Reproduced from "Messenger of Mathematics", v. 41: 155-160 (1912)

1

ON AN ABSOLUTE CRITERION FOR FITTING FREQUENCY CURVES.

By *R. A. Fisher*, Gonville and Caius College, Cambridge.

1. If we set ourselves the problem, in its essence one of frequent occurrence, of finding the arbitrary elements in a function of known form, which best suit a set of actual observations, we are met at the outset by an arbitrariness which appears to invalidate any results we may obtain. In the general problem of fitting a theoretical curve, either to an observed curve, or to an observed series of ordinates, it is, indeed, possible to specify a number of different standards of conformity between the observations and the theoretical curve, which definitely lead to different though mutually approximate results. This mutual approximation, though convenient in practice in that it allows a computer to make a legitimate choice of the method which is arithmetically simplest, is harmful from the theoretical standpoint as tending to obscure the practical discrepancies, and the theoretical indefiniteness which actually exist.

2. Two methods of curve fitting may first be noted, in which we shall use a sign of summation when the observations comprise a finite number of ordinates only, and an integral sign when the curve itself is observed, even though the integrals may in practice be estimated by a process of summation.

Consider f a function of known form, involving arbitrary elements $\theta_1, \theta_2, \ldots, \theta_r$ and x the abscissa; let y be the observed ordinate corresponding to a given x. Then a natural method of getting suitable values for $\theta_1, \theta_2, \ldots, \theta_r$, that is of fitting the observations, is to make $\int_{-\infty}^{+\infty} (f-y)^2 dx$ a minimum for variations of any θ; or if the ordinate is observed at finite and equal intervals of the abscissa, we should substitute $\Sigma (f-y)^2$ for the integral.

This method will obviously give a good result to the eye in cases where a good result is possible; the equations to which it gives rise are, however, often practically insoluble, a difficulty which renders the method less useful than the simplicity of its principle would suggest.

The method of moments is possibly of more value, though its arbitrary nature is more apparent. If we solve the first r equations of the type

$$\int_{-\infty}^{+\infty} f\,dx = \int_{-\infty}^{+\infty} y\,dx \qquad \text{or } \Sigma f = \Sigma y,$$

$$\int_{-\infty}^{+\infty} xf\,dx = \int_{-\infty}^{+\infty} xy\,dx \qquad \text{or } \Sigma xf = \Sigma x, \quad *$$

$$\int_{-\infty}^{+\infty} x^2 f\,dx = \int_{-\infty}^{+\infty} x^2 y\,dx, \text{ etc. or } \Sigma x^2 f = \Sigma x^2 y, \text{ etc.,}$$

we may obtain values for the r unknowns, which will give a curve to the eye about as good as that of least squares, by a method which for some purposes is found to be more convenient.

3. The first of the above methods is obviously inapplicable to frequency curves, even if we wished to accept its standard of "goodness of fit." If we suppose that the observations comprise a complete and continuous curve, an arbitrariness arises in the scaling of the abscissa line, for if ξ, any function of x, were substituted for x, the criterion would be modified. While, if a finite number of observations are grouped about a series of ordinates, there is an additional arbitrariness in choosing the positions of the ordinates and the distances between them.

For a finite number, n, of observations the method of moments really gives the equations

$$\Sigma f = n, \qquad \Sigma xf = \sum_1^n x, \qquad \Sigma x^2 f = \sum_1^n x^2, \qquad \text{etc.,}$$

against which the above objections cannot be urged; still a choice has been made without theoretical justification in selecting this set of r equations of the general form

$$\Sigma x^p f = \sum_1^n x^p.$$

But we may solve the real problem directly.

If f is an ordinate of the theoretical curve of unit area, then $p = f\delta x$ is the chance of an observation falling within the range δx; and if

$$\log P' = \sum_1^n \log p,$$

* For Σx, read Σxy.

then P' is proportional to the chance of a given set of observations occurring. The factors δx are independent of the theoretical curve, so the probability of any particular set of θ's is proportional to P, where

$$\log P = \sum_1^n \log f.$$

The most probable set of values for the θ's will make P a maximum.

If a continuous curve is observed—*e.g.*, the period during which a barometer is above any level during the year is a continuous function from which may be derived the relative frequency with which it stands at any height—we should use the expression

$$\log P = \int_{-\infty}^{\infty} y \log f \, dx.$$

4. For example, let us take the normal curve of frequency of errors

$$f = \frac{h}{\sqrt{\pi}} e^{-h^2(x-m)^2},$$

where h and m are to be determined to fit a set of n observations. Our criterion gives, neglecting a constant term,

$$\begin{aligned}\log P &= n \log h - h^2 \Sigma (x-m)^2 \\ &= n \log h - h^2 n (m-\bar{x})^2 - h^2 \Sigma (x-\bar{x})^2,\end{aligned}$$

where $n\bar{x} = \Sigma x$.

