\,,$$

so kann dieser vorteilhafteste Wert von z (analog dem Verfahren der Methode der kleinsten Quadrate) der Bedingung

$$E(x_i - z\Delta^{2k}x_{i-k})^2 = \text{minimum}$$

unterworfen werden. Das symbol E bedeutet hier mathematische Erwartung.

Wenn x eine zufällige Variable ist, so daß $Ex_i = \text{Const}$, $Ex_i^2 = \text{Const}$, $Ex_i\, x_j = Ex_i\, Ex_j$ $(j \neq i)$, so verwandelt sich obige Gleichung[2] in

$$Ex_i^2 - 2zE(x_i\Delta^{2k}x_{i-k} + z^2E(\Delta^{2k}x_{i-k})^2$$

oder

$$Ex^2 - 2z\,C_{2k}^{\ k}\,E(x_i - Ex)^2 + z^2\,C_{4k}^{\ 2k}\,E(x_i - Ex)^2.$$

Die erste Ableitung nach z ergebt

$$-2\,C_{2k}^{\ k}\,E(x_i - Ex)^2 + 2z\,C_{4k}^{\ 2k}\,E(x_i - Ex)^2.$$

Setzt man sie gleich Null, so erhält man daraus

$$z = [(2k)!]^3/[(4k)!k!k!].$$

Da die 2te Ableitung positiv ist, so ist der hier gefundene Wert von z ein Minimum. Es ist also

$$u_i' = u_i - \{[(2k)!]^3/[(4k)!k!k!]\}\Delta^{2k}x_{i-k} = u_i - \{[(2k)!]^3/[(4k)!k!k!]\}\Delta^{2k}u_{i-k},$$

oder, *in zentralen Differenzen ausgedrückt,*

$$u_o' = u_o - \{[(2k)!]^3/[(4k)!k!k!]\}\delta^{2k}u_o.$$

Das wäre also diejenige Ausgleichungsformel, welche der Anwendung der Variate Difference Methode entspricht. Setzt man hier $k = 1, 2, 3, \ldots$ ein, so erhält man unmittelbar

bei $k = 1$, $u_o' = (1/3)[u_o + (u_1 + u_{-1})]$,
bei $k = 2$, $u_o' = (1/35)[17u_o + 12(u_1 + u_{-1}) - 3(u_2 + u_{-2})]$,
bei $k = 3$, $u_o' = (1/231)[131u_o + 75(u_1 + u_{-1}) - 30(u_2 + u_{-2}) + 5(u_3 + u_{-3})]$,
u.s.w.

Wir haben also für die "*Variate-Difference*"*-Ausgleichungsmethode genau dieselben Koeffizienten gefunden, welche die Sheppard'sche Ausgleichung in dem Falle ergiebt, wenn man sein n der Ordnung seiner Parabel* (*oder der Hälfte der Ordnung unserer Differenz*) *gleich setzt*!

Dieses Ergebnis kam für mich seinerzeit recht unerwartet, obwohl es ja ohne besondere Schwierigkeiten aus Sheppard's Formeln abgeleitet werden kann. Ob es sonst bekannt ist, kann ich nicht beurteilen, da für mich leider die Sheppard'sche Arbeit bis jetzt unerreichbar geblieben ist, und ich überhaupt hier in Varna in sehr wenige Werke der engl. und amerik. wissenschaftlichen Litteratur Einsicht bekommen kann.

Jedenfalls, glaube ich erwiesen zu haben, dass es nicht angeht, der Var. Diff. Methode die Bestimmung der *zufälligen Residualen* x_i gerade nach dem Sheppard'schen Verfahren gegenüberzustellen. Entweder sind bei letzterem genau dieselben Korrektionen, wie bei ersterer, anzubringen, oder aber muß man die von der evolutorischen Komponente befreiten $2k$-ten Differenzen als mit einem gewissen konstanten Faktor multiplizierte *wahre* Werte von x ansehen und folglich an der Existenz einer mehr oder weniger beträchtlichen Korrelation zwischen $\Delta^{2k}x_i$ und $\Delta^{2k}x_{i+j}$ (oder bzw. zwischen $\Delta^{2k}x_i$ und $\Delta^{2k}y_{i+j}$) keinen Anstoß nehmen. Ich für meinen Teil entscheide mich natürlich für erstere Alternative.

Leider konnte ich mir hier die Rhodes'sche Arbeit ebenfalls nicht verschaffen und bin daher außerstande sein Verfahren von meinem Standpunkte aus zu kontrollieren. Ich halte es aber für sehr wahrscheinlich, daß auch letzteres zur selben Klasse der Ausgleichungsmethoden gehört, die ich am Anfang dieses Brief erwähnte. Ich möchte daher annehmen daß

1) Ihre $R_{x_p y_p}$ auf S. 308 von Pearson & Elderton (1923) Korrektionen zu erhalten haben, die Sie, wahrscheinlich, den durch die *Var. Diff. Meth.* erbrachten Resultaten näherbringen würden; daß

2) Ihre Korrelationskoeffizienten sowie für X_p, X_{p+i} (ebenda S. 303), als auch für X_p, $Y_{p\pm i}$ (S. 305) sehr wohl *spurious* sein *können*, und daher ohne vorherige Korrektion nicht der Kommentare auf S. 306 – 307 bedürfen und nicht gegen die Anwendbarkeit der Var. Diff. Methode zeugen können; und daß

3) Wenigstens bei Anwendung einer anderen Sheppard'schen Ausgleichungsformel Sie auch zu mit der Var. Diff. Methode identischen Resultaten gelangen könnten.

Bedürfen letztere nicht der Korrektionen, die Sie auf S. 309 andeuten? Ich bin imstande, auch auf diese Frage zu antworten, möchte aber hier Ihre Zeit nicht zu lange mit meinem Brief in Anspruch nehmen. Nur soviel sei gesagt, daß ich den genauen Ausdruck für die *Stand. dev.* der Differenz

$$\sigma_k^2 - \sigma_{k+1}^2 = \sum_{i=1}^{N-k} \frac{\Delta^k x_i^2}{C_{2k}{}^k (N-k)} - \sum_{i=1}^{N-k-1} \frac{\Delta^{k+1} x_i^2}{C_{2k+2}{}^{k+1}(N-k-1)},$$

bestimmt habe. Bei kleinerem N (also auch in Ihrem Falle) ist letztere doch derart, daß man nicht anzunehmen braucht, diese Differenz sei gleich Null. Auch für eine reine *random series* kann dann die Reihe σ_k^2, σ_{k+1}^2, σ_{k+2}^2, ... sehr wohl allmählich ansteigen oder, im Gegenteil, langsam abfallen. Desgleichen, natürlich auch die Reihe p_k, p_{k+1}, p_{k+2}, ... wenn hier p_j für

$$\sum_{i=1}^{N-j} \Delta^j x_i \Delta^j y_i / C_{2j}{}^j (N-j)$$

steht. Wenn aber die Reihe p_i, von einem gewissen k angefangen, wirklich stabil geworden ist, so kann hierauf dasjenige Theorem angewandt werden, welches ich in einer Fußnote auf S. 146 meines Artickels (1923) andeutete, d.h. je länger die *konstante* Reihe p_k, p_{k+1}, p_{k+2}, … desto wahrscheinlicher kann man annehmen, daß alle Korrelationskoeffizienten zwischen x_i und $y_{i\pm j}$ dem Nullpunkt recht nahe zu stehen kommen (natürlich, ausgenommen $x_i y_i$).

Um zum Schlusse zu gelangen. Meine Ansicht über die Variate Difference Methode geht jetzt dahin, daß diese auch als ein regelrechtes Ausgleichungs-verfahren angesehen werden kann, welches zum Teil mit dem Sheppard'schen zu identifizieren ist. Den einen Vorteil der ersteren haben Sie sehr richtig auf S. 284 [Pearson & Elderton (1923)] angegeben: die Leichtigkeit, mit der man feststellen kann, ob die evolutorische Komponente wirklich schon als eliminiert angesehen zu werden vermag. Dazu kommt die Sicherheit der Kontrollen, da für alle wichtigeren Mittelwerte, Momente und Kriterien jetzt die *Stand. Deviationen* bestimmt sind. Ferner erlaubt diese Methode tiefer in die gegenseitigen Verhältnisse der einzelnen Komponenten einer evolutorischen oder säkularen Reihe einzudringen, als es bei anderen Verfahren bis jetzt der Fall gewesen ist. Was bei ersterer offen zu Tage liegt, wird bei den anderen Verfahren manchmal nur im Stillen hineingeschmuggelt (siehe erste und zweite Seite dieses Briefes) [3].

Obige Ausführungen bilden eine gedrängte und unvollkommene Darstellung einiger Ergebnisse meiner neuen Untersuchung über die Var. Diff. Methode, an der ich im Sommer gearbeitet habe und die ich zwischen Weinachten und Ostern zu schließen hoffe (ca. 3 – 4 Druckboden). Ich werde sehr durch die Berechnung von verschiedenen numerischen Beispielen aufgehalten, welche ich ohne fremde Hilfe auszuführen gezwungen bin. Ich habe einige hypothetische Beispiele konstruiert, aber auch eine Anzahl konkreten Reihen untersucht, wobei ich manchmal zu recht interessanten Ergebnissen gelangt bin. Unter anderem, ist es mir bis jetzt, trotz aller Bemühungen, nicht gelungen, eine wirklich *zackige* Reihe zu finden, die den kurzperiodischen Sinus-Reihen Yule's (1921) einigermassen entsprochen hätte. Ich habe aber dennoch einige Methoden ausgearbeitet, die es erlauben, eine wirklich vorhandene schädliche z-Komponente zu eliminieren.

Ferner habe ich die recht interessanten Zusammenhänge untersucht, die zwischen dem Lexis – Bortkiewicz'schen Divergenzkoeffizienten [4] Q^2 und der Var.-Diff. Methode bestehen, und dabei auch die Frage über die Konstruktion von aus verschiedenen Komponenten zusammengesetzten evolutorischen Reihen beleuchtet, speziell über den Zusammenhang zwischen meinen z-Reihen und der Lexis'schen übernormalen Stabilität.

Ich hoffe, die Arbeit irgendwo mit Hilfe unseres Verbandes russischer akademischer Lehrer im Auslande (dessen Vorsitzender ist Prof. Vinogradoff, Oxford) unterbringen zu können, doch sind die Aussichten noch ungewiss. Das Manuskript in Bulgarischer Sprache zu drucken hat, natürlich, keinen Sinn.

Ich bitte sehr um Entschuldigung, daß ich Ihre Zeit mit diesem Briefe in Anspruch nehme. Ich möchte aber dadurch einerseits mein Prioritätsrecht auf die hier dargelegten wissenschaftlichen Ideen sichern und, andererseits, tröste ich mich damit, dass Sie, Herr Professor, in den behandelten Fragen teilweise auch engagiert sind und daher vielleicht für Sie einiges Interesse Sich abgewinnen könnten.

Mit vorzüglicher Hochachtung, Ihr Oskar Anderson

Letter No. 2

Prof. Oskar Anderson, Ecole Supérieure de Commerce, Varna, Bulgarie
Varna, den 10. Dezember 1925

Sehr geehrter Herr Professor,

Soeben erhielt ich Ihren w. Brief vom 4.XII und beeile mich ihn zu beantworten. Meines Wissens, ist es dem Vorsitzenden des Verbandes Russischer Akadem. Lehrer im Auslande, Prof. Vinogradoff (Oxford), gelungen, von einigen Akademien der Wissenschaften (Oslo, Rom, London, ...) zu erreichen, daß sie sich bereit erklärt haben, gemäß seiner Rekommandation auch Schriften russischer Autoren zum Drucke anzunehmen. Eben diese Möglichkeit hatte ich im Sinn, als ich letztens an Sie schrieb. Da jedoch Prof. Vinogradoff möglicherweise (was ihm niemand verargen kann) die Schriften seiner Kollegen vom Fach, also Historiker und Juristen, bevorzugen könnte, und ich auch sonst befürchte, daß dank dem großen Andrang von anderen *Konkurrenten* der Druck meiner Arbeit Gefahr liefe, sich stark zu verzögern, so wäre ich natürlich sehr froh, wenn Sie dieselbe wirklich für die *Biometrika* annehmen wollten. Diese Zeitschrift hat Weltruhm und zudem dürfte ihr Auditorium mit den Grundideen der *Variate Difference*-Methode wohl bekannt sein. Das würde mir erlauben, mich kürzer zu fassen. Ob ich meine Arbeit in jene 15 – 20 Seiten, die Sie mir eventuell anbieten, hineinzwängen kann, – weiß ich noch nicht, will es aber versuchen. Jedenfalls, werde ich das ganze Material in kurze und möglichst selbstständige Abschnitte einteilen und es Ihnen überlassen, das eine oder das andere davon zu streichen.

Das Gestrichene werde ich dann trachten, in Norvegen oder Italien unterzubringen. Mit Ihrer Publikation Pearson & Elderton (1915) will ich bei Raummangel mich auch möglichst wenig befassen und werde mich nur mit einigen kurzen Bemerkungen begnügen. Ich hoffe, dass Sie, Herr Professor, es vielleicht für angebracht halten werden, in einer Vor- oder Schluss-Bemerkung zu meinem Artickel ihren jetzigen Standpunkt in der Frage darzulegen. Sind Sie mit mir einverstanden, so stehen Ihnen ja zu Korrektions-Berechnungen die Kräfte Ihres Laboratoriums zur Verfügung! Desgleichen werde ich auch keine Korrektionen zum Rhodes'schen oder anderen *Ausgleichungsmethoden* deduzieren.

Ich hoffe, noch bis Anfang Februar Ihnen meine Schrift in neuer Redaktion zuschicken zu können und werde dann mit Ungeduld auf Ihre endgiltige Entscheidung warten. Letzten Endes bleibt mir ja noch die von Ihnen zugestandene Möglichkeit im Mai-Heft der *Biometrika* nur den Text meines Briefes vom 27. November zu veröffentlichen.

Zum Schluss möchte ich Ihnen hier noch eine Frage vorlegen, welche Ihnen vielleicht gänzlich verfrüht erscheinen wird, für mich aber schon jetzt von einem gewissen aktuellen Interesse ist, die Frage über die Sprache nämlich. Fast jeder Russische oder Deutsche Gelehrte liest Englisch, aber relativ nur wenige Angel-Sachsen und Franzosen verstehen deutsch oder wollen deutsch verstehen. Daher wäre es für mich sehr wichtig, wenn mein zukünftiger Artickel (falls er wirklich von Ihnen angenommen werden sollte) in englischer Sprache erschiene. Wenn ich die Übersetzung selber besorge, so kostet es mich verhältnißmäßig viel Zeit und – die Hauptsache – werde ich doch nicht sicher sein, ob mir nicht irgendwo etliche lächerliche Russismen untergelaufen sind. Auf der 4. Seite eines jeden *Biometrika*-Heftes steht:

Russian contributors may use either Russian or German but their papers will be translated into English before publication.

Was muss ich nun tun, um eine solche Übersetzung zu veranlassen? Ist das eine Geld-Frage?

Mit vorzüglicher Hochachtung Ihr O. Anderson

Notes

1.{*Evolutionary* series is a term likely coined by Lexis (1879, §1). Nowadays, we would say, a series possessing a trend.}

2.{Anderson obviously had in mind the *Bedingung* formulated above.}

3.{The second page of Anderson's manuscript ended after the formula for $(u_o - u_o')$.}

4.{Bortkiewicz (and, even more, Markov and Chuprov) had indeed studied the Lexian theory, but the coefficient Q (later replaced by Q^2) was due to Lexis alone. In his *Biometrika* paper Anderson (1926 – 1927) did not anymore mention Bortkiewicz in that connection. A few lines below, Anderson once more refers to Lexis, – to his theory of the stability of statistical series. It is strange, however, that neither here, nor in the published paper did he mention Chuprov's refutation (1918 – 1919) of the Lexian reasoning.}

References

Anderson, O. (1923), Über eines neues Verfahren bei Anwendung der "Variate-Difference" Methode. *Biometrika*, vol. 15, pp. 134 – 149 and 423. Reprint: Anderson (1963, Bd. 1, pp. 12 – 27).

--- (1926 – 1927), Über die Anwendung der Differenzenmethode (variate difference method) bei Reihenausgleichungen, Stabilitätsuntersuchungen und Korrelationsmessungen. *Biometrika*, vols 18, pp. 293 – 320 and 19, pp. 53 – 86. Reprint with Anderson's handwritten corrections: Anderson (1963, Bd. 1, pp. 39 – 100).

--- (1963), *Ausgewählte Schriften*, Bde 1 – 2. Tübingen. Hrsg. H. Strecker.
Chuprov, A.A. (1918 – 1919), Zur Theorie der Stabilität statistischer Reihen. *Skand. Aktuarietidskr.*, Bd. 1, pp. 199 – 256; Bd. 2, pp. 80 – 133. Signed Tschuprov.
Lexis, W. (1879), Über die Theorie der Stabilität statistischer Reihen. *Abh. zur Theorie der Bevölkerungs- und Moralstatistik.* Jena, 1903, pp. 170 – 212.
Pearson, K., Elderton, E.M. (1915), Further evidence of natural selection in man. *Biometrika*, vol. 10, pp. 488 – 506.
--- (1923), On the variate difference method. *Biometrika,*vol. 14, pp. 281 – 310.
Sheynin, O. (1996), *Chuprov.* Göttingen.
Whittaker, E.T., Robinson, G. (1924), *The Calculus of Observations.* London.
Yule, G.U. (1921), On the time-correlation problem, with especial reference to the variate-difference correlation method. *J. Roy. Stat. Soc.*, vol. 84, pp. 497 – 526.

11. Ya. Mordukh.
On Connected Trials Corresponding to the Condition of Stochastic Commutativity

Unpublished; kept at University College London, Pearson Papers, NNo. 442 and 627/2 …

Foreword by Translator

Hardly anything is known about Mordukh, a former student of Chuprov, who emigrated from Russia. Chuprov [7, pp. 60 – 61] mentioned him most favorably (*our economist* […] *graduated from Uppsala University and moved to Dresden* {where Chuprov then lived} *to be my student*) and noted his mathematical talent. However, it also followed from Chuprov's correspondence (Ibidem) that Mordukh had not found a position for himself. Answering my inquiry, Anders Molander (Archives of that University) stated on 21 Jan. 2000 that Jacob Mordukh, born 4 July 1895, was matriculated there on 20 Jan. 1919 and graduated (apparently, from the Philosophical Faculty) on 14 Sept. 1921 as Bachelor of Arts. His subjects were Slavonic languages, mathematics and statistics. Nothing else is known about him.

Below, in the article now translated, Mordukh enlarges on Chuprov's discovery of exchangeability, as it is called now; see [6], where the latter's contribution is highly appraised (pp. 246 and 253 – 255). I adduce a few remarks about terminology. First, Mordukh used the terms *disjunction* and *conjunction* as well as the appropriate verbs and participles. Second, he wrote *law of distribution of the values* … just as Chuprov did. Third, Mordukh preferred *random variable* to the now standard Russian term *random quantity*. Here also, he apparently took after Chuprov who nevertheless wavered between these two terms [7, §15.4]. Fourth and last, I replaced Mordukh's unusual *parameters-h* by *parameters* $\{h\}$.

* * *

1. Among the constructions of the mathematical theory of probability applied for justifying the methods of statistical investigations, two patterns claim to be considered as the cornerstones: the scheme of independent trials usually illustrated by extracting balls from an urn with their immediate return; and the pattern of dependent trials corresponding to the case of unreturned balls. The exposition below keeps to a more general formulation. Suppose that a closed urn contains S tickets, s_1 of them marked with number $x^{(1)}$, s_2 balls with $x^{(2)}$, …; and s_k, with the number $x^{(k)}$. Suppose also that N tickets are extracted. If the drawn ticket, after its number is recorded, is returned to the urn before the next extraction takes place, we have the layout of the returned ticket; otherwise, we speak about the scheme of the unreturned ticket.

The transition from the second pattern and its appropriate formulas to the first one is known to present no difficulties. The scheme of the returned ticket can formally be considered as the limiting case of the other one: the greater is the number S of the tickets in the urn, the weaker is the connection between the separate trials so that in the limit, as $S = \infty$ (and with a finite number of trials, N), the pattern of dependent trials passes on to that of independent trials. Suppose for example that the urn contains s white and s black balls. The probability of drawing in two consecutive extractions, first a white ball and then a black one, will be equal to $P = (1/2)\cdot(1/2) = 1/4$ when the ball is returned; and to $P_s = (s/2s)\cdot[s/(2s-1)] = (1/2)\cdot\{1/[2-(1/s)]\}$ when the ball is not returned, and we find that $P = \lim P_s$ as $s = \infty$.

This derivation for the pattern of the returned ticket is not really interesting because the establishment of the formulas for the other scheme is much more difficult. And, on the contrary, it would be somewhat beneficial to have the possibility of an inverse transition from the formulas of the arrangement of the returned ticket to the more complicated relations valid for the other pattern. This problem was not apparently posed before although it admits of a very simple solution interesting not only because it simplifies the derivation of formulas for the layout of the non-returned ticket, but also because it leads to important theoretical constructions.

2. N trials are made on a random variable x that can take values $x^{(1)}, x^{(2)}, \ldots, x^{(k)}$, with probabilities $p_1, p_2, \ldots, p_k$ respectively and the law of distribution of this variable remains fixed all the time. Considering all these trials together as a single whole, we shall call them a *system* if the separate trials somehow depend one on another, and a *totality* if they are mutually independent.

Denote the random empirical values taken by the variable x at the first, the second, …, the N-th trial by $x_1, x_2, \ldots, x_N$. Let $\mathrm{E}x$ be the expectation of x and $\mathrm{E}\,x_1^{h_1}x_2^{h_2}\ldots x_N^{h_N}$, the expectation of the product $x_1^{h_1}x_2^{h_2}\ldots x_N^{h_N}$. An indefinite number of parameters can characterize the connections between the trials. Among them, those of the type of $\mathrm{E}\,x_1^{h_1}x_2^{h_2}\ldots x_N^{h_N}$ are known to be most widely used. When applying them, let us stipulate that the system (or the totality) of the trials is called *uniform* if this expectation persists under any permutation of the indices h [1].

Supposing that

$$\mathrm{E}\,x_1^{h_1}x_2^{h_2}\ldots x_N^{h_N} = m_{h_1h_2\cdots h_N}$$

and denoting a permutation of the indices by $h_{i1}, h_{i2}, \ldots, h_{iN}$, we may consequently characterize a uniform system (totality) by the condition that

$$m_{h_1h_2\cdots h_N} = m_{h_{i1}h_{i2}\ldots h_{iN}} \qquad (1)$$

takes place for any permutation. In particular, when agreeing to write $m_{h00\ldots0} = m_h$, we shall have

$$\mathrm{E}x_1^h = \mathrm{E}x_2^h = \ldots = \mathrm{E}x_N^h = \mathrm{E}x^h = m_h. \qquad (2a;\ 2b)$$

In accord with the theorem about the expectation of a product for mutually independent trials we have

$$\mathrm{E}\,x_1^{h_1}x_2^{h_2}\ldots x_N^{h_N} = Ex_1^{h_1}Ex_2^{h_2}\ldots Ex_N^{h_N} \text{ or } m_{h_1h_2\cdots h_N} = m_{h_1}m_{h_2}\ldots m_{h_N}.$$

We therefore easily convince ourselves that the transition from a uniform system of trials and its appropriate formulas expressed through parameters $m_{h_1h_2\cdots h_N}$ to a totality and its formulas (the *disjunction* of a uniform system into a totality, as we shall call this operation) presents no formal mathematical difficulties and is always justified. When replacing expressions of the type of $m_{h_1h_2\cdots h_N}$ in the formulas pertaining to a uniform system by the corresponding expressions $m_{h_1}m_{h_2}...m_{h_N}$ we may always consider the relations obtained as belonging to the case of a totality.

On the contrary, such simple rules for transforming the formulas for an inverse transition from a totality to a uniform system (for a *conjunction* of a totality into a uniform system) can be indicated not for all cases at all. It is not, however, difficult to establish one very simple and general condition under which the conjunction of the formulas, – that is, a transition from the expressions of the type of $m_{h_1}m_{h_2}...m_{h_N}$ to expressions $m_{h_1h_2\cdots h_N}$, – means a conjunction of a totality into a uniform system.

3. When considering a totality as a disjuncted system, and, inversely, understanding a system as a conjuncted totality, we introduce the following notation for the operations of disjunction and conjunction, respectively:

$$]\,m_{h_1h_2\cdots h_N}\,[\; = m_{h_1}m_{h_2}...m_{h_N}\,,\; [\,m_{h_1}m_{h_2}...m_{h_N}\,] = m_{h_1h_2\cdots h_N}\,.$$

Formulating the symbolic substance of these formulas in words, we shall say that the operation of disjunction comes to the *factorization* of the *product* $m_{h_1h_2\cdots h_N}$ with the factors being $m_{h_1}, m_{h_2}, ..., m_{h_N}$; and, inversely, the operation of conjunction is as though a *multiplication* of these factors resulting in the *product* $m_{h_1h_2\cdots h_N}$.

We are considering integral rational functions $G(x_1; x_2; \ldots, x_N)$; that is, polynomials of the type

$$G(x_1; x_2; \ldots, x_N) = \sum c_{h_1h_2...h_N}\, x_1^{h_1} x_2^{h_2} ... x_N^{h_N}$$

where $c_{h_1h_2...h_N}$ are the coefficients of the terms (of the products) $x_1^{h_1} x_2^{h_2} ... x_N^{h_N}$ and the sum is extended over the various systems of the exponents h.

Owing to the theorem about the expectation of a sum, that applies both to totalities and systems of trials to the same extent, we have

$$\text{E}G(x_1; x_2; \ldots, x_N) =$$
$$\text{E}\sum c_{h_1h_2...h_N}\, x_1^{h_1} x_2^{h_2} ... x_N^{h_N} = \sum c_{h_1h_2...h_N}\, \text{E}\, x_1^{h_1} x_2^{h_2} ... x_N^{h_N}\,.$$

Consequently, denoting EG for a totality by E]G[and by E[G] for the case of a uniform system, we shall have

$$\text{E}]G[= \sum c_{h_1h_2...h_N}\, m_{h_1}, m_{h_2}, ..., m_{h_N}\,,\; \text{E}[G] = \sum c_{h_1h_2...h_N}\, m_{h_1h_2\cdots h_N}\,.$$

We distinguish between two cases: Either none of the coefficients c includes the parameters $\{m\}$; or, at least one of them is their function. *In the first instance* we have

$$]\,c_{h_1h_2...h_N}\, m_{h_1h_2\cdots h_N}\,[\; = c_{h_1h_2...h_N}\, m_{h_1}m_{h_2} \ldots m_{h_N} \qquad (3)$$

and, inversely,

$$[c_{h_1h_2...h_N}\ m_{h_1}m_{h_2}...m_{h_N}] = c_{h_1h_2...h_N}\ m_{h_1h_2\cdots h_N}\ . \qquad (4)$$

Accordingly, by disjuncting the expression for E[G] and conjuncting the expression for E]G[we find that [2]

$$]\sum c_{h_1h_2...h_N}\ m_{h_1h_2\cdots h_N}[= \sum c_{h_1h_2...h_N}\ m_{h_1}m_{h_2}...m_{h_N} = \text{E}]G[,$$
$$[\sum c_{h_1h_2...h_N}\ m_{h_1}m_{h_2}...m_{h_N}] = \sum c_{h_1h_2...h_N}\ m_{h_1h_2\cdots h_N} = \text{E}[G].$$

Consequently, we satisfy ourselves that by disjuncting the formulas concerning a uniform system and by conjuncting those pertaining to a totality, we shall indeed turn them into each other. *In the second instance* formula (3) persists but the inverse relation (4) does not take place because here, in addition to the *mulipliers* $m_{h_1}, m_{h_2}, ..., m_{h_N}$, the factors m_h included in the coefficients c are also *multiplied* ; consequently, instead of the *product* sought, $m_{h_1h_2\cdots h_N}$, we obtain a different *product*, $m_{h_1h_2...h_Nh...}$ lacking the appropriate meaning. It indeed follows that in this second case a conjunction applied to the formulas pertaining to a totality does not lead to the formulas concerning a uniform system.

