ZASTOSOWANIA MATEMATYKI
VI (1963)

H. STEINHAUS (Wrocław)

# *PROBABILITY, CREDIBILITY, POSSIBILITY**

The subject of the present work is associated with the name of Thomas Bayes(1). It is an essential association because Bayes not only discovered a way of calculating the probabilities of causes whose consequences are being observed but also was the first to have doubts regarding his own discovery.

**1.** Let us recall Bayes' classical formula. Let $A, B, C, \ldots, N$ be a series (set) of occurrences excluding each other. We call them the *causes* of a phenomenon $Z$ which we have observed. We are concerned with the *posterior probability* $P_Z(A)$ that it was actually $A$ (and not $B, C, \ldots$ or $N$) that had preceded $Z$. Let us assume that the *prior probabilities* $P(A), P(B), \ldots, P(N)$, i.e. the probabilities of each of the occurrences $A, B, \ldots, N$, had been known before the consequence $Z$ was observed, and that $P_A(Z), P_B(Z), \ldots, P_N(Z)$, the so-called *conditional probabilities*, had also been known beforehand. Here $P_A(Z)$ denotes the probability of the fact that $Z$ follows $A$, $P_B(Z)$ — the probability that $Z$ follows $B$, etc. In Bayes' postulate $P_A(Z)$ is expressed by the well-known prior and conditional probabilities:

$$P_Z(A) = \frac{P(A)P_A(Z)}{P(A)P_A(Z)+P(B)P_B(Z)+\ldots+P(N)P_N(Z)}. \tag{1}$$

* Niniejsza praca była ogłoszona w naszym piśmie po polsku w tomie 1 (1954), str. 159-172. Obecnie ogłaszamy ją po angielsku, aby jej treść uczynić dostępniejszą obcym czytelnikom. *Redakcja.*

Данная работа была опубликована в нашем журнале на польском языке в томе 1 (1954), стр. 159-172. В настоящее время публикуем её на английском языке для того, чтобы содержание работы сделать доступным читателям не знакомым с польском языком. *Редакция.*

This paper appeared in our periodical in Polish in vol. 1 (1954), pp. 159-172. We are now publishing it in English in order to make it easier for foreign readers to get acquainted with its contents. *Editors.*

(1) T. Bayes, *An essay towards solving a problem in the doctrine of chances*, Philosophical Transactions 53 (1763), p. 370.

Formally formula (1) is correct, but it is rarely applicable to practical problems, since prior probabilities are seldom known. The legal establishment of paternity is an interesting exception. Owing to Hirszfeld's investigations of more than 2000 cases, the prior probability that the man sued for alimony is really the father has been found to be about 70% (in pre-war Poland)(2). Here, prior probability refers to the moment after instituting serological investigations but before reading their result $Z$. When the result is known, we can compute from formula (1) the posterior probability $P_Z(A)$ that the man sued is the father of the child in whose name he has been summoned. It is precisely this posterior probability that interests the court.

However, here we are not concerned with the exceptions, but with everyday situations in which we do not know the prior probabilities. Such situations arise in practically any technical or scientific investigations, so that the question of retaining formula (1) or replacing it by another one is a fundamental problem of scientific induction and statistical inference.

Such typical questions as: how many TB cases there are in a locality where of 10 random persons examined 2 were found afflicted; or how many steel rods in a warehouse are suitable for building if of 10 checked 2 were too weak – differ only superficially; their mathematical substance is identical and may be properly classified among the problems dealt with in the present paper.

In such cases Bayes employed a *postulate* which we shall call by his name and denote by $\mathcal{B}$. Namely, he assumes that the distribution of prior probabilities is *uniform*, i.e. that the numbers $P(A), P(B), \ldots, P(N)$ are equal. In the instance of the paternity claim, there are only 2 possibilities, $A$ and $B$ (the defendant either is or is not the father). The Bayes postulate would thus state that $P(A) = P(B) = 0.5$.

Some anthropologists have indeed been using formula (1), assuming $P(A) = 0.5$ but they have not realized the fact that they are employing the Bayes rule, and they are even less likely to have realised the essential nature of the problem.

The Bayes postulate applied to the paternity question is downright false, because $P(A) = 0.70$, as we have mentioned before.

In the steel rod case the Bayes postulate would presume the occurrence with equal frequency of sets of rods of quality 1%, 2%, 3%, ..., 100%. Such a hypothesis is undefendable either theoretically or experimentally.

(2) H. Steinhaus, *On establishing paternity* (in Polish). Zastosowania Matematyki 1 (1953), pp. 67-82.

Let us concede that the Bayes postulate would not be opposed if it were only applied to cases where uniformity is empirically or theoretically justified, or where that postulate can be used as a working hypothesis until the moment, when the actual prior distribution is found by experience.

Instead, this postulate is usually accepted as a dogma — it is most needed in the very cases where there is no hope at all that it will prove either in agreement with or in opposition to reality.

Let us take a new drug as an example. What would be the meaning of assuming a uniform distribution of the drug's efficacy? It would mean that drugs of 10% or 15% efficacy are produced in laboratories just as frequently as those of 55% or 60% efficacy. Such a hypothesis would never be either disproved or verified, because there are too few drugs (statistically) for a given disease; what is more, each belongs to a different scientific age, which makes it artificial and unreasonable to deduce the efficacy of modern synthetic drugs, based on entirely new theories, from the quality of the old serum medicines.

2. The dogma of uniformity, i.e. the Bayes postulate, has been vigorously opposed by the English school (R. A. Fisher, K. Pearson). R. A. Fisher the renovator of statistics, introduced the idea of fiducial probability and the Polish scientist Jerzy Spława-Neyman, who propagated new ideas in statistics in the USA, introduced the concept of "confidence interval" (which is a literal translation of the Polish term: "przedział ufności"). Even before the second world war it was generally accepted that only these new concepts, free from the Bayes postulate, are correct. Hence W. Feller, a celebrity in the field of the calculus of probability, compares the adherents of the Bayes rule and formula ("who are using it for the reason of its logical admissibility and its agreement with our way of thinking") to "Plato, who used the same type of arguments to prove the existence of Atlantis...". And he concludes: "...the contemporary theory of statistical tests and estimations is less intuitive but more realistic. One can not only defend it but also apply" it ([3]). It is characteristic that in Feller's text-book the Bayes rule is printed in brevier type.

