Publications

B14-1214-750-7138

BULLETIN

OF THE

UNIVERSITY OF TEXAS

1915 : No. 4.

JANUARY 15 1915

The Arithmetic Mean as Approximately the Most Probable Value a Posteriori Under the Gaussian Probability Law

BY

EDWARD L. DODD

Published by the University six times a month and entered as second class matter at the postoffice at Austin, Texas

Publications of the University of Texas

The University publishes bulletins six times a month. These comprise the official publications of the University, publications on humanistic and scientific subjects, bulletins prepared by the Department of Extension and by the Bureau of Municipal Research and other bulletins of general educational interest. With the exception of special numbers, any bulletin will be sent to a citizen of Texas free on request. All communications about University publications should be addressed to the Editor of University Publications, University of Texas, Austin.

BULLETIN

OF THE

UNIVERSITY OF TEXAS

1915 : No. 4.

JANUARY 15 1915

The Arithmetic Mean as Approximately the Most Probable Value a Posteriori Under the Gaussian Probability Law

BY

EDWARD L. DODD

Published by the University six times a month and entered as second class matter at the postoffice at Austin, Texas

The benefits of education and of useful knowledge, generally diffused through a community, are essential to the preservation of a free government.

Sam Houston.

Cultivated mind is the guardian genius of democracy.....It is the only dictator that freemen acknowledge and the only security that freemen desire.

Mirabeau B. Lamar.

THE ARITHMETIC MEAN AS APPROXIMATELY THE MOST PROBABLE VALUE A POSTERIORI UNDER THE GAUSSIAN PROBABILITY LAW.

BY EDWARD L. DODD.

§1. OBJECT OF PAPER.

The object of the present paper is to harmonize as much as possible the Principle of the Arithmetic Mean and the Gaussian Probability Law, viewed from the standpoint of probability *a posteriori*. By the principle of the arithmetic mean is understood the statement that, if measurements are made of a magnitude under like circumstances, the most probable value of the magnitude is the arithmetic mean or "average" of the measurements. The Gaussian law will be explained later.

That a lack of harmony exists has been known for a long time. * But experience has seemed to substantiate both principle and law to such a high degree that an analysis of the discrepancy between the two has not invited the serious attention of many mathematicians.

In fact, a casual reader of many books on Least Squares or the Theory of Measurements might get the idea that the principle could be deduced from the law or vice versa. It is not the intention here to condemn the suppression of intricate details in an elementary text-book. But the separate presentation of the principle and the law appears preferable to an attempt to unite the two by pseudo-logic.

Assuming the validity of the Gaussian Law, I have compared the arithmetic mean with several functions of the measurements.† These comparisons show the arithmetic mean superior

*Bertrand: Calcul des Probabilités, Paris (1889), p. 180. "La règle des moyennes, il importe d'insister sur ce point, n'est ni demontrée ni exacte."

†The probability of the arithmetic mean compared with that of certain other functions of the measurements; Annals of Mathematics, June, 1913, pp. 186-198.

The error-risk of certain functions of the measurements; Monatshefte fuer Mathematik und Physik, XXIV Yahrgang, 1913, pp. 268-276.

The error-risk of the median compared with that of the arithmetic mean; Bulletin of the University of Texas, No. 323, March 15, 1914.

to most of the functions considered, but not to all. To conduct these comparisons, direct probability was used. In this paper, on the contrary, probability *a posteriori* will be used.

Poincarè discusses * the relation between the law and the principle, making use of probability *a posteriori* and the so-called "probability of causes." He finds that the "proof" given by Gauss involves an assumption that the *a priori* probability is a constant.

This assumption, while it may seem somewhat unwarranted, appears at first sight as the most simple and perhaps the most reasonable assumption to make. But the assumption happens to be at variance † with one of the most fundamental principles of the theory of probability; viz., that the sum of the probabilities of the possible events shall be equal to *one* (unity), the symbol for certainty. On that ground we can not entertain such an assumption.