We thus conclude that the condition for the conjunction of the formulas expressing the expectation of integral rational functions $G(x_1; x_2; ..., x_N)$ to conjunct a totality into a uniform system, – that this condition is, that none of the coefficients c includes the parameters $\{m\}$. In this, and only in this instance, we are always justified, when replacing the expression of the type $m_{h_1}m_{h_2}...m_{h_N}$ in the formulas pertaining to a totality by the corresponding expressions $m_{h_1h_2\cdots h_N}$, in considering the obtained relations as belonging to the case of a uniform system.

A few examples. Supposing that

$$x_{(N)} = (1/N)\ (x_1 + x_2 + \ldots + x_N),\ m_{r(N)} = \text{E}x^r_{(N)}$$

we shall have for the case of a totality of trials [3]

$$m_{r]N[} = m_1^r + (1/N)\,C_r^2\,(m_1^{r-2}\,m_2 - m_1^r) +$$
$$(1/N^2)\{3\,C_r^4\,m_1^{r-4}m_2^2 + C_r^3\,[m_1^{r-3}m_3 - \frac{3(r-1)}{2}m_1^{r-2}m_2 + \frac{3r-1}{4}m_1^r]\} + \ldots$$

In particular,

$$m_{1]N[} = m_1;\ m_{2]N[} = m_1^2 + (1/N)\ (m_2 - m_1^2);$$
$$m_{3]N[} = m_1^3 + (3/N)\ (m_1m_2 - m_1^3) + (1/N^2)\ (m_3 - m_1m_2 + 2m_1^3);$$
$m_{4]N[}$ = {I omit this formula}.

Applying conjunction to these formulas, we directly obtain the formulas concerning a uniform system [4]:

$$m_{r[N]} = m_{11...1} + (1/N)\,C_r^2\,(m_{11...12} - m_{11...1}) +$$
$$(1/N^2)[3\,C_r^4\,m_{11...122} + C_r^3\,(m_{11...13} - \frac{3(r-1)}{2}m_{11...12} + \frac{3r-1}{4}m_{11...1})].$$

In particular,

$m_{1[N]} = m_1$; $m_{2[N]} = m_{11} + (1/N)(m_2 - m_{11})$;
$m_{3[N]} = m_{111} + (3/N)(m_{12} - m_{111}) + (1/N^2)(m_3 - m_{12} + 2\,m_{111})$;
$m_{4[N]}$ = {I omit this formula.}

4. Parameters {μ} which are the expectations of the products

$$(x_1 - m_1)^{h_1}(x_2 - m_1)^{h_2}...(x_N - m_1)^{h_N} \quad (5)$$

are very often used along with the parameters $\{m\}$ as a characteristic of a studied system or totality of trials. Supposing that

$$\mathrm{E}\,(x_1 - m_1)^{h_1}(x_2 - m_1)^{h_2}...(x_N - m_1)^{h_N} = \mu_{h_1 h_2 ... h_N}$$

we may, similar to what we did before, describe a uniform system (or totality) of trials by the condition that the equality

$$\mu_{h_1 h_2 ... h_N} = \mu_{h_{i1} h_{i2} ... h_{iN}}$$

takes place for any permutation of the indices $h_1, h_2, \ldots, h_N$. In particular, agreeing to write μ_h instead of $\mu_{h0\ldots0}$, we shall have

$$\mathrm{E}(x_1 - m_1)^h = \mathrm{E}(x_2 - m_1)^h = \ldots = \mathrm{E}(x_N - m_1)^h = \mathrm{E}(x - m_1)^h = \mu_h.$$

Noting that for mutually independent trials

$$\mathrm{E}\,(x_1 - m_1)^{h_1}(x_2 - m_1)^{h_2}...(x_N - m_1)^{h_N} =$$
$$E(x_1 - m_1)^{h_1}E(x_2 - m_1)^{h_2}...E(x_N - m_1)^{h_N},$$

or $\mu_{h_1 h_2 ... h_N} = \mu_{h_1}\mu_{h_2}...\mu_{h_N}$, we easily become convinced that, when applying the parameters {μ}, a disjunction of a uniform system into a totality is carried out as simple as in the case in which parameters $\{m\}$ are being used.

On the contrary, when transferring from a totality to a uniform system, identical rules for conjunction cannot be established for both sets of parameters because the coefficients of integral rational functions (5) include the powers of the parameter m_1, and, consequently, on the strength of the above, a conjunction of the parameters $\{m\}$ is not here justified at all. As an illustration, let us consider the relation between μ_{22} and μ_2^2. Noting that

$$\mathrm{E}(x_1 - m_1)^2(x_2 - m_1)^2 =$$
$$\mathrm{E}[x_1^2x_2^2 - 2x_1x_2(x_1 + x_2)m_1 + (x_1^2 + x_2^2)m_1^2 + 4x_1x_2m_1^2 - 2(x_1 + x_2)m_1^3 + m_1^4]$$

we find for the case of a uniform system

$$\mu_{22} = m_{22} - 4m_{12}m_1 + 2m_2m_1^2 + 4m_{11}m_1^2 - 3m_1^4$$

whereas for a totality

$\mu_2^2 = m_2^2 - 2m_2m_1^2 + m_1^4$.

Since

$]m_{22} - 4m_{12}m_1 + 2m_2m_1^2 + 4m_{11}m_1^2 - 3m_1^4[= m_2^2 - 2m_2m_1^2 + m_1^4$,

we convince ourselves that, when considering μ_{22} as a function of the parameters $\{m\}$, we may extend the rules of disjunction to cover it: $]\mu_{22}[= \mu_2^2$. The inverse relation does not, however, exist because

$[m_2^2 - 2m_2m_1^2 + m_1^4] \neq m_{22} - 4m_{12}m_1 + 2m_2m_1^2 + 4m_{11}m_1^2 - 3m_1^4$.

Consequently, the rules of conjunction cannot be extended to cover μ_2^2 as a function of the parameters $\{m\}$, and $[\mu_2^2] \neq \mu_{22}$. It is of course possible to formulate the rules of conjunction directly for the parameters $\{\mu\}$ when issuing not from the functions $G(x_1; x_2; \ldots, x_N)$ of the variables $x_1; x_2; \ldots, x_N$ themselves but from functions of the *deviations* of the variables from their expectation m_1: $G(X_1; X_2; \ldots, X_N) = G(x_1 - m_1; x_2 - m_1; \ldots; x_N - m_1)$, – and then, accordingly and analogous to the above, to determine

$]\mu_{h_1h_2\ldots h_N}[= \mu_{h_1}\mu_{h_2}\ldots\mu_{h_N}$, and, inversely, $[\mu_{h_1}\mu_{h_2}\ldots\mu_{h_N}] = \mu_{h_1h_2\ldots h_N}$.

For example, considering $E(x_1 - m_1)^2(x_2 - m_1)^2$ as $E(X_1^2 X_2^2)$ and replacing m by μ when passing on from variables x to variables X, we shall have for a totality $E(X_1^2 X_2^2) = E(X_1^2)$ $E(X_2^2) = \mu_2^2$. And, for a uniform system, $E(X_1^2 X_2^2) = \mu_{22}$ and we may consequently write not only $]\mu_{22}[= \mu_2^2$ but also inversely (as distinct from the previous example in which μ_2^2 and μ_{22} were considered as functions of the parameters $\{m\}$): $[\mu_2^2] = \mu_{22}$. Saying nothing about the violation of the unity of conjunction with respect to the parameters $\{\mu\}$ and to their expression through the parameters $\{m\}$ by the extension of the previous rules of this operation onto the parameters $\{\mu\}$, the conjunction of the $\{\mu\}$ parameters thus defined is unrealizable in any of the cases in which a transition to the parameters of the type $\mu_{1h_2\ldots h_N}$ is intended. Indeed, $\mu_1 = 0$ and the formulas to be conjuncted cannot therefore include terms of the type μ_1,
$\mu_{h_2}, \ldots, \mu_{h_N}$ containing the *factor* μ_1 and necessary for obtaining the *product*
$\mu_{1h_2\ldots h_N}$.

At the same time, it is indeed this property, $\mu_1 = 0$, that underlies the supplementary calculational significance of the parameters μ_h. In a number of cases we manage to replace the parameters $\{m\}$ by the corresponding parameters $\{\mu\}$; owing to this (and because of the property $\mu_1 = 0$) we are able to represent awkward and hardly surveyable formulas in a simpler and more obvious form. Such a replacement of the parameters $m_{h_1h_2\ldots h_N}$ by parameters $\mu_{h_1h_2\ldots h_N}$ is always possible if only the studied function $G(x_1; x_2; \ldots, x_N)$ whose expectation is being determined, satisfies the condition

$$G(x_1; x_2; \ldots, x_N) = G(x_1 - a; x_2 - a; \ldots; x_N - a) \quad (6)$$

for any value of a and, consequently, for $a = m_1$ as well. As an example of a function satisfying this condition we adduce the deviation of the random empirical value of the variable x from the mean of its values, $x_{(N)}$. Namely, we have

$$x_i - x_{(N)} = (x_i - a) - (1/N)\sum_{i=1}^{N} (x_i - a)$$

and we may therefore express the expectation $E(x_i - x_{(N)})^h$ in terms of both the parameters $\{m\}$ and only of the parameters $\{\mu\}$. For example,

$$E(x_i - x_{(N)})^2 = [(N-1)/N](m_2 - m_1^2) = [(N-1)/N](\mu_2 - \mu_1^2) = [(N-1)/N]\ \mu_2.$$

We consider functions of a special kind obeying the condition (6) and, namely, the expressions

$$M'_h = (x_1 - x_2)\ (x_1 - x_3)\ \dots\ (x_1 - x_{1+h})$$

and in general

$$M'_{h_1 h_2 \dots h_N} =$$
$$(x_i - x_{i+1})\ (x_i - x_{i+2})\ \dots\ (x_i - x_{i+h_1})\ (x_j - x_{j+1})\ (x_j - x_{j+2})\ \dots (x_j - x_{j+h_2})\ \dots$$
$$(x_k - x_{k+1})\ \dots\ (x_k - x_{k+h_N})$$

where the x's are the random empirical values of the variable x taken by it in different trials [5].

Denoting by M''_h, $M^{(3)}_h$, …, $M^{(i)}_h$, etc the expressions similar to M'_h so that for example

$$M''_h = (x_2 - x_1)\ (x_2 - x_3)\ \dots\ (x_2 - x_{1+h}),\ M^{(i)}_h = (x_i - x_{i+1})\ (x_i - x_{i+2})\ \dots\ (x_i - x_{i+h})$$

and in the same way denoting by $M''_{h_1 h_2 \dots h_N}$, $M^{(3)}{}_{h_1 h_2 \dots h_N}$, $M^{(i)}{}_{h_1 h_2 \dots h_N}$ etc the expressions similar in the same sense to $M'_{h_1 h_2 \dots h_N}$, we shall have, for the case of a uniform system or totality of trials,

$$EM'_h = EM''_h = \dots = EM^{(i)}_h,\ E\,M'_{h_1 h_2 \dots h_N} = E\,M''_{h_1 h_2 \dots h_N} = \dots = E\,M^{(i)}{}_{h_1 h_2 \dots h_N}.$$

Supposing that

$$E(x_i - x_{i+1})\ \dots\ (x_i - x_{i+h_1})\ (x_j - x_{j+1}) \dots (x_j - x_{j+h_2}) \dots (x_k - x_{k+1}) \dots (x_k - x_{k+h_N}) =$$
$$M_{h_1 h_2 \dots h_N}$$

we may describe, analogous to what we did before, a uniform system (or totality) by the condition that the equality

$$M_{h_1 h_2 \dots h_N} = M_{h_{i1} h_{i2} \dots h_{iN}}$$

takes place for any permutations of the indices. In particular, we shall have

$$E(x_1 - x_2)\ (x_1 - x_3)\ \dots\ (x_1 - x_{1+h}) = E(x_2 - x_1)\ (x_2 - x_3)\ \dots\ (x_2 - x_{1+h}) = \dots =$$
$$E(x_i - x_{i+1})\ (x_i - x_{i+2})\ \dots\ (x_i - x_{i+h}) = M_h.$$

Noting that for independent trials

$$E(x_i - x_{i+1})\ (x_i - x_{i+2})\ \dots\ (x_i - x_{i+h_1})\ (x_j - x_{j+1})\ (x_j - x_{j+2})\ \dots (x_j - x_{j+h_2})\ \dots$$

$(x_k - x_{k+1}) \ldots (x_k - x_{k+h_N}) = \mathrm{E}(x_i - x_{i+1})(x_i - x_{i+2}) \ldots (x_i - x_{i+h_1}) \cdot$
$\mathrm{E}(x_j - x_{j+1})(x_j - x_{j+2}) \ldots (x_j - x_{j+h_2}) \ldots \cdot \mathrm{E}(x_k - x_{k+1}) \ldots (x_k - x_{k+h_N})$

or $M_{h_1 h_2 \ldots h_N} = M_{h_1} M_{h_2} \ldots M_{h_N}$ and that

$\mathrm{E}(x_1 - x_2)(x_1 - x_3) \ldots (x_1 - x_{1+h}) = \mathrm{E}\, x_1^h - \mathrm{E}(x_2 + x_3 + \ldots + x_{1+h})\, x_1^{h-1} +$
$\mathrm{E}(x_2x_3 + x_2x_4 + x_3x_4 + \ldots + x_hx_{1+h})\, x_1^{h-2} - \mathrm{E}(x_2x_3x_4 + x_2x_3x_5 + \ldots)\, x_1^{h-3} + \ldots +$
$(-1^h)\mathrm{E}(x_2x_3x_4 \ldots x_{1+h})$

or

$$M_h = \sum_{i=0}^{h} (-1)^i C_h^i m_{h-i} m_1^i = \sum_{i=0}^{h} (-1)^i C_h^i \mu_{h-i} \mu_1^i = \mu_h, \qquad (7)$$

we satisfy ourselves without difficulties that a disjunction of a uniform system into a totality is always accomplished by a disjunction of the parameters $\{M\}$ and their simple replacement by parameters $\{\mu\}$. Let us write low-case $\mathcal{M}$ instead of capital M both for a uniform system and a totality [6]. In the first instance only we may define, in accord with the above, the operation of disjunction of the parameters $\{\mathcal{M}\}$:

$]\mathcal{M}_{h_1 h_2 \ldots h_N}[= \mu_{h_1} \mu_{h_2} \ldots \mu_{h_N}$ and in particular $]\mathcal{M}_h[= \mu_h$.

Inversely, when defining the operation of conjunction of the parameters $\{\mu\}$as

$[\mu_{h_1} \mu_{h_2} \ldots \mu_{h_N}] = \mathcal{M}_{h_1 h_2 \ldots h_N}$ and in particular $[\mu_h] = \mathcal{M}_h$,

it is not difficult to show that, if the initial function, whose expectation is being determined, satisfies the condition (6), and if the coefficients of the variables do not include the parameters $\{\mu\}$, the conjunction of a totality into a uniform system is brought about by a simple replacement of these parameters by parameters $\{\mathcal{M}\}$ and a *multiplication*, *i.e.*, a conjunction, of the latter.

When writing out

$(x_1 - x_2)^2 = [(x_1 - x_3) + (x_3 - x_2)][(x_1 - x_4) + (x_4 - x_2)] =$
$(x_1 - x_3)(x_1 - x_4) + (x_1 - x_3)(x_4 - x_2) + (x_3 - x_2)(x_1 - x_4) + (x_2 - x_3)(x_2 - x_4) =$
$M'_2 + M'_{11} + M''_{11} + M''_2,$

$(x_1 - x_2)^3 = [(x_1 - x_3) + (x_3 - x_2)][(x_1 - x_4) + (x_4 - x_2)][(x_1 - x_5) + (x_5 - x_2)] =$
$(x_1 - x_3)(x_1 - x_4)(x_1 - x_5) + (x_1 - x_3)(x_1 - x_4)(x_5 - x_2) + \ldots =$
$M'_3 + M'_{12} + M''_{12} + M^{(3)}_{12} + M^{(4)}_{12} + M^{(5)}_{12} + M^{(6)}_{12} - M''_3, \ldots$

$(x_1 - x_2)^m =$
$[(x_1 - x_3) + (x_3 - x_2)][(x_1 - x_4) + (x_4 - x_2)] \ldots [(x_1 - x_{m+2}) + (x_{m+2} - x_2)] = \ldots$

we convince ourselves that $(x_1 - x_2)^m$ can be identically represented as an algebraic sum of expressions $M_{h_1 h_2 \ldots h_N}^{(i)}$. In absolutely the same way we show that, in general, any expression $(x_1 - x_2)^m (x_1 - x_3)^n \ldots (x_1 - x_N)^p$ can be represented as

$$\sum_{(i), h_1, h_2, \ldots, h_N} \pm M_{h_1 h_2 \ldots h_N}^{(i)}$$

where the sum extends over the corresponding values of the indices h and i.

On the other hand, we note that functions satisfying the condition (6) can be represented as functions of the differences of the variables; for example, as
$F(x_1 - x_2; x_1 - x_3; \dots; x_1 - x_N)$. Namely, supposing that $a = x_1$, we find that

$$G(x_1; x_2; \dots, x_N) = G(0; x_2 - x_1; x_3 - x_1; \dots; x_N - x_1)$$

so that

$$G(x_1; x_2; \dots, x_N) = \sum_{m_1, m_2, \dots, m_N} c_{m_1 m_2 \dots m_N} \; x_1^{m_1} x_2^{m_2} \dots x_N^{m_N} =$$
$$\sum_{m_1, m_2, \dots, m_N} c_{0 m_2 \dots m_N} \, (x_2 - x_1)^{m_2} (x_3 - x_1)^{m_3} \dots (x_N - x_1)^{m_N} .$$

And, since

$$(x_2 - x_1)^{m_2} (x_3 - x_1)^{m_3} \dots (x_N - x_1)^{m_N} = \sum_{(i), h_1, h_2, \dots, h_N} \pm M_{h_1 h_2 \dots h_N}{}^{(i)},$$

we indeed satisfy ourselves that any integral rational function obeying the condition (6) can be represented as

$$G(x_1; x_2; \dots, x_N) = \sum_{(i), h_1, h_2, \dots, h_N} c_{h_1 h_2 \dots h_N}{}^{(i)} \, M_{h_1 h_2 \dots h_N}{}^{(i)} .$$

Noting that

$$\mathrm{E}\, M'_{h_1 h_2 \dots h_N} = \mathrm{E}\, M''_{h_1 h_2 \dots h_N} = \dots = \mathrm{E}\, M^{(i)}{}_{h_1 h_2 \dots N} = M_{h_1 h_2 \dots h_N}$$

we find that, in accord with the theorem about the expectation of a sum,

$$\mathrm{E} G(x_1; x_2; \dots, x_N) = \sum_{h_1, h_2, \dots, h_N} c_{h_1 h_2 \dots h_N} \; M_{h_1 h_2 \dots h_N} .$$

In particular, in the case of a totality we have

$$\mathrm{E}]G[= \sum_{h_1, h_2, \dots, h_N} c_{h_1 h_2 \dots h_N} \; \mu_{h_1} \mu_{h_2} \dots \mu_{h_N}$$

and, for a uniform system,

$$\mathrm{E}[G] = \sum_{h_1, h_2, \dots, h_N} c_{h_1 h_2 \dots h_N} \; \mathcal{M}_{h_1 h_2 \dots h_N} .$$

Repeating step by step the course of reasoning in §2, we convince ourselves that, with respect to parameters $\{M\}$ as well, the disjunction of the formulas concerning a uniform system, and a conjunction of the formulas belonging to a totality, bring about a transition from one set to another one. We are therefore justified, in all the cases in which $G(x_1; x_2; \dots, x_N)$ satisfies the indicated conditions, in replacing the expressions of the type $\mu_{h_1} \mu_{h_2} \dots \mu_{h_N}$ in the formulas concerning totalities by the corresponding expressions $\mathcal{M}_{h_1 h_2 \dots h_N}$ and to consider the thus obtained relations as pertaining to a uniform system. On the other hand, noting that $M_{1h_2...h_N} = 0$ [7], we become convinced, that the abovementioned inconvenience, that occurs when conjuncting the parameters $\{\mu\}$ because of the vanishing of μ_1, disappears. The supplementary calculational importance of the parameters $\{M\}$ is indeed based on the vanishing of $M_{1h_2...h_N}$. [8]

Example. Supposing that

$$U_r^{(N)} = \mathrm{E}(1/N)\sum_{i=1}^{N} (x_i - x_{(N)})^r \quad (8)$$

and that, in general,

$$U_{r_1 r_2 ... r_m}^{(N)} = (1/N^m)\mathrm{E}\sum_{i=1}^{N} (x_i - x_{(N)})^{r_1} \sum_{i=1}^{N} (x_i - x_{(N)})^{r_2} \ldots \sum_{i=1}^{N} (x_i - x_{(N)})^{r_m}, \quad (9)$$

we shall have for a totality of trials [9]

$$U_r^{]N[} = \mu_r - (1/N)[r\mu_r - (1/2)r^{[-2]}\mu_{r-2}\mu_2] + (1/N^2)\{(1/2)r^{[-2]}\mu_r - [(r-1)/2]\, r^{[-2]}\mu_{r-2}\mu_2 - (1/6)\, r^{[-3]}\mu_{r-3}\mu_3 + (1/8)\, r^{[-4]}\mu_{r-4}\mu_2^2\} + \ldots,$$

$$U_{rr}^{]N[} = \mu_r^2 + (1/N)[\mu_{2r} - (2r+1)\mu_r^2 - 2r\mu_{r+1}\mu_{r-1} + r^2\mu_{r-1}^2\mu_2 + r(r-1)\mu_r\mu_{r-2}\mu_2] - (1/N^2)[2r\mu_{2r} - r(2r-1)\mu_{2r-2}\mu_2 - 4r^2\mu_{r+1}\mu_{r-1} - 3(r+1)\mu_r^2 - r(r-1)\mu_{r+2}\mu_{r-2} + 3r^2\mu_{r-1}^2\mu_2 + r(r-1)(4r+1)\mu_r\mu_{r-2}\mu_2 + r^{[-3]}\mu_{r+1}\mu_{r-3}\mu_2 + r^2(r-1)\mu_{r-1}\mu_{r-2}\mu_3 + (1/3)r^{[-3]}\mu_r\mu_{r-3}\mu_3 - (3/4)r^2(r-1)^2\mu_{r-2}^2\mu_2^2 - r^2(r-1)(r-2)\mu_{r-1}\mu_{r-3}\mu_2^2 - (1/4)r^{[-4]}\mu_r\mu_{r-4}\mu_2^2] + \ldots$$

Applying the operation of conjunction to these formulas, we directly obtain the formulas concerning a uniform system:

$$U_r^{[N]} = \mathcal{M}_r - (1/N)[r\mathcal{M}_r - (1/2)\, r^{[-2]}\, \mathcal{M}_{r-2,2}] - (1/N^2)\{(1/2)\, r^{[-2]}\, \mathcal{M}_r - [(r-1)/2]\, r^{[-2]}\, \mathcal{M}_{r-2,2} - (1/6)\, r^{[-3]}\, \mathcal{M}_{r-3,3} + (1/8)\, r^{[-4]}\, \mathcal{M}_{r-4,2,2}] + \ldots$$

$$U_{rr}^{[N]} = \mathcal{M}_{rr} + (1/N)[\mathcal{M}_{2r} - (2r+1)\mathcal{M}_{rr} - 2r\mathcal{M}_{r+1,r-1} + r^2\mathcal{M}_{r-1,r-1,2}] + r(r-1)\mathcal{M}_{r,r-2,2}] -$$

$$(1/N^2)[2r\mathcal{M}_{2r} - r(2r-1)\mathcal{M}_{r-2,2} - 4r^2\mathcal{M}_{r+1,r-1} - 3(r+1)\mathcal{M}_{rr} - r(r-1)\mathcal{M}_{r+2,r-2} + 3r^{[-3]}\mathcal{M}_{r-1,r-1,2} + r(r-1)(4r+1)\mathcal{M}_{r,r-2,2} + r^{[-3]}\mathcal{M}_{r+1,r-3,2} + r^2(r-1)\mathcal{M}_{r-1,r-2,3} + (1/3)r^{[-3]}\mathcal{M}_{r,r-3,3} - (3/4)r^2(r-1)^2\mathcal{M}_{r-2,r-2,2,2} - r^2(r-1)(r-2)\mathcal{M}_{r-1,r-3,2,2} - (1/4)\, r^{[-4]}{}_{r,r-4,2,2}] + \ldots$$

Supposing that, for example, $r = 2$ and 3, we get from the first formula

$$U_2^{[N]} = \mathcal{M}_2 - (1/N)\mathcal{M}_2 = [(N-1)/N]\,\mathcal{M}_2,$$
$$U_3^{[N]} = \mathcal{M}_3 - (3/N)\mathcal{M}_3 + (2/N^2)\mathcal{M}_3 = [(N-1)(N-2)/N^2]\mathcal{M}_3. \quad (11)$$

The second formula, as well as the first one for $r > 3$, only provides approximate expressions containing terms to within the order $(1/N^2)$ inclusively.

5. We return to the patterns of the returned and unreturned tickets. Let the urn still contain S tickets with s_1 of them marked by the number $x^{(1)}$; s_2 of them, by $x^{(2)}$; …; and s_k, by the number $x^{(k)}$. Considering these numbers as the values of some random variable x having probabilities s_1/S, s_2/S, …, s_k/S, respectively and let us agree to denote the numbers appearing in the first, the second, …, the N-th extraction, by $x_1, x_2, …, x_N$. The scheme of the returned ticket deals, as it is immediately clear, with a *totality* of trials on a variable whose law of distribution is always fixed. In the case of the unreturned ticket we have a *system* of trials. It is not difficult to show that the system thus obtained is uniform [2, pp. 216 – 219].

Indeed, if the extracted ticket is not returned, it is absolutely indifferent whether the N tickets are drawn in turn, one by one, or all at once. It follows that the order of the appearance of the separate tickets does not matter at all and that we may enumerate them as we please considering any of them as being extracted at the first, the second, … drawing. The expectation $\mathrm{E}\, x_1^{h_1} x_2^{h_2} ... x_N^{h_N}$ remains therefore invariable under any permutation $x_1, x_2, …, x_N$; or, which is the same, under any permutation of the indices $h_1, h_2, …, h_N$. Consequently, the thus obtained system of trials is uniform.