Obviously the dilemma of "pro or contra Bayes" is not a negligible one; it comprises a wide field of applications from geodesy to mathematics. Do not let us forget that both the question of outlining a geodetic area for a future city and the question of giving permission to use peniciline of Polish manufacture require drawing conclusions concern-

---

([3]) W. Feller, *An Introduction to Probability Theory and its Applications*, New York 1950, p. 85.

ing real quantities from data burdened with errors (in the first instance such quantities are the true coordinates of the orientation points, in the second — the strength of the peniciline). What makes the chaos in this field still greater is the fact that scientists, doctors and engineers were forbidden to use the Bayes rule before they had time to learn how to apply it. Many of them are under the impression that it is just an argument about mathematical terms without any practical significance. Wald's sequential analysis in the field of statistical quality control disregards the Bayes rule entirely, and the advantages of his method have been attributed precisely to this rejection and consequently used as an argument against the Bayes rule(4).

In effect practically all serious mathematicians have left Bayes' camp(5); the only ones that have remained are those who have not understood the objections of the new school. Let us for short call this new doctrine "the theory of credibility". Only Norbert Wiener(6) has had the temerity to call it "a terminological trick". Independently, J. Oderfeld has called attention to the similarity of the rules of procedure in the statistical quality control arising from the Bayes rule and postulate to those of the new theory(7). This has induced me to study the principles of the statistical quality control; I have found that none of the existing methods is definitely superior to the others — they only differ in a more or less effective masking of arbitrary hypotheses(8). In the present publication I wish to discuss the relation between the Bayes rule and postulate and the corresponding methods of the new theory.

Here a difficulty arises from the fact that text-books use various terms, such as: "likelihood", "verisimilitude", "fiducial probability" "confidence interval" etc., none of which is precisely the right one. Therefore we shall have to dispense with the exactness of translation or the agreement with the accepted terminology and ourselves define the concept of credibility, which quantity in Fisher's theory replaces the probability of causes and constitutes the fundamental idea of the new doctrine. We shall explain it with the aid of an example.

---

(4) H. Steinhaus, *Quality control by sampling*, Colloquium Mathematicum 2 (1951), pp. 98-108. See § 7 on pp. 105-107.

(5) One of those who remained in it is the well-known astronomer H. Jeffreys, author of the *Theory of Probability*, Oxford 1948.

(6) An expression of Wiener, cited in H. Steinhaus's *Quality control by sampling*, Colloquium Mathematicum 2 (1951), p. 104. It is taken from Wiener's book *Cybernetics*, New York 1948, pp. 109-110.

(7) J. Oderfeld, *On the dual aspect of sampling plans*, Colloquium Mathematicum 2 (1951), pp. 89-97.

(8) H. Steinhaus, *The principles of statistical quality control* (in Polish), Zastosowania Matematyki 1 (1953), pp. 4-27.

Let us assume that a certain typical reaction of a guinea-pig depends on an unknown content $x$ of a certain substance in 1 $cm^3$ of blood, taken from the veins of a patient, and injected to the guinea-pig. Let us further assume that the expected reaction does occur and that it rises with the increase of $x$. If the reaction occurs in $m$ guinea-pigs out of a total of $n$ guinea-pigs while it does not occur in $n-m$ guinea-pigs, then *ex definitione* the probability that the content $a$ of this substance in the patient's blood will in $n$ experiments yield the above reaction more than $m$ times is what we shall call the *credibility* of the hypothesis that for the patient in question $x < a$. This is the essence extracted from various texts of the ruling school of thought — and not easily extracted at that.

3. Let us discuss first the concept of credibility on another example where observation may give any real quantities (and not only a finite number of them as in the above experiment). Such an example is the position of a material point on a marked material straight line. The true position of the point corresponds to mark $x$ on the scale, but $x$ is unknown and the observation gives the result $\xi$; $\xi - x$ is the error of observation. We know the probabilities of errors, namely the function $p(x, \xi)$, which enables us to calculate the probability that position $x$ gives observation $\xi$ belonging to the interval $\langle \xi_0, \xi_0 + \Delta\xi \rangle$; this probability is

$$p(x, \xi_0)\Delta\xi + o(\Delta\xi)$$ (9)

We could make this example more striking by a specialization of function $p(x, \xi)$, assuming for instance

$$p(x, \xi) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\xi)^2/2\sigma^2}, \tag{2}$$

which corresponds to the normal distribution of errors. This specialization is unnecessary; for our aims it is sufficient to assume that density $p$ is a non-negative and continuous function of the error $\xi - x$,

$$p(x, \xi) \mp f(\xi - x), \tag{3}$$

and that

$$\int_{-\infty}^{\infty} p(x, \xi)\, d\xi = 1 \tag{4}$$

for every $x$.

Assumption (4) is a mathematical equivalent of the certainty that for every quantity $x$ we shall obtain some observation $\xi$, which is obvious.

(9) The symbol $o(x)$ denotes a magnitude which, when divided by $x$, tends to zero together with $x$.

Notice that density (2), corresponding to the normal distribution of error (the mean error being independent of the $x$ measured) satisfies the above assumptions. They are also satisfied by the density corresponding to observation $\bar{\xi}$, which is the mean of $k$ independent observations $\xi_i$ ($i = 1, 2, \ldots, k$) if to each of them corresponds one and the same $p$ satisfying the assumption.

A typical problem now is this: if the observation of $x$ has given $\xi = b$, what is the probability that $x < a$? We shall denote this probability by $P(x < a;\ \xi = b)$. In order to apply the Bayes rule (1) one ought to know the prior probability $F(a)$ that $x < a$, i.e. the probability of this inequality at the time before the observation. $F(a)$ may be called the *prior distribution* of the random variable $x$.