A natural course is then to try to modify this assumption and to deal with an *a priori* probability which is *nearly* constant, taking care to give this word "nearly" some mathematical precision. Again, it is natural to enquire if there are other assumptions concerning the *a priori* probability which will appeal to us as in any sense reasonable. And in view of the intimate relation which seems to exist between the Gaussian law and the arithmetic mean, it is natural to try to combine with the Gaussian law some postulates, as broad as possible, concerning the *a priori* probability, which will lead logically to the arithmetic mean as at least a very natural and close approximation to the most probable value of the unknown. *The object of this paper is, then, to investigate the nature of the a priori probability which permits a close relation between the Gaussian probability law and the principle of the arithmetic mean.*

The meaning of probability *a priori* in a problem of this kind is a matter for reflection. I shall not attempt to define it. We are forced continually to deal with problems in which the ultimate concepts are undefined and perhaps susceptible of considerable latitude of interpretation. In geometry we deal with

*Calcul des Probabilités (1912), p. 169.

†Bulletin of the American Mathematical Society, June, 1913, pp. 479-482.

the straight line. But what is a straight line? We may try to shift the difficulty of defining a straight line to algebra and make use of the linear function; but this does not define a straight line as geometrically conceived. Again a "stretched string" is a good description; but it is not a definition.

The notion of probability *a priori* will be developed in descriptive fashion in the section which follows, and certain definitions and postulates given, also certain fallacies mentioned. Following this will be the formal statement and proof of four theorems, involving various hypotheses concerning the *a priori* probability. This will be followed by a section on defective hypotheses, those which are inadequate to bring the Gaussian law and the arithmetic mean into close relation. And a short discussion will follow this.

§1. INTRODUCTORY CONCEPTS AND POSTULATES.

A physicist learns that a meter-rod, that he has ordered, has been shipped to him. Supposing for the sake of simplicity that a meter is just 39.37 inches, what is the probability that the rod will be 39.37 inches long? What is the probability that it will be 39.38 inches long? Or 39.35 inches long? This problem may illustrate in a general way how an *a priori* probability may be conceived. The physicist has made no measurement of this rod; in fact, he has not even seen it. His order may have been misunderstood, and something altogether wrong may have been sent him. Nevertheless, it would be generally admitted that he would be more likely to receive a rod between 39.30 and 39.40 inches in length than a rod between 40.00 and 40.10 inches. And it is natural to attempt to get an expression for this *a priori* probability, rough though the approximation may be.

It frequently happens that some hypothesis about the *a priori* probability seems almost necessary, to make a start in certain problems. But fortunately the influence of this probability often becomes ultimately negligible; so that, even though it has been poorly represented, the harm done is of a vanishing nature.

Poincarè * introduces an essentially unknown function to represent a certain probability in a problem on the roulette wheel

*Poincaré, loc. cit., pp 148-152; see also p. 277.

and in a problem on the distribution of planets; and the influence of the function is practically nil.

In view of the fact that the distribution of errors of measurements and the deviations from the normal in biological observations follow with more or less approximation the Gaussian probability law, it is not unnatural to assume that the *a priori* probability may likewise be approximately Gaussian; for example, to assume that the probability *a priori* that the length of the meter-bar lies between α and β is

$$(1) \qquad p = \frac{k}{\sqrt{\pi}} \int_{\alpha}^{\beta} e^{-k^2 (a-z)^2} dz,$$

in which e=2.718..., a=39.37, and k, the measure of precision, depends upon the reputation of the firm for accuracy in construction. The Gaussian law, as is well known, makes large errors less likely than small errors, and very large errors well nigh impossible.

Many authors favor the use of a constant *a priori* probability. In the present problem, however, it is obvious that the probability that a rod between 1000 and 1001 inches long will be sent is not as great as the probability that a rod between 39 and 40 inches will be sent. It would be rash to assert that the present problem is typical of all problems that arise. But the use of a constant *a priori* probability in certain cases is highly objectionable, especially when it refers to an unknown magnitude for which all real numbers are assumed possible. To show this, let the probability *a priori* that the unknown true value lies between α and β be the integral of $\Psi(z)dz$ from α to β. Now the symbol for certainty is unity. And so, as the unknown certainly lies between $-\infty$ and $+\infty$,

$$(2) \qquad \int_{-\infty}^{+\infty} \Psi(z) dz = 1.$$

As I have already * pointed out, this equation can not be satisfied if $\Psi(z)$ is a constant. The failure to recognize this, lies

*Bulletin of the American Mathematical Society, loc. cit.

at the basis of a fallacious deduction of the Gaussian law from the so-called principle of the arithmetic mean. The same fallacy underlies an argument for the reverse process, attempting to get the principle of the arithmetic mean from the Gaussian law.