The analytical connection between the separate extractions made without returning the ticket can be defined by the following main property of this arrangement: if $N = S$, the tickets marked by the numbers $x_1, x_2, …, x_N$ exhaust the urn; they therefore represent the s_1 tickets marked by the number $x^{(1)}$, s_2 of them, by $x^{(2)}$; …; and s_k, by the number $x^{(k)}$, all of them taken in some order. Issuing from this property, we shall have

$$m_h = \mathrm{E}x^h = \sum_{i=1}^{k} (s_i/S)x^{(i)h} = (1/S)\sum_{j=1}^{S} x_j^h$$

and, in general,

$$m_{h_1 h_2 \cdots h_N} = \mathrm{E}\, x_1^{h_1} x_2^{h_2} ... x_N^{h_N} = (1/S^{[-N]})\sum x_{i_1}^{h_1} x_{i_2}^{h_2} ... x_{i_N}^{h_N},$$

cf. notation (10). The sum extends over all the combinations with repetitions of the indices i_1, i_2, …, i_N taken from numbers 1, 2, …, S, N at a time. Noting that this sum is a symmetric function of the variables $x_1, x_2, …, x_S$ and can therefore be expressed by elementary symmetric functions of the type

$$\sum_{i=1}^{S} x_i^h,$$

we convince ourselves that for the pattern of the unreturned ticket the parameters $m_{h_1 h_2 \cdots h_N}$ can be rationally expressed through the parameters m_h. We have for example [5]

$$m_{11} = [1/S(S-1)]\sum_{i,j} x_i x_j =$$
$$[1/S(S-1)]\,[(x_1 + x_2 + … + x_S)^2 - (x_1^2 + x_2^2 + + … + x_S^2)] =$$
$$[1/S(S-1)]\,(S^2 m_1^2 - S m_2),$$

$$m_{12} = [1/S(S-1)]\sum_{i,j} x_i x_j^2 = [1/S(S-1)][\sum_{i=1}^{S} x_i \sum_{j=1}^{S} x_j^2 - \sum_{i=1}^{S} x_i^3] =$$
$$[1/S(S-1)]\,(S^2 m_1 m_2 - S m_3),$$

$$m_{111} = [1/S(S-1)(S-2)]\sum_{i,j,k} x_i x_j x_k =$$

$$[1/S(S-1)(S-2)][(\sum_{i=1}^{S} x_i)^3 - 3\sum_{i=1}^{S} x_i \sum_{j=1}^{S} x_j^2 + 2\sum_{i=1}^{S} x_i^3] =$$
$$[1/S(S-1)(S-2)]\,(S^3 m_1^3 - 3S^2 m_1 m_2 + 2S m_3).$$

And in the same way (Ibidem)

$m_{13} = [1/S(S-1)]\,(S^2 m_1 m_3 - S m_4)$, $m_{22} = [1/S(S-1)]\,(S^2 m_2^2 - S m_4)$,
$m_{112} = [1/S(S-1)(S-2)]\,(S^2 m_1^2 m_2 - 2S^2 m_1 m_3 - S^2 m_2^2 + 2S m_4)$,
$m_{1111} =$
$[1/S(S-1)(S-2)(S-3)]\,(S^4 m_1^4 - 6S^3 m_1^2 m_2 + 8S^2 m_1 m_3 + 3S^2 m_2^2 - 6S m_4)$.

Without providing the proof itself, we also indicate the general formula:

$$m_{h_1 h_2 \cdots h_N} \frac{(-1)^N}{S^{[-N]}} \sum_{r=1}^{N} \{(-S)^r \sum_{k_1 k_2 \ldots k_r} [\underline{|k_1 - 1}\ \underline{|k_2 - 1} \ldots \underline{|k_r - 1} \cdot \sum_{h_1 h_2 \ldots h_N} m_{h_1+h_2+\ldots+h_\alpha}\, m_{h_{\alpha+1}+h_{\alpha+2}+\ldots+h_\beta} \ldots m_{h_\gamma+\ldots+h_\delta}]\} \quad (12)$$

where the second sum extends over all the values of $k_1, k_2, \ldots, k_r$ satisfying the conditions $1 \le k_1 \le k_2 \le \ldots \le k_r$ and $k_1 + k_2 + \ldots + k_r = N$ and the third one, over all the groups of all the combinations without repetitions of the indices $h_1, h_2, \ldots, h_N$ taken $k_1, k_2, \ldots, k_r$ at a time. Then {in my own notation}, $\alpha = k_1$, $\beta = k_1 + k_2$, $\gamma = k_1 + k_2 + \ldots + k_{r-1} + 1 + \ldots$ and $\delta = k_1 + k_2 + \ldots + k_r$. {The author had not explained the meaning of symbols such as $\underline{|a}$ which also appear in §8}

Supposing for example that in this general formula $N = 2$, 3 or 4, we find that (Ibidem)

$$m_{h_1 h_2} = \{1/[S(S-1)]\}(S^2 m_{h_1} m_{h_2} - S\, m_{h_1+h_2}),$$

$$m_{h_1 h_2 h_3} = \{1/[S(S-1)(S-2)]\}[S^3 m_{h_1} m_{h_2} m_{h_3} -$$
$$S^2(m_{h_1} m_{h_2+h_3} + m_{h_2} m_{h_1+h_3} + m_{h_3} m_{h_1+h_2}) + 2S\, m_{h_1+h_2+h_3}],$$

$m_{h_1 h_2 h_3 h_4} =$ {I omit this formula.}

6. Considering

$$\mu_{h_1 h_2 \ldots h_N} = \mathrm{E}\,(x_1 - m_1)^{h_1}(x_2 - m_1)^{h_2} \ldots (x_N - m_1)^{h_N}$$

as the expectation $\mathrm{E}\, X_1^{h_1} X_2^{h_2} \ldots X_N^{h_N}$ with $X_i = x_i - m_1$, $i = 1, 2, \ldots, N$, we satisfy ourselves that, when replacing m by μ everywhere in the formulas above, we obtain formulas expressing the parameters $\mu_{h_1 h_2 \ldots h_N}$ through the parameters μ_h. Noting that $\mu_1 = 0$, we find that, for example,

$\mu_{11} = -[1/(S-1)]\mu_2$, $\mu_{12} = -[1/(S-1)]\mu_3$, $\mu_{111} = [2/(S-1)(S-2)]\mu_3$,
$\mu_{13} = -[1/(S-1)]\mu_4$, $\mu_{112} = [2/(S-1)(S-2)]\mu_4 - [S/(S-1)(S-2)]\mu_2^2$,
$\mu_{22} = -[1/(S-1)]\mu_4 + [S/(S-1)]\mu_2^2$,
$\mu_{1111} = -[6/(S-1)(S-2)(S-3)]\mu_4 + [3S/(S-1)(S-2)(S-3)]\mu_2^2$.

When expressing the parameters $M_{h_1 h_2 \ldots h_N}$ through the parameters $\mu_{h_1 h_2 \ldots h_N}$ [10] and replacing the latter by their expression in terms of parameters μ_h, we obtain formulas determining the former quantities as special parameters in the pattern of the unreturned ticket. We have for example

$$M_2 = \mu_2 - \mu_{11} = [S/(S-1)]\mu_2,\ M_3 = \mu_3 - 3\mu_{12} + 2\mu_{111} = [S^2/(S-1)(S-2)]\mu_3,$$
$$M_4 = \mu_4 - 4\mu_{13} + 6\mu_{112} - 3\mu_{1111} =$$
$$\frac{S^3}{(S-1)(S-2)(S-3)}\mu_4 - \frac{S(2S-3)}{(S-1)(S-2)(S-3)}(\mu_4 + 3\mu_2^2),$$
$$M_{22} = \mu_{22} - 2\mu_{112} + \mu_{1111} = \frac{S^3}{(S-1)(S-2)(S-3)}\mu_2^2 -$$
$$\frac{S}{(S-2)(S-3)}(\mu_4 + 3\mu_2^2) \text{ etc.}$$

These formulas might be considered as describing the transition from a uniform system in general to its special case, to the arrangement of the unreturned ticket. The transition from the layout of the returned ticket to the other one is thus brought about in two stages: First, by conjuncting the totality into a uniform system, *i.e.*, by a transition from parameters μ_{h_i}, $i = 1, 2, \ldots, N$, concerning the scheme of the returned ticket to parameters $M_{h_1 h_2 \ldots h_N}$ characterising a uniform system; and, second, by an inverse transition from these latter parameters to parameters μ_h accomplished by the formulas for the transfer from a uniform system to its special case, to the scheme of the unreturned ticket. The uniform system and the parameters $\{M\}$ describing it are thus an intermediate link, a changing station of sorts on the way from the pattern of the returned ticket to the other one.

A few examples will explain this. Issuing from (9) and (8) we shall have for the arrangement of the returned ticket [3, p. 186, formula (7) and p. 192, formula (19)]

$$U_2^{]N[} = [(N-1)/N]\mu_2,\ U_3^{]N[} = [(N-1)(N-2)/N^2]\mu_3, \text{ r}$$
$$U_4^{]N[} = [(N-1)(N-2)(N-3)/N^3]\mu_4 + [(N-1)(2N-3)/N^3](\mu_4 + 3\mu_2^2),$$
$$U_{22}^{]N[} = [(N-1)(N-2)(N-3)/N^3]\mu_2^2 + [(N-1)^2/N^3](\mu_4 + 3\mu_2^2).$$

Applying the operation of conjunction to these formulas we pass over to the formulas concerning a uniform system (for $U_2^{[N]}$ and $U_3^{[N]}$ see formulas (11))

$$U_4^{[N]} = [(N-1)(N-2)(N-3)/N^3]M_4 + [(N-1)(2N-3)/N^3](M_4 + 3M_{22}),$$
$$U_{22}^{[N]} = [(N-1)(N-2)(N-3)/N^3]M_{22} + [(N-1)(2N-3)/N^3](M_4 + 3M_{22}).$$

Substituting the above expressions for the parameters M_2, M_3, M_4 and M_{22} into these formulas and denoting $U_{r_1 r_2 \ldots r_m}^{[N]}$ in the special case of a uniform system, – in the pattern of the unreturned ticket, – by $U_{r_1 r_2 \ldots r_m}^{[N/S]}$, we find that, finally,

$$U_2^{[N/S]} = \frac{(N-1)S}{N(S-1)}\mu_2,\ U_3^{[N/S]} = \frac{(N-1)(N-2)S^2}{N^2(S-1)(S-2)}\mu_3,$$

$$U_4^{[N/S]} = \frac{(N-1)(N-2)(N-3)S^3}{N^3(S-1)(S-2)(S-3)}\mu_4 +$$

$$\frac{(N-1)((S-N)(2NS-3S-3N+3)S}{N^3(S-1)(S-2)(S-3)}(\mu_4 + 3\mu_2^2),$$

$$U_{22}^{[N/S]} = \frac{(N-1)(N-2)(N-3)S^3}{N^3(S-1)(S-2)(S-3)}\mu_2^2 +$$
$$\frac{(N-1)(S-N)(NS-S-N-1)S}{N^3(S-1)(S-2)(S-3)}(\mu_4 + 3\mu_2^2).$$

Introducing, as we did before, the notation (10), we can also represent these formulas in the following way:

$$U_2^{[N/S]} = \frac{N^{[-2]}S^2}{N^2S^{[-2]}}\mu_2,\ U_3^{[N/S]} = \frac{N^{[-3]}S^3}{N^3S^{[-3]}}\mu_3,$$

$$U_4^{[N/S]} = \frac{N^{[-4]}S^4}{N^4S^{[-4]}}\mu_4 + \frac{N^{[-2]}S^2(S-N)(2NS-3S-3N+3)}{N^2S^{[-2]}N^2(S-2)(S-3)}(\mu_4 + 3\mu_2^2),$$
$$U_{22}^{[N/S]} = \frac{N^{[-4]}S^4}{N^4S^{[-4]}}\mu_2^2 + \frac{N^{[-2]}S^2(S-N)(NS-S-N-1)}{N^2S^{[-2]}N^2(S-2)(S-3)}(\mu_4 + 3\mu_2^2).$$

It is useful to note the following relation between the parameters $\{м\}$ and $\{\mu\}$:

$$м_4 + 3м_{22} = [S/(S-1)]\,(\mu_4 + 3\mu_2^2), \qquad (13)$$

or, which is the same,

$$S^{[-2]}(м_4 + 3м_{22}) = S^2(\mu_4 + 3\mu_2^2).$$

7. Such is the general method that enables us to throw a bridge from the formulas concerning the scheme of the returned ticket to those describing the other pattern. In some cases, however, the formulas sought can be also obtained in a shorter way. Thus, in the examples above, the following reasoning will rapidly lead us to our goal. And the supplementary calculational importance of the parameters $\{M\}$ will reveal itself with special clearness. We issue from the remark that for $N = S$ and denoting

$$x_{(S)} = (1/S)(x_1 + x_2 + \ldots + x_S) = m_1$$

we have

$$U_r^{[S/S]} = \text{E}(1/S)\sum_{i=1}^{S}(x_i - x_S)^r = \text{E}(1/S)\sum_{i=1}^{S}(x_i - m_1)^r = \text{E}\mu_r = \mu_r,$$
$$U_{r_1 r_2 \ldots r_m}^{[S/S]} = \text{E}\{[(1/S)\sum_{i=1}^{S}(x_i - m_1)^{r_1}]\ \ldots\ [(1/S)\sum_{i=1}^{S}(x_i - m_1)^{r_m}]\} =$$
$$\mu_{r_1}\mu_{r_2}\ldots\mu_{r_m}\ .$$

Taking this into account and, on the other hand, expressing $U_{r_1 r_2 \ldots r_m}^{[S/S]}$ through the parameters $\{м\}$, we shall have

$$\mu_2 = U_2^{[S/S]} = [(S-1)/S]\mathcal{M}_2,\ \mu_3 = U_3^{[S/S]} = [(S-1)(S-2)/S^2]\mathcal{M}_3,$$
$$\mu_4 = U_4^{[S/S]} = \frac{(S-1)(S-2)(S-3)}{S^3}\mathcal{M}_4 + \frac{(S-1)(2S-3)}{S^3}(\mathcal{M}_4 + 3\mathcal{M}_{22}),$$
$$\mu_2^{\,2} = U_{22}^{[S/S]} = \frac{(S-1)(S-2)(S-3)}{S^3}\mathcal{M}_{22} + \frac{(S-1)^2}{S^3}(\mathcal{M}_4 + 3\mathcal{M}_{22})$$

so that

$$\mathcal{M}_2 = [S/(S-1)]\mu_2,\ \mathcal{M}_3 = \frac{S^2}{(S-1)(S-2)}\mu_3,$$
$$\mathcal{M}_4 = \frac{S^3}{(S-1)(S-2)(S-3)}\mu_4 - \frac{S(2S-3)}{(S-1)(S-2)(S-3)}(\mu_4 + 3\mu_2^{\,2}),$$
$$\mathcal{M}_{22} = \frac{S^3}{(S-1)(S-2)(S-3)}\mu_2^{\,2} - \frac{S}{(S-2)(S-3)}(\mu_4 + 3\mu_2^{\,2}).$$

We thus derive all the already known to us formulas for the transition from a uniform system to its special case, the layout of the unreturned ticket. Making use of the relation (13), we obtain without difficulties the sought expressions for $U_2^{[N/S]}$, $U_3^{[N/S]}$, $U_4^{[N/S]}$ and $U_{22}^{[N/S]}$.

Another example. In the scheme of the returned ticket we have (cf. [3, p. 186, formula (7)]

$$U_5^{]N[} = \frac{(N-1)(N-2)(N-3)(N-4)}{N^4}\mu_5 + \frac{5(N-1)(N-2)^2}{N^4}(\mu_5 + 2\mu_2\mu_3)$$

and, as it follows, for a uniform system,

$$U_5^{[N]} = \frac{(N-1)(N-2)(N-3)(N-4)}{N^4}\mathcal{M}_5 + \frac{5(N-1)(N-2)^2}{N^4}(\mathcal{M}_5 + 2\mathcal{M}_{23}).$$

When applying the expectation of $U_{23}^{(N)}$ we obtain for the totality of trials (cf. Note 11)

$$U_5^{]N[} + 2\,U_{23}^{]N[} = \frac{(N-1)(N-2)}{N^2}(\mu_5 + 2\mu_2\mu_3)$$

and, consequently, for a uniform system,

$$U_5^{[N]} + 2\,U_{23}^{[N]} = \frac{(N-1)(N-2)}{N^2}(\mathcal{M}_5 + 2\mathcal{M}_{23}).$$

Noting that, on the other hand, when $N = S$,

$$U_5^{[S/S]} = \mu_5,\ U_{23}^{[S/S]} = \mu_2\mu_3,$$

we obtain, analogous to the above,

$$\mu_5 = U_5^{[S/S]} = \frac{(S-1)(S-2)(S-3)(S-4)}{S^4}\mathcal{M}_5 + 5\frac{(S-1)(S-2)^2}{S^4}(\mathcal{M}_5 + 2\mathcal{M}_{23}),$$

$$\mu_5 + 2\mu_2\mu_3 = U_5^{[S/S]} + 2U_{23}^{[S/S]} = \frac{(S-1)(S-2)}{S^2}(m_5 + 2m_{23})$$

so that

$$m_5 + 2m_{23} = \frac{S^2}{(S-1)(S-2)}(\mu_5 + 2\mu_2\mu_3),\ S^{[-3]}(m_5 + 2m_{23}) = S^3(\mu_5 + 2\mu_2\mu_3),$$

$$m_5 = \frac{S^4}{(S-1)(S-2)(S-3)(S-4)}\mu_5 -$$
$$5\frac{S^2(S-2)}{(S-1)(S-2)(S-3)(S-4)}(\mu_5 + 2\mu_2\mu_3).$$

Substituting these values of m_5 and $(m_5 + 2m_{23})$ in the above expressions for $U_5^{[N]}$, we finally arrive at

$$U_5^{[N/S]} = \frac{(N-1)(N-2)(N-3)(N-4)S^4}{N^4(S-1)(S-2)(S-3)(S-4)}\mu_5 +$$
$$5\frac{(N-1)(N-2)(S-N)(NS-2S-2N+2)S^2}{N^4(S-1)(S-2)(S-3)(S-4)}(\mu_5 + 2\mu_2\mu_3),$$

or, in another notation,

$$U_5^{[N/S]} = \frac{N^{[-5]}S^5}{N^5S^{[-5]}}\mu_5 + 5\ \frac{N^{[-3]}S^3(S-N)(NS-2S+2N+2)}{N^3S^{[-3]}N^2(S-3)(S-4)}(\mu_5 + 2\mu_2\mu_3).$$

8. As an example of a uniform system of trials we have until now considered the layout of the unreturned ticket. The scheme of an attached ticket can serve as another illustration, formally very much resembling the first one but at the same time contrary to it in a certain sense [12]. We arrive at the latter when, after each extraction, not only the drawn ticket is returned back, but a new ticket with the same number is also put in the urn. Thus, N consecutively added tickets correspond to N consecutive extractions. Without dwelling on this pattern [13], we adduce a few formulas describing it. Making use of notation (10), we obtain a formula similar to the known to us relation (12) for the pattern of an unreturned ticket:

$$m_{h_1h_2\cdots h_N}\ (1/S^{[+N]})\sum_{r=1}^{N}\ \{S^r\sum_{k_1k_2\ldots k_r}[|\underline{k_1-1}\ |\underline{k_2-1}\ \ldots\ |\underline{k_r-1}\cdot$$
$$\sum_{h_1h_2\ldots h_N} m_{h_1+h_2+\ldots+h_\alpha}\ m_{h_{\alpha+1}+h_{\alpha+2}+\ldots+h_\beta}\ldots m_{h_\gamma+\ldots+h_\delta}\ ]\}$$

where the sums extend over the same domains.

Passing on to the parameters $\{\mu\}$ we find that, in particular, as in (12),

$$\mu_{11} = \frac{1}{S+1}\mu_2,\ \mu_{12} = \frac{1}{S+1}\mu_3,\ \mu_{111} = \frac{2}{(S+1)(S+2)}\mu_3,\ \mu_{13} = \frac{1}{S+1}\mu_4,$$
$$\mu_{112} = \frac{2}{(S+1)(S+2)}\mu_4 + \frac{S}{(S+1)(S+2)}\mu_2^2,\ \mu_{22} = \frac{1}{S+1}\mu_4 + \frac{S}{S+1}\mu_2^2,$$

$$\mu_{1111} = \frac{6}{(S+1)(S+2)(S+3)}\mu_4 + \frac{3S}{(S+1)(S+2)(S+3)}\mu_2^2,$$

so that

$$m_2 = \mu_2 - \mu_{11} = \frac{S}{S+1}\mu_2,\ m_3 = \mu_3 - 3\mu_{12} + 2\mu_{111} = \frac{S^2}{(S+1)(S+2)}\mu_3,$$
$$m_4 = \mu_4 - 4\mu_{13} + 6\mu_{112} - 3\mu_{1111} =$$
$$\frac{S^3}{(S+1)(S+2)(S+3)}\mu_4 + \frac{S(2S+3)}{(S+1)(S+2)(S+3)}(\mu_4 + 3\mu_2^2),$$

$$m_{22} = \mu_{22} - 2\mu_{112} + \mu_{1111} = \frac{S^3}{(S+1)(S+2)(S+3)}\mu_2^2 +$$
$$\frac{S}{(S+2)(S+3)}(\mu_4 + 3\mu_2^2).$$

In the same way we obtain

$$m_4 + 3m_{22} = \frac{S}{S+1}(\mu_4 + 3\mu_2^2) \text{ or } S^{[+2]}(m_4 + 3m_{22}) = S^2(\mu_4 + 3\mu_2^2),$$
$$m_5 + 2m_{23} = \frac{S^2}{(S+1)(S+2)}(\mu_5 + 2\mu_2\mu_3) \text{ or } S^{[+3]}(m_5 + 2m_{23}) = S^3(\mu_5 + 2\mu_2\mu_3).$$

Denoting $U_r^{[N]}$ for the pattern of an attached ticket by $U_r^{[S/N]}$ we shall have finally

$$U_2^{[S/N]} = \frac{(N-1)S}{N(S+1)}\mu_2,\ U_3^{[S/N]} = \frac{(N-1)(N-2)S^2}{N^2(S+1)(S+2)}\mu_3,$$

$$U_4^{[S/N]} = \frac{(N-1)(N-2)(N-3)S^3}{N^3(S+1)(S+2)(S+3)}\mu_4 +$$
$$\frac{(N-1)(S+N)(2NS-3S+3N-3)S}{N^3(S+1)(S+2)(S+3)}(\mu_4 + 3\mu_2^2),$$

$$U_5^{[S/N]} = \frac{(N-1)(N-2)(N-3)(N-4)}{N^4}\frac{S^4}{(S+1)(S+2)(S+3)(S+4)}\mu_5 +$$
$$\frac{5(N-1)(N-2)(S+N)(NS-2S+2N-2)S^2}{N^4(S+1)(S+2)(S+3)(S+4)}(\mu_5 + 2\mu_2\mu_3).$$

Noting that all these formulas pass on to the corresponding formulas for the scheme of an unreturned ticket when S is replaced by $-S$, we convince ourselves that the pattern of an attached and the unreturned tickets formally represent one and the same type of uniform systems.

9. In §2 we offered an indication of the concept of uniform system based on the relation (1) persisting under any permutation of the indices. Noting that for independent trials relation (2b) held when the factors in its right side changed places, we satisfy ourselves that, in this case, the same property is true with respect to the indices in the left side of (1) and that, therefore, if the law of distribution of the variable remains fixed, the totality of the trials possesses the property of uniformity.

The uniformity of a totality is thus founded on the property of a product not to change its value when the order of its factors is changed, – on the commutativity of the product. When considering the issue under a somewhat different *stochastic* [1, p. 3] point of view, we may say: The uniformity of a totality is based on the persistence of the expectation in the left side of (2a) under any order of the trials. It is this property that we call *stochastic commutativity*.

It is not difficult to convince ourselves that, for a system, the property of uniformity is also founded on the stochastic commutativity of the trials: here also, the constancy of the value of $m_{h_1 h_2 \ldots h_N}$ for any permutation of its indices, and the independence of the expectation mentioned above of the order of the trials on the random variable x, are only different verbal formulations of one and the same proposition.

The concept of stochastic commutativity can also be presented in a somewhat different form. Denoting, as we did before, the probabilities that the variable x takes values $x^{(1)}$, $x^{(2)}$, …, $x^{(k)}$, by $p_1, p_2, \ldots, p_k$ respectively, let us agree to designate the probability that x takes values $x^{(i1)}$, $x^{(i2)}$, …, $x^{(iN)}$ at the first, the second, …, the N-th trial **14**, by $p_{i1\ i2\ \ldots\ iN}$. Noting that for mutually independent trials this probability equals $p_{i1} \cdot p_{i2} \ldots p_{iN}$ and that the latter product does not change when its factors change places, we become convinced that in case of a totality of trials the probability by $p_{i1\ i2\ \ldots\ iN}$ remains constant for any permutation of its indices. In other words, the probability that in N trials the variable x will take the values $x^{(i1)}$, $x^{(i2)}$, …, $x^{(iN)}$ does not depend on the order in which these values appear.

For a totality of trials this proposition is naturally trivial, because, if the trials are mutually independent, their order is absolutely indifferent and we may enumerate them as we please. However, the main indication of stochastic commutativity of trials (as we, in conformity with the above, are calling the property of independence of the quantity $p_{i1} \cdot p_{i2} \ldots p_{iN}$ of the order of the trials) is not their mutual independence at all, but the constancy of the law of distribution of the appropriate variable. Abandoning therefore the supposition of independence of the trials, we obtain stochastically commutative *systems* of trials; that is, *systems* for which the order of the trials is stochastically indifferent. The patterns of the unreturned and an attached tickets can illustrate this. Making use of notation (10), we find that the probability of drawing h_1 tickets with number $x^{(1)}$, h_2 tickets with number $x^{(2)}$, …, h_k tickets with number $x^{(k)}$ from the urn in a definite order will be, for the various arrangements and having $N = h_1 + h_2 + \ldots + h_k$, is, for

A returned ticket: $s_1^{h_1} s_2^{h_2} \ldots s_k^{h_k} / S^N$;

An unreturned ticket: $s_1^{[-h_1]} s_2^{[-h_2]} \ldots s_k^{[-h_k]} / S^{[-N]}$;

An attached ticket: $s_1^{[+h_1]} s_2^{[+h_2]} \ldots s_k^{[+h_k]} / S^{[+N]}$.

Neither is it difficult to show that, assuming such a definition of stochastic commutativity, that its notion coincides with the concept of uniformity. Noting that

$$m_{h_1 h_2 \ldots h_N} = \sum_{i1=1}^{k} \sum_{i2=1}^{k} \ldots \sum_{iN=1}^{k} p_{i1} \cdot p_{i2} \ldots p_{iN}\ x^{(i1)^{h1}} x^{(i2)^{h2}} \ldots x^{(iN)^{hN}}$$

we satisfy ourselves that the constancy of $p_{i1} \cdot p_{i2} \ldots p_{iN}$ for any permutation of its indices leads to the same property for $m_{h_1 h_2 \ldots h_N}$, *i.e.*, that a system of trials obeying the condition of stochastic commutativity is always uniform. Inversely, an algebraic analysis can show that a constancy of $m_{h_1 h_2 \ldots h_N}$ for any permutation of its indices leads to the same property for $p_{i1} \cdot p_{i2}$

… p_{iN}; that is, that a uniform system always satisfies the condition of stochastic commutativity. Both concepts thus coincide.