Those opposing the Bayes rule say that:

1. We do not know $F(a)$ and we are not allowed to accept as $F$ some arbitrary function, such as the uniform distribution which, furthermore, does not exist when the random variable $x$ is not bounded;

2. The Bayes postulate of uniform distribution leads to a contradiction;

3. Since $x$ is not a random variable, formula (1) has no equivalent in the law of great numbers to give it a statistical meaning;

4. The Bayes formula is applicable only when the conditional probabilities $P_A(Z), P_B(Z), \ldots$, are known, which requires — according to the classical definition of conditional probability — the knowledge of probabilities $P(AZ)$, $P(BZ)$ etc., i.e. the knowledge of the joint distribution of the pair of random variables $(X, Y)$, of which $X$ runs over all the causes, such as $A, B, C, \ldots$ etc., and $Y$ runs over all the effects (such as $Z, Z', Z''$). In that case, however, formula (1) is not necessary, because the required left-hand side ($P_Z(A)$) is simply the quotient $P(AZ)/\sum P(XZ)$;

5. The Bayes formula is erroneous because if the effect $Z$ takes place the cause $A$ either has or has not occurred; therefore its posterior probability $P_Z(A)$ can only be 1 or 0, contrary to formula (1);

6. The Bayes formula is unnecessary, because there are other methods, universal and free from drawbacks.

One of the other methods mentioned in point 6 is based on the concept of credibility. It gives no answer at all to the question how much is $P(x < a;\ \xi = b)$ but defines instead the *credibility* $C(x < a;\ \xi = b)$ of $x < a$ if $\xi = b$ by the relation

$$C(x < a;\ \xi = b) = P(\xi > b;\ x = a). \tag{5}$$

The calculation of the right side does not require the knowledge of the function $F(a)$, because

$$P(\xi > b;\ x = a) = \int_b^\infty p(a, t)\,dt, \tag{6}$$

which follows immediately from the definition of the function $p$. The new method thus defines uniquely a certain statistical parameter $W$ as the function of $a$ and $b$, as has been seen in (5) and (6)

$$C(x < a;\ \xi = b) = \int_b^\infty p(a, t)\,dt. \tag{7}$$

We have called this parameter the credibility of $x < a$ for $\xi = b$. This is not the probability of this fact at all but the probability of another fact (which has not taken place) under another condition, which is not said to have been satisfied. It is therefore a conventional parameter measuring the confidence degree which may be accorded to the hypothesis that $x < a$ if $\xi = b$ has been observed.

This convention can aspire to usefulness only if the function $f(a, b) = P(\xi > b;\ x = a)$ is for each $b$ a non-decreasing function of the variable $a$; we want the credibility of the consequence to be at least the same as the credibility of the reason; because for $a_1 < a_2$ the inequality $x < a_1$ implies $x < a_2$, we want to have $W(x < a_2;\ \xi = b) \geqslant W(x < a_1;\ \xi = b)$, and this together with (5) compels us to accept the above mentioned condition concerning $f(a, b)$; this condition is not at all an automatic consequence of $P$ being a probability.

Now let us compute $P(x < a;\ \xi = b)$ according to the classical Bayes rule with postulate $\mathscr{B}$, i.e. with the uniform prior distribution of the random variable $x$. In view of the infinity of the interval, we shall use approximation; we shall first assume uniformity in the interval $|x| \leqslant T$ and then in the Bayes formula we shall pass from $T$ to $\infty$. Therefore let the prior probability that $x$ is in the interval $\langle x_0, x_0 + dx\rangle$ be $g(x)dx$ and let

$$g(x) = \begin{cases} 1/2T & \text{for} \quad |x| \leqslant T, \\ 0 & \text{for} \quad |x| > T. \end{cases} \tag{8}$$

The Bayes rule gives

$$P(x < a;\ \xi = b) = \frac{\int_{-\infty}^{a} \cdot g(x)p(x, b)\,db \cdot dx}{\int_{-\infty}^{\infty} \cdot g(x)p(x, b)\,db \cdot dx} = \frac{\int_{-\infty}^{a} g(x)p(x, b)\,dx}{\int_{-\infty}^{\infty} g(x)p(x, b)\,dx}. \tag{9}$$

The limit of the fraction (9) for $T \to \infty$ is — for (8)

$$\frac{\int_{-\infty}^{a} p(t, b)\,dt}{\int_{-\infty}^{\infty} p(t, b)\,dt},$$

hence by (4) the expression

$$\int_{-\infty}^{a} p(t, b)\,dt.$$

Let us denote by $P_T$ the left-hand side of the formula (9) and by $P_{\mathscr{B}}$ its limiting value limit for $T \to \infty$ and we shall obtain according to the preceding sentence

$$P_{\mathscr{B}}(x < a;\ \xi = b) = \int_{-\infty}^{a} p(t, b)\,dt,$$

from which, by substituting $u = a + b - t$ and owing to (3), we obtain

$$P_{\mathscr{B}}(x < a;\ \xi = b) = \int_{b}^{\infty} p(a, u)\,du. \tag{10}$$

Comparing relations (7) and (10) we obtain

$$C(x < a;\ \xi = b) = P_{\mathscr{B}}(x < a;\ \xi = b). \tag{11}$$

Let us call the probability $P_{\mathscr{B}}$ on the right side of (11) "a possibility". This new term is to remind us that it is a probability calculated not without hypotheses but on the basis of the Bayes postulate $\mathscr{B}$. Instead of the symbol $P_{\mathscr{B}}$ let us use the letter $M$ and we shall have

$$M(x < a;\ \xi = b) = C(x < a;\ \xi = b). \tag{12}$$

Hence in the present instance the *possibility of* $x < a$ for $\xi = b$ equals the *credibility* of $x < a$ for $\xi = b$. We have defined a new concept, the possibility $M(x < a;\ \xi = b)$ by

$$M(x < a;\ \xi = b) \overset{\text{df}}{=} P_{\mathscr{B}}(x < a;\ \xi = b). \tag{13}$$

In our theory this is an equivalent of credibility, which was determined by relation (5). Possibility is only a term, but its introduction gives an equal footing to both theories: the classical Bayes theory and the new theory of his opponents. At present the situation is as follows: We give up the calculation of $P(x < a;\ \xi = b)$ because the opponents of the classical theory are right when they say that without the knowledge of the prior distribution the probability $P$ is impossible to calculate and we do not want to use the arbitrary postulate $\mathscr{B}$. But the expression "possibility" leads us out of the dilemma because we now call $P$ (calculated as if the postulate $\mathscr{B}$ were satisfied) the possibility $M(x < a;\ \xi = b)$ of the fact that $x < a$ for $\xi = b$.

What course do the opponents of the Bayes rule and the creators of "fiducial probability" take? They calculate a different probability from the one they were asked about and they call it the credibility $W(x < a;\ \xi = b)$ of $x < a$ when $\xi = b$.

