On Gauss's proof of the normal law of errors. By Mr HAROLD JEFFREYS, St John's College.

[*Received* 3 February, *read* 6 March 1933.]

Gauss gave a well-known proof that under certain conditions the postulate that the arithmetic mean of a number of measures is the most probable estimate of the true value, given the observations, implies the normal law of error. I found recently that in an important practical case the mean is the most probable value, although the normal law does not hold*. I suggested an explanation of the apparent discrepancy, but it does not seem to be the true one in the case under consideration.

If we take the true value to be x and the probability of an observed value between y and $y + dy$ to be $\phi(y - x)\,dy$, Gauss assumes the probability, given x, of a set of observed values between x_1 and $x_1 + dx_1$, x_2 and $x_2 + dx_2$, ... to be

$$\phi(x_1 - x)\,\phi(x_2 - x)\dots\phi(x_n - x)\,dx_1 dx_2\dots dx_n, \tag{1}$$

and infers that, if the prior probability of x is uniformly distributed, the most probable value of x is given by

$$\frac{\phi'(x_1 - x)}{\phi(x_1 - x)} + \dots + \frac{\phi'(x_n - x)}{\phi(x_n - x)} = 0. \tag{2}$$

According to the postulate of the arithmetic mean this must be equivalent to

$$(x_1 - x) + \dots + (x_n - x) = 0, \tag{3}$$

whence, provided that the numbers x_1 to x_n are capable of taking all values, Gauss infers that each of the fractions in (2) is proportional to its argument, and proceeds to the normal law by integration.

In a case which I considered, the interval to be measured is found by reading both ends to the nearest multiple of the finite step of a measuring instrument. If the true value is $n + x$, where $0 \leqslant x \leqslant 1$, the probability of making the measure n is found to be $1 - x$, and that of the measure $n + 1$ is x. The probability in m trials of l times making the measure n and $m - l$ times making the measure $n + 1$ is

$${}^mC_l\,(1 - x)^l\,x^{m-l}, \tag{4}$$

and choosing x to make this a maximum easily gives

$$x = \frac{m - l}{m}. \tag{5}$$

* *Scientific Inference*, 70.

Thus the most probable value on the data is the arithmetic mean, although the probability of the measures is not distributed according to the normal law.

It may be noticed that (1) is the probability, given x and the law, of obtaining the observations *in the given order*. If we are considering the probability of obtaining them at all, in *any* order, (1) must be multiplied by $n!$. The corresponding condition is assumed in (4), in which the factor ${}^mC_l$ should be dropped if we are considering a particular order. But since the factors do not involve x they do not affect either argument.

Let us then assume that two observed values are possible; they may without loss of generality be taken to be n and $n+1$. We can now repeat Gauss's argument. For a given x the probability of reading n is taken to be $\phi(x)$, and that of reading $n+1$ is $1-\phi(x)$. Then the probability, given x, of obtaining the measures in their actual order is

$$\{\phi(x)\}^l \{1-\phi(x)\}^{m-l}, \tag{6}$$

which is a maximum if

$$\frac{l\phi'(x)}{\phi(x)} - \frac{(m-l)\phi'(x)}{1-\phi(x)} = 0. \tag{7}$$

This is to be equivalent to the mean value postulate

$$lx - (m-l)(1-x) = 0 \tag{8}$$

for all values of m, l and x. Neglecting the case of $\phi(x)$ constant, which leads to no preference between different values of x, we have

$$x\phi(x) = (1-x)\{1-\phi(x)\}, \tag{9}$$

whence

$$\phi(x) = 1-x. \tag{10}$$

This agrees with the form already given, and shows it to be unique.

It appears therefore that the difference between Gauss's result and mine arises from the fact that in (2) the observations are supposed capable of an indefinite range of values. In my problem only two observed values are possible, this being the condition that arises in the type of measurement there dealt with.

Keynes* has shown that, if the probability density of an observation is a function not only of the magnitude of the error, but of the true value itself, the postulate of the arithmetic mean leads to the law

$$\phi(y-x, x) = \exp\{f'(x)(x-y) - f(x) + g(y)\},$$

* Keynes, *Treatise on Probability*, 197.

where f and g are two arbitrary functions. This rests on an argument similar to Gauss's. Mr M. S. Bartlett, in an unpublished paper, points out that it can represent the present type of law if we take

$$f(x) = (x-1)\log(1-x) - x\log x; \quad g(x) = 0; \quad y = 0 \text{ or } 1.$$

My previous explanation was that there is some doubt whether the joint probability of the set of errors is the product of the probabilities of the errors independently, given the true value. It assumes that, given x, the occurrence of any observation is irrelevant to the probability of any other. This is clearly untrue in many cases; thus, if the normal law holds but the standard error σ is unknown, a large error in the nth observation is clearly more likely if the first $n-1$ observations show a large scatter than if they show a small one. In such a case we must allow for the distribution of the prior probability of σ, and the posterior probability of x is not distributed according to the probability of the observations given x. But when our estimate of the scale of the distribution of the errors is not to be affected by the deviations observed, independence may be assumed. This condition is satisfied in my problem. Thus for the applicability of (1) it is necessary that the form of the law of error shall be supposed determined by considerations external to the observations actually made; it is not necessary that it should be a *known* function of the error, but it is necessary that it should be known to be some single function and not one of a group of functions to be compared by means of the observations. Thus it would be valid if the law was

$$\phi(\xi) = \psi(\xi, h),$$

where h is a parameter initially known, but not if it is initially unknown and capable of a range of values. Given such previous knowledge, (1) holds and (2) follows from it provided the prior probability of x is uniformly distributed. Then (3) enables us to select which of such laws are consistent with the principle of the arithmetic mean. The existence of any such laws however seems to require the introduction of causal theories as in Hagen's treatment and its generalizations.

It should be mentioned that if the standard error is initially unknown, the resulting failure of (1) does not imply any error of the postulate of the mean. For if the normal law holds and the prior probability density of x and h is $f(h)$, the probability of the observed values, given x and h, is

$$\left(\frac{h}{\sqrt{\pi}}\right)^n \exp\{-h^2 \Sigma (x_r - x)^2\} \, \Pi (dx_r),$$

and the posterior probability density of x and h, given the observations, is proportional to this quantity multiplied by $f(h)$. But this is a maximum for variations of x, whatever the form of $f(h)$, if

$$\Sigma (x_r - x) = 0.$$

It is easy to see, however, that (1) in this case becomes

$$\prod_{r=1}^{n} \int_0^\infty \frac{h}{\sqrt{\pi}} f(h) \exp\{-h^2 (x_r - x)^2\} \, dh$$

$$= \int_0^\infty \left(\frac{h}{\sqrt{\pi}}\right)^n f(h) \exp\{-h^2 \sum_{r=1}^{n} (x_r - x)^2\} \, dh,$$

which is untrue.