10. We have brought a uniform system and a totality under a common concept of *trials obeying the condition of stochastic commutativity*. The formation of such a common notion seems all the more expedient because it covers various types of interrelations between the trials which cannot be empirically distinguished one from another. No tests can be provided which could have, for example, establish, without taking into account prior data, {only} by studying some numbers marked on the extracted tickets, whether the drawings were made according to the scheme of a returned, of an unreturned, or an attached ticket.

It was Chuprov [4] who asked himself whether totalities from the various types of uniform systems can be empirically distinguished from each other. Indicating that the Lexian test for the normal stability of a series (the coefficient of dispersion Q^2 should be approximately equal to 1) is necessary but not sufficient for the assumption of normal stability (*i.e.*, for the constancy of the law of distribution of the variable in all the trials and for their mutual independence) to hold, he discovered that the other methods provided by him were no better. On the basis of our constructions and in particular by applying the parameters $\{M\}$ introduced by us, Chuprov's findings can be represented even more obviously and, besides, in a generalized form. And at the same time it is revealed with an absolute clearness exactly why the different types of a uniform system cannot be empirically distinguished one from another or from a totality.

A detailed mathematical analysis of the issue would have demanded more place than we have at our disposal and we shall therefore restrict our attention to brief general indications. The methods by which an investigator attempts to establish whether a series of empirical values considered by him conforms to some stochastic assumptions may be separated into two groups in which

1) A number of functions of the empirical values of the variable, whose expectations are equal to each other provided that the stochastic suppositions are valid, is constructed.

2) Such a function of these empirical values of the variable whose expectation under the given assumptions takes a definite numerical value, is constructed. Thus, the Lexian test of normal stability, $EQ^2 = 1$.

For these methods of either group to lead to the solution {of the problem}, the suggested tests should be not only necessary but also sufficient. This means that the initial equalities should only take place when the given assumptions are fulfilled, and not to be valid under other suppositions from which the researcher attempts to separate himself. When pondering over the operation of conjunction of a totality into a uniform system as studied above, we are at once convinced that the methods of the first group cannot lead to the isolation of the case in which we deal with a totality of trials from those instances where we have some modification of a uniform system before us. Indeed, denoting by $F(x_1; x_2; \ldots; x_N)$ and $\Phi(x_1; x_2; \ldots; x_N)$ two integral rational functions of the empirical values of the variable x, whose expectations are equal to each other under the assumptions of a totality, we see at once, when mentally carrying out the operation of conjunction, that the equality $EF = E\Phi$ should also be valid for the general case of a uniform system.

As to the methods that we attributed to the second group, we may in essence extend the same course of reasoning to these also, and to consider the numerical constant as that second function of the empirical values of the variable, whose expectation is being compared with the expectation of the constructed coefficient. A more precise mathematical analysis which we do not here reproduce, leads to the same conclusion. It shows that for any coefficient whose expectation is equal to some numerical constant under the assumptions that the variable obeys a fixed law of distribution and its values are mutually independent, we obtain for this expectation the same value as in the general case of a uniform system.

It follows that the question about the empirical possibility of distinguishing between a normal and a non-normal dispersion should be answered in the negative.

Notes

1. See [5] where the author establishes the notion of uniform connection to which our concept of uniform system is indeed adjoined. Note that a uniform system of trials can only exist if the law of distribution of the variable remains fixed during all the trials. This directly follows from the definition if we set all the h's excepting one of them equal to zero.

2. We suppose that $]a + b + c + \ldots[=]a[+]b[+]c[+ \ldots$ and that, in the same ways, $[a + b + c + \ldots] = [a] + [b] + [c] + \ldots$; that is, a disjunction (a conjunction) of a *sum* is defined as a disjunction (a conjunction) of its *terms*.

3. See [3, p. 151, formulas (10) and (11)]. We replace μ_2, μ_3 and μ_4 in these formulas by their expressions through m_1, m_2, m_3 and m_4 (same source, p. 148, formula (2)). For a totality of trials we write $m_{r]N[}$; and for a uniform system $m_{r[N]}$ instead of {Chuprov's notation} $m_{r(N)}$.

4. Cf. [5]. Note that when conjuncting a power, we write it down as a product. For example, $m_1^2 = m_1m_1$ and $[m_1^2] = [m_1m_1] = m_{11}$ etc.

5. We always suppose that $j > i + h_1$, $k > \ldots > j + h_2$, etc, *i.e.*, that not a single value from among $x_j, x_{j+1}, \ldots, x_{j+h_2}$ coincides with any of the values $x_i, x_{i+1}, \ldots, x_{i+h_1}$ and in the same way that not a single value from among $x_k, x_{k+1}, \ldots, x_{k+h_N}$ coincides with any of the previous x's, etc.

6.{Here and in the appropriate instances below the author uses the Russian low-case *м*.}

7. Since

$$M_{1h_2\ldots h_N} = \mathrm{E}[(x_1 - x_2)(x_3 - x_4)(x_3 - x_5)\ldots(x_3 - x_{3+h_2})\ldots] =$$
$$\mathrm{E}[x_1(x_3 - x_4)\ldots(x_3 - x_{3+h_2})\ldots] - \mathrm{E}[x_2(x_3 - x_4)\ldots(x_3 - x_{3+h_2})\ldots] = 0.$$

8. Absolutely in the same way as in the case of a totality in which we have, because of $\mu_1 = 0$,

$$\sum_{i=0}^{h}(-1)^i C_h^i m_{h-i} m_1^i = \sum_{i=0}^{h}(-1)^i C_h^i \mu_{h-i}\mu_1^i = \mu_h,$$

here, in the case of a uniform system, we obtain

$$\sum_{i=0}^{h}(-1)^i C_h^i m_{h-i,\,1,1,\ldots,1} = \sum_{i=0}^{h}(-1)^i C_h^i \mu_{h-i,\,1,1,\ldots,1} =$$
$$\sum_{i=0}^{h}(-1)^i C_h^i M_{h-i,\,1,1,\ldots,1} = M_h$$

because of $M_{1h_2\ldots h_N} = 0$.

9. See [3, p. 186, formula (6) and p. 189, formula (12)]. There,

$$r^{[-k]} = r(r-1)\ldots(r-k+1). \qquad (10)$$

10. Thus (cf. Note 8), we have

$$M_h = \sum_{i=0}^{h}(-1)^i C_h^i \mu_{h-i,\,1,1,\ldots,1}.$$

11. It can be considered as a particular case of a more general identity,

$$U_4^{]N[} + 3U_{22}^{]N[} = [(N-1)/N]\,(\mu_4 + 3\mu_2^2)$$

or, when passing on to a uniform system, of

$$U_4^{[N]} + 3U_{22}^{[N]} = [(N-1)/N]\,(M_4 + 3M_{22}).$$

Supposing that $N = S$, we indeed find that

$$(\mu_4 + 3\mu_2^2) = U_4^{[S/S]} + 3U_{22}^{[S/S]} = [(S-1)/S]\,(M_4 + 3M_{22}),$$
$$S^{[-2]}(M_4 + 3M_{22}) = S^2(\mu_4 + 3\mu_2^2).$$

12. The pattern of an unreturned ticket is contrary to that of an attached ticket just as, in the Lexian terminology, a scheme of a supernormally stable statistical series is opposed to the arrangement of a series with a subnormal stability.

13. The uniformity of the system of trials obtained on the basis of the new layout is very simply revealed by means of a test established in the next section.

14. Magnitudes $x^{(i1)}, x^{(i2)}, \ldots, x^{(iN)}$ represent some of the values $x^{(1)}, x^{(2)}, \ldots, x^{(k)}$; some or all of them may be identical. It is assumed here as it was before that the law of distribution of the variable remains fixed.

References

1. Bortkiewicz, L., von (1917), *Die Iterationen*. Berlin.

2. Chuprov, A.A. (1918 – 1919), Zur Theorie der Stabilität statistischer Reihen. *Skand. Aktuarietidskr.*, t. 1, pp. 199 – 256; t. 2, pp. 80 – 133.

3. --- (1918 – 1919 and 1921), On the mathematical expectation of the moments of frequency distributions. *Biometrika*, vol. 12, pp. 140 – 169 and 185 – 210; vol. 13, pp. 283 – 295.

4. --- (1922), Ist die normale Dispersion empirisch nachweisbar? *Nordisk Stat. Tidskr.*, t. 1, pp. 369 – 393.

5. --- (1923), On the mathematical expectation of the moments of frequency distributions in the case of correlated observations. *Metron*, t. 2, pp. 461 – 493 and 646 – 683.

6. Seneta, E. (1987), Chuprov on finite exchangeability, expectation of ratios and measures of association. *Hist. Math.*, vol. 14, pp. 243 – 257.

7. Sheynin, O. (1990, in Russian), *Chuprov*. Göttingen, 1996.

12. A.N. Kolmogorov. Determining the Center of Scattering and the Measure of Precision Given a Restricted Number of Observations

Izvestia Akademii Nauk SSSR, ser. Math., vol. 6, 1942, pp. 3 – 32

Foreword by Translator

This paper was apparently written hastily, and, at the time, its subject-matter did not perhaps belong to the author's main scientific field. He mixed up Bayesian ideas and the concept of confidence intervals and in §4 he showed that the posterior distribution of a parameter was asymptotically normal without mentioning that this was due to the well-known Bernstein- von Mises theorem. Points of more general interest are Kolmogorov's debate with Bernstein on confidence probability and, in Note 11, a *new axiom of the theory of probability.* The author apparently set high store by artillery (even apart from ballistics) as a field of application for probability theory. Indeed, this is seen from Gnedenko's relevant statement [1, p. 211] which he inserted even without substantiating it, a fact about which I then expressed my doubts). And Gnedenko certainly attempted to remain in line with his former teacher.

In both of Kolmogorov's papers here translated, apparently for the benefit of his readers, the author numbered almost all the displayed formulas whether mentioned in the sequel or not. I only preserved these numbers in the paper just below.

1. Gnedenko, B.V., Sheynin, O.B. (1978, in Russian), Theory of probability. A chapter in *Mathematics of the 19th Century* (pp. 211 – 288). Editors, A.N. Kolmogorov, A.P. Youshkevich. Basel, 1992 and 2001.

* * *

This paper is appearing owing to two circumstances. First, intending to explicate his viewpoint and investigations on the stochastic justification of mathematical statistics in several later articles, the author considers it expedient to premise them by a detailed critical examination of the existing methods carrying it out by issuing from a sufficiently simple classical problem of mathematical statistics. For this goal it is quite natural to choose the problem of estimating the parameters of the Gaussian law of distribution given n independent observations.

Second, the author was asked to offer his conclusion about the differences of opinion existing among artillery men on the methods of estimating the measure of precision by experimental data, see for example [10 – 12]. The author became therefore aware of the desirability of acquainting them with the results achieved by Student and Fisher concerning small samples. Exactly these issues definitively determined the concrete subject-matter of this article.

It is clear now that the article only mainly claims to be methodologically interesting. The author believes that the new factual information is represented here by the definition of sufficient statistics and sufficient systems of statistics (§2) and by the specification of the remainder terms in limit theorems (§4). The need to compare critically the various approaches to the studied problems inevitably led to a rather lengthy article as compared with the elementary nature of the problems here considered.

Introduction. Suppose that random variables

$$x_1, x_2, \ldots, x_n \qquad (1)$$

are independent and obey the Gaussian law of distribution with a common center of scattering a and common measure of precision h. In this case the n-dimensional law of distribution of the x_i's is known to be determined by the density

$$f(x_1; x_2; \ldots; x_n|a; h) = (h^n/\pi^{n/2})\exp(-h^2S^2), \qquad (2)$$

$$S^2 = (x_1 - a)^2 + (x_2 - a)^2 + \ldots + (x_n - a)^2. \qquad (3)$$

Instead of formula (2) it is sometimes convenient to apply an equivalent formula

$$f(x_1; x_2; \ldots; x_n \mid a; h) = (h^n/\pi^{n/2})\exp[-h^2S_1^2 - nh^2(\bar{x} - a)^2], \qquad (4)$$
$$\bar{x} = (x_1 + x_2 + \ldots + x_n)/n,\ S = (x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2. \quad (5; 6)$$

All courses in probability theory for artillery men consider the following three problems.

1. Assuming h to be known, approximately estimate a by the observed values of (1).
2. Assuming that a is known, approximately estimate h by the same values.
3. Again issuing from (1), approximately determine both a and h.

Practically this means that it is required to indicate functions $\bar{a}$ and $\bar{h}$ of the known magnitudes, – of (1) and *h* in Problem 1; of (1) and *a* in Problem 2; and of only (1) in Problem 3, – which should be most reasonably chosen as the *approximate values* of the estimated magnitudes.

In addition, it is required to estimate the mean precision attained by applying the approximate $\bar{a}$ and $\bar{h}$. And it is sometimes additionally needed to indicate such functions a' and a'', h' and h'' of the magnitudes given in the problem under consideration, that it would be possible to state, without the risk of making wrong conclusions too often, that $a' \le a \le a''$ and, respectively, $h' \le h \le h''$. Here, a' and a'' are called the *confidence limits* of *a*, and h' and h'', the confidence limits of *h*.

1. The Classical Method. The classical method of solving the formulated problems is based on the assumption that, before observing the values of (1), the estimated magnitudes (*a*, in Problem 1; *h*, in Problem 2; and both *a* and *h*, in Problem 3) obey some *prior* law of distribution. Supposing that this law is known, it is possible to calculate the conditional (*posterior*) law of distribution of the estimated parameters if the results of observation (1) are known. For the sake of simplicity we restrict our attention to the case of continuous prior laws given by the appropriate densities.

In Problem 1, applying the Bayes theorem, we shall obtain the following expression for the conditional density of *a*, given the values of (1):

$$\varphi_1(a \mid x_1; x_2; \ldots; x_n) = \frac{\exp[-nh^2(a-\bar{x})^2]\varphi_1(a)}{\int_{-\infty}^{\infty}\exp[-nh^2(a-\bar{x})^2\varphi_1(a)da}. \qquad (7)$$

Here, $\varphi_1(a)$ is the unconditional (prior) density of *a* before observation.

In Problems 2 and 3 the corresponding formulas are

$$\varphi_2(h \mid x_1; x_2; \ldots; x_n) = \frac{h^n \exp(-h^2S^2)\varphi_2(h)}{\int_0^{\infty} h^n \exp(-h^2S^2)\varphi_2(h)dh}, \qquad (8)$$

$$\varphi_3(a; h \mid x_1; x_2; \ldots; x_n) = \frac{h^n \exp[-h^2S_1^{\,2} - nh^2(a-\bar{x})^2]\varphi_3(a;h)}{\int_{-\infty}^{\infty}\int_0^{\infty} h^n \exp[-h^2S_1^{\,2} - nh^2(a-\bar{x})^2]\varphi_3(a;h)dhda} \qquad (9)$$

where $\varphi_2(h)$ and $\varphi_3(a; h)$ are the respective unconditional densities.

Formulas (7) – (9) are unfit for direct application not only because they are involved, but, mainly, since the prior densities included there are usually unknown. In addition, it should be clearly understood that the very assumption about the *existence* of any certain prior distribution of *a* and *h* can only be justified for some separate and sufficiently restricted classes of cases. For example, it is quite reasonable to consider the law of distribution of the measure of precision when shooting from a rifle under some definite conditions and for a randomly chosen soldier [1] from a given regiment [2]. It is senseless, however, to discuss the prior distribution of this measure for shooting in general (under any conditions and from any firearms belonging to times past or to future years).

2. Sufficient Statistics and Systems of Sufficient Statistics. In this section, we assume that the prior densities $\varphi_1(a)$, $\varphi_2(h)$ and $\varphi_3(a; h)$ exist. No practically useful estimates of *a* or *h* can be obtained by this assumption taken by itself, but some sufficiently instructive general corollaries may be elicited from it.

For Problem 1, formula (7) shows that the conditional density $\varphi_1(a \mid x_1; x_2; \ldots; x_n)$ is completely determined by the prior density $\varphi_1(a)$, the measure of precision h (which is assumed here to be given beforehand) and the mean value $\bar{x}$ of the observed magnitudes (1). It follows that, for any prior distribution of a and a given h, all that (1) additionally contributes to the estimation of a is included in only one magnitude, $\bar{x}$. It is therefore said that, in Problem 1, $\bar{x}$ is a *sufficient statistic* for estimating a.

The general definition of a sufficient statistic may be thus formulated [3]. Let the observed magnitudes (1) have a law of distribution depending on parameters $\theta_1, \theta_2, \ldots, \theta_s$ whose values are unknown. Any function of the observed magnitudes (1) is called a *statistic* [4]. A statistic χ is called sufficient for parameter θ_j if the conditional distribution of the parameter, given (1), is completely determined by the prior distribution of $\theta_1, \theta_2, \ldots, \theta_s$ and the value of the statistic χ. Formula (8) shows that S in Problem 2 is a sufficient statistic for estimating h.

The definition of a sufficient statistic is generalized as follows. A system of functions

$$\chi_i(x_1; x_2; \ldots; x_n),\ i = 1, 2, \ldots, m$$

is called a *sufficient system of statistics* for the system of parameters $\theta_1, \theta_2, \ldots, \theta_k$ (where $k \le s$, so that, generally speaking, this system only constitutes a part of the complete system of parameters) if the conditional k-dimensional distribution of $\theta_1, \theta_2, \ldots, \theta_k$ given (1) is completely determined by the prior distribution of $\theta_1, \theta_2, \ldots, \theta_s$ and the values of the statistics $\chi_1, \chi_2, \ldots, \chi_m$.

Formula (9) shows that in Problem 3 the magnitudes $\bar{x}$ and S_1 constitute a sufficient system of statistics for the system of parameters a and h. For each of these parameters taken separately the system ($\bar{x}$; S_1) is obviously also sufficient [5].

Following Fisher, the results obtained above may be thus summarized: *All the information* [6] *included in the observations* (1) *with respect to a and h is determined, in Problem* 1, *by the value of* $\bar{x}$; *in Problem* 2, *by S*; and, in Problem 3, by the values of $\bar{x}$ and S_1.

Consequently, it should be resolved that, under our assumptions, the search for the most perfect methods of estimating a and h may be restricted to such that only make use of the observed magnitudes (1) by calculating the appropriate values of $\bar{x}$ (in Problem 1), S (in Problem 2), and $\bar{x}$ and S_1 (in Problem 3). For example, we may refuse to consider the estimation of a by the median of the observations or by the mean of the extreme observational results

$$d = (x_{\max} + x_{\min})/2. \qquad (10)$$

Such methods can only be interesting in that they are very simple [7].

In particular, when searching for a most sensible type of functions $\vec{a}, \bar{h}$, a', a'', h', h'' (Introduction), it is natural to restrict our efforts by functions $\bar{a}$, a', a'' only depending on $\bar{x}$ and $\bar{h}$ (Problem 1); by $\bar{h}$, h' and h'', only depending on S and a (Problem 2); and by $\vec{a}, \bar{h}$, a', a'', h', h'' only depending on $\bar{x}$ and S_1 (Problem 3).

This conclusion, quite conforming to the general opinion of the practitioners, and of artillery men in particular, could have been justified by other methods as well, not at all depending on the assumption about the existence of the prior distributions of a and h.

3. The Hypothesis of a Constant Prior Density and Its Criticism. Many treatises intended for artillery men assume that the prior densities in formulas (7) – (9) are constant, *i.e.*, that it is possible to consider that

$$\varphi_1(a) = \text{Const},\ \varphi_2(h) = \text{Const},\ \varphi_3(a; h) = \text{Const}.$$

Strictly speaking, this assumption is not only arbitrary, it is also certainly wrong since it contradicts the demands that

$$\int_{-\infty}^{\infty} \varphi_1(a)\, da = 1, \int_{-\infty}^{\infty} \varphi_2(h)\, dh = 1, \int_{-\infty}^{\infty} \int_0^{\infty} \varphi_3(a; h)\, dh\, da = 1$$

which follow from the main principles of the theory of probability.

In some cases, however, the approximate constancy of the prior distributions can persist within a sufficiently large range of the arguments, a and h. In such instances, it may be hoped that the formulas that follow from (7) – (9) when replacing the functions φ_1, φ_2 and φ_3 by constants, will be approximately correct. In §4 we shall see that it is possible to reckon on this when the number of observations n is sufficiently large; and, in Problem 1, also when n is small if only the mean square deviation

$$\sigma = 1/(h\sqrt{2}) \qquad (11)$$

is sufficiently small as compared with the a priori admissible range of a. Replacing the functions φ_1, φ_2 and φ_3 in formulas (7) – (9) by constants, we obtain

$$\varphi_1(a \mid x_1; x_2; \ldots; x_n) = (h\sqrt{n/\pi})\exp[-nh^2(a - \bar{x})^2], \qquad (12)$$

$$\varphi_2(h \mid x_1; x_2; \ldots; x_n) = \frac{2S^{n+1}}{\Gamma[(n+1)/2]}\, h^n \exp(-h^2S^2), \qquad (13)$$

$$\varphi_3(a; h \mid x_1; x_2; \ldots; x_n) = \frac{2\sqrt{n}S_1^{\,n}}{\sqrt{\pi}\,\Gamma(n/2)}\, h^n \exp[-h^2S_1^{\,2} - nh^2(a - \bar{x})^2]. \qquad (14)$$

For Problem 1 formula (12) leads to the conclusion that the conditional distribution of a given (1) will be Gaussian with parameters $\bar{x}$ and $h\sqrt{n}$. Being simple, this result is quite satisfactory. However, as indicated above, it is only established by unfoundedly assuming that $\varphi_1(a)$ = Const. In §4 we shall see that this result can {nevertheless} be justified as approximately correct under some sufficiently natural assumptions.

Formulas (13) and (14) are also obtained by means of unfounded assumptions, $\varphi_2(h)$ = Const and $\varphi_3(a; h)$ = Const. Indeed, their reliable substantiation is only possible in the form of a limit theorem establishing that, under some natural assumptions, and a *sufficiently large number of observations n*, these formulas are approximately correct. Nevertheless, we shall see in §4 that, given the same assumptions and large values of n, these formulas can be fairly approximated by much more simple formulas (18) and (19). Thus it occurs that for small values of n formulas (13) and (14) are unjustified, and superfluous otherwise; and that they have no practical significance at all.

To illustrate the arbitrariness of the results arrived at by issuing from the hypothesis of a constant prior density when n is not too large, we shall consider Problem 2 from a somewhat different viewpoint. The assumption that the mean square deviation σ possesses a constant prior density is not less natural than the same assumption about h. And this alternative assumption leads to

$$\varphi_2(h) = \text{Const}, \; \varphi_2(h) = \text{Const}/h^2,$$

$$\varphi_2(h \mid x_1; x_2; \ldots; x_n) = \frac{2S^{n-1}}{\Gamma[(n-1)/2]}\, h^{n-2} \exp(-h^2S^2). \qquad (13\text{bis})$$

Let us calculate now, by means of formulas (13) and (13bis), the conditional expectation of h given (1). We have, respectively,

$$\bar{h}^{*} = \frac{\Gamma[(n+2)/2]}{S\,\Gamma[(n+1)/2]},\quad \bar{h}^{**} = \frac{\Gamma(n/2)}{S\,\Gamma[(n-1)/2]}. \qquad (15;\ 15\text{bis})$$

It is easy to determine that

$$\bar{h}^{*} = \bar{h}^{**}[1 + (1/n)]. \qquad (16)$$

We see that for large values of n the difference between $\bar{h}^{*}$ and $\bar{h}^{**}$ is not large but that it can be very considerable otherwise.

4. Limit Theorems. Here, we will establish that, under some natural assumptions and a sufficiently large number of observations n (and in some cases, for Problem 1, for small values of n as well), it is possible to apply the approximate formulas

$$\varphi_1(a|x_1; x_2; \ldots; x_n) \sim (h\sqrt{n/\pi})\exp[-nh^2(a-\bar{x})^2], \qquad (17)$$
$$\varphi_2(h|x_1; x_2; \ldots; x_n) \sim (s\sqrt{2/\pi})\exp[-2S^2(h-\bar{h})^2], \qquad (18)$$
$$\varphi_3(a; h|x_1; x_2; \ldots; x_n) \sim (S_1\bar{h}_1\sqrt{2n}/\pi)\exp[-n\bar{h}_1^2(a-\bar{x})^2 - 2S_1(h-\bar{h}_1)^2] \qquad (19)$$

where

$$\bar{h} = (1/S)\sqrt{n/2},\quad \bar{h}_1 = (1/S_1)\sqrt{(n-1)/2}. \qquad (20,\ 21)$$

Introducing

$$\alpha = h\sqrt{n}\,(a-\bar{x}),\ \chi = S\sqrt{2}(h-\bar{h}),\ \alpha_1 = h_1\sqrt{n}\,(a-\bar{x}),\ \chi_1 = S_1\sqrt{2}(h-\bar{h}_1) \qquad (22-25)$$

whose conditional densities, given (1), are, respectively,

$$\psi_1(\alpha|x_1; x_2; \ldots; x_n) = (1/h\sqrt{n})\,\varphi_1(a|x_1; x_2; \ldots; x_n), \qquad (26)$$
$$\psi_2(\chi|x_1; x_2; \ldots; x_n) = (1/S\sqrt{2})\,\varphi_2(h|x_1; x_2; \ldots; x_n), \qquad (27)$$
$$\psi_3(\alpha_1; \chi_1|x_1; x_2; \ldots; x_n) = (1/S_1\bar{h}_1\sqrt{2n})\,\varphi_3(a; h|x_1; x_2; \ldots; x_n), \qquad (28)$$

we may rewrite formulas (17) – (19) as

$$\psi_1(\alpha|x_1; x_2; \ldots; x_n) \sim (1/\sqrt{\pi})\exp(-\alpha^2), \qquad (29)$$
$$\psi_2(\chi|x_1; x_2; \ldots; x_n) \sim (1/\sqrt{\pi})\exp(-\chi^2), \qquad (30)$$
$$\psi_3(\alpha_1; \chi_1|x_1; x_2; \ldots; x_n) \sim (1/\pi)\exp(-\alpha_1^2 - \chi^2). \qquad (31)$$

When considering Problem 1 the following limit theorem justifies the applicability of the approximate formula (29), – or, which is the same, of its equivalent, the formula (17).

Theorem 1. *If the prior density* $\varphi_1(a)$ *has a bounded first derivative and* $\varphi_1(\bar{x}) \neq 0$, *then, uniformly with respect to* α,

$$\psi_1(\alpha|x_1; x_2; \ldots; x_n) = (1/\sqrt{\pi})\exp(-\alpha^2)\{1 + O[1/(h\sqrt{n})](1 + |\alpha|)\}. \qquad (32)$$

Here, as $nh^2 \to \infty$ and for constant $\varphi_1(a)$ and $\bar{x}$, $O[1/(h\sqrt{n})]$ is a magnitude having an order not higher than $[1/(h\sqrt{n})]$, uniformly with respect to α.