Relation (12) $M \equiv C$ therefore breaks the myth about freeing the calculus of probability from difficulties by the introduction of the credibility concept and vindicates the Bayes rule and postulate, because it shows (even if, for the time being, only as an example) that credibility is simply a probability calculated by the use of this rule and postulate, all scruples being dispensed with by giving this quantity a special name. Thus the first accusation mentioned in § 3 is disposed of.

The question arises why we introduce a new name "possibility" if the relation $M \equiv C$ proves that "credibility" is sufficient. We do that because, though in the example discussed here this relation does exist, there are several instances where it does not: there are problems, where the conditional probabilities do not satisfy assumptions (3) and (4). In those cases, however, one could just as well apply the method of possibility as that of credibility. The parallelism of both methods is obvious and in a certain particularly large class of problems we have shown that they are identical. In that class the verisimilitude method will still be useful in instances where it is easier to calculate by formula (5) than by the Bayes formula, which is often the case.

Just as we have defined the possibility that $x < a$ when $\xi = b$, we may, by analogy, define the possibility that $a_1 \leqslant x < a_2$ when $\xi = b$, namely

$$(14) \qquad M(a_1 \leqslant x < a_2;\ \xi = b) \overset{\text{df}}{=} P_{\mathscr{B}}(a_1 \leqslant x < a_2;\ \xi = b) \qquad (a_1 < a_2).$$

Here the symbols are self-explanatory. The left side expresses the possibility of a double inequality under the condition $\xi = b$; the right hand side expresses the probability of the same inequality under the same condition, calculated from the rule and postulate $\mathscr{B}$. The classical probability calculation gives directly

$$(15) \quad P_{\mathscr{B}}(a_1 \leqslant x < a;\ \xi = b) = P(x < a_2;\ \xi = b) - P_{\mathscr{B}}(x < a_1;\ \xi = b),$$

and (13), (14) and (15) imply

$$(16) \quad M(a_1 \leqslant x < a_2;\ \xi = b) = M(x < a_2;\ \xi = b) - M(x < a_1;\ \xi = b).$$

Property (16) may be taken as a definition of the left hand side instead of (14). This suggests an analogous definition in the theory of credibility:

$$(17)\quad W(a_1 \leqslant x < a_2;\ \xi = b) \overset{\text{df}}{=} W(x < a_2;\ \xi = b) - W(x < a_1;\ \xi = b)$$

for $a_1 \leqslant a_2$. This useful generalization of the credibility concept allows us to define, for instance, the credibility of the fact that with a certain result for a sample the quality of the lot lies in the interval $(a_1, a_2)$; without the statement that the credibility is the probability (though of another fact) this generalization would not suggest itself to us.

Let us pass now to a different example, comprising an enormous field of technical and scientifical investigations.

4. Suppose that quantity $x$ is not directly measurable and that the observations consist of independent trials; at each trial let $x$ be the probability that the phenomenon $Z$ will arise. Thus, for instance, $x$ may be a characteristic of a drug and $Z$ the test of this drug. Notice that in this instance the expression "probability" plays one rôle more: the quantity measured is itself a probability.

We write the result of observations as $(\xi = m)/n$, which is to signify that there have been $n$ independent trials in which result $Z$ has been obtained $m$ times (and has not been obtained $n-m$ times); the integers $m, n$ fulfil the inequalities $0 \leqslant m \leqslant n$.

The question is how to calculate $P\big(x < a; (\xi = m)/n\big)$, i.e. the probability that $x$ is smaller than $a$, if in $n$ experiments $Z$ has been observed $m$ times.

The theory of credibility gives an evasive answer, calculating the ordinary probability $P$ of a different fact in different conditions and calling it the credibility $W$ of the original fact in the original condition. We thus have

$$(18)\qquad C\left(x < a; \frac{\xi = m}{n}\right) \overset{\text{df}}{=} P\left(\frac{\xi > m}{n}; x = a\right);$$

on the right-hand side, $P$ is the probability that, for $x = a$, $Z$ will occur more than $m$ times in $n$ experiments. One can criticize definition (18) on the ground that for $m = n$ and for every $a$ we have $C = 0$. In particular, we have $C = 0$ for $a = 1$. Thus, if for instance in 5 experiments with a drug all five are successful, the credibility of the hypothesis $x < 1$, i.e. the hypothesis that the medicine is not safe, will be zero. This statement will be even stranger with only one successful experiment. One could avoid this paradox by writing on the right side of (18) the inequality $(\xi \geqslant m)/n$ instead of $(\xi > m)/n$, but then another paradox will occur: the changed definition gives $C = 1$ for $m = 0$ with every $a$, even very small, because for $m = 0$ the corrected inequality undoubtedly occurs for $n = 1$ as well as for $n = 100$; we shall infer hence that the credibility of the hypothesis that the drug once tried with a negative result is worthless is 1, exactly the same as the credibility of the same hypothesis regarding the medicine tried 100 times with a negative result.

An easy way out of this inconvenience is the change of the denominator on the right side of (18) from $n$ to $n+1$:

$$(19) \qquad C\left(x < a; \frac{\xi = m}{n}\right) \overset{\text{df}}{=} P\left(\frac{\xi > m}{n+1}; x = a\right)$$

$$(m = 0, 1, \ldots, n; \; 0 \leqslant a \leqslant 1).$$

Here, there is no paradox either for $m = 0$ or for $m = n$. $W$ assumes the value 0 if and only if $a = 0$, which is in agreement with the natural postulate of the zero credibility of the impossible relation $x < a$ when $a = 0$. $C$ equals 1 if and only if $a = 1$, which is expressed by the no less natural postulate of a unitary (therefore maximal) credibility of the relation $x < a$ when $a = 1$. Whatever the result of a finite number of experiments, we can assert with the maximal credibility that the medicine is not absolutely efficacious. For $0 < a < 1$ we always have $0 < C < 1$.

Incidentally, let us observe that even practical physicians have noticed the difficulties involved in the extreme experiments of $m = 0$, $m = n$, which make it impossible to assume naively that the strength of the drug is equal to $m/n$.