A still more objectionable fallacy, presented to accomplish this end, consists in confusing two distinct probabilities. The expression,

$$(3) \qquad \Phi(z) = \left(\frac{h}{\sqrt{\pi}}\right)^n e^{-h^2[(z-m_1)^2+\ldots+(z-m_n)^2]}$$

is first set up as the probability that if z is the true value the measurements $m, m_2, \ldots m_n$ *will be made;* and then the attempt is made to regard this expression (3) as the probability that *z is the true value,* the measurements having been made. Then by setting the first derivative equal to zero, the result,

$$(4) \qquad z = \frac{m_1+m_2+\ldots+m_n}{n}$$

is obtained; and it is asserted that the average or arithmetic mean is the most probable value of the unknown true value.

It is the object of this paper to arrive at the conclusion that the average is *approximately* the most probable value by assuming that $\Psi(z)$ satisfies the requirement (2) and certain other natural conditions of a general nature, somewhat analogous to the conditions placed by Poincaré upon the arbitrary functions in his roulette and planet problems. Different hypotheses will be made for $\Psi(z)$. It may seem that the only condition needed in addition to (2) is that $\Psi(z)$ be continuous, so that it would be *practically constant in small intervals.* But, as will be shown, *this condition in no wise guarantees that the arithmetic mean will even approximate the most probable value.*

We do not here undertake to define *a priori* probability or indeed any kind of probability. A useful description of a probability may be an *ideal frequency.* In the preceding illustration, perhaps the firm has the reputation of making its meter-rods a trifle too long. In place of *a* in (1), the physicist may put his guess, *g,* which may or may not be 39.37. In thinking of p in (1) as an ideal frequency, we may have in mind that if upon the *n* occasions the physicist guesses that the length of the coming rod will be g, then in about pn cases it is to be expected that the length of the rod will be between α and β.

Concerning each measurement, $m_1, m_2 \ldots m_n$ it will be assumed that it is subject to the Gaussian law, with measure of precision, h. That is, if a is the true value, and $x=a-m$, the probability that the error of a measurement to be made will lie between x_1 and x_2 is

$$(5) \qquad \frac{h}{\sqrt{\pi}}\int_{x_1}^{x_2} e^{-h^2x^2}\, dx.$$

Or, stated in the differential form, * the probability that the error x will be made,—that is, an error between x and $x+dx$,—is

$$(6) \qquad \frac{h}{\sqrt{\pi}} e^{-h^2x^2} dx.$$

Then the probability that the n errors, $x_1, x_2 \ldots x_n$ *will be* made is

$$(7) \qquad \left(\frac{h}{\sqrt{\pi}}\right)^n e^{-h^2[x_1^2+x_2^2+\ldots+x_n^2]} dx_1 \ldots dx_n$$

This resembles $\Phi(z)$ in (3) somewhat: but in (7), x_1 is the *actual error*, $a-m_1$; whereas, in (3), $z-m_1$ is merely a *residual* with respect to z, where z is a candidate for recognition as the true value.

Now let $\Psi(z)\,dz$ be the probability *a priori* that z is the true value, and let

$$(8) \qquad F(z)=\Psi(z)\Phi(z).$$

Then, by Bayes' theorem, the probability *a posteriori* that z is the unknown true value, after the measurements *have been* made, is

$$(9) \qquad \frac{1}{c}F\ z)\,dz,$$

where c is the integral of $F(z)\,dz$ from $-\infty$ to $+\infty$. We seek

*For the sake of brevity, the differential form will be sometimes used. From an inspection of Theorem III or Theorem IV the reader will be able to restate the first two theorems in the more cumbrous but more precise form.

the value Z of z which will make $F(z)$ the maximum. Then, *a posteriori*, the most probable value of the unknown is Z. By this is meant that it will be more probable *a posteriori* that the true value will differ from Z by less than a small ϵ than that the true value will differ from any other real number by less than ϵ.

§2. FIRST HYPOTHESIS CONCERNING THE A PRIORI PROBABILITY.