To prove this theorem we note that, on the strength of (7), (22) and (26),

$$\psi_1(\alpha|x_1; x_2; \ldots; x_n) = \frac{\exp(-\alpha^2)\varphi_1[\bar{x}+(\alpha/h\sqrt{n})]}{\int_{-\infty}^{\infty}\exp(-\alpha^2)\varphi_1[\bar{x}+(\alpha/h\sqrt{n})]d\alpha}. \quad (33)$$

Since the first derivative of $\varphi_1(a)$ is bounded, we have, uniformly with respect to α,

$$\varphi_1[\bar{x} + (\alpha/h\sqrt{n})] = \varphi_1(\bar{x}) + \alpha\, O[1/h\sqrt{n}].$$

If $\varphi_1(\bar{x}) \neq 0$, inserting this estimate into (33), we obtain, after some transformations, the estimate (32).

Without assuming that the first derivative of $\varphi_1(a)$ is bounded and only demanding that $\varphi_1(a)$ be continuous at point $a = \bar{x}$, it would have been possible to obtain a weaker result

$$\psi_1(\alpha|x_1; x_2; \ldots; x_n) = (1/\sqrt{\pi}) \exp(-\alpha^2) + R$$

where $R \to 0$ as $nh^2 \to \infty$, uniformly with respect to α on any finite interval $\alpha' \le \alpha \le \alpha''$. On the contrary, when strengthening the assumption about the *smoothness* (bounded higher derivatives, analyticity, etc) of the function $\varphi_1(a)$, it would not have been possible to replace the factor $O[1/h\sqrt{n}]$ in the estimate (32) by any other one of a lower order. Indeed, when assuming a bounded second derivative of $\varphi_1(a)$, we may apply the estimate

$$\varphi_1[\bar{x} + (\alpha/h\sqrt{n})] = \varphi_1(\bar{x}) + \varphi'_1(\bar{x})\, [\alpha/(h\sqrt{n})] + \alpha^2\, O[1/nh^2].$$

Inserting it in formula (33) we shall obtain, after some transformations,

$$\psi_1(\alpha|x_1; x_2; \ldots; x_n) = (1/\sqrt{\pi})\exp(-\alpha^2)[1 + \frac{\varphi_1'(\bar{x})}{\varphi_1(\bar{x})}\frac{\alpha}{h\sqrt{n}} + (1 + \alpha^2)O[1/nh^2]. \quad (34)$$

This formula shows that for $\varphi'_1(\bar{x}) \neq 0$ and with a bounded second derivative $\varphi''_1(a)$ the correction term $O[1/h\sqrt{n}](1 + \alpha^2)$ in formula (32) has indeed order $1/h\sqrt{n}$ for any fixed $\alpha \neq 0$.

For the conditional expectation of a given (1)

$$E(a|x_1; x_2; \ldots; x_n) = \bar{x} + (1/h\sqrt{n})\, E(\alpha|x_1; x_2; \ldots; x_n) =$$
$$\bar{x} + (1/h\sqrt{n})\int_{-\infty}^{\infty} \alpha\psi_1(\alpha|x_1; x_2; \ldots; x_n)\, d\alpha$$

formula (32) leads to the estimate

$$E(a|x_1; x_2; \ldots; x_n) = \bar{x} + O(1/nh^2). \quad (35)$$

Here, the order of the correction term $O(1/nh^2)$ cannot be lowered by any additional assumptions about the *smoothness* of the function $\varphi_1(a)$ since for a bounded second derivative $\varphi''_1(a)$ it follows from (34) that

$$E(a|x_1; x_2; \ldots; x_n) = \bar{x} + \frac{\varphi_1'(\bar{x})}{2\varphi_1(\bar{x})}\frac{1}{nh^2} + O[1/(h\sqrt{n})^3]. \quad (36)$$

Formula (36) shows that for $\varphi'_1(\bar{x}) \neq 0$ and with a bounded $\varphi''_1(a)$ the correction term in formula (35) has indeed order $1/nh^2$.

When issuing from (35) and neglecting the terms of the order of $1/nh^2$, it is natural to assume $\bar{x}$ as the approximate value of the center of scattering a given (1). Because of (17) the measure of precision of this approximate value can be roughly considered to be $h\sqrt{n}$.

More precisely, the estimate [8]

$$E(a - \bar{x})^2|x_1; x_2; \ldots; x_n) = (1/2nh^2)[1 + O(1/h\sqrt{n})] \quad (37)$$

follows for the conditional expectation of the square of the deviation $(a - \bar{x})$ given (1), *i.e.*, for

$$E(a - \bar{x})^2|x_1; x_2; \ldots; x_n) = (1/nh^2)\, E(\alpha^2|x_1; x_2; \ldots; x_n) =$$
$$(1/nh^2)\int_{-\infty}^{\infty} \alpha^2\psi_1(\alpha|x_1; x_2; \ldots; x_n)\, d\alpha.$$

On the strength of (37) and supposing that the measure of precision of an approximate value $\bar{\theta}$ of some parameter θ given (1) is the magnitude

$$h(\bar{\theta}|x_1; x_2; \ldots; x_n) = \frac{1}{\sqrt{2E[(\theta-\bar{\theta})^2|x_1;x_2;\ldots;x_n]}}, \quad (38)$$

we obtain for $\bar{x}$ considered as an approximation of a

$$h(\bar{x}|x_1; x_2; \ldots; x_n) = h\sqrt{n}[1 + O(1/h\sqrt{n})]. \quad (39)$$

To obtain the confidence limits for a given (1) it is natural to consider the conditional probabilities

$$P[|a - \bar{x}| \leq (c/h\sqrt{n})|x_1; x_2; \ldots; x_n] = P(|\alpha| \leq c|x_1; x_2; \ldots; x_n) =$$
$$\int_{-c}^{c} \psi_1(\alpha|x_1; x_2; \ldots; x_n)\, d\alpha.$$

Their estimate follows [9] from (32):

$$P[|a - \bar{x}| \leq (c/h\sqrt{n})|x_1; x_2; \ldots; x_n] = (2/\sqrt{\pi})\int_0^c \exp(-\alpha^2)\, d\alpha + O(1/h\sqrt{n}). \quad (40)$$

Here, $O(1/h\sqrt{n})$ is a magnitude having order not higher than $(1/h\sqrt{n})$ uniformly with respect to c. Thus, neglecting $O(1/h\sqrt{n})$, we may say that, given (1), a is situated within the limits

$$\bar{x} - (c/h\sqrt{n}) \leq a \leq \bar{x} + (c/h\sqrt{n}) \quad (41a)$$

with probability

$$\omega = (2/\sqrt{\pi})\int_0^c \exp(-\alpha^2)\, d\alpha. \quad (41b)$$

The applicability of the approximate formula (30), or, which is the same, of its equivalent, the formula (18), to Problem 2 is justified by

Theorem 2. *If the prior density* $\varphi_2(h)$ *has a bounded first derivative and* $\varphi_2(\bar{h}) \neq 0$, *then, uniformly with respect to* χ,

$$\psi_2(\chi|x_1; x_2; \ldots; x_n) = (1/\sqrt{\pi})\exp(-\chi^2)[1 + (1 + |\chi|)\,O(1/\sqrt{n})]. \qquad (42)$$

We do not prove this theorem. The proof is somewhat more complicated than the proof of Theorem 1 but the ideas underlying both are quite similar. As before, the expression $O(1/\sqrt{n})$ in (42) stands for a magnitude having order $(1/\sqrt{n})$ uniformly with respect to χ for constant $\varphi_2(h)$ and $\bar{h}$ and as $n \to \infty$.

The estimate

$$E(h|x_1; x_2; \ldots; x_n) = \bar{h}\,[1 + O(1/n)] \qquad (43)$$

follows from (42) for the conditional expectation

$$E(h|x_1; x_2; \ldots; x_n) = \bar{h} + (1/S\sqrt{2})\,E(\chi|x_1; x_2; \ldots; x_n) =$$
$$\bar{h}\,[1 + (1/\sqrt{n})\,E(\chi|x_1; x_2; \ldots; x_n) = \bar{h}\,[1 + (1/\sqrt{n})\int_{-\infty}^{\infty} \chi\psi_2(\chi|x_1; x_2; \ldots; x_n)\,d\chi].$$

On the strength of (43), neglecting a relative error of order $1/n$, it is natural to assume $\bar{h}$ as the approximate value of h. If the precise expression for the prior density $\varphi_2(h)$ is unknown, it is unavoidable to neglect such relative errors when keeping to the viewpoint adopted in this section concerning the choice of an approximate value for h. Indeed, we have seen (end of §3) that, when replacing $\varphi_2(h) = \text{Const}$ by $\varphi_2(h) = \text{Const}/h^2$, the relative change in $E(h|x_1; x_2; \ldots; x_n)$ is $1/n$. It would be easy to adduce such examples of the same change in $E(h|x_1; x_2; \ldots; x_n)$ caused by the change of $\varphi_2(h)$ where these functions would have obeyed all the demands necessary for densities and possessed bounded derivations of any high order.

Owing to(18) we may approximately assume that the measure of precision of the approximate value $\bar{h}$ of h is $(S/\sqrt{2}) = \sqrt{n}/\bar{h}$. More precisely

$$E[(h - \bar{h})^2|x_1; x_2; \ldots; x_n] = (\bar{h}^2/2n)[1 + O(1/\sqrt{n})], \qquad (44)$$
$$h(\bar{h}|x_1; x_2; \ldots; x_n) = \frac{1}{\sqrt{2E[(h-\bar{h})^2|x_1; x_2; \ldots; x_n]}} = (\sqrt{n}/\bar{h})[1 + O(1/\sqrt{n})]. (45)$$

Finally, it easily follows from (42) that

$$P[|h - \bar{h}| \leq (c\bar{h}/\sqrt{n})|x_1; x_2; \ldots; x_n] = (2/\sqrt{\pi})\int_0^c \exp(-\chi^2)\,d\chi + O(1/\sqrt{n}). \ (46)$$

Neglecting $O(1/\sqrt{n})$, we may therefore say that, given (1), h is situated within the limits

$$\bar{h}\,(1 - c\sqrt{n}) \leq h \leq \bar{h}\,(1 + c\sqrt{n})$$

with probability (41b).

For Problem 3 we have

Theorem 3. *If the prior density* $\varphi_3(a; h)$ *has bounded first derivatives with respect to a and h, and* $\varphi_3(a; \bar{h}) \neq 0$, *then, uniformly with respect to* α_1 *and* χ_1,

$$\psi_3(\alpha_1; \chi_1|x_1; x_2; \ldots; x_n) = (1/\pi)\exp(-\alpha_1^2 - \chi_1^2)[1 + (1 + |\alpha_1| + |\chi_1|)O(1/\sqrt{n})]. \quad (47)$$

As in Theorem 2, $O(1/\sqrt{n})$ denotes a magnitude having order $(1/\sqrt{n})$ uniformly with respect to α_1 and χ_1 if $\varphi_3(a; h)$, $\bar{x}$ and $\bar{h}_1$ are constant and $n \to \infty$. The proof is quite similar to that of Theorem 2 and we do not adduce it.

It follows from (47) that

$$E(a|x_1; x_2; \ldots; x_n) = \bar{x} + (1/\bar{h}_1)\, O(1/n), \quad (48)$$

$$E[(a - \bar{x})^2|x_1; x_2; \ldots; x_n] = (1/2n\bar{h}_1^{\,2})[1 + O(1/\sqrt{n})], \quad (49)$$

$$P[|(a - \bar{x})| \leq (c/\bar{h}_1\sqrt{n})|x_1; x_2; \ldots; x_n] = (2/\sqrt{\pi})\int_0^c \exp(-\alpha^2)d\alpha + O(1/\sqrt{n}), \quad (50)$$

$$E(h|x_1; x_2; \ldots; x_n) = \bar{h}_1[1 + O(1/\sqrt{n})], \quad (51)$$

$$E[(h - \bar{h})^2|x_1; x_2; \ldots; x_n] = (\bar{h}_1^{\,2}/2n)[1 + O(1/\sqrt{n})], \quad (52)$$

$$P[|h - \bar{h}_1| \leq (c\bar{h}_1/\sqrt{n})|x_1; x_2; \ldots; x_n] = (2/\sqrt{\pi})\int_0^c \exp(-\chi^2)\, d\chi + O(1/\sqrt{n}). \quad (53)$$

Neglecting magnitudes of the order of $1/n$ in Problem 3, it is natural to assume, on the strength of formulas (48) and (51), that $\bar{x}$ is an approximate value of a, and $\bar{h}_1$, an approximate value of h. And, issuing from formulas (49) and (52), we may approximately consider $\bar{h}_1\sqrt{n}$ *and* $\sqrt{n}/\bar{h}_1$ as the measures of precision of these approximations respectively.

Formulas (50) and (53) allow us to determine the confidence limits for a and h corresponding to within magnitudes of the order of $1/\sqrt{n}$ with a given probability ω.

As in Problem 2, the conditional expectation $E(h|x_1; x_2; \ldots; x_n)$ is only determined by formula (51) to within factor $[1 + O(1/\sqrt{n})]$. Therefore, in keeping with the viewpoint adopted in this section, discussions about choosing $\bar{h}_1$ or, for example, $\bar{h}_2 = \sqrt{n}/S_1\sqrt{2}$ as the approximate value for h are meaningless.

From the practical point of view, Theorems 1, 2 and 3 are not equally important. According to Theorem 1, the precision of the approximate formulas (29) and (17) increases for a constant $\varphi_1(a)$ not only with an increasing n, but also with the increase in the measure of precision h. Therefore, if the mean square deviation $\sigma = 1/h\sqrt{2}$ is small as compared with the a priori admissible region of a, we are somewhat justified in applying these formulas for small values of n (and even for $n = 1$) as well. However, in the case of Theorems 2 and 3 the remainder terms of formulas (42) and (47) only decrease with an increasing n so that they do not offer anything for small values of n.

5. The Fisherian Confidence Limits and Confidence Probabilities. As stated in the Introduction, the problem of approximately estimating parameter θ given (1) can be formulated in particular thus: It is required to determine such *confidence limits* $\theta'(x_1; x_2; \ldots; x_n)$ and $\theta''(x_1; x_2; \ldots; x_n)$ for θ that it would be practically possible to neglect the case in which θ is situated beyond the interval (the *confidence interval*) $[\theta'; \theta'']$.

In order to judge whether certain confidence limits θ'; θ'' for parameter θ are suitable for a *given* (1), it is natural to consider the conditional probability

$$P(\theta' \leq \theta \leq \theta''|x_1; x_2; \ldots; x_n). \quad (54)$$

If it is close to unity (for example, if it is 0.99 or 0.999), we will be inclined to assume, without considerable hesitation, that $\theta' \le \theta \le \theta''$. Consequently, when the conditional probabilities (54) are known for any θ' and θ'', it is natural to assume some probability ω sufficiently close to unity and to choose values of $\theta'(x_1; x_2; \dots; x_n)$ and $\theta''(x_1; x_2; \dots; x_n)$ for each system (1) such that

$$P(\theta' \le \theta \le \theta'' | x_1; x_2; \dots; x_n) = \omega; \qquad (55)$$

and, in addition, that, under this condition, the length of the interval $[\theta'; \theta'']$ will be the least possible.

For example, in Problem 1, assuming that formula (17) is correct, the shortest confidence interval for a obeying restriction (55) is given by formulas (41a; 41b). Note, however, concerning this example, that formula (17) may only be justified (even as an approximation) under rather restrictive assumptions specified in §4. As to the strict expression (7) for the conditional probability $\varphi_1(a|x_1; x_2; \dots; x_n)$, it includes the prior density $\varphi_1(a)$ which is usually unknown.

The same situation exists in most of the other problems of estimating parameters. The strict expression of the conditional probabilities (54) usually includes an unknown distribution of the parameters.

There exists an opinion, upheld in the Soviet Union by Bernstein (see, for example, [5]), that in cases in which the prior distribution of the parameters is unknown, the theory of probability cannot offer the practitioner anything excepting limit theorems similar to those indicated in §4. According to this point of view, if the prior distribution of the parameters is unknown, and given a restricted number of observations, an objective scientific approach to the most sensible choice of confidence limits for the estimated parameters is simply impossible.

Here, it is certainly true that the conditional distribution of the parameters, given the results of the observations, depends on the prior distribution of the same parameters, and we cannot disregard it. But the opinion, that the indication of sensible confidence limits for the estimated parameters is inseparably linked with considering conditional probabilities (54), is wrong.

In most practical (in particular, artillery) problems the matter concerns the establishment of general rules for estimating parameters to be recommended for systematic application to some vast category of cases. In this section devoted to confidence intervals, we are concerned with rules such as:

Under certain general conditions it is recommended to consider, whatever be the observational results (1), *that the value of parameter* θ *is situated within the boundaries* $\theta'(x_1; x_2; \dots; x_n)$ *and* $\theta''(x_1; x_2; \dots; x_n)$. When recommending such a rule for future mathematical application without knowing the values (1) in each separate case, there is no reason to consider the conditional probabilities (54). Instead, it is natural to turn to the *unconditional probability*

$$P[\theta'(x_1; x_2; \dots; x_n) \le \theta \le \theta''(x_1; x_2; \dots; x_n)] \qquad (56)$$

that no error will occur when applying the rule.

Given the type of the functions $\theta'(x_1; x_2; \dots; x_n)$ and $\theta''(x_1; x_2; \dots; x_n)$, the unconditional probability (56) is generally determined by the distribution of the magnitudes (1) which depends on the parameters $\theta, \theta_1, \theta_2, \dots, \theta_s$ and by the unconditional (prior) distribution of these parameters. Denoting the conditional probability of obeying the inequalities $\theta' \le \theta \le \theta''$ when the values of the parameters are given by

$$P(\theta' \le \theta \le \theta'' | \theta, \theta_1, \theta_2, \ldots, \theta_s) \qquad (57)$$

and assuming that the prior distribution of the parameters has density $\varphi(\theta, \theta_1, \theta_2, \ldots, \theta_s)$, we will obtain the following expression for the unconditional probability (56):

$$P(\theta' \le \theta \le \theta'') = \int \int \ldots \int P(\theta' \le \theta \le \theta'' | \theta, \theta_1, \ldots, \theta_s) \varphi(\theta, \theta_1, \ldots, \theta_s) d\theta d\theta_1 \ldots d\theta_s. \qquad (58)$$

A particular case of such rules, when the conditional probability (57) remains constant at all possible values of $\theta, \theta_1, \ldots, \theta_s$, is especially important for practice. If this conditional probability is constant and equals ω, then, on the strength of (59),

$$P(\theta' \le \theta \le \theta'') = \int \int \ldots \int \omega\varphi(\theta, \theta_1, \ldots, \theta_s) d\,\theta d\,\theta_1 \ldots d\,\theta_s = \omega.$$

This means that *the unconditional probability* (56) *does not depend on the unconditional distribution of the parameters* [10].

We have already indicated in §1 that the very hypothesis on the existence of a prior distribution of the parameters is not always sensible. However, if the conditional probability (57) does not depend on the values of the parameters and is invariably equal to one and the same number ω then it is natural to consider that the unconditional probability (56) exists and is equal to ω even in those cases in which the hypothesis on the existence of a prior distribution of the parameters is not admitted [11].

If the conditional probability (57) is equal to ω for all the possible values of the parameters (so that, consequently, the same is true with respect to the unconditional probability (56) for any form of the prior distribution of the parameters), we shall say, following Fisher [1; 2], that our rule has a certain *confidence probability* equal to ω. It is easy to see that for Problem 1 the rule that recommends to assume that a is situated within the boundaries

$$a' \le a \le a'' \qquad (59)$$

where

$$a' = \bar{x} + c'/h\sqrt{n}, \; a'' = \bar{x} + c''/h\sqrt{n} \qquad (60)$$

has a certain confidence probability

$$\omega = (1/\sqrt{\pi}) \int_{c'}^{c''} \exp(-\alpha^2)\, d\alpha. \qquad (61)$$

Indeed, for any a and h,

$$P(a' \le a \le a'' | a; h) = P(\bar{x} + c'/h\sqrt{n} \le a \le \bar{x} + c''/h\sqrt{n} | a; h) =$$

$$P(a - c''/h\sqrt{n} \le \bar{x} \le a - c'/h\sqrt{n} | a; h) = (1/\sqrt{\pi}) \int_{-c''}^{-c'} \exp(-\alpha^2)\, d\alpha =$$

$$(1/\sqrt{\pi}) \int_{c'}^{c''} \exp(-\alpha^2)\, d\alpha.$$

For example, if $c' = -2$ and $c'' = 2$,

$$\omega = (2/\sqrt{\pi}) \int_0^2 \exp(-\alpha^2)\, d\alpha = 0.9953.$$

Thus, the rule that recommends to assume that

$$|a - \bar{x}| \le 2/h\sqrt{n} \qquad (62)$$

has confidence probability $\omega = 0.9953$. In order to ascertain definitively the meaning and the practical importance of the notion of confidence probability, let us dwell on this example. Suppose that we want to apply the rule (62) in some sequence of cases $E_1, E_2, \ldots, E_n$. The values of a_k, h_k, and n_k correspond to each of the cases E_k. However, absolutely independently of these values, the probability of the inequality

$$|\bar{x}_k - a_k| \le 2/h_k\sqrt{n_k} \qquad (62_k)$$

in this case is $\omega = 0.9953$. If the systems of x_i's which correspond here to the different E_k's are independent one from another, then the events A_k consisting in that the appropriate inequalities (62_k) are valid, are also independent. Owing to the Bernoulli theorem, given this condition and a sufficiently large N, the frequency M/N of these inequalities being obeyed in the sequence of cases E_k will be arbitrarily close to $\omega = 0.9953$.Consequently, in any sufficiently long series of independent cases E_k the rule (62) will lead to correct results in about 99.5% of all cases, and to wrong results in approximately 0.5%. For justifying this conclusion it is only necessary that the set of the considered cases $E_1, E_2, \ldots, E_N$ be determined beforehand independently of the values of the x_i's obtained by observation. Bernstein indicated a clever example of a misunderstanding that is here possible if no attention is paid to this circumstance [12].

After all this, a warning against a wide-spread mistake [13] seems almost superfluous. Namely, the equality for the unconditional probability

$$P(|a - \bar{x}| \le 2/h\sqrt{n}) = 0.9953$$

follows if

$$P(|a - \bar{x}| \le 2/h\sqrt{n} | a;\, h) = 0.9953 \qquad (63)$$

for all possible values of a and h. However, it does not at all follow from (63) that for any fixed values of (1)

$$P(|a - \bar{x}| \le 2/h\sqrt{n} | x_1;\, x_2;\, \ldots;\, x_n) = 0.9953.$$

In concluding this section, I note that it is sometimes necessary to consider the rules for establishing confidence limits for an estimated parameter θ which do not possess any definite confidence probability. In such cases, the part similar to that of confidence probability is played by the lower bound

$$\omega = \inf P(\theta' \le \theta \le \theta'' | \theta, \theta_1, \theta_2, \ldots, \theta_n)$$

of the conditional probability for the validity of the inequalities $\theta' \le \theta \le \theta''$ at various combinations of the values of the parameters $\theta, \theta_1, \theta_2, \ldots, \theta_n$. Following Neyman, this lower bound has been called the *coefficient of confidence* of the given rule [14].

6. A Sensible Choice of Confidence Limits Corresponding to a Given Confidence Probability. After what was said in §5, the following formulation of the problem of estimating a parameter θ given (1) becomes understandable. For each ω (0 < ω < 1) it is required to determine, as functions θ'_ω and θ''_ω of (1), and, if necessary, of parameters which are assumed to be known in the given problem, such confidence limits for θ that the rule recommending to assume that $\theta'_\omega \le \theta \le \theta''_\omega$ has confidence probability equal to ω.

The problem thus expressed is not always solvable. When its solution is impossible, we have to turn to rules of estimating the parameter θ lacking a certain confidence probability and to apply the concept of coefficient of confidence indicated at the end of §5. On the other hand, in many cases the formulated problem admits, for each ω, not one, but many solutions. From among these, it is natural to prefer such that lead to shorter confidence intervals [θ'_ω; θ''_ω]. I intend to devote another paper to considering, in a general outline, the problem of discovering such *most effective* rules possessing a given confidence probability (or a given coefficient of confidence).

For Problems 1 – 3 the following simplifications in formulating the issue about discovering which sensible confidence limits for *a* and *h* are natural.

1. It is natural to restrict our attention to considering confidence limits depending, when *n* and ω are given, in addition to the parameters supposed to be known, only on the corresponding sufficient statistics [15] or sufficient systems of statistics. We will therefore assume that, in Problem 1, the confidence limits a' and a'' only depend on *h* and $\bar{x}$; in Problem 2, the confidence limits h' and h'' only depend on *a* and *S*; and, in problem 3, a' and a'', h' and h'' only depend on $\bar{x}$ and S_1.

2. It is natural to wish [16] that the rules for determining confidence limits be invariant with respect to change of scale; of the origin; and of the choice of the positive direction along the Ox axis, *i.e.*, with respect to transformations

$$x^* = kx + b \tag{64}$$

where *b* is an arbitrary real number and *k* is an arbitrary real number differing from zero. Under this transformation, *a*, *h*, $\bar{x}$, *S*, and S_1 are replaced by

$$a^* = ka + b,\ h^* = h/|k|,\ \bar{x}^* = k\bar{x} + b,\ S^* = |k|S,\ S_1^* = |k|S_1.$$

This demand of invariance is reduced to the fulfilment of the following relations, given fixed *n* and ω, for any real $k \ne 0$ and *b* and a', a'', h' and h'' being functions of the arguments indicated above in Item 1:

Problem 1: $a'(h^*, \bar{x}^*) = k\,a'(h; \bar{x}) + b$, $a''(h^*, \bar{x}^*) = ka''(h; \bar{x}) + b$.
Problem 2: $h^*(a^*; S^*) = h'(a; S)/|k|$, $h''(a^*; S^*) = h''(a; S)/|k|$.
Problem 3: $a'(\bar{x}^*; S_1^*) = ka'(\bar{x}; S_1) + b$, $a''(\bar{x}^*; S_1^*) = ka''(\bar{x}; S_1) + b$,
$h'(\bar{x}^*; S_1) = h'(\bar{x}; S_1)/|k|$, $h''(\bar{x}^*; S_1^*) = h''(\bar{x}; S_1)/|k|$.

Issuing from Demands 1 and 2, we may conclude that the confidence limits should have the form

$$a' = \bar{x} - A_0/h,\ a'' = \bar{x} + A_0/h;\ h' = B'/S,\ h'' = B''/S, \tag{65, 66}$$
$$a_1' = \bar{x} - C_0S_1,\ a_1'' = \bar{x} + C_0S_1;\ h_1' = B_1'/S_1;\ h_1'' = B_1''/S_1 \tag{67, 68}$$

for Problems 1, 2 and 3 respectively. Here, for a fixed *n*, A_0, B', B'', C_0, B_1', and B_1'' only depend on ω. If

$$A = h(a - \bar{x}),\ B = hS,\ C = (a - \bar{x})/S_1,\ B_1 = hS_1 \qquad (69-72)$$

then, as it is easily seen, the inequalities

$$a' \le a \le a'',\ h' \le h \le h'',\ a_1' \le a \le a_1'',\ h_1' \le h \le h''$$

are equivalent to inequalities

$$-A_0 \le A \le A_0,\ B' \le B \le B'',\ -C_0 \le C \le C_0,\ B_1' \le B \le B_1''$$

respectively.