Let us now determine the *possibility* $M$ that $x < a$ if $Z$ happened $m$ times in $n$ experiments; analogically to (13)

$$(20) \qquad M\left(x < a; \frac{\xi = m}{n}\right) \overset{\text{df}}{=} P\left(x < a; \frac{\xi = n}{n}\right).$$

To obtain the relation $M = C$ analogous to (12) it is necessary to prove the equality of the right-hand sides in (19) and (20). According to the Bayes rule, we have

$$P\left(x < a; \frac{\xi = m}{n}\right) = \int_0^a x^m (1-x)^{n-m} dx \Big/ \int_0^1 x^m (1-x)^{n-m} dx,$$

$$P\left(\frac{\xi > m}{n+1}; x = a\right) = \sum_{k=m+1}^{n+1} P\left(\frac{\xi = k}{n+1}; x = a\right) = \sum_{k=m+1}^{n+1} \binom{n+1}{k} a^k (1-a)^{n+1-k}$$

$$= \int_0^a x^m (1-x)^{n-m} dx \Big/ \int_0^1 x^m (1-x)^{n-m} dx,$$

which gives the very equality that is required and allows us to write the theorem

$$(21) \qquad M\left(x < a; \frac{\xi = m}{n}\right) = C\left(x < a; \frac{\xi = m}{n}\right).$$

The basic relation (21), exactly as in the continuous case of relation (12), vindicates the Bayes rule, demonstrating that this rule leads to the same number as credibility([10]).

5. There remains, however, a serious objection 2°, relating both to the continuous case and to our recent discussion. This objection states that $P_{\mathscr{B}}$, and therefore also $M$, do not satisfy the condition of invariance. Speaking more precisely: When we observe $x$ and obtain $\xi$, some other investigator, interested in $x^3$ for instance, reads, simultaneously with our reading $\xi$, a number on a different scale connected with the scale $\xi$ in the same manner as two scales side by side on the same ruler are connected. Let us mark them $X = x^3$, $\Xi = \xi^3$, $A = a^3$, $B = b^3$. The second investigator's question is

$$P(X < A;\ \Xi = B) = ?$$

Now, propositions $X < A$ and $x < a$ are equivalent and propositions $\Xi = B$ and $\xi = b$ are likewise equivalent. Consequently, the inequality $X < A$ with the condition $\Xi = B$ occurs if and only if the inequality $x < a$ occurs with the condition $\xi = b$, and the possibilities of both conditional events should be equal:

$$M(X < A;\ \Xi = B) = M(x < a;\ \xi = b); \tag{22}$$

when both investigators calculate the possibilities of the relations they are interested in, they ought to obtain identical results; definition (13) and this demand give

$$P_{\mathscr{B}}(X < A;\ \Xi = B) = P_{\mathscr{B}}(x < a;\ \xi = b). \tag{23}$$

Postulate $\mathscr{B}$ for $x$, however, denotes something else than postulate $\mathscr{B}$ for $x^3$, because the uniform distribution of variable $x$ is inconsistent with the uniform distribution of $x^3$. By the transformation from the variables $x$, $\xi$ to $X$, $\Xi$ the conditional probabilities change also, so that relation (23) may either hold or not. It is easy to show, however, that the main rôle here is played by the conditional probabilities. Namely, if we denote by $P(X, \Xi)$ a function which is analogous to the one denoted already by $p(x, \xi)$ and which may be calculated from the expression $p(x, \xi)d\xi$ by substituting $x = \gamma(X)$, $\xi = \gamma(\Xi)$ (here $\gamma(X) = \sqrt[3]{X}$), then it will be suficient to verify whether $P(X, \Xi)$ has the form $P = F(X - \Xi)$. If so, then

([10]) Cf. J. Oderfeld, *On the dual aspect of sampling plans*, Colloquium Mathematicum 2 (1951), pp. 89-97. Some of the formulas from that paper have been checked by B. Kowalczyk.

we have relation (23) and therefore also (22). Indeed, by (11) and (5), we obtain

$$(24)\qquad \begin{aligned} P_{\mathscr{B}}(X < A;\ \Xi = B) &= P(\Xi > B;\ X = A), \\ P_{\mathscr{B}}(x < a;\ \xi = b) &= P(\xi > b;\ x = a), \end{aligned}$$

where the right-hand sides are equal, being the probabilities of the equivalent relations, whence follows (23). The conclusion is the following:

The calculation of possibilities according to our definition (13) gives always the same results if we use as the objects of observation variables for which the law of error agrees with relation (23). Since that law is not hypothetical but can be found with arbitrary precision on the basis of sufficiently numerous observations, and can also be calculated mathematically for the transformed variable if it is known for the initial variable, it is never doubtful whether the proper variable is used. However, it may happen that neither the variable itself nor its transformations satisfies relation (23); then the concept of probability loses its proper significance.

Here the adherents of credibility may raise the objection that credibility can always be calculated without investigating the character of the variable observed, and that a transformation of the variables does not change it. We may answer that in cases where it is impossible to use the concept of possibility there sometimes occur functions of the error for which it is possible to choose a suitable prior distribution in such a way that the Bayes rule together with that distribution gives a posterior probability equal to credibility. Obviously, to rely on absolute invariance of credibility (which invariance is indisputable) is to impose unconsciously upon the variable observed a prior distribution, every time completely different and at times completely fantastic. If the Bayes postulate has been condemned for imposing every time a uniform distribution, often inconsistent with the nature of the problem, then the unconscious application of the Bayes rule to various prior distributions and, moreover, without specifying them and comparing them with nature, seems to be even more risky.

In the continuous case an instructive example is that of a variable $x$ which is bounded in a finite interval $(c, d)$ and has a uniform prior distribution in that interval. Let the function $p(x, \xi)$ be defined by (2). If we know the prior distribution, we may use the classical Bayes rule and calculate, for instance, the probability that if the measurement has given $\xi = c$, $x$ is smaller than $c$; then the posterior probability is zero and the Bayes rule will give it without any hypotheses. If, however, we do not know the prior distribution (which is actually such as has been stated above), then we must resort either to the concept of credibility or to that of possibility in order to solve the problem. As we have proved in § 3, probability and possibility are equal in this case. It is therefore sufficient to calculate the credibility $C(x < c;\ \xi = c)$; for which expression (7) may be used with the result

$$C(x < c;\ \xi = c) = \int_c^\infty p(c, t)\,dt = \frac{1}{\sigma\sqrt{2\pi}} \int_c^\infty e^{-(c-t)^2/2\sigma^2}\,dt = 0.5 .$$

Both methods give $50\,^0/_0$, which in, view of the actual impossibility of the relation $x < c$, is false information. The only difference is that the adherent of credibility would supply this information more freely than the believer in the possibility concept, because the latter would have to verify beforehand whether function (2) has the form (3); in our example it has.