Let g be a guess at the true value, and let the *a priori* probability function be

$$\Psi(z)=\frac{k}{\sqrt{\pi}}e^{-k^2(g-z)^2}. \tag{10}$$

Then from (8),

$$F(z)=\frac{k}{\sqrt{\pi}}\left(\frac{h}{\sqrt{\pi}}\right)^n e^{-[k^2(g-z)^2+h^2\Sigma(z-m)^2]}.$$

Setting the first derivative equal to zero and solving gives the value Z of z, making $F(z)$ the maximum.

$$Z=\frac{k^2g+h^2(m_1+\ldots+m_n)}{k^2+nh^2} \tag{11}$$

A maximum actually occurs here; since the bracket above is a quadratic with the coefficient of z^2 positive. Let M designate the arithmetic mean or average of the measurements; and divide the numerator and denominator in (11) by nh^2. Then

$$Z=\frac{M+\eta_1}{1+\eta_2}=(M+\eta_1)(1+\omega)=M(1+\omega)+\sigma, \tag{12}$$

where η_1, η_2, ω and σ approach zero with increasing n. In fact, if it be postulated that M does not increase indefinitely in numerical value—and in practice this is usually the case—then $M\omega$ is an infinitesimal; and thus

$$Z=M+\eta, \tag{13}$$

where η approaches zero with increasing n.

Theorem I. Let the probability a priori that the unknown true value is z be *

$$\frac{k}{\sqrt{\pi}} e^{-k^2(g-z)^2} dz,$$

where g is any guess at the unknown. Let the probability that the error of a measurement to be made will be x be

$$\frac{h}{\sqrt{\pi}} e^{-h^2x^2} dx.$$

Then, a posteriori, after n measurements with arithmetic mean M have been made, the most probable value of the unknown is

$$M(1+\omega)+\sigma,$$

where ω and σ approach zero with increasing n.

Discussion. If the tangent of an angle near 90° is being measured, the condition for (13) may not be satisfied. To illustrate further by an example, suppose that the measurements turn out to be the odd integers in natural order: $m_1=1$, $m_2=3$, $m_3=5$, etc. Then in (11), $m_1+m_2+ \ldots +m_n=n^2$; and $M=n$. Suppose furthermore that $g=-1$, and that $h=1=k$. Then

$$Z=\frac{-1+n^2}{1+n}=n-1=M-1.$$

Thus $M-Z=1$; and so this difference is not an infinitesimal.

In general, k and h in (11) are not equal. In place of the guess g in (10) a perliminary measurement m may be made; and thus m would replace g in (10). This may seem at first sight to make k equal to h, so that by (11) Z would become the exact average of the $(n+1)$ measurements. But even with a measurement m there is no justification for making k in (10) equal to h. The probability that *z is the true value* after a measurement m has been made, is entirely distinct from the probability that *the measurement m will be made* if z is the true value. The failure to recognize the distinction between a "probability of cause" and a direct probability has been the source of many fallacies. The distinction can be brought out

*See (5) and (6) and coresponding foot-note.

clearly by urn problems where the exact probability can be computed, under certain hypotheses. The probability that an *urn contained* two white balls and two black balls if a white ball and a black ball have been drawn, is quite different from the probability that from an urn containing two white balls and two black balls a white ball and a black ball *will be* drawn.

§3. SECOND HYPOTHESIS CONCERNING THE A PRIORI PROBABILITY.

By differentiating $F(z)$ in (8), we obtain

$$(14) \qquad F'(z)=\Phi(z)\left[\Psi'(z)+\Psi(z)\left\{-2h^2(nz-\Sigma m)\right\}\right]$$

Let $f(z)$ be the function obtained by dividing the bracket in (14) by $-2h^2n\Psi(z)$. Then

$$(15) \qquad f(z)=z-M-\frac{\Psi'(z)}{2h^2n\Psi(z)}.$$

Now $F'(z)=0$ provided $f(z)=0$. To make $f(z)$ vanish when z is nearly M, we naturally impose some condition to make the last term in (15) negligible with increasing n.

THEOREM II. *Let the probability * a priori that the unknown true value is z be $\Psi(z)dz$,* where

$$|\Psi'(z)|<K\Psi(z)$$

for all values of z, K being some constant. Let the probability that the error of a measurement to be made will be x be

$$\frac{h}{\sqrt{\pi}}e^{-h^2x^2}\,dx.$$

Then, a posteriori, after n measurements with arithmetic mean M have been made, the most probable value of the unknown is

$$M+\sigma,$$

where σ approaches zero with increasing n.

The hypothesis excludes the possibility of $\Psi'(z)$ becoming infinite; and thus $\Psi(z)$ is continuous. Furthermore by hypoth-

*It is further assumed in all the theorems of this paper that $\Psi(z)$ must satisfy (2) to properly represent probability; and that $\Psi(z)$ is not negative, for we do not recognize negative probability.

esis $\Psi(z)$ can not become zero; thus the denominator in (15) is not zero.