The factor, decisive for the success of all the following, is that *the laws of distribution of* A, B, C, and B_1 *calculated for fixed a and h by issuing from formula* (1), *are independent of these parameters.* The densities of these laws are

$$f_1(A) = \sqrt{n/\pi}\,\exp(-nA^2),\ f_2(B) = \frac{B^{n-1}}{\Gamma(n/2)}\exp(-B^2),$$
$$f_3(C) = \frac{\Gamma(n/2)}{\Gamma[(n-1)/2]}\sqrt{n/\pi}\,(1+nC^2)^{-n/2}, \qquad (73-76)$$
$$f_4(B_1) = \frac{B_1^{n/2}}{\Gamma[(n-1)/2]}\exp(-B_1^2).$$

Therefore, the probabilities

$$P(a' \le a \le a''|a;h) = P(-A_0 \le A \le A_0|a;h),$$
$$P(h' \le h \le h''|a;h) = P(B' \le B \le B''|a;h),$$
$$P(a_1' \le a \le a_1''|a;h) = P(-C_0 \le C \le C_0|a;h),$$
$$P(h_1' \le h \le h_1''|a;h) = P(B_1' \le B \le B_1''|a;h)$$

are independent of *a* and *h* so that they may be considered as confidence probabilities in the sense of the definition of §5. Calculating these probabilities in accord with formulas (73) – (76) and equating them to ω, we will have

$$\omega = 2\sqrt{n/\pi}\int_0^{A_0}\exp(-nA^2)dA = (2/\sqrt{\pi})\int_0^{\alpha_0}\exp(-\alpha^2)d\alpha, \qquad (77)$$

$$\omega = \frac{1}{\Gamma(n/2)}\int_{B'}^{B''} B^{n-1}\exp(-B^2)dB = \frac{1}{2^{(n-2)/2}\Gamma(n/2)}\int_{\chi'}^{\chi''}\chi^{n-1}\exp(-\chi^2/2)d\chi, \qquad (78)$$

$$\omega = 2\sqrt{n/\pi}\,\frac{\Gamma(n/2)}{\Gamma[(n-1)/2]}\int_0^{C_0}(1+nC^2)^{-n/2}dC =$$
$$\frac{2\Gamma(n/2)}{\sqrt{\pi(n-1)}\,\Gamma[(n-1)/2]}\int_0^{\gamma_0}\{1+[\gamma^2/(n-1)]\}^{-n/2}d\gamma, \qquad (79)$$

$$\omega = \frac{1}{\Gamma[(n-1)/2]}\int_{B_1'}^{B_1''} B_1^{n-2}\exp(-B_1^2)dB_1 =$$
$$\frac{1}{2^{(n-3)/2}\Gamma[(n-1)/2]}\int_{\chi_1'}^{\chi_1''}\chi^{n-2}\exp(-\chi^2/2)d\chi. \qquad (80)$$

For the sake of ensuring the possibility of directly using the existing tables we have introduced here

$\alpha_0 = A_0\sqrt{n}$, $\chi' = B'\sqrt{2}$, $\chi'' = B''\sqrt{2}$, $\gamma_0 = C_0\sqrt{n(n-1)}$, $\chi'_1 = B'_1\sqrt{2}$, $\chi''_1 = B''_1\sqrt{2}$.

When joining formulas (65) – (68) to these relations we obtain

$$a' = \bar{x} - \alpha_0/h\sqrt{n},\ a'' = \bar{x} + \alpha_0/h\sqrt{n},\ h' = \chi'/S\sqrt{2},\ h'' = \chi''/S\sqrt{2},$$
$$a_1' = \bar{x} - \gamma_0 S_1/\sqrt{n(n-1)},\ a_1'' = \bar{x} + \gamma_0 S_1/\sqrt{n(n-1)}, \quad (81-84)$$
$$h_1' = \chi_1'/S_1\sqrt{2},\ h_1'' = \chi_1''/S_1\sqrt{2}.$$

In this notation the confidence limits of a and h are determined by the following inequalities:

$$|a - \bar{x}| \le \alpha_0/h\sqrt{n},\ \chi'/S\sqrt{2} \le h \le \chi''/S\sqrt{2},$$
$$|a - \bar{x}| \le \gamma_0 S_1/\sqrt{n(n-1)},\ \chi_1'/S_1\sqrt{2} \le h \le \chi_1''/S_1\sqrt{2} \quad (85-88)$$

for Problems 1 – 3 respectively. Here, α_0, χ', χ'',γ_0, χ_1', χ_1'' should be chosen in a way satisfying relations (77 – 80); α_0 and γ_0 are thus uniquely determined given the confidence probability, whereas the choice of the other four magnitudes remains to some extent arbitrary, see §8.

7. Practical Conclusions about Estimating the Center of Scattering. The rules below follow from the deliberations of §6.

1. *In Problem* 1 *we may assume, with confidence probability* ω, *that*

$$|a - \bar{x}| \le \alpha_0/h\sqrt{n}$$

where α_0 *is determined from*

$$\omega = (2/\sqrt{\pi})\int_0^{\alpha_0} \exp(-\alpha^2)d\alpha.$$

2. *In Problem* 3 *we may assume, with confidence probability* ω, *that*

$$|a - \bar{x}| \le \gamma_0 S_1/\sqrt{n(n-1)}$$

where γ_0 *is determined from* (79). I adduce a table showing the dependence of γ_0 on ω for various values of n; it is extracted from a more complete table included, for example, in [1]. The last line of the table shows the limiting values of γ_0 as $n \to \infty$ calculated by the formula

$$\omega = \sqrt{2/\pi}\int_0^{\gamma_0} \exp(-\gamma^2/2)d\gamma. \quad (89)$$

Note that formula (50) derived in §4 is equivalent to

$$P(|a - \bar{x}| \le \gamma_0 S_1/\sqrt{n(n-1)}\,|x_1; x_2; \ldots; x_n) =$$
$$\sqrt{2/\pi}\int_0^{\gamma} \exp(-\gamma^2/2)d\gamma + O(1/\sqrt{n}). \quad (90)$$

We can now estimate how dangerous it is to neglect the remainder term $O(1/\sqrt{n})$ when having small values of n. If neglecting it, we would have concluded, for example, as it is often done in textbooks, that we may expect the validity of the inequalities

$$|a - \bar{x}| \le 2.576 S_1/\sqrt{n(n-1)} \qquad (91)$$

with probability 0.99. Actually, however, if, for example, $n = 5$, this inequality will be violated in about 6% of all cases; for ensuring {not more than} 1% of violations, we ought to apply, as shown in our table, equality (91) with factor 4.604 instead of 2.576. Only when $n > 30$ is it quite admissible, as considered in practical applications, to make use of the limiting formula (89).

When applying the rules formulated in this section, it is naturally useful to remember the remarks made at the end of §5 about the meaning of the concept of confidence probability.

8. The Choice of Confidence Limits for the Measure of Precision. Conditions (78) and (80) still leave some arbitrariness in the choice of χ', χ'', χ_1', χ_1''. Given ω, it is natural to choose these magnitudes so that, allowing for the abovementioned conditions, the intervals $[\chi'; \chi'']$ and $[\chi_1'; \chi_1'']$ become as short as possible. Under this additional restriction χ', χ'', χ_1', and χ_1'' are uniquely determined by ω and n.

However, since the appropriate tables are lacking, the actual calculation of the four magnitudes as determined by this condition is rather difficult. Therefore, instead of deriving the shortest confidence intervals $[\chi'; \chi'']$ and $[\chi_1'; \chi_1'']$ corresponding to the given confidence probability ω, practitioners achieve the validity of relations (78) and (80) by calculating χ', χ'', χ_1', and χ_1'' from the equalities

$$\frac{1}{2^{(n-2)/2}\Gamma(n/2)} \int_0^{\chi'} \chi^{n-1}\exp(-\chi^2/2)d\chi = (1-\omega)/2, \qquad (92a)$$

$$\frac{1}{2^{(n-2)/2}\Gamma(n/2)} \int_{\chi''}^{\infty} \chi^{n-1}\exp(\chi^2/2)d\chi = (1-\omega)/2, \qquad (92b)$$

$$\frac{1}{2^{(n-3)/2}\Gamma[(n-1)/2]} \int_0^{\chi_1'} \chi^{n-2}\exp(\chi^2/2)d\chi = (1-\omega)/2, \qquad (93a)$$

$$\frac{1}{2^{(n-3)/2}\Gamma(n/2)} \int_{\chi_1''}^{\infty} \chi^{n-2}\exp(\chi^2/2)d\chi = (1-\omega)/2 \qquad (93b)$$

and applying the available tables (see, for example, [7]) showing the dependence between χ^2 and

$$P_k(\chi) = \frac{1}{2^{(k-2)/2}\Gamma(k/2)} \int_{\chi}^{\infty} \chi^{k-1}\exp(\chi^2/2)d\chi.$$

For $n > 30$ it is possible to use the limiting formulas

$$\chi' = \sqrt{n} - c,\ \chi'' = \sqrt{n} + c,\ \chi_1' = \sqrt{n-1} - c,\ \chi_1'' = \sqrt{n-1} + c \qquad (94-95)$$

where c is determined from the condition

$$\omega = (2/\sqrt{\pi}) \int_0^c \exp(-z^2)dz.$$

Note also that (95),(96), (83),(85), (19) and (20) lead to

$$h' = \bar{h}\,[1 - (c/\sqrt{n})],\ h'' = \bar{h}\,[1 + (c/\sqrt{n})], \quad (96)$$
$$h_1' = \bar{h}_1\,[1 - (c/\sqrt{n-1})],\ h_1'' = \bar{h}_1\,[1 + (c/\sqrt{n-1})]. \quad (97)$$

When comparing these formulas with (46) and (53) we see that here also for large values of n the confidence limits obtained in accordance with the Fisherian method essentially coincide with those determined on the basis of the limit theorems of §4.

9. A Sensible Choice of Approximate Values of the Estimated Parameters. Instead of the confidence limits for the estimated parameter θ corresponding to a given confidence probability ω, it is often desirable to have one approximate value $\bar{\theta}$ of this parameter. The problem of the most sensible choice of such an approximate value corresponding to the given observations (1) can be formulated in many different ways.

From the viewpoint of the classical method (§§1 and 4), the most natural way is to assume as the approximate value the conditional expectation

$$\bar{\theta} = \mathrm{E}(\theta|x_1; x_2; \ldots; x_n). \quad (98)$$

Indeed, it is easy to show that this choice leads to the minimal value of the conditional expectation $\mathrm{E}[(\theta - \bar{\theta})^2|x_1; x_2; \ldots; x_n]$ of the square of the deviation $(\theta - \bar{\theta})$, *i.e.*, that it provides the maximal possible value of the measure of precision $h(\bar{\theta}|x_1; x_2; \ldots; x_n)$ as determined by formula (36).

According to this viewpoint, and under some natural assumptions about the prior distributions of a and h for large values of n (or of nh^2 in Problem 1), we may consider $\bar{x}$ as the approximate value of a; in Problem 2, $\bar{h}$ as the approximate value of h; and, in Problem 3, $\bar{h}_1$ {rather than $\bar{h}$ }(§4). However, keeping to the approach of §4, any other magnitudes, $\bar{x}^*$, $\bar{h}^*$ and $\bar{h}_1^*$ obeying the following demands may be assumed as the approximate values of a and h instead of $\bar{x}$, $\bar{h}$ and $\bar{h}_1$:

$$\bar{x}^* = \bar{x} + O(1/nh^2),\ \bar{h}^* = \bar{h}\,[1 + O(1/n)],$$
$$\bar{x}^* = \bar{x} + (1/\bar{h}_1)\,O(1/n),\ \bar{h}_1^* = \bar{h}_1\,[1 + O(1/n)]$$

in Problems 1 – 3 respectively.

This indefiniteness lies at the heart of the matter when only qualitative assumptions of sufficient *smoothness* are made with regard to the prior distributions of a and h. If the main problem when estimating a parameter θ is an indication for each ω $(0 < \omega < 1)$ of confidence limits θ'_ω and θ''_ω corresponding to the confidence probability (or to the coefficient of confidence) ω, then these limits are usually obtained in such a way that, for $\omega' > \omega$, $\theta'_\omega \le \theta'_{\omega'} \le \theta''_{\omega'} \le \theta''_\omega$. Under this condition, it usually occurs that, as $\omega \to 0$, the lower and the upper bounds θ'_ω and θ''_ω tend (from below and from above, respectively) to a common limit $\bar{\theta}$. In this case it is natural to assume $\bar{\theta}$ as the approximate value of θ; indeed, only such a rule can ensure the inequalities $\theta'_\omega \le \bar{\theta} \le \theta''_\omega$ for any ω. From this viewpoint, choosing the confidence limits for a and h as it was done in §§7 and 8, we should assume $\bar{x}$ as the approximate value of a; and, in Problems 2 and 3, the approximate value of h will be $\bar{h}^*$ determined by equalities

$$\bar{h}* = S\bar{\chi}*\sqrt{2},\quad \frac{1}{2^{(n-2)/2}\Gamma(n/2)}\int_0^{\bar{\chi}*}\chi^{n-1}\exp(-\chi^2/2)d\chi = 1/2, \tag{99}$$

and $\bar{h}_1*$ determined by equalities

$$\bar{h}_1* = S_1\bar{\chi}_1*\sqrt{2},\quad \frac{1}{2^{(n-3)/2}\Gamma(n/2)}\int_0^{\bar{\chi}_1*}\chi^{n-2}\exp(-\chi^2/2)d\chi = 1/2 \tag{100}$$

respectively. I adduce the values of $(\bar{\chi}*)^2$ for $n \le 20$. {The table, actually for $n = 1(1)10$, is omitted here.} For $n > 10$ we may consider that, with an error less than 0.01,

$$(\bar{\chi}*)^2 \sim n - 2/3. \tag{101}$$

For n observations $\bar{\chi}_1*$ is equal to $\bar{\chi}*$ corresponding to $(n - 1)$ observations [17]. If the choice of the approximate values of a and h is considered in itself, then it is nevertheless natural to restrict it to the values satisfying the following two conditions [18].

1. In each problem, the approximate values depend, in addition to the parameters supposed to be known, only on the appropriate sufficient statistics, or sufficient systems of statistics.

2. The approximate values are invariant with respect to the transformations of the Ox axis of the type (64).

It is possible to conclude from these demands that *only* $\bar{x}$ *may be taken as an approximate value of a; and, as an approximate value of h, we ought to assume, in Problem* 2, $\bar{h} = \bar{B}/S$; *and in Problem* 3, $\bar{h}_1 = \bar{B}_1/S$ *where* $\bar{B}$ *and* $\bar{B}_1$ *only depend on n.*

We may apply various additional conditions for determining the most sensible values of the factors $\bar{B}$ and $\bar{B}_1$. For example, it is possible to demand that, for any values of a and h, the conditions $E(\bar{h}|a; h) = h$, $E(\bar{h}_1|a; h) = h$ be satisfied. These demands can only be met if

$$\bar{B} = \bar{B}_1,\quad \bar{B} = \frac{\Gamma[(n+2)/2]}{\Gamma[(n+1)/2]},\quad \bar{B}_1 = \frac{\Gamma[(n+1)/2]}{\Gamma(n/2)},$$

i.e., if, assuming these values for $\bar{B}$ and $\bar{B}_1$, we will have

$$\bar{h} = \bar{B}/S \text{ and } \bar{h}_1 = \bar{B}_1/S. \tag{102}$$

We shall soon see that the finality of this last result should not be overestimated.

The demand made use of above that *the systematic error be absent*, can be formulated with respect to the approximate value $\bar{\theta}$ of any parameter θ under consideration. In a general form, this demand is expressed thus: The equality

$$E(\bar{\theta}|\theta; \theta_1; \theta_2; \ldots; \theta_n) = \theta \tag{103}$$

should hold for all possible values of the parameters $\theta, \theta_1, \theta_2, \ldots, \theta_s$ of a given problem. The approximation $\bar{x}$ for the center of scattering a satisfies this demand both in Problem 1 and Problem 3. Now we will determine approximations devoid of systematic error for the mean square deviation (9). It is natural to restrict our attention here to approximations of the form $\bar{\sigma} = kS$ (Problem 2) and $\bar{\sigma}_1 = k_1S_1$ (Problem 3) [19]. The demand that the systematic error be absent, $E(\bar{\sigma}|a; h) = \sigma$, $E(\bar{\sigma}_1|a; h) = \sigma$, leads then to the necessity of assuming

$$k = (1/\sqrt{2})\frac{\Gamma(n/2)}{\Gamma[(n+1)/2]},\ k_1 = (1/\sqrt{2})\frac{\Gamma[(n-1)/2]}{\Gamma(n/2)}, \qquad (104)$$

i.e., of choosing $\overline{\sigma}$ and $\overline{\sigma}_1$ (see above) in accord with these magnitudes. However, after that it is natural to take as an approximation to h

$$\overline{h} = (1/S)\frac{\Gamma[(n+1)/2]}{\Gamma(n/2)},\ \overline{h}_1 = (1/S_1)\frac{\Gamma(n/2)}{\Gamma[(n-1)/2]} \qquad (102\text{bis})$$

in Problems 2 and 3 respectively.

If the lack of systematic error is demanded in the determination of σ^2 (of the variance), and if the approximation to σ^2 is being determined in the form $\overline{\sigma}^2 = qS^2$ (Problem 2) and $\overline{\sigma}_1^{\ 2} = q_1S_1^{\ 2}$ (Problem 3), then we have to assume

$$q = 1 - n,\ q_1 = 1/(n-1),\ \overline{\sigma}^2 = S^2/n,\ \overline{\sigma}_1^{\ 2} = S_1^{\ 2}/(n-1). \qquad (105)$$

To achieve conformity with formula (103), it is then natural to assume the approximation to h as

$$\overline{h} = \sqrt{n/2S},\ \overline{h}_1 = \sqrt{(n-1)/2S_1} \qquad (98\text{ter})$$

in Problems 2 and 3 respectively. The last approximations are the most generally accepted in modern mathematical statistics. We have applied them in §4, where, however, the choice between (98), (98bis) and (98ter) was not essential since the limit theorems of §4 persisted anyway.

From the practitioner's viewpoint, the differences between the three formulas are not essential when only once determining the measure of precision h by means of (1). Indeed, the differences between these approximations have order h/n, and the order of their deviations from the true value [20], h, is higher, $h/\sqrt{n}$. Therefore, since we are only able to determine h to within deviations of order $h/\sqrt{n}$, we may almost equally well apply any of these approximations which differ one from another to within magnitudes of the order h/n.

The matter is quite different if a large number of various measures of precision (for example, corresponding to various conditions of gunfire) has to be determined, each time only by a small number of observations. In this case, the absence of a systematic error in some magnitude, calculated by issuing from the approximate value of the measure of precision, can become very essential. Depending on whether this magnitude is, for example, h, σ or σ^2, the approximate values of h should be determined by formulas (98), (98bis) or (98ter) respectively.

In particular, from the point of view of an artillery man, according to the opinion of Prof. Gelvikh [10; 11] [21], it is most essential to determine without systematic error the expected expenditure of shells required for hitting a target. In the most typical cases (two-dimensional scattering and a small target as compared with the scattering) this expected expenditure, according to him, is proportional to the product $\sigma^{(1)}\sigma^{(2)}$ of the mean square deviations in the two directions. Suppose that we estimate $\sigma^{(1)}$ and $\sigma^{(2)}$ by their approximations $\overline{\sigma}^{(1)}$ and $\overline{\sigma}^{(2)}$ derived from observations $(x_1^{(1)}, x_2^{(1)}, \ldots, x_n^{(1)})$ and $(x_1^{(2)}, x_2^{(2)}, \ldots, x_m^{(2)})$ respectively. If the $x_i^{(1)}$ are independent of $x_j^{(2)}$, then, for any $\sigma^{(1)}$, $\sigma^{(2)}$, $a^{(1)}$ and $a^{(2)}$ (where the last two magnitudes are the centers of scattering for $x_i^{(1)}$ and $x_j^{(2)}$ respectively), we have

$$\mathrm{E}(\overline{\sigma}^{(1)}\overline{\sigma}^{(2)}|a^{(1)}; a^{(2)}; \sigma^{(1)}; \sigma^{(2)}) = \mathrm{E}(\overline{\sigma}^{(1)}|a^{(1)}; \sigma^{(1)})\ \mathrm{E}(\overline{\sigma}^{(2)}|a^{(2)}; \sigma^{(2)})$$

for the product $\bar{\sigma}^{(1)}\bar{\sigma}^{(2)}$ of the approximations $\bar{\sigma}^{(1)}$ and $\bar{\sigma}^{(2)}$. Therefore, to obtain, identically for all possible values of $a^{(1)}$, $a^{(2)}$, $\sigma^{(1)}$ and $\sigma^{(2)}$,

$$E(\bar{\sigma}^{(1)}\bar{\sigma}^{(2)}|a^{(1)}; a^{(2)}; \sigma^{(1)}; \sigma^{(2)}) = \sigma^{(1)}\sigma^{(2)}$$

it is sufficient to choose $\bar{\sigma}^{(1)}$ and $\bar{\sigma}^{(2)}$ satisfying the conditions

$$E(\bar{\sigma}^{(1)}|a^{(1)}; \sigma^{(1)}) = \sigma^{(1)},\ E(\bar{\sigma}^{(2)}|a^{(2)}; \sigma^{(2)}) = \sigma^{(2)}.$$

Thus, to obtain without a systematic error the estimated expected expenditure of shells under the conditions specified by Gelvikh, we ought to make use of estimates (104) leading to $\sigma^{(1)}$ and $\sigma^{(2)}$ devoid of systematic error. Accordingly, issuing from his demands, the estimate (98bis) is naturally preferable for h [22]. As a rule, in accord with Gelvikh's point of view, the final choice of the most expedient form of the approximate values of h for a small number of observations is determined not by some general demands of probability theory, but by additional conditions which may differ in various practical problems.

Notes

1. This means that the probability of being chosen is one and the same for each soldier.
2. The distribution depends on many factors. If, for example, the regiment consists of soldiers belonging to two different drafts, it can possess two peaks.
3. Fisher introduced this notion otherwise, see [1 – 3].
4. A statistic can depend on some parameters whose values are assumed to be known (for example, on h in Problem 1 or on a in Problem 2). It is only essential that it does not depend on $\theta_1, \theta_2, \ldots, \theta_s$ which are here supposed unknown.
5. Note that in this problem the magnitude $\bar{x}$ taken alone is not anymore a sufficient statistic for a. Indeed, the conditional density of a, given (1), is expressed by the formula

$$\varphi(a|x_1, x_2, \ldots, x_n) = \frac{\int_0^\infty h^n \exp[-h^2 S_1^{\,2} - nh^2(a-\bar{x})^2]\varphi_3(a;h)dh}{\int_{-\infty}^{\infty}\int_0^\infty h^n \exp[-h^2 S_1^{\,2} - nh^2(a-\bar{x})^2\varphi_3(a;h)dhda}.$$

It is easy to find out that the left side is not uniquely determined by $\varphi_3(a; h)$ and $\bar{x}$ but in addition essentially depends on S_1. It can be shown that, in general, there does not exist any continuous function $\chi(x_1, x_2, \ldots, x_n)$ that, under the conditions of Problem 3, would be a sufficient statistic for a. The same is true here with respect to h in Problem 3.

For the sake of simplicity of the calculations, we have assumed that the prior distributions of a and h are continuous and expressed by densities. However, all the conclusions about sufficient statistics in Problems 1 and 2 and about the sufficient system of statistics in Problem 3 persist even without this restriction.

6. Elsewhere, I intend to offer a precise definition of the term *information* conforming to the Fisherian use of the word.
7. All this is valid under normality of the x_i's which we assumed from the very beginning. The situation will change once we abandon this assumption.

To explain this fact, let us consider the following example. Suppose that the random variables (1) are independent and have a common uniform distribution on interval $[a - 1/2; a + 1/2]$. Then their n-dimensional density will be

$f(x_1, x_2, \ldots, x_n) = 1$ if $x_{max} - 1/2 < a < x_{min} - 1/2$ and $= 0$ otherwise.

Denote the prior distribution of a by $\varphi(a)$, then, by the Bayes theorem, the conditional density of a, given (1), will be

$$\varphi(a|x_1, x_2, \ldots, x_n) = \int_{x_{max}-1/2}^{x_{min}+1/2} \varphi(a)da \text{ if } x_{max} - 1/2 < a < x_{min} - 1/2 \text{ and}$$
$$= 0 \text{ otherwise.}$$

The system of two statistics, x_{max} and x_{min}, will obviously be a sufficient system for a. And it would be quite natural to assume (10) as the appropriate value of a. It would be possible to show that for a large n the difference $(a - d)$ will be here, as a rule, considerably less than the difference $(a - \bar{x})$.

8. A more precise estimate

$$E[(a - \bar{x})^2|x_1, x_2, \ldots, x_n] = (1/2nh^2)[1 + O(1/nh^2)] \qquad (37')$$

can be obtained from (34). We cannot lower the order of the remainder term here by further strengthening the demands on the *smoothness* of the function $\varphi_1(a)$.

9. A more precise estimate

$$P(|a - \bar{x}| \le c/h\sqrt{n}|x_1, x_2, \ldots, x_n) = (2/\sqrt{\pi})\int_0^c \exp(-\alpha^2)d\alpha + O(1/nh^2) \qquad (40')$$

can be obtained from (34). We cannot lower the order of the remainder term here by further strengthening the demands on the *smoothness* of the function $\varphi_1(a)$.

10. We only obtained this result for unconditional distributions of the parameters given by their density $\varphi(\theta; \theta_1, \theta_2, \ldots, \theta_s)$. It is easy to see, however, that it persists in the general case as well.

11. In this case the matter consists in assuming the following new axiom of the theory of probability: If the conditional probability $P(A|\theta_1, \theta_2, \ldots, \theta_s)$ of some event A exists for all the possible values of the parameters $\theta_1, \theta_2, \ldots, \theta_s$ and is equal to one and the same number ω, then the unconditional probability $P(A)$ of event A exists and is equal to ω.

Note that when comparing the discussed method with the classical approach the issue about the acceptability of this new axiom does not arise since the classical method is necessarily based on admitting the existence of a prior distribution of the parameters so that the new axiom becomes superfluous.

12. See [5, §7]. A curious problem arises in connection with this example: To formulate such a rule for selecting boxes which will guarantee the buyer, with a sufficiently low risk of error, the purchase of not less than 95% of the boxes satisfying his demand that $|a_i - a| < 2$. This problem admits of a quite proper interpretation from the viewpoint of confidence probabilities (or, more precisely, of coefficients of confidence, see the end of this section). I intend to return elsewhere to this problem and to some similar problems having a serious practical importance.

13. Commited by Fisher in some of his writings.

14. See an example [6] of an elementary problem in which the application of rules having no definite confidence probability and only possessing a certain coefficient of confidence is unavoidable.