Objection 2° is much easier to refute in the non-continuous case. Here the phenomenon $Z$, or any other phenomenon equivalent to it, is given directly, while $x$ is the probability that $Z$ will occur. This uniquely determines variable $x$, which we seek on the basis of experiments whose effect can only be: "$Z$ has occurred", or "$Z$ has not occurred". No variable $X$ expressed by $x$ (e.g. $X = x^3$) (with the exception of $X \equiv x$) is a probability and therefore $x$ is marked and pointed out by the problem itself. This disposes also of objection (5) in an enormous class of random inspection of quality by attributes.

**6.** In this class we can – as in the instance of the continuous result – define the possibility of the two-sided relation $a_1 \leqslant x \leqslant a_2$. We obtain formulae analogous to (14)-(17), replacing everywhere $\xi = b$ by $(\xi = m)/n$

$$M\left(a_1 \leqslant x < a_2;\ \frac{\xi = m}{n}\right) \overset{\text{df}}{=} P\left(a_1 \leqslant x < a_2;\ \frac{\xi = m}{n}\right), \tag{25}$$

$$M\left(a_1 \leqslant x < a_2;\ \frac{\xi = m}{n}\right) = M\left(x < a_2;\ \frac{\xi = m}{n}\right) - M\left(x < a_1;\ \frac{\xi = m}{n}\right), \tag{26}$$

$$C\left(a_1 \leqslant x < a_2;\ \frac{\xi = m}{n}\right) \overset{\text{df}}{=} C\left(x < a_2;\ \frac{\xi = m}{n}\right) - C\left(x < a_1;\ \frac{\xi = m}{n}\right), \tag{27}$$

$$C\left(a_1 \leqslant x < a_2;\ \frac{\xi = m}{n}\right) = M\left(a_1 \leqslant x < a_2;\ \frac{\xi = m}{n}\right). \tag{28}$$

Theorem (26) follows from definitions (25) and (20) and hence theorem (28) is a consequence of definition (27). These formulae facilitate the understanding of the problems of statistical quality control. If two values $a_1, a_2$ $(a_1 < a_2)$ are given, then it is possible to determine a sequential acceptance plan by a regulation stating that a certain lot is to be accepted if the result $(\xi = m)/n$ is such that

$$\text{(I)} \qquad M\left(a_1 \leqslant x < 1;\ \frac{\xi = m}{n}\right) \geqslant 95\,\%,$$

and rejected if the result $(\xi = m)/n$ is such that

$$\text{(II)} \qquad M\left(0 \leqslant x < a_2;\ \frac{\xi = m}{n}\right) \geqslant 95\,\%,$$

and that inspection should be continued until either (I) or (II) is satisfied. Such a scheme differs from the sequential analysis of A. Wald by the fact that it always gives the solution in a finite number of steps if $a_1 < a_2$. Expression (28) allows us to determine our sequence also in terms of the theory of credibility, but it arose from a retrospective way of thinking, namely from the regulation that investigations should be continued until the result of the sample reduces below $5\,\%$ the a posteriori probability that the lot accepted (rejected) is worse (better) than $a_1(a_2)$.

7. The objection that $x$ is not a random variable ought to be understood to mean that the values of $x$ do not form a set where a distribution could be defined. It is sufficient, however, for each statement referring to probability to permit a verification by an experiment (even if only in thought) and the counting up of the relative frequency of successful experiments. In our example, however, such a verification is actually unnecessary, because, as we have demonstrated, in the continuous case we have by (5) and (11)

$$\text{(29)} \qquad P_{\mathfrak{B}}(x < a;\ \xi = b) = P(\xi > b;\ x = a).$$

Therefore, to compare the quantity $P_{\mathfrak{B}}$ with experience it is sufficient to compare the right side of expression (29), i.e. to find the frequency of the results $\xi > b$ for the accepted $x = a$; it will simply be a verification by the ordinary law of great numbers, which is indisputable. On the other hand, the direct frequency interpretation of the left-hand side of the formula (29) is also possible, particularly in the example of § 4.

The variable $x$ is given in turn such values $x_1, x_2, \ldots, x_k, \ldots$, that the sequence $\{x_n\}$ obtains equipartition in the interval $(0,1)$; this means that the relative frequency of its terms smaller than $t$ is equal to $t$ for each $t$ of the $(0,1)$ interval. For each $x_k$ a result $(\xi = m_k)/n$ is determined, expressed by the number $m_k$ of the occurrences of the phenomenon $Z$ in $n$ independent consecutive trials (with $n$ determined once for all), each of them having the probability of success equal to $x_k$. From the sequence $\{m_k\}$ the subsequence $\{m_{k_j}\}$, defined by the condition $m_k = m$, is extracted. For brevity let us denote number $x_{k_j}$ by $x'_j$ and compute in the sequence $\{x'_j\}$ the relative frequency of terms for which $x_j < a$. This frequency should be equal to $P(x < a; (\xi = m)/n)$. It is a special form of the law of great numbers, though this particular statement is seldom found in text-books of the calculus of probability. One of the causes of the aversion against the Bayes rule is probably the conviction (perhaps unconscious) that either this concept is impossible or its proof transgresses the limits of the calculus of probability. How fallacious is this opinion is proved by the elementary formula (1). Namely, this formula may be interpreted outside the calculus of probability as a theorem on relative frequencies in the sequence of pairs ( , ) in which the first place is occupied by $A, B, \ldots, N$ and the second by $Z$ or non-$Z$. According to this interpretation, $P(A)$ for instance should be read as the relative frequency in the whole sequence of pairs which have $A$ in the first place, and $P_Z(A)$ as the relative frequency of those pairs among other pairs having $Z$ in the second place, etc. According to the ordinary law of great numbers it is possible to replace everywhere on the right side the frequencies by the probabilities. Such a frequency is equal to the a posteriori probability $P_Z(A)$, calculated from the Bayes formula (1). Q.E.D.