The function given in (10) does not satisfy the requirements of Theorem II; since

$$\Psi'(z) = [2k^2(g-z)]\Psi(z). \tag{16}$$

There are, however, simple functions which satisfy these requirements and which have a graph closely resembling the common probability curve—see (6) and (10)—and which satisfy (2) For example:

$$\Psi(z) = \frac{1}{\pi(1+l^2z^2)}, \quad l > o. \tag{17}$$

§4 Third Hypothesis Concerning the A Priori Probability.

In the theorem about to be stated, the special condition to be placed upon $\Psi'(z)$ is suggested by (16). The differential form of statement will now be laid aside.

Theorem III. *Let the probability a priori that the true value lies between* a *and* $a+\delta$ *be*

$$\int_a^{a+\delta} \Psi(z)\,dz,$$

where $\Psi(z)$ *is positive, except perhaps at isolated points; and*

$$|\Psi'(z)| \leqq [c_1|z| + c_2]\Psi(z), \tag{18}$$

c_1 *and* c_2 *being positive constants. Let the probability that the error of a measurement to be made will lie between* x_1 *and* x_2 *be*

$$\frac{h}{\sqrt{\pi}} \int_{x_1}^{x_2} e^{-h^2x^2}\,dx$$

Then after n *measurements with arithmetic mean* M *have been made, the probability a posteriori that the true value lies between* a *and* $a+\delta$ *is greatest when*

$$a = M(1+\omega) + \sigma, \tag{19}$$

where $\lim_{n=\infty,\ \delta=o} \omega = o$ *and* $\lim_{n=\infty,\ \delta=o} \sigma = o$ *uniformly.*

Proof. By Bayes' theorem, the probability *a posteriori* that the true value lies between a and $a+\delta$ is

$$(20)\qquad P=\frac{1}{c}\int_{a}^{a+\delta} F(z)\,dz.$$

The abridged form of this statement was given in (9) For brevity set

$$(21)\qquad \epsilon_1=\frac{c_1}{2h^2n}\quad,\quad \epsilon_2=\frac{c_2}{2h^2n}$$

These approach zero with increasing n.

First, suppose that $\Psi(z)>0$. Then from (15), $f(z)<0$ provided

$$(22)\qquad z-M+\epsilon_1|z|+\epsilon_2<0;$$

because of (18) and (21). This may be written

$$(23)\qquad z+\epsilon_1|z|<M-\epsilon_2.$$

Now if $M>0$ and the ϵ's are small, this will be satisfied if

$$(24)\qquad z<\frac{M-\epsilon_2}{1+\epsilon_1};$$

whereas if $M\leq 0$, (23) will be satisfied if

$$(25)\qquad z<\frac{M-\epsilon_2}{1-\epsilon_1}$$

In place of (24) and (25) we may write simply

$$(26)\qquad z<(M-\epsilon_2)(1+\epsilon_3),$$

where ϵ_3 is an infinitesimal. Hence if z satisfies (26), $f(z)$ will be negative, $F'(z)$ will be positive by (3), (14) and (15), and $F(z)$ will be an increasing function of z. Hence P in (20) can not take its largest value, for a given δ, if

$$(27)\qquad a<(M-\epsilon_2)(1+\epsilon_3)-\delta.$$

This δ can be made as small as we please. Likewise it can be shown that P can not take its largest value when

(28) $$\alpha > (M+\epsilon_2)(1+\epsilon_4),$$

where ϵ_4 is an infinitesimal. But with δ fixed, P is a continuous function of α and hence takes on its maximum. Hence (19) follows from (27) and (28). Here ω and σ approach zero *uniformly;* for from (21), the rapidity with which ϵ_1 and ϵ_2 approach zero does not depend upon the magnitude of *M*. In Theorems I. and II., ω and σ likewise approach zero uniformly.

Suppose now that $\Psi(z)=0$ at some isolated points. Then by (18) and (14) $F'(z)=0$ at these isolated points. This will not affect the character of F(z) as an increasing function or as a decreasing function.