15. {Here, for the first time, the author translated the English term *sufficient* by the appropriate Russian equivalent. In §2 he used a Russian term tantamount to the English *exhaustive* which had not stood the test of time.}

16. I borrow this requirement from Borovitsky [8; 9]. To avoid misunderstanding, I consider it necessary to add that I believe that the main substance of his work is erroneous.

17. This choice of the approximate values of $\bar{h}$ * and $\bar{h}_1$ * for *h* is connected with the method of determining the confidence limits for *h* as indicated in §8. It is not the only possible one, and neither is it even the best one as considered from any sufficiently justified viewpoint.

18. Cf. the conditions imposed on *a′*, *a″*, *h′* and *h″* in §6.

19. Such a form of the approximations for σ follows from the conditions 1 and 2 above.

20. {Note the author's use of this term of the classical error theory.}

21. {Gelvikh later served time in a labor camp as a *saboteur* and his books had been banned but I am unable to supply a reference.}

22. This conclusion is made in R.E. Sorkin's unpublished work which he kindly gave me. The same result naturally persists for the cases of one- and three-dimensional scatter. Gelvikh himself mistakenly believes that the problem as formulated by him leads to bounds (100ter).

References

1. Fisher, R.A. (1925), *Statistical Methods for Research Workers*. London.

2. --- (1922), On the mathematical foundations of theoretical statistics. *Phil. Trans. Roy. Soc*., vol. A222, pp. 309 – 368.

3. Neyman, J., Pearson, E.S. (1936), Sufficient statistics and uniformly most powerful tests of statistical hypotheses. *Stat. Res. Mem. Univ. College Lond*., vol. 1, pp. 113- 137.

4. Kolmogorov, A.N. (1931, in Russian), The method of median in the theory of errors. *Sel. Works*, vol. 2. Dordrecht, 1992.

5. Bernstein, S.N. (1941), On the Fisherian "confidence" probabilities. Translated in this book.

6. Neyman, J. (1935), On the problem of confidence intervals. *Annals Math. Stat*., vol. 6, pp. 111 – 116.

7. Romanovsky, V.I. (1938), *Математическая статистика* (Math. Stat.). M. – L.

8. Brodovitsky, K.V. (1939), On the necessary and sufficient conditions for the prior probabilities to have sense. *Trudy Sredneaziatsk. Gosudarstven. Univ*., ser. 5a (math.), No. 19. (R)

9. --- (1939), Formulating the issue of similarity in the theory of statistical samples. Ibidem, No. 20. (R)

10. Gelvikh, P.A. (1941), *Курс теории вероятностей* (Course in Theory of probability). M.

11. --- (1927), On a wrong derivation in the theory of probability. *Izvestia Voenno-Technich. Akad*., pp. 128 – 140. (R)

12. Sorkin, R.E. (1936), On the most probable values of the measure of precision as obtained by observations. *Izvestia Artill. Akad.*, vol. 20, pp. 81 – 92. (R)

13. A.N. Kolmogorov. The Number of Hits after Several Shots and the General Principles of Estimating the Efficiency of a System of Firing

Trudy [Steklov] *Matematich. Institut Akademii Nauk SSSR*, No. 12, 1945, pp. 7 – 25 …

Introduction

In the sequel, we shall consider a group of *n* shots each of which can either *hit*, or *miss* the target. The number of hits μ will evidently only take values $m = 0, 1, 2, \ldots, n$, and we denote the probability of achieving exactly *m* hits by $P_m = P(\mu = m)$. By means of these probabilities *the expected number of hits* Eμ can be written down as [1]

$$E\mu = P_1 + 2P_2 + \ldots + nP_n \qquad (2)$$

and the probability of achieving *not less than m hits* is

$$R_m = P(\mu \geq m) = P_m + P_{m+1} + \ldots + P_n = 1 - (P_0 + \ldots + P_{m-1}). \quad (3)$$

In particular, the probability of *at least one hit* is $R_1 = 1 - P_0$. The expectation Eμ and the probabilities R_m are the main indicators for estimating the efficiency of a system of firing adopted in the current military literature [2].

Both Eμ and the probabilities R_m are determined by the probabilities P_m in accord with formulas (2) and (3), *i.e.*, by the *distribution* of the random variable μ. Therefore, in principle it would be essential to examine, above all, under what conditions is the knowledge of this distribution (that is, of the totality of probabilities P_m for $m = 0, 1, 2, \ldots, n$) sufficient for estimating the efficiency of a system of firing.

Generally speaking, this knowledge is not at all always sufficient. For example, when shelling a target occupying a large area, it is often essential not only to achieve a sufficiently large number of hits, but to distribute the hit-points over the area of the target in a way guaranteeing the hitting of a sufficiently large portion of this area. Here, however, we leave such cases aside.

In restricting the estimation of the efficiency of the system of firing to considering the distribution of the probabilities of the number of hits, that is, of the probabilities $P_0, P_1, P_2, \ldots, P_n$, it is natural to question whether it is possible to replace this totality by some single magnitude depending on them, $W = f(P_0; P_1; P_2; \ldots, P_n)$, and to declare it *the indicator of the efficiency* of the system of firing.

The expectation Eμ or the probability R_m of achieving not less than m hits, with m being *the number of hits necessary for destroying the target*, is often assumed as such a single indicator of efficiency. Considerations widely disseminated in the literature on the comparative benefits and shortcomings of estimating *by expectation* and *by probability* often do not possess sufficient clearness which compels me to devote §1 to this issue.

In §§2 and 3 I discuss the purely mathematical topics of precise and approximate calculation of the probabilities P_m and R_m and adduce tables of the Poisson distribution with corrections which can be rather widely applied, as I think, when solving various problems of the theory of firing. In §4, issuing from the deliberations explicated in §1, and the formulas derived in §3, I formulate the problem of *firing with an artificial scattering* in a general way, and, in particular, I define the very notion of artificial scattering. I shall develop the ideas of this somewhat abstract section elsewhere. In §5 I consider a particular case of determining the probability of hitting the target, and, by introducing a *reduced* target, exonerate to some extent the wide-spread method of reducing the problem on the probability of destroying the target by several hits to that of the probability of at least one hit.

1. The Choice of the Indicator of the Efficiency of Firing

I begin by considering two typical cases between which exist many intermediate instances. *The first case*. The firing is carried out to achieve a quite definite goal (to sink a ship, to shoot down an aircraft, etc) which can only be *either accomplished or not*; and we are only interested in the probability $P(A)$ of success [3]. Denote the conditional probability of success by $P(A|m)$ if there will be exactly m hits. Then, according to the theorem on the total probability, and assuming that $P(A|0) = 0$,

$$P(A) = P_1P(A|1) + P_2P(A|2) + \ldots + P_nP(A|n). \quad (5)$$

If, in particular, success is certain when $\mu \geq m$ and impossible otherwise, *i.e.*, if

$P(A|r) = 1$ if $r \geq m$ and 0 otherwise, (6)

the general formula (5) becomes $P(A) = R_m$. For example, when success is already certain after achieving at least one hit, then, obviously, $P(A) = R_1$. For $m > 1$ the assumption (6) becomes rather artificial: it is difficult to imagine such a concrete situation when success is guaranteed by *ten* hits, but would have still been absolutely impossible to achieve in *nine* hits. It is more natural to suppose that the probability $P(A|m)$ *gradually* heightens with the number of hits m. In this connection I consider in §5 the case of

$$P(A|m) = 1 - e^{\alpha m} \quad (9)$$

where α is some constant. I have chosen this type of dependence of $P(A|m)$ on m because in many cases it allows to derive sufficiently simple expressions for the probability $P(A)$ which is the main studied magnitude.

At the same time, formula (9) taken with various values of the constant α seems to be not less suitable for approximately depicting the different relations which it is possible to encounter in real life than formula (6). It is quite natural to assume, however, that the conditional probabilities $P(A|m)$ do not decrease with an increasing m; *i.e.*, that

$$D_m = P(A|m) - P(A|m - 1) \geq 0.$$

Consequently, it is convenient to rewrite the general formula (5) as [4]

$$P(A) = D_1P_1 + D_2P_2 + \ldots + D_nP_n. \quad (11)$$

If, for example, success is impossible when achieving *less than three hits*; if it has probability 1/3 after *three* hits, 2/3 after *four* hits; and is certain when achieving *more than four hits*, then, for $n = 10$, formula (5) provides

$$P(A) = (1/3)P_3 + (2/3)P_4 + P_5 + P_6 + \ldots + P_{10}$$

whereas formula (11) furnishes a simpler expression

$$P(A) = (1/3)R_3 + (1/3)R_4 + (1/3)R_5.$$

The above is sufficient for ascertaining the importance of the probabilities R_m.

The second case. The firing is only one of many similar mutually independent firings and we are only interested in ascertaining the mean damage inflicted on the enemy. Here, it is sufficient to know, with regard to each separate firing, the expectation Eξ of damage ξ.

Denote the conditional expectation of ξ by $E(\xi|m)$ when assuming that exactly m hits were achieved. In accord with the well-known formula for the total expectation, and assuming that $E(\xi|0) = 0$, we have

$$E\xi = P_1E(\xi|1) + P_2E(\xi|2) + \ldots + P_nE(\xi|n). \quad (12)$$

Supposing that the expectation of damage $E(\xi|m)$ is proportional to the number of hits; that is, that

$$E(\xi|m) = km \quad (13)$$

where k is a constant factor, formula (12) reduces to

$E\xi = kE\mu$. (14)

In many particular cases this assumption, and, consequently, formula (14) which is its corollary, may be considered sufficiently justified. Then the efficiency of the firing from the viewpoint of the expected damage is indeed determined by the expected number of hits. It seems, however, that such particular cases, when a noticeable deviation from proportionality (13) is observed, are encountered not less often; when, consequently, the replacement of the general formula (12) by (14) is inadmissible.

It is quite natural to assume that in any case $E(\xi|m)$ does not *decrease* with the increase in m; *i.e.*, that

$$c_m = E(\xi|m) - E(\xi|m-1) \geq 0.$$

Accordingly, it is convenient to rewrite the general formula (12) as

$$E\xi = c_1R_1 + c_2R_2 + \ldots + c_nR_n. \quad (16)$$

The above is sufficient for understanding when we may apply the expected number of hits $E\xi$ and when should we turn to the general formula (16) for estimating the expected damage inflicted on the enemy by shelling. In each of the two typical instances discussed above it was possible to estimate the efficiency of the system of firing by one single magnitude which we may call the *indicator of its efficiency*. In the first case it was the probability $P(A)$; in the second instance, the expectation $E\xi$; and in both cases the expression of the type

$$W = c_1R_1 + c_2R_2 + \ldots + c_nR_n \quad (17)$$

with c_m being some non-negative coefficients actually served as the indicator of the efficiency of the system of firing. It may be thought that the expression (17) is also sufficiently flexible for covering a number of intermediate cases not discussed above.

Once the indicator of the efficiency of the system of firing conforming to the concrete situation is chosen, it becomes natural to choose the most advantageous system of firing from among those possible (and consisting of a *stipulated number of shots* n), – the one for which this indicator takes the largest possible value. In §5, we shall consider such problems on determining the systems of firing having the largest indicator of efficiency when the *expenditure of ammunition* is given beforehand.

In concluding this section, we note that

$$E\mu = R_1 + R_2 + \ldots + R_n \quad (18)$$

and that the magnitudes R_m are connected with $E\mu$ by the well-known *Chebyshev inequality* $R_m \leq E\mu/m$.

2. The Case of Mutually Independent Hits after Single Rounds

Suppose that the numbers $i = 1, 2, \ldots n$ are assigned to the n shots. Denote the random event of the i-th shot *hitting/missing* the target by B_i/C_i and the corresponding probabilities by $p_i = P(B_i)$ and $q_i = P(C_i) = 1 - p_i$. The expectation can be expressed by the well-known formula

$$E\mu = p_1 + p_2 + + \ldots + p_n. \quad (21)$$

This is the formula that attaches especial simplicity to the expression Eμ. Unlike this magnitude, the probabilities P_m and R_m cannot be uniquely represented in the general case by the probabilities $p_i = P(B_i)$. For determining these, it is necessary, generally speaking, to know, in addition to the probabilities p_i of each B_i, the essence of the dependence between the events B_i. Here, we consider the case of *mutually independent* random events $B_1, B_2, \ldots, B_n$ so that the probability that *hits* occur as a result of shots $i_1 < i_2 < \ldots < i_m$ and *do not occur* in any other shot is [5]

$$\frac{p_{i_1} p_{i_2} \cdots p_{i_m}}{q_{i_1} q_{i_2} \cdots q_{i_m}} \prod_j q_j.$$

Adding all such products corresponding to a given m, we will indeed determine the probability

$$P_m = \left(\sum \sum \ldots \sum \frac{p_{i_1} p_{i_2} \cdots p_{i_m}}{q_{i_1} q_{i_2} \cdots q_{i_m}}\right) \prod_j q_j,\ i_1 < i_2 < \ldots < i_m \quad (22)$$

of achieving exactly m hits after n shots. In particular,

$$P_0 = \prod_j q_j,\ P_1 = \left(\sum_i \frac{p_i}{q_i}\right) \prod_j q_j,\ P_2 = \left(\sum \sum \frac{p_{i_1} p_{i_2}}{q_{i_1} q_{i_2}}\right) \prod_j q_j,\ i_1 < i_2.$$

The number of terms in brackets in (22) is equal to C_n^m. Therefore, if

$$p_1 = p_2 = \ldots = p_n = p, \quad (23)$$

we arrive at the well-known formula $P_m = C_n^m p^m q^{n-m}$, $q = 1 - p$.

The general formula (22) is too complicated for application without extreme necessity. When the probabilities p_i of hit for each separate shot are *sufficiently low*, it can be replaced by much more convenient appropriate formulas. These are the more exact, the lower is the upper bound $\lambda = \max(p_1; p_2; \ldots; p_n)$ of the probabilities p_i. The simplest [6] among them is the *Poisson formula*

$$P_m = (a^m/m!)e^{-a} + O(\lambda),\ a = p_1 + p_2 + + \ldots + p_n = E\mu. \quad (27)$$

This formula represents the probability P_m to within the remainder term $O(\lambda)$. The published tables [8] of the function $\psi(m; a) = (a^m/m!)e^{-a}$ make its practical application simpler.

A more complicated but at the same time a more precise formula [7] can be recommended:

$$P_m = \psi(m; a) - (b/2)\nabla^2\psi(m; a) + O(\lambda^2) \quad (30)$$

where

$$b = p_1^2 + p_2^2 + + \ldots + p_n^2, \quad (31)$$
$$\nabla^2\psi(m; a) = \psi(m; a) - 2\psi(m-1; a) + \psi(m-2; a).$$ [9]

The remainder term in (30) has order λ^2.

When all the probabilities are identical (23),

$a = np$, $b = np^2 = a^2/n$, $\lambda = a/n$

and formulas (27) and (30) are reduced to [10]

$$P_m = \psi(m; a) + O(1/n),\ P_m = \psi(m; a) - (a^2/2n)\nabla^2\psi(m; a) + O(1/n^2). \quad (33; 34)$$

To show that the refined formulas (31) and (34) are advantageous as compared with (27) and (33), let us consider the following example: $n = 5$, $p_1 = p_2 = \ldots = p_5 = 0.3$, $a = 1.5$ and $b = 0.45$. The corresponding values of $P'_m = \psi(m; a)$, $P''_m = \psi(m; a) - (b/2)\nabla^2\psi(m; a)$ and P_m are given in Table 1 {omitted}.

Since the probabilities p_i are here still rather large, the Poisson formula (27) or the formula (33) only offer a very rough approximation P'_m to the true values of P_m but the refined formulas (30) and (34) already provide the approximations P''_m which differ from P_m less than by 0.015.

Let us also consider the case $n = 50$, $p_1 = p_i = 0.03$, $i = 1, 2, \ldots, 50$, $a = 1.5$, $b = 0.045$. The corresponding values of P_m, P'_m and P''_m are adduced in Table 2 {omitted} to within 0.00001. Here, even the usual Poisson formula provides admissible results (the deviations of P'_m from P_m are less than 0.005) whereas the refined formulas (30) and (34) furnish P_m with an error less than 0.0001.

For the probabilities R_m of achieving *not less than m* hits, formulas (27), (30), (33) and (34), when issuing from (3), lead to

$$R_m = H(m; a) + O(\lambda),\ R_m = H(m; a) - (b/2)\nabla^2 H(m; a) + O(\lambda^2), \quad (35; 36)$$
$$R_m = H(m; a) + O(1/n),\ R_m = H(m; a) - (a^2/2n)\nabla^2 H(m; a) + O(1/n^2) \quad (38a; b)$$

where, assuming that $H(m; a) = 1$ for $m < 1$,

$$H(m; a) = 1 - \psi(0; a) - \psi(1; a) - \ldots - \psi(m - 1; a).$$

The tables of

$$H(1; a) = 1 - e^{-a} \text{ and } H(2; a) = 1 - (1 + a)e^{-a}$$

are available in a number of treatises on the theory of firing [11]. The values of $H(m; a)$ and

$$\nabla^2 H(m; a) = -\nabla\psi(m - 1; a) \quad (41)$$

for $m \le 11$ are adduced in Tables 1 and 2 of the Supplement to this book {to the original Russian source} and also there see Fig. 1. By means of these tables the probabilities R_m are determined in accord with formulas (36) and (38) with a very small work input. Suppose for example that $n = 24$ and

$p_1 = p_2 = \ldots = p_6 = 0.20$, $p_7 = p_8 = \ldots = p_{12} = 0.10$, $p_{13} = p_{14} = \ldots = p_{18} = 0.15$ and $p_{19} = p_{20} = \ldots = p_{24} = 0.05$.

It is required to determine the probability R_3 of achieving not less than three hits. We have $a = \Sigma p_i = 3.00$, $b = \Sigma p_i^2 = 0.45$. According to the tables, for $m = 3$ and $a = 3$, $H = 0.577$ and $\nabla^2 H = 0.075$. Substituting these values in formula (36) results in

$$R_3 \approx 0.577 + (0.45/2)\cdot 0.075 = 0.594.$$

To compare, we adduce the elementary calculation {omitted} by means of formula (22):

$R_3 = 1 – (P_0 + P_1 + P_2) = 0.59503$.

Let us return now to justifying the formulas (27) and (30). They are derived from the expansion of the probabilities into Charlier series

$$P_m = \sum_{k=0}^{\infty} A_k \nabla^k \psi(m; t) \qquad (42)$$

where

$$A_k = e^t \sum_m \nabla^k \psi(m; t)\, P_m,\ \nabla^0 \psi(m; t) = \psi(m; t), \qquad (43; 44a)$$

$$\nabla^{k+1}\psi(m; t) = \nabla^k \psi(m; t) - \nabla^{k-1}\psi(m; t). \qquad (44b)$$

The coefficients A_k can also be represented as

$$A_k = \sum_{i=0}^{k} (-1)^i (t^i/i!)[F_{k-1}/(k-1)!] \qquad (45)$$

where F_s are the *factorial moments* of the random variable μ, *i.e.*,

$$F_0 = \sum_m P_m = 1,\ F_1 = \sum_m mP_m = a,\ F_s = \sum_m m(m-1)\ \dots\ (m-s+1)P_m. \qquad (46)$$

Formulas (42) – (46) are applicable not only to our special case in which the probabilities P_m are represented by (22) but to any probabilities $P_m = P(\mu = m)$ for an arbitrary random variable μ only taking a finite number of integer non-negative values $m = 0, 1, 2, \dots, n$ [12].

The parameter t is here arbitrary. It is usually assumed that $t = a$, and then the formulas for the first coefficients become somewhat more simple. Namely, if $t = a$, we will have

$A_0 = 1, A_1 = 0, A_2 = (F_2/2) – aF_1 + (a^2/2)$,
$A_3 = (F_3/6) – aF_2/2 + (a^2/2)F_1 – (a^3/6)$.

For our particular case, assuming that

$a = p_1 + p_2 + + \dots + p_n, b = p_1^2 + p_2^2 + + \dots + p_n^2$,
$c = p_1^3 + p_2^3 + + \dots + p_n^3, d = p_1^4 + p_2^4 + + \dots + p_n^4$,

we have, again for $t = a$,

$$A_0 = 1, A_1 = 0, A_2 = - b/2, A_3 = - c/3, A_4 = - d/4 + (b^2/8). \qquad (50)$$

Since $b, c, d, \dots$ are magnitudes of the order not higher than $\lambda, \lambda^2, \lambda^3, \dots$ respectively, $A_2 = O(\lambda)$, $A_3 = O(\lambda^2)$, $A_4 = O(\lambda^2)$. It can be shown that, for any $k \geq 1$,

$A_{2k-1} = O(\lambda^k), A_{2k} = O(\lambda^k)$.

A more subtle analysis shows that not only the terms of the series (42) having numbers $(2k – 1)$ and $2k$ are, in our case [13], of an order not higher than λ^k; the same is true with regard to the sum of all the following terms. Thus, in our case, when curtailing our series by the term with number $(2k - 2)$, we commit an error of the order not higher than λ^k. Assuming $k = 1$ and 2, we arrive at formulas (27) and (30) by means of (50).

In applications, the order of error with respect to some parameter (in our case, to λ) selected as the *main infinitesimal* only offers most preliminary indications about the worth of some approximate formula. To appraise the practical applicability of an approximate formula for *small but finite* values of the parameter, it is necessary either to calculate a sufficient number of typical examples, or to derive estimates of the errors in the form of *inequalities* satisfied even for finite values of the parameter.

Until now, I have only obtained sufficiently simple estimates for $m = 1$, and I adduce them without proof [14] directly for formulas (35) and (38). For $m = 1$ they provide [15]

$$R_1 = 1 - e^{-a} + O(\lambda),\ R_1 = 1 - [1 - (b/2)]e^{-a} + O(\lambda^2). \qquad (52;\ 53)$$

The respective estimates in the form of inequalities are

$$1 - e^{-a} \le R_1 \le (1 - e^{-a})\{1 + [\lambda/2(1 - \lambda)]\},$$
$$1 - [1 - (b/2)]e^{-a} \le R_1 \le \{1 - e^{-a}[1 - (b/2)]\}\{1 + [\lambda^2/3(1 - \lambda)]\}.$$

Table 3 {omitted} shows the values of the expressions $[\lambda/2(1 - \lambda)]$ and $[\lambda^2/3(1 - \lambda)]$ for some values of λ. It is seen for example that for $\lambda = 0.2$, when determining R_1 in accord with formula (52), we run the risk of being mistaken not more than by 12.5%, and not more than by 1.7% when applying formula (53). It would be very desirable to obtain equally compact estimates in the form of inequalities for the probabilities R_m at $m > 1$.

3. Classification of the Factors Conditioning the Results of Firing and the Importance of the Dependence between Hits after Single Rounds

When considering some problems concerning the choice of a sensible system of firing, it is convenient to separate the factors from which the result of firing depends into the following four groups.

1) *Factors assumed to be known beforehand.*

2) *Factors at our disposal*. These are usually the number of shots and their distribution over time (between the limits allowed by the number of artillery pieces available, their rate of fire, the stock of shells), and, mainly, the aim of the pieces at each shot (azimuth; sight; time of detonation for explosive shells).

For the sake of definiteness we will now consider the case of percussion firing with a fixed number of shots n fired at fixed moments of time. In this case we still generally have at our disposal the choice of two parameters for each shot, the azimuth and the sight (the range), and we denote these, for the i-th shot, by α_i and β_i respectively. Mathematically speaking, the choice of a sensible system of firing under these conditions is then reduced to selecting the most advantageous combination of the values of $2n$ parameters, $\alpha_1, \alpha_2, \ldots, \alpha_n$ and $\beta_1, \beta_2, \ldots, \beta_n$.

3) *Random factors influencing the results of all the shots, or of some of them.* Such are, for example, the errors in determining the location of the target if the aiming of the piece depends on them for several shots [16]; the manner of the maneuvering of a moving target, etc. When mathematically studying the issues of firing, it is admitted that all the factors of this group are determined by the values of some parameters $\theta_1, \theta_2, \ldots, \theta_s$ obeying a certain *law of distribution* usually defined by the appropriate *density* $f(\theta_1; \theta_2; \ldots; \theta_s)$. For the sake of brevity we denote the totality of the parameters θ_r $(r = 1, 2, \ldots, s)$ by a single letter θ.

4) *Random factors not depending one on another or on the factors of the third group with each of them only influencing some single shot*. These are the factors leading to the so-called

technical scattering and to errors in training a piece when this is done independently for each separate shot.

The probabilities P_m, R_m and p_i and the magnitudes Eμ and W connected with them and considered in §§1 and 2 are calculated under the assumption that all the factors of the first two groups are fixed. It is senseless to discuss, for example, the probability of achieving three hits when the conditions of firing are absolutely unknown or when the external conditions are known but the aiming of the piece is not. We shall therefore suppose that these probabilities and magnitudes are functions of the parameters α_i and β_i (whose dependence on the factors of the first group is not necessary to indicate because these factors are always assumed constant).

In addition to *unconditional probabilities* P_m, R_m and p_i and expectations Eμ we will consider the *conditional* probabilities

$$P_m(\theta) = P(\mu = m|\theta_1; \theta_2; \dots; \theta_s),\ R_m(\theta) = P(\mu \geq m|\theta_1; \theta_2; \dots; \theta_s),$$
$$p_i(\theta) = P(B_i|\theta_1; \theta_2; \dots; \theta_s)$$

and the *conditional expectations* $E(\mu|\theta_1; \theta_2; \dots; \theta_s)$ for fixed values of the parameters θ_1, θ_2, …, θ_s. All these magnitudes are connected with the unconditional probabilities and expectations by the well-known formulas

$$P_m = \int \int \dots \int P_m(\theta) f(\theta)\, d\theta_1 d\theta_2 \dots d\theta_s), \quad (60)$$
$$R_m = \int \int \dots \int R_m(\theta) f(\theta)\, d\theta_1 d\theta_2 \dots d\theta_s),$$
$$p_i = \int \int \dots \int p_i(\theta) f(\theta)\, d\theta_1 d\theta_2 \dots d\theta_s),$$
$$E\mu = \int \int \dots \int E(\mu|\theta) f(\theta)\, d\theta_1 d\theta_2 \dots d\theta_s).$$

For the magnitudes

$$W(\theta) = c_1R_1(\theta) + c_2R_2(\theta) + \dots + c_nR_n(\theta),$$
$$W = c_1R_1 + c_2R_2 + \dots + c_nR_n \quad (65)$$

we have a similar formula

$$W = \int \int \dots \int W(\theta) f(\theta)\, d\theta_1 d\theta_2 \dots d\theta_s. \quad (66)$$

From the viewpoint of the here adopted classification of the factors determining the results of firing, it may be said that, when assuming, in §2, that the events B_i (achievement of hits after single rounds) were independent, we neglected the existence of the factors of the third group. Generally speaking, however, the random events B_i should only be considered *conditionally independent for fixed values of the parameters* θ_1, θ_2, …, θ_s. Therefore, *generally speaking, all the formulas of* §2 *should be applied not to the unconditional probabilities* P_m, R_m *and* p_i *but to the conditional probabilities* $P_m(\theta)$, $R_m(\theta)$ *and* $p_i(\theta)$. For example, when applying formula (22) to conditional probabilities $P_m(\theta)$ and $p_i(\theta)$ and calculating the unconditional probability P_m in accord with formula (60), and denoting $1 - p_i(\theta) = q_i(\theta)$, we obtain for it ($i_1 < i_2 < \dots < i_m$)

$$\int \int \dots \int f(\theta)\left(\sum\sum\dots\sum \frac{p_{i_1}(\theta)p_{i_2}(\theta)\dots p_{i_m}(\theta)}{q_{i_1}(\theta)q_{i_2}(\theta)\dots q_{i_m}(\theta)}\right)\prod_j q_j(\theta)\, d\theta_1 d\theta_2 \dots d\theta_s$$

(67)

which is suitable even without the special assumption about the unconditional independence of the events B_i.