This verification is based on the fictitious sequence $\{x_n\}$ provided with equipartition. It refutes, therefore, the argument that statements about a posteriori probabilities are statistically unverifiable even by an imagined experiment but, instead, it stands open to two new accusations. The first concerns the introduction, in another form, of the Bayes postulate concealed under the assumption of equipartition. The second concerns the necessity of counting the frequencies of a certain subsequence in another subsequence of trials, which makes any real experiment almost illusory. For that reason we shall now verify our theory of possibility in a different way and by another example, i.e. that from § 3. There possibility was defined by (13). Suppose that in the $j$-th measurement ($j = 1, 2, 3, \ldots$) $X_j$ denotes the true value of the measured quality $x_j$. In this manner we permit not only an absolute variety of observations but also a variety of functions $p_j$ playing the rôle of density $p$ (defined in § 3) in the consecutive measurements, taking care only that the postulate of independence of measurements be satisfied. Let $\xi_j$ denote generally the result of the $j$-th observation. Let us assume that the results in an actual series were $\xi_j = b_j$. Let us fix apriori the number $P$ ($0 \leqslant P \leqslant 1$; e.g. $P = 0.95$). Each time we calculate $a_j$ from the condition

$$M(x_j < a_j;\ \xi_j = b_j) = 0 \qquad (\text{ex. } M = 0.95). \tag{30}$$

This calculation naturally does not require the knowledge of $X_j$. The relation $X_j \leqslant a_j$, however, is objectively either true or false.

THEOREM. *The frequency of those measurements in which* $X_j < a_j$ *is — with respect to the sequence of all measurements — equal to* $P$ (*in the given example* 95 %) *with probability* 1.

Before we prove the theorem let us notice that it perfectly justifies the postulate of possibility from the practical point of view, because it does not assume anything at all with regard to the sequence of true values $X_j$; neither does it require them to be drawn by random out of some population. In spite of that it allows us, on the basis of measurements, to judge these values with the frequency of error determined a priori (here this frequency is $1-P$; for instance 5 %).

Proof. In view of (5) and (11) relation (30) is equivalent to

$$\mathscr{P}(\xi_j > b_j;\, x_j = a_j) = P; \tag{31}$$

the letter $\mathscr{P}$ denotes here the conditional probabilities hither to marked $P$. In view of the non-negativity of function $p$, $\mathscr{P}$ is a non-decreasing function of $a_j$, whence and by (31) it follows that for $X_j < a_j$ (and only for such $X$) we have

$$\mathscr{P}(\xi_j > b_j;\, x_j = X_j) \leqslant P. \tag{32}$$

But the theorem states that the frequency of cases $X_j \leqslant a_j$ is $P$ (from § 1). Since in these cases (and in these only) (32) occurs, the assertion is synonymous with the fact that (with probability 1) the frequency of cases $\mathscr{P} \leqslant P$ is $P$. In other words the theorem may be said to mean the equipartition of the sequence $\mathscr{P}\{\xi_j > b_j;\, x_j = X_j\}$ has. This is a fairly general lemma, which it is sufficient to understand to be able immediately to verify it. Its meaning is as follows: from the measurements of quantity $x_j$ (the true values $X_j$ of which are unknown to us) we obtain the results $b_j$, and then, having learned the values of $X_j$, we compute $\mathscr{P}_j = \mathscr{P}(\xi > b_j;\, x_j = X_j\}$; therefore $\mathscr{P}_j$ is the probability that a repetition of the measurement of quantity $x_j$ will give a greater result than the actual result $b_j$. This probability has been calculated owing to the knowledge of the true value $X_j$ of the measured quantity $x_j$. Therefore $\{\mathscr{P}_j\}$ is a sequence of random variables. It is about this sequence that we assert that it has equipartition.

Proof of the lemma. The lemma will result from the law of great numbers when we prove that each random variable $\mathscr{P}_j$ taken separately has a uniform distribution. Dropping the index $j$, we have to demonstrate that the random variable $\mathscr{P}$ defined by

$$\mathscr{P} = \mathscr{P}(\xi > b;\, x = X) \tag{33}$$

has a uniform distribution. From (33) it is obvious that

$$\mathscr{P} = \int_{b}^{\infty} p(\xi, X)\, d\xi. \tag{34}$$

Let $u, v$ belong to the interval $(0,1)$ and let $u \leqslant v$. Now let us find $b_u, b_v$ from the relations

$$\int_{b_u}^{\infty} p(\xi, X)\, d\xi = u, \qquad \int_{b_v}^{\infty} p(\xi, X)\, d\xi = v. \tag{35}$$

In view of (34) and (35) it is obvious that relation $u \leqslant \mathscr{P} \leqslant v$ is equivalent to the relation $b_u \geqslant b \geqslant b_v$, whence the probability of the first relation is equal to the probability of the second one. But this second probability $\Gamma(b_u, b_v)$ is

$$\int_{b_v}^{b_u} p(\xi, X)\, d\xi$$

whence, according to (35),

$$\Gamma(b_u, b_v) = v - u.$$

This implies that the probability $\Pi(u, v)$ of the relation $u \leqslant \mathscr{P} \leqslant v$ is also $v - u$, which is exactly the meaning of the lemma.

It should be stressed here that there are obviously two modes of verification of the possibility rule. The first is objectionable to the statistician because it requires, in the case of the acceptance of a lot, an artificial assumption of equipartition of the incoming true qualities, and in the continuous case it is totally out of the question. The second one is free from these difficulties and feasible in practice. As a rule, the second one is used jointly with the theory of credibility. It must be remembered, however, that when the experiment has given a certain result $R$, from which we deduce inequality $N$ with probability $P$, we are allowed to understand the practician's question as to the sense of $P$ not only to mean that we are to determine $R$ and $N$ (according to the first verification) and then to investigate whether indeed the result $R$ is accompanied by the cause $N$ in the fraction $P$ of the sequence of consecutive experiments, but also that in every experiment we match the result $R_i$ — according to theoretical computations — with such an $N_i$ that the corresponding $P_i$ is equal to $P$, and then we test, taking into account the whole sequence of experiments, whether it gives a fraction $P$ of cases in which $R_i$ actually is associated with $N_i$ calculated theoretically. It is this second verification that is of a paramount interest to the practician, because it gives him the means of reducing at will the fractions of erroneous statements.