To see that (18) is not a superfluous condition let

(29) $$\Psi(z)=Ce^{-|z^3|},$$

where C is chosen to satisfy (2). The graph of (29) is a curve symmetrical with respect to the "*Y* axis," has its maximum at z=0, has just one point of inflection on each side of the *Y axis*, and otherwise resembles the usual probability curve (10), with g=0. Then by (8), if h=1,

(30) $$F(z)=C\left(\frac{1}{\sqrt{\pi}}\right)^n e^{-z^3-\Sigma(z-m)^2}$$

If now the measurements turn out to be the odd integers, 1, 3, 5, ... then $\Sigma m=n^2$, and F(z) takes its maximum when $z=(1/3)M(\sqrt{7}-1)$. Thus (19) is not satisfied; nor is (18).

It will be noticed that Theorem I. is a special case of Theorem III.; but Theorem I. was given first, because of simplicity of development.

§5 Fourth Hypothesis Concerning the A Priori Probability.

There are certain cases in which measurements must lie between two constants. If we accept the most elementary conception of an angle, the angle must lie between 0° and 180° The tangent of this angle, however, may have any real value positive or negative. But even if we are measuring the tangent of an angle, there would usually be an interval, from b_1 to b_2, in which the average M would in practice be. To postulate that M must lie in (b_1, b_2) would be *contradictory to the Gaussian law.* And so the theorem to be given applies in strictness to the case where M *does* lie in (b_1, b_2), rather than to the case where M must lie in (b_1, b_2).

THEOREM IV. *Let the probability a priori that the true value lies between* a *and* $a+\delta$ *be*

$$\int_a^{a+\delta} \Psi(z)dz,$$

where $\Psi(z)$ *is limited for all values of* z*; and has a positive minimum in some interval* (b_1, b_2), *or at least in that part of* $(b_1+\epsilon, b_2-\epsilon)$ *which remains when a finite number of sub-intervals of the form* $(\xi-\epsilon, \xi+\epsilon)$ *are removed, with* ϵ *small at pleasure. Let the probability that the error of a measurement to be made will lie between* x_1 *and* x_2 *be*

$$\frac{h}{\sqrt{\pi}}\int_{x_1}^{x_2} e^{-h^2x^2}dx$$

Then after n *measurements with arithmetic mean* M *have been made, the probability a posteriori that the true value lies between* a *and* $a+\delta$ *is greatest when*

$$a=M+\sigma$$

where $\lim\limits_{n=\infty,\ \delta=o} \sigma=o$ *uniformly in* (b_1, b_2) *provided that* M *continues to lie in* (b_1, b_2).

Proof. Let $v_1, v_2, \ldots v_n$ be the residuals of the measurements with respect to M; that is, let $v_1=M-m_1$, etc. Then $z-m_1=z-M+v_1$, etc.; and, since $\Sigma v=o$, (3) becomes

$$\Phi(z)=\left(\frac{h}{\sqrt{\pi}}\right)^n e^{-h^2\Sigma v^2}\left[e^{-nh^2(z-M)^2}\right] \tag{31}$$

This shows, even more simply than (3), that $\Phi(z)$ takes its maximum when $z=M$, and is an increasing function when $z<M$ and a decreasing function when $z>M$. We wish to show that, under the conditions of the hypothesis, $\Phi(z)$ can, when n is large enough, force its own point of maximum M upon the product $F(z)$ in (8) and so upon P in (20), to as close an approximation as we please. From (31) it follows that

$$(32)\qquad \frac{\Phi(z+\frac{\epsilon}{2})}{\Phi(z+\frac{\epsilon}{4})}=e^{-nh^2\left[\frac{\epsilon}{2}(z-M)+\frac{3\epsilon^2}{16}\right]}.$$

In particular,

$$(33)\qquad \frac{\Phi\left(M+\frac{\epsilon}{2}\right)}{\Phi\left(M+\frac{\epsilon}{4}\right)}=e^{\frac{-3nh^2\epsilon^2}{16}}.$$

Now, by hypothesis, $\Psi(z)$ is limited; that is, there is a constant K such that $\Psi(z)<K$. By hypothesis also, if z lies in the interval from $b_1+\frac{\epsilon}{2}$ to $b_2-\frac{\epsilon}{2}$ with sub-intervals of the form $(\xi-\frac{\epsilon}{2}, \xi+\frac{\epsilon}{2})$ removed, there is a positive number T such that $\Psi(z)>T$.