All the conditional probabilities $P_m(\theta)$, $R_m(\theta)$ and $p_i(\theta)$ and the expectation $E(\mu|\theta)$ here introduced certainly depend, in addition to the explicitly indicated parameters $\theta_1, \theta_2, \ldots, \theta_s$, on the parameters α_i and β_i which is also true for the unconditional probabilities P_m, R_m and p_i and for the unconditional expectation $E\mu$.

4. The General Formulation of the Problem about Artificial Scattering

We turn now to considering the problem already sketched at the end of §1: *Determine the system of firing that, given the number of shots n, leads to the maximal value of the indicator of the efficiency W*. This problem consists in determining the maximal value of the function

$$W = W(\alpha_1; \alpha_2; \ldots; \alpha_n; \beta_1; \beta_2; \ldots; \beta_n)$$

and the corresponding combination ($\alpha_1^*, \alpha_2^*, \ldots, \alpha_n^*$ and $\beta_1^*, \beta_2^*, \ldots, \beta_n^*$) of the values of the parameters α_i and β_i.

Supposing that the probability p_i of achieving a hit at the i-th shot only depends on the corresponding azimuth and range, but not on these magnitudes for the other shots [17], *i.e.*, that p_i is only a function of two variables, α_i and β_i, $p_i = p_i(\alpha_i; \beta_i)$. Assume also that, as it most often happens, this function attains its maximal value $\max p_i = p_i(\overline{\alpha}_i; \overline{\beta}_i)$ for some quite definite, single combination $(\overline{\alpha}_i; \overline{\beta}_i)$ of the values of the azimuth α_i and the range β_i. Then a natural question is, whether or not

$$\max W = W(\overline{\alpha}_1; \overline{\alpha}_2; \ldots; \overline{\alpha}_n; \overline{\beta}_1; \overline{\beta}_2; \ldots; \overline{\beta}_n), \quad (71)$$

i.e., whether it would be sufficient, for achieving maximal efficiency of firing, to attain maximal probability of hitting at each separate shot.

In two special cases the answer is yes. First, *equality* (71) *is valid if* $W = E\mu$. Second, *it is valid for indicators of efficiency of the type* (65) *with non-negative coefficients* c_m (*in particular, if* $W = R_m$ *for any m*) *if the events* B_i *are mutually independent*; that is, when assuming that the factors of the third group may be neglected.

The first proposition directly follows from formula (21). To become convinced in the correctness of the second one, it is sufficient to notice here that, for independent events B_i, the probabilities R_m will, on the strength of formulas (3) and (22), be quite definite single-valued functions

$$R_m = F_m(p_1; p_2; \ldots; p_n)$$

of the probabilities p_i; and it is easy to prove that these functions will be *non-decreasing* with respect to each of their arguments.

In the next section we will see, however, that no less important are those cases in which

$$W = W(\overline{\alpha}_1; \overline{\alpha}_2; \ldots; \overline{\alpha}_n; \overline{\beta}_1; \overline{\beta}_2; \ldots; \overline{\beta}_n)$$

is considerably less than the max W; *i.e.*, that the system of firing leading to the maximum probability of hits for each single round will not be the most sensible. In such cases, to attain maximal efficiency of the firing as a whole, the aiming of separate shots should deviate from those for which maximal probability of hitting is attained at each shot. Such firing is called *firing with artificial scattering*. The two special cases described above can now be formulated thus:

1) Artificial scattering cannot be useful if the measure of the efficiency of firing is the expected number of hits $E\mu$.

2) Artificial scattering is useless if the hits by single rounds are mutually independent events.

As a rule, in both these cases artificial scattering is not only useless, but also damaging, that is, leading to a decrease in the efficiency of firing. A typical situation when artificial scattering can be advantageous is such when

1) It is most essential to achieve even a small number of hits, – considerably smaller than the total number of shots, n.

2) From among the random factors conditioning the achievement of some number of hits most important are those that influence all the firing (*i.e.*, those of the third group).

The first condition is realized in an especially sharp form in such cases in which *one hit* is sufficient for fulfilling the formulated aim; that is, in cases in which it is natural to assume that $W = R_1$.

5. The Probability of a Hit in the Case in Which $P(A|m) = 1 - e^{-m}$

Suppose that, as $m \to \infty$, $\lim P(A|m) = 1$. Then, as the number of hits increases unboundedly, the target will certainly (with probability = 1) be hit sooner or later. If the hits occur one after another rather than several at once, the expected number of hits after which the target is destroyed is

$$\omega = \sum_{r=1}^{\infty} rD_r = \sum_{r=1}^{\infty} r[P(A|r) - P(A|r-1)]$$

where D_r is the probability of destroying the target at exactly the r-th hit. For the deliberations below, it is also useful to write this down as

$$\omega = \sum_{r=0}^{\infty} [1 - P(A|r)]. \qquad (77)$$

In brief, we will call ω *the mean necessary number of hits*. Obviously, $\omega \geq 1$ and $\omega = 1$ is only possible if

$$P(A|r) = 0 \text{ at } r = 0 \text{ and } 1 \text{ at } r \geq 1,$$

that is, when the target is certainly destroyed by the first hit. For an integer m, $\omega = m$,

$$P(A|r) = 0 \text{ at } r < m \text{ and } 1 \text{ otherwise}. \qquad (80)$$

Above, I indicated, however, that for $\omega > 1$ the case (80) is exceptional: more often $P(A|r)$ *gradually* increases with r. Here, we consider the case of

$$P(A|r) = 1 - e^{-\alpha r} \qquad (81)$$

where α is some positive constant. The assumption (81) is not less arbitrary (but not more either!) than (80) but it *leads to considerably simpler results*.

From (81), because of (77), it follows that

$\omega = 1/[1 - e^{-\alpha}]$, $\alpha = -\ln [1 - (1/\omega)]$. (82; 83)

Let us dwell on the case of mutually independent hits by single rounds considered in §2. Here, for the probability

$$P(\bar{A}) = 1 - P(A) = \sum p_i [1 - P(A|i)],$$

we obtain in accord with formulas (22) and (81) and after some transformations the relation

$$P(\bar{A}) = \prod_i [p_i e^{-\alpha} + q_i]. \quad (85)$$

Denoting

$$p_i' = p_i/\omega = p_i[1 - e^{-\alpha}],\ q_i' = 1 - p_i'$$

we have from (85)

$$P(\bar{A}) = \prod_i q_i' = \prod_i [1 - (p_i/\omega)],\ P(A) = 1 - \prod_i [1 - (p_i/\omega)]. \quad (87; 88)$$

Formula (88) shows that, *under the assumptions made, the probability of hitting the target* $P(A)$ *is equal to the probability* $P(A)$ *existing had the destruction of the target been already certainly achieved by one hit, but with the probabilities* p_i *replaced by* $p_i' = p_i/\omega$. The same conclusion is also true when the hits after single rounds depend one on another in a way considered in §3 if we take $p_i'(\theta) = p_i(\theta)/\omega$.

Notes

1. {Kolmogorov numbered almost all of the displayed formulas; I deleted the numbers which were niot really necessary.}

2. Here and below I understand *firing* as a group of shots made in pursuing some common aim. It can be done either in one volley, in several volleys, or in a sequence of single rounds. Then, *a system of fire* is its order established in advance; it can envisage a fixed number of shots (as assumed in this paper) or a ceasefire after achieving the contemplated goal, the distribution of shots among different ranges; the order of the ranging fire, etc.

The result of each actual firing is to a considerable extent *random*. Hence, the efficiency of a system of firing cannot be characterized by the result of some isolated firing carried out in accord with it. *The estimation of the efficiency of a system of firing can only depend on the distribution of probabilities of the possible results of separate firings carried out in accord with it.* Any magnitude that is uniquely determined by this distribution may be considered as some *characteristic* of the given system of firing. In our case, when the firing consists of n shots, each of them resulting in a *hit* or a *miss*, such characteristics include, in particular, magnitudes $E\mu$, P_m and R_m. On the contrary, the number of hits μ may only be considered as such a characteristic in the case which does not interest us, when one from among the probabilities P_m is unity and all the other ones are zero.

3. Here, A is the *random event* signifying success.

4. Formula (11) is derived from (5) by means of the Abel transformation that I apply here in the following slightly unusual form: *If* $a_0 = 0$, $\nabla_m = a_m - a_{m-1}$, $R_m = P_m + P_{m+1} + \ldots + P_n$, *then*

$$a_1P_1 + a_2P_2 + + \ldots + a_nP_n = \nabla_1R_1 + \nabla_2R_2 + \ldots + \nabla_nR_n.$$

In a similar way, (16) will be derived from (12) and (18) from (17).

5. In this section, if additional restrictions are not indicated, symbols $\sum$ and $\prod$ extend over all the integer values of the appropriate subscript from 1 to n.

6. The proof of this generalization of the Poisson theorem is provided below in this section. *As usual*, O(λ) *denotes a magnitude of the same order as* λ.

7. Pearson, K., Editor (1914), *Tables for Statisticians and Biometricians.* London, 1924.

8. The proof is offered below in this section.

9. For $m < 0$ we assume that $\psi(m; a) = 0$ so that

$$\nabla^2\psi(0; a) = \psi(0; a),\ \nabla^2\psi(1; a) = \psi(1; a) - \psi(0; a).$$

10. The estimates of the remainder term here indicated, $O(1/n)$ and $O(1/n^2)$, are correct for a *constant a*, or for a *bounded a* changing with n.

11. See, for example, Gelvikh, P.A., *Стрельба* (Firing), vol. 1.

12. If μ takes an infinite number of values $m = 0, 1, 2, \ldots$, then, for the applicability of formulas (42) – (46), it is necessary to impose some restrictions on the probabilities P_m. It is sufficient, for example, to demand that the series

$$\sum_m \frac{P_m}{(m+1)\sqrt{\psi(m;t)}}$$

converges.

13. That is, the case in which P_m is represented by formulas (22) and $t = a$.

14. Cf. Mises, R. *Wahrscheinlichkeitsrechnung,* etc. Leipzig – Wien, 1931, p. 149.

15. Because, for $m < 1$, $H(m; a) = 1$ and we have

$$\nabla^2 H(a; 1) = H(a; 1) - 1.$$

16. And, in general, the so-called repeated errors.

17. This assumption can be wrong, if, for example, the maneuvers of the target depend on the system of firing.

15. A.N. Kolmogorov. The Main Problems of Theoretical Statistics. An Abstract

. *Второе всесоюзное совещание по математической статистике. Ташкент, 1948* (Second All-Union Conference on Mathematical Statistics. Tashkent, 1948). Tashkent, 1948, pp. 216 – 220 …

Foreword by Translator

Only the Abstract, translated below, of Kolmogorov's report at a statistical conference (see this book) was published. The author's views about theoretical statistics were not being generally accepted, see Sheynin, O. (1999), Statistics, definitions of, in Kotz, S., Editor, *Enc. of Stat. Sciences*, Update vol. 3, pp. 704 – 711.

* * *

1. It is customary to separate mathematical statistics into its descriptive and theoretical parts. Taking account of the structure of courses in this discipline, we will, however, distinguish, as a rule, not two, but rather three components: a descriptive part; an exposition of the necessary body of facts on probability theory; and theoretical statistics.

Such structures are entirely appropriate. If mathematical statistics, as justifiably accepted, is a science of the mathematical methods of studying mass phenomena, then the theory of probability should be considered its organic part. Since this theory took shape as an independent science much earlier than the two other parts of mathematical statistics did, and since many of its applications do not demand any developed tools belonging to these other components, I do not at all intend to make any practical conclusions, for example about the expediency of mathematical statistics taking over probability. For me, this remark was only necessary for ascertaining the position of theoretical statistics among other sciences.

2. Descriptive statistics is mostly an auxiliary discipline studying technical issues of systematizing mass data, rational methods of computing their various characteristics (means, moments, etc) and the relations existing among these. Problems about a rational choice of summary characteristics of statistical populations are somewhat more interesting, but they can only seldom be solved in the context of descriptive statistics itself [1]. Be that as it may, the subject-matter of this narrow discipline can be considered more or less formed. Its subsequent development will probably consist by small additions to its arsenal, insofar as these become necessary for the branches of science serviced by descriptive statistics, and, above all, for theoretical statistics.

3. An *explanation* of mass regularities and a rational *monitoring* of mass phenomena are of course those exclusive aims, for whose attainment statistical data are collected and treated according to the rules of descriptive statistics. The solution of both these problems in each special field of mass phenomena (physical, biological or social) is based on concrete investigations of the specific properties of the given field and entirely belongs to the appropriate special science. In those cases, however, in which the notion of *probability* can be applied to the mass phenomena under study (which is not at all their unconditional property) [2], the mathematical machinery for solving these problems is provided by probability theory. It is not at all accidental that the two higher sections of mathematical statistics, overstepping the boundaries of descriptive statistics, issue from this notion: Leaving aside the elementary, purely arithmetical properties of large populations studied by descriptive statistics, the rich in content general regularities of mass phenomena of any nature (from *physical* to social) are as yet only stochastically known to us.

4. The theory of probability teaches us, in its main chapters, how to calculate, issuing from some *assumptions* about the nature of the law of distributions, the probabilities of some compound events whose probabilities are not directly given in the initial suppositions. For example, the initial assumptions in the modern theory of Markov random processes relate to the probabilities of transitions from one state to another one during small elementary intervals of time, whereas the probabilities wanted relate to integral characteristics of the behavior of the studied system over large intervals of time.

5. The testing of stochastic assumptions (hypotheses), like the testing of scientific hypotheses in general, is achieved by observing the conformity of inferences following from them with reality. In the case of very large populations, where the law of large numbers sometimes acts with a precision even exceeding the possibilities afforded by observation, such testing does not demand the development of any special *theory of hypothesis testing*. This is the situation existing, for example, in the kinetic theory of gases, and, in general, in most of the physical applications of probability theory; thus (in the sense interesting us at this conference) the so-called *physical statistics* is weakly connected with *theoretical statistics*.

6. When the stock of observational and experimental data is restricted, the problem of their sufficiency for making theoretical conclusions, or for rationally directing our practical activities, becomes critical. Exactly here lies the domain of theoretical statistics, the third section of mathematical statistics. Understood in this sense, it has developed from the part of probability theory devoted to *inverse problems* that were being solved on the basis of the Bayes theorem. Soon, however, theoretical statistics developed in volume into a vast independent science. It became still more independent when it was found that one of its main aims consisted in discovering *invariant* solutions of its problems equally suitable for any *prior distribution* of the estimated parameters [3]. This fact does not tear theoretical statistics away from the general sections of probability theory, but creates for it, all at once, a very peculiar field of research.

The system of the main concepts of theoretical statistics is still, however, in the making. Only gradually does this discipline cease to be the *applied theory of probability* in the sense of consisting of a collection of separate applied stochastic problems unconnected with each other by general ideas.

7. Of special importance for the shaping of theoretical statistics as an independent science were

a) The classical theory of estimating hypotheses on the basis of the Bayes theorem.

b) The Fisherian notion of *information* contained in statistical data and the connected concepts of an *exhaustive* [4] system of statistical characteristics, etc.

c) The modern theory of statistical estimators based on the notion of the *coefficient of confidence* in a statistical *rule*.

8. From among the latest directions of research in mathematical statistics, the following are of a fundamental importance for understanding the general logical nature of its problems: the extension of the concept of confidence limits [5] onto estimating random variables rather than only parameters of the laws of distribution; works on *sequential analysis*; and, especially, many contributions on statistical methods of product control and regulation (in particular, on rejection of defective articles).

9. These new directions do not conform in a natural way to the idea, prevalent in the contemporary Anglo-American school, according to which theoretical statistics is a theory of *purely cognitive* estimators (hypotheses, parameters, or, in the case of *non-parametric* problems, distributions themselves) determined by a stock of statistical data. It occurs that the search for a rational direction of practical activity on the basis of a given set of data should not be split into two problems, those of extracting theoretical information from the statistical data, and of its further practical usage. To ascertain directly the most rational line of practical activities by means of the data is often more advantageous, since the separate solution of the two abovementioned problems involves a peculiar loss of some information contained in the statistical data. In addition, the very collection of data should in many cases be flexibly planned so as to depend on the results being obtained. Neither of these remarks can claim to be fully original; they completely conform to long-standing traditions of practical work. However, the means for their organic inclusion into a systematic construction of a mathematical theory of statistical methods of investigating and regulating industrial processes are only now being outlined.

10. The *tactical* point of view expounded in §9 will perhaps also help in terminating the discussion on the justification of the statistical usage of *fiducial probabilities* and of their generalization, the *coefficients of confidence*.

11. As a result of its historical development which is sketched above, the position of theoretical statistics among other sciences is perceived by me in the following way. Both the investigation of any real phenomena, and a practical influence on their course is always based on constructing hypotheses about the regularities governing them which can be either rigorously causal or *stochastic*. The accumulation of a vast statistical material is only essential in the case of regularities of at least a partly stochastic nature [6].

Problems involving direct calculations of probabilities of some course of phenomena that interest us, given hypothetically assumed regularities, do not yet constitute the subject-matter of probability theory. The problem of theoretical statistics, however, consists in developing methods in the required direction, viz, in

a) The use of a restricted stock of observations for
testing the assumed hypotheses;
determining the parameters contained in the hypotheses;
rationally monitoring the processes under study;
subsequently evaluating the results thus achieved.

b) The rational planning of the collection of observational data for attaining the aims listed under a) in a most successful and economical way.

12. I believe that, according to the point of view indicated in §11, it is possible to construct a harmonious system of theoretical statistics embracing all its main sections, both developed and appearing at present.

The decisive importance of the Russian science during the first stages of its development (Chebyshev, Markov); Bernstein's fundamental work on the applications of probability theory; the enormous work done by Romanovsky and his school in mathematical statistics; the creation of statistical methods of studying stationary series of connected variables by Slutsky; and the founding of the elements of the theory of non-parametric estimators by Glivenko, Smirnov and myself, pledge that this aim is quite up to Soviet scientists.

Notes

1. A.Ya. Boiarsky, in his theory of means [1], indicated an interesting case in which the problem of a rational choice of a summary characteristic admits a simple and quite general solution.

2. See for example my paper [2].

3. See my paper mentioned in Note 2.

4. {This term is not anymore used in Russian literature. Cf. Note 15 to the author's previous work on p. 000 in this book.}

5. In this instance, the American school is using a new term, *tolerance* rather than *fiducial limit.* {The author thus translated *fiducial* as *confidence,* see just above. Cf. my Foreword to the same previous work (see Note 4).}

6. Sometimes, however, such stochastic regularities are only superimposed on rigorously causal ones as *errors of observation*.

References

1. Boiarsky, A.Ya. (1929), The mean. In author's *Теоретические исследования по статистике* (Theoretical Investigations in Statistics). M., 1974, pp. 19 – 49.

2. Kolmogorov, A.N. (1942), Determining the center of scattering, etc. Translated in this book.

15. A.N. Kolmogorov. His Views on Statistics: His Report at a Statistical Conference

Foreword by Translator

In 1954, a nation-wide statistical conference was held in Moscow. It was organized by the Academy of Sciences, the Ministry for Higher Education and the Central Statistical Directorate and its main aim was to seal a Marxist definition of statistics, see Kotz (1965) and Sheynin (1998, pp. 540 – 541). Below, I translate two accounts of Kolmogorov's report at that Conference, also see an abstract of his report at a conference in 1948 translated in this book.

15a. Anonymous. Account of the All-Union Conference on Problems of Statistics (Extract)

Moscow, 1954 (extract). *Vestnik Statistiki*, No. 5, 1954, pp. 39 – 95 (pp. 46 – 47)

At first Kolmogorov dwells on the causes that led to the discussion of the problems of statistics. It became necessary, he said, in the first place, to reject sharply those manifestations of the abuse of mathematics in studying social phenomena that are so characteristic of the bourgeois science. Its representatives make up, for example, differential equations which allegedly ought to predict the course of economic phenomena; apply without any foundation hypotheses of stationarity and stability of time series, etc. The discussion was also called forth by the need to surmount in the statistical science and practice, once and for all, the mistaken aspiration of some {Soviet} statisticians to guide themselves by chaotic processes and phenomena [1]. And, finally, the last fact that makes the sharp discussion necessary, Kolmogorov indicated, consists in that we have for a long time cultivated a wrong belief in the existence, in addition to mathematical statistics and statistics as a social-economic science, of something like yet another non-mathematical, although universal *general* theory of statistics [2] which essentially comes to mathematical statistics and some technical methods of collecting and treating statistical data. Accordingly, mathematical statistics was declared a part of this *general theory of statistics*. Such views, also expressed at this Conference, are wrong.

It cannot be denied that there exists a certain set of methods and formulas united under the name of *mathematical statistics*, useful and auxiliary for each concrete science such as biology or economics. Mathematical statistics is a mathematical science, it cannot be abolished or even made into an applied theory of probability. Not all of it is based on this theory. The contents of mathematical statistics is described in detail in Kolmogorov's article [2].

The study of the quantitative relations in the real world, taken in their pure form, is generally the subject of mathematics. Therefore, all that, which is common in the statistical methodology of the natural and social sciences, all that which is here indifferent to the specific character of natural or social phenomena, belongs to a section of mathematics, to mathematical statistics. […] [3]

15b. Anonymous. On the Part of the Law of Large Numbers in Statistics (Extract)

Uchenye Zapiski po Statistike, vol. 1, 1955, pp. 153 – 165 (pp. 156 – 158) …

While dwelling on the law of large numbers in statistics, Kolmogorov indicates that attempts were made in our {Soviet} literature to declare this law altogether senseless or pseudoscientific. However, considering, for the time being, indisputable examples bearing no relation to social science, the fact that this lamp remains motionless, and does not fly up to the ceiling, is the result of the action of the law of large numbers. Air molecules move according to the kinetic theory of gases with great velocities, and if their collisions were not equalized according to this law, we probably would have been unable to confer here. This exaggeration towards a total denial of the theory of probability possibly belongs to the past. I did not hear such pronouncements at our Conference that the theory is not needed at all.

Kolmogorov then dwelt on the role of the theory of probability and mathematical statistics in social-economic sciences. He considers it undoubtedly true that the more complex is the studied sphere of phenomena, the more it is qualitatively diverse, the less applicable becomes the mathematical method. He referred to his previous article [1] where he ordered sciences beginning with astronomy (where everything sufficiently obeys mathematics), going over to the flight of projectiles (where everything also seems to obey mathematics sufficiently well, but where, actually, the situation is opposite), then to biology and to social-economic sciences (where mathematics remains subordinate). As to stability, it is indeed true that the concept of a certain stability, and, more precisely, of the stability of frequencies, underpins the very concept of probability. It is required that, when experiments are repeated many times over, the frequencies tend to one and the same number, to the probability.

Stability of this kind indeed occurs in inorganic nature, although even there this is not exactly so. The probability of radioactive decay, of the emission of one or another particle from an atom during a given interval of time, was until recently believed to be absolutely stable. Only lately was it discovered that even this is not exactly so, that even for a spontaneous decay this probability is not a completely stable magnitude either. Here, the matter depends on the degree of stability, but the qualitative difficulty of applying this concept also depends on this degree.

Kolmogorov offers an example. It is impossible to formulate the concept of climate without mentioning stability since climate is the very frequency of the repetition of different individual kinds of weather. This year, there was little snowfall, but the climate in Moscow did not yet change. Although climate consists of a series of probabilities (to have so many fine days in March, etc), it nevertheless changes gradually; however, during a restricted period of time, it is possible to apply conditionally the concept of stability. Otherwise the concept of climate will disappear (just as temperature will disappear in physics).

The further we advance towards more animated and more complex phenomena, the more restricted is the applicability of the concept of stability, and this is especially true in the case of social phenomena. However, the applicability of the concept of statistical, of stochastic stability is not completely done away with here either. Recall, for example, Kolmogorov went on to say, one of the fields of the work of Soviet mathematicians, where the technical results are indisputably good, the stochastic methods of calculating {predicting?} the work of automatic telephone networks. We are completely justified in considering that the distribution of calls is a chaotic phenomenon. No-one prohibits any citizen to ring up on personal business {anyone else} at any hours of the day or night. From the social point of view, the cause of the calls are random, but, nevertheless, during a usual day a definite statistical distribution of calls establishes itself here in Moscow. Normally, it is stable from day to day (of course, during a restricted interval of time). This is the foundation of a workable technical science closely adjoining the phenomena of social life, and stochastic calculations are applied here with unquestionable success.

The role of mathematics, of the theory of probability and mathematical statistics in social-economic sciences proper is the least significant, but it does not disappear altogether. All the machinery of the so-called descriptive statistics, the technical methods of statistical calculus, remain intact. Sampling, which, after all, belongs to mathematical statistics, also remains. Its mathematical aspect is the same in all fields and we apply it with great success.

Investigations of stochastic chaotic processes are much less important, especially for us, in our planned State system. Nevertheless, there exist certain fields, for example insurance, where we have to encounter chaotic phenomena. It is for overcoming the {effects of} chaotic, unordered {disordered} circumstances of life that insurance exists; and studies of these circumstances are only possible by means of the theory of probability.

The theory of probability is now also applicable to the sphere of servicing the mass needs of the population. For a given city, the provision of goods, the requirement for which is conditioned by various circumstances of personal life, will be stable for a restricted period of time, whereas a small store should have some surplus of stocks. This is a typical area for practically applying the law of large numbers and investigating the deviations from this law in insufficiently large collectives[4].

Notes

1. {This term, chaotic, also appears several times below. Had Kolmogorov himself really applied it?}

2. {Kolmogorov himself had kept to this viewpoint, see p. 000 of this book.}

3. {I omitted the last paragraph whose subject-matter is included in more detail in the second account, see below.}

4. {When listing several applications of the statistical method in social-economic sciences, Kolmogorov omitted demography. This subject was dangerous: the census of 1937 (proclaimed worthless and followed by a decimation of the Central Statistical Directorate) revealed a demographic catastrophe occasioned by arbitrary rule, uprooting of millions of people, mass hunger and savage witch-hunts, see Sheynin (1998, pp. 534 – 535). And the war losses of population during 1941 – 1945 were being hushed up.}

References

1. Kolmogorov, A.N. (a later version: 1974, in Russian), Mathematics. *Great Sov. Enc*. (English edition), vol. 15, 1977, pp. 573 – 585.

2. Kolmogorov, A.N., Prokhorov, Yu.V. (1954, in Russian), Mathematical statistics. *Enc. Math.*, vol. 6, 1990, pp. 138 – 142.

3. Kotz, S. (1965), Statistics in the USSR. *Survey*, vol. 57, pp. 132 – 141.

4. Sheynin, O. (1998), Statistics in the Soviet epoch. *Jahrbücher Nat.-Ökon. u. Statistik*, Bd. 217, pp. 529 – 549.