**8.** The question arises whether the example from § 4 lends itself to a similar verification. The answer is affirmative, but for small $n$ analogous statements are only approximately verifiable. This failure of single plans of $m//n$ type is caused by the fact that from the discrete random variable (such as $m$ is) we draw conclusions regarding the continuous variable $x$. It is possible that if one were to accept for $m$ all the values from 0 to $n$ (fractional as well) without changing the classical formula for the probability of the result $m/n$, the above inconvenience could be removed, but the present author has not investigated this possibility.

Such a widening of the concept of the number of good items (for instance "in the investigated sample of 10 items there are 9.7 good items") would also have practical uses. One could for instance consider as good only the cable sustaining the weight of 100 kg or more, and regard that breaking at 70 kg as 0.7 good item and 0.3 defective item.

The second question would be whether one can verify the example from § 3 like that from § 4, namely by using the sequence $\{x_n\}$ provided with equipartition. To do this we would have to overcome two difficulties. The first one — less formidable — is the lack of definition of an equipartition sequence in an infinite interval (such is the $x$ interval). A more serious difficulty would be involved in extracting from the sequence of measurements the subsequence composed of those measurements which resulted in $b$. Since in each measurement the probability of obtaining the result $b$ is zero, the probability that there will be no such measurements is 1. Possibly this difficulty is one of the more or less consciously felt obstacles which discourage the mathematicians from using the Bayes rule. As we have seen in the example from § 3, one could resort to a verification considerably more effective and free from theoretical and practical objections. But the above mentioned difficulty exists not only in statistical verification. It has its analogy in the interpretation of the Bayes rule itself, namely of the continuous variant of that formula. This is the variant (9) in § 3. The function $g(x)$ should now be interpreted an an arbitrary non-negative function with integral 1 in the interval $(-\infty, \infty)$. We do not postulate $\mathscr{B}$ at the moment. The classical calculus of probability demands the computing of the conditional probabilities $p(x, b)$ as quotients; namely $p(x, b)$ is the quotient of the prior probability that "$x$ and $b$" by the prior probability that "$x$". Because both these probabilities are usually zero, therefore $p(x, b) = 0/0$. I disregard here the way out of the difficulty by the limit passage. More important is that the knowledge of the probability that "$x$ and $b$" (or more precisely that $x$ is placed between $x_0$ and $x_0 + dx$ while at the same time $b$ is between $b_0$ and $b_9 + db$) presumes the knowledge of the joint distribution function of the pair of variables ($x$ and $b$), and in such a case the Bayes formula is completely superfluous. This is objection 4°. But in practice there are many instances of conditional probabilities given directly — just as in the problem of measurements in § 3. Also the problem of paternity may be formulated in this manner (though it is not necessary). These remarks suffice to remove all doubts which might arise in reading the preceding paragraphs.

**9.** There is still the question of the relation of the present paper to all that is being said about the Bayes formula and postulate by the present-day authorities in the field of the calculus of probability and mathematical statistics. Undoubtedly a great deal of what the (non-specialist) Polish reader is being told here, perhaps for the first time, may be found in serious foreign text-books. Unfortunately they do not help one to understand the point of view of contemporary science, not only because of the contradictory opinions they present, but also because of their obscure reasoning. This is caused by the fact that most authors apply our problem to particular cases. Some of them observe that *prior* distribution is of little practical importance; others say that in typical problems with the urn such distribution cannot be used. On the other hand, exagerated generalization is also out of place if, for instance, one wants to define *a posteriori* two unknown parameters at once. It impedes the comprehension of the reasoning, which should first of all interpret the principal problem. J. V. Uspensky believes that the lack of prior distribution in the urn problems seems obvious only at first sight. J. L. Coolidge, using as an example Bertrand's suit for the faulty rulette-table, states that though the Bayes formula is defective, there is nothing better. A British astronomer H. Jeffreys found a uniform distribution for the infinite interval (see § 3 formulae (8)-(12) and relation (11)) but for a special $p$, namely for the $p$ of formula (2) and not for the general one, given by (3). He is also interested in invariance (mentioned here in § 5) but he does not notice criterion (23). He stresses the fact that no probability, whether prior, conditional or posterior, is simply a frequence. This could possibly mean the relinquishing of statistical verification, which the present work considers as the only reasonable criterion of the correctness of statements. That he does not see a possibility of such verification is probably the result of his inability to understand the "equal possibility" of Laplace's in the frequency sense. Indeed, none of these authors uses our equipartition model. Even such a historical detail as the question of agreement or non-agreement between Jeffreys and R. A. Fisher, as represented by the first of these authors, arouses some doubts in the reader. According to Jeffreys, Charles Pearson was the only man to believe in the Bayes rule and the frequential definition of probabilities. Even such a serious author as M. G. Kendall says that the assertion which we have denoted by $x < a$; $\xi = b$ cannot have other probability than 0 or 1, because $x$ either is or is not smaller than $a$ — *tertium non datur*. The present author considers this objection (5° in § 5) as unfounded because it can be applied to nearly every case in the calculus of probability, if one accepts the determinism of physical phenomena (which indeed is not denied by any of the authors quoted). When we ask if the card taken out of the pack is an ace of spades, the probability — in view of the fact that even before taking the card out the answer was predetermined — would be 1 or 0. An even better counter example is a card already taken out of the pack but kept face down. Each player would apply here the calculus of probability, though the assertion "the card is an ace of spades" does not at all differ from "$x < a$" because both are applied to facts already accomplished. None of the above-mentioned authors uses our concept of possibility. It is common knowledge that R. A. Fisher's papers are not particularly lucid. Therefore even now the differences of opinion between his theory of "fiducial probability" and Jerzy Neyman's "confidence interval" are not quite clear to all mathematicians.

The continually reviving controversy about the Bayes rule and the theory of "inductive reasoning" is characterised by quotations found in the texts of the above-mentioned authors or at the tops of the chapters as mottoes. Shakespeare's "Hamlet", Kipling's "Captain Courageous", "Through the Looking Glass" by L. Carroll (author of "Alice in Wonderland"), the stories of "One and a Thousand Nights" and other tales based on the theory of the improbabilities have been called to bear witness to the assertions of serious and eminent students of the calculus of probability. May this remark save the present author from the accusation that he has been dealing with problems upon which science has already pronounced its last word, and which have consequently ceased tob e problems, and have passed to the chapters of generally accepted text-books.