Suppose, first, that M differs from b_1, b_2 and every ξ by at least ϵ. Now take n in (33) large enough so that

$$(34)\qquad \frac{\Phi\left(M+\frac{\epsilon}{2}\right)}{\Phi\left(M+\frac{\epsilon}{4}\right)}<\frac{T}{K}\ ,\quad \frac{\Phi\left(M-\frac{\epsilon}{2}\right)}{\Phi\left(M-\frac{\epsilon}{4}\right)}<\frac{T}{K}\ .$$

Then

$$(35)\qquad K\Phi\left(M-\frac{\epsilon}{2}\right)<T\Phi\left(M-\frac{\epsilon}{4}\right).$$

But Φ (z) is an increasing function when $z<M$. Hence by (35, if z_1 is any value of $z<M-\frac{\epsilon}{2}$ and z_2 is any value of z between $M-\frac{\epsilon}{4}$ and M, then

$$\Psi\left(z_1\right)\Phi\left(z_1\right)<K\Phi\left(M-\frac{\epsilon}{2}\right)<T\Phi\left(M-\frac{\epsilon}{4}\right)<\Psi\left(z_2\right)\Phi\left(z_2\right).$$

Thus $F(z_1)<F(z_2)$. Hence, if we take any particular $\delta<\frac{\epsilon}{4}$, the maximum value of P in (20) can not occur when $a<M-\epsilon$. Likewise from (34) it follows that the maximum of P can not occure when $a>M+\epsilon$. But with a chosen δ, the integral P is a continuous function of a; and hence P takes on its largest value when a lies between $M-\epsilon$ and $M+\epsilon$.

If, in particular, $M = \xi + \epsilon$, then by (34),

$$\frac{\Phi\left(\xi + \epsilon + \frac{\epsilon}{2}\right)}{\Phi\left(\xi + \epsilon + \frac{\epsilon}{4}\right)} < \frac{T}{K} \tag{36}$$

But from (32), whatever be the value of M,

$$\frac{\Phi\left(\xi + \epsilon + \frac{\epsilon}{2}\right)}{\Psi\left(\xi + \epsilon + \frac{\epsilon}{4}\right)} = e^{-nh^2\left[\frac{\epsilon}{2}\left(\xi + \epsilon - M\right) + \frac{3}{16}\epsilon^2\right]},$$

and decreases with decreasing M. Hence (34), being satisfied when $M = \xi + \epsilon$, is satisfied *a fortiori*, when $M < \xi + \epsilon$. Hence, when $M < \xi + \epsilon$, and n and δ are chosen as specified above, P can not take its maximum when $a > \xi + 2\epsilon$. And likewise when $M > \xi - \epsilon$, P can not take its maximum when $a < \xi - 2\epsilon$. Thus if M falls between $\xi - \epsilon$ and $\xi + \epsilon$, P takes its maximum when a lies between $\xi - 2\epsilon$ and $\xi + 2\epsilon$; and $|a - M| < 3\epsilon$. The reasoning is analogus if M falls near b_1 or b_2.

§6. Four Defective Hypotheses Concerning the A Priori Probability.

Four hypotheses will now be mentioned which are untenable or artificial or inadequate.

1. Each real number is equally likely *a priori* to be the true value.

This makes $\Psi(z)$ a constant, and (2) can not be satisfied.

2. Each real number in a certain interval $(b_1\ b_2)$ is *a priori* equally likely to be the true value, and it is impossible for the true value to lie outside this interval.

I have given an example * for which this hypothesis is natural. But in the general case it appears artificial to postulate that the *a priori* probability drops suddenly to zero at the ends of an interval, when these ends can be at best only hazily imagined.

Even if we adopt this hypothesis, it would not follow that the arithmetic mean M is the most probable value of the unknown, without the addition proviso that M lies in (b_1, b_2). For if $\Psi(z) = 0$ outside (b_1, b_2), then by (8) the probability *a pos-*

*Bulletin of the American Mathematical Society, loc. cit., p. 481.

teriori that the true value lies outside (b_1, b_2) is also zero. Thus the most probable value of the unknown could not be outside (b_1, b_2); whereas the Gaussian law permits M to have *any value whatever*.

3. The *a priori* probability is practically constant in small intervals,—or, as we may wish to express it, $\Psi(z)$ is a continuous function of z,—and all real numbers are possible values of the unknown true value.

These conditions are satisfied by $\Psi(z)$ in (29); and hence are inadequate.

4. The *a priori* probability is continuous, and is zero outside a certain interval.

The supposition that $\Psi(z)=0$ outside (b_1, b_2) leads to the defect mentioned under No. 2.

§7. Four Types of A Priori Probability.

In a given case there may be a strong probability that the arithmetic mean M will not increase indefinitely, even though something like the tangent of an angle is being measured. Nevertheless instruments may be subject to a progressive change due to a change in temperature. Or, indeed, an increasing set of measurements may be the result of mere chance as we usually understand the term. If a gambler loses in one night as much money as he has won previously in a year, he may well suspect that the dice are loaded. And an experimenter may well suspect that his instruments are suffering from some ailment, if his measurements persist in increasing. But in both cases it may be simply a *run of bad luck.*

It may seldom be clear just what type of *a priori* probability to postulate; but the theorems just given and subsequent discussion permit us to distinguish four types of *a priori* probability in accordance with the ease with which this probability allows itself to be eliminated when M increases indefinitely, thus departing indefinitely from any value which *a priori* may be the most probable value of the unknown.

1. The weakest $\Psi(z)$, that has been mentioned as permissible, is given by (17). This has no power of resistance, and its influence is evanescent with increasing n.

2. The function (10) has a greater power of resistance. An example was given after Theorem I., in which $g=-1$ and $Z=M-1$. The function (10) can clip off a constant—in the example, unity—from the arithmetic mean, before this is presentable as the most probable value of the unknown.

3. The function (29) is still stronger. In the example given, it permits a number only about 55% of the arithmetic mean to come forth as the most probable value. But even this function is not absolutely prohibitive of a large most probable value. In spite of its strong preference for zero as the value of the unkonwn, it acknowledges the possibility of any value; and the persistent increase of M forces up the most probable value.

4. A function absolutely prohibitive of a most probable value outside an interval (b_1, b_2) can be formed by making $\Psi(z)=0$ outside (b_1, b_2). This follows from (8), (9), and (20).

As long as M remains in some interval (b_1, b_2), both functions (10) and (29) and in general the $\Psi(z)$ just mentioned exert a vanishing influence upon the most probable value, by Theorem IV.

With large values of M, it is easy to see why (29) should be more stubborn than (10), since when $z>k^2$,

$$e^{-z^3} < e^{-k^2z^2}$$

and diminishes much more rapidly as z increases.

§8 Discussion of the Principle of the Arithmetic Mean.

The four theorems of this paper present the arithmetic mean as a natural approximation for the most probable value, from the standpoint of probability *a posteriori,* under certain broad conditions. It has not been proved that the arithmetic mean is the best approximation. Actual comparison of the arithmetic mean with the median and other functions of the measurements from the standpoint of probability *a posteriori* may perhaps be made; but these comparisons conditioned by an unknown *a priori* probability would seem less conclusive than comparisons made from the standpoint of direct probability.

One of the strongest rivals of the arithmetic mean just now is the median; because of its frequent use in statistical work. The median is more easily ascertained than the arithmetic mean. We have simply to arrange the measurements in the order of their magnitude and pick out the middle one,—the number of measurements being odd. If X is an approximation for the true value; and $\Sigma(X-m)^2$ is a minimum, then X is the arithmetic mean; whereas if $\Sigma|X-m|$ is a minimum, then X is the median. The median has some of the same characteristics as the arithmetic mean. Under the Gaussian law, the probability that the median of n measurements will differ from the true value by less than any preassigned positive ϵ, approaches unity (certainty) when the number of measurements increases indefinitely.* *Furthermore, it has not been substantiated empirically that the arithmetic mean is better than the median and other functions of measurements that have been used.*

When the theory of probability was in its infancy, the arithmetic mean had the support of tradition; and it was natural enough for mathematicians to try to inject into the theory of probability the postulate that the arithmetic mean is the most probable value. With the growth of the theory of probability—based largely upon the careful study of distributions suggested by games of chance—the postulate that the arithmetic mean is the most probable value comes to seem more and more gratuitous. To set up such a postulate now appears to be putting the cart before the horse. Much better it is to try to use the information gathered from the study of probability to discover what is the best function, if there be a best function.

It has not been the object of this paper to compare functions nor to enthrone the arithmetic mean. But the object has been to show that the arithmetic mean has at least the distinction of issuing from the Gaussian law as a very *natural* approximation for the most probable value *a posteriori*, under broad and tenable hypotheses concerning the *a priori* probability.

---

*This may be proved by using equation (14), Annals of Mathematics, loc. cit., p. 195, and showing that the probability that the median will differ from the true value by more than ϵ approaches zero.