Differentiating with respect to m, we get

$$-2h^2 n (m-\bar{x}) = 0,$$

and with respect to h

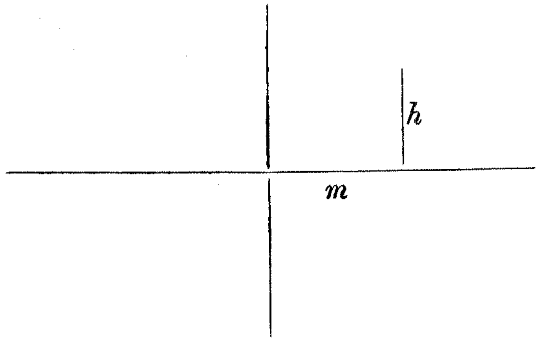
$$\frac{n}{h} = 2h \{n (m-\bar{x})^2 + \Sigma (x-\bar{x})^2\};$$

giving $m = \bar{x}$ $\qquad 2h^2 = \dfrac{n}{\Sigma v^2},$

where v is written for $x - \bar{x}$; neglecting the solution $h = 0$, $m = \infty$, when P is a minimum. Since the value usually accepted is

$$2h^2 = \frac{n-1}{\Sigma v^2},$$

it will be necessary to examine one or two of the methods by which this answer is obtained.



5. Corresponding to any pair of values, m and h, we can find the value of P, and the inverse probability system may be represented by the surface traced out by a point at a height P above the point on a plane, of which m and h are the coordinates.

The actual maximum of P occurs, as we have shown, at the point

$$m = \bar{x},$$

$$2h^2 = \frac{n}{\Sigma v^2}.$$

(a) In an interesting investigation* Mr. T. L. Bennett takes the maximum value of

$$\int_{-\infty}^{+\infty} P\,dm,$$

for variations of h, *i.e.*, of

$$h^n e^{-h^2\Sigma(x-\bar{x})^2} \int_{-\infty}^{+\infty} e^{-h^2 n(m-\bar{x})^2}\,dm,$$

or of

$$\frac{\sqrt{\pi}}{h\sqrt{n}} h^n e^{-h^2\Sigma v^2},$$

whence

$$(n-1)\,h^{n-2} = 2h^n \Sigma v^2$$

$$2h^2 = \frac{n-1}{\Sigma v^2},$$

a determination which gives the section perpendicular to the axis of h, the area of which is a maximum, though it does not pass through the actual maximum point.

* *Errors of Observation*, Technical Lecture, No. 4, 1907-08, Survey Department, Egypt.

We shall see (in § 6) that the integration with respect to m is illegitimate and has no definite meaning with respect to inverse probability.

(*b*) The usual text-book discussion* of the relation between h^2 and μ^2, where $n\mu^2 = \Sigma v^2$, assumes that the observed value of μ^2 is the same as the average value for a large number of sets of n observations each; thus the average value of $(x-m)^2$ being $\frac{1}{2h^2}$, the average value of $(\bar{x}-m)^2$—that is of

$$\frac{1}{n^2}(x_1 - m + x_2 - m \ldots x_n - m)^2$$

equals the average value of $\frac{1}{n^2}\overset{1}{\underset{n}{\Sigma}}(x-m)^2$, since the product terms go out—is

$$\frac{1}{n^2}\frac{n}{2h^2} = \frac{1}{2nh^2},$$

and the average value of $n\mu^2 = \Sigma(\bar{x}-x)^2$ is that of

$$\Sigma(m-x)^2 - n(\bar{x}-m)^2,$$

that is
$$\frac{n}{2h^2} - \frac{1}{2h^2} = \frac{n-1}{2h^2};$$

and if the most probable value for h was such as to make the observed quantity μ^2 take up its average value we should have

$$h^2 = \frac{n-1}{2n\mu^2}.$$

The basis of the above method becomes less convincing when we consider that the frequencies with which different values of μ^2 occur, for a given value of h, cannot give a normal distribution, since μ^2 can only vary from 0 to $+\infty$; and that a frequency distribution might easily be constructed to have a zero at its mean, in which case the above basis would give us perhaps the only value for h, which could not possibly have given rise to the observed value of μ^2.

The distinction between the most probable value of h, and the value which makes μ^2 take up its average value, is illustrated by our treatment of the quantity $(\bar{x}-m)^2$, the average value of which is $\frac{1}{2nh^2}$, but the most probable value being zero, we say that the most probable value of m is $\bar{x}$, not $\bar{x} \pm \frac{1}{h\sqrt{(2n)}}$.

* Chauvenet, *Spherical Astronomy*, Note II., Appendix § 17.

If a frequency curve of unit area were drawn, showing the frequencies with which different values of μ^2 occur, for a given h, and if b were the ordinate corresponding to the observed μ^2, then we should expect the equation

$$\frac{\partial b}{\partial h} = 0$$

to give the most probable value of h. It is sufficient here, however, to point out the incorrectness of the assumption upon which some writers on the Theory of Errors have based their results.

6. We have now obtained an absolute criterion for finding the relative probabilities of different sets of values for the elements of a probability system of known form. It would now seem natural to obtain an expression for the probability that the true values of the elements should lie within any given range. Unfortunately we cannot do so. The quantity P must be considered as the relative probability of the set of values $\theta_1, \theta_2, \ldots, \theta_r$; but it would be illegitimate to multiply this quantity by the variations $d\theta_1, d\theta_2, \ldots, d\theta_r$, and integrate through a region, and to compare the integral over this region with the integral over all possible values of the θ's. P is a relative probability only, suitable to compare point with point, but incapable of being interpreted as a probability distribution over a region, or of giving any estimate of absolute probability.

This may be easily seen, since the same frequency curve might equally be specified by any r independent functions of the θ's, say $\phi_1, \phi_2, \ldots, \phi_r$, and the relative values of P would be unchanged by such a transformation; but the probability that the true values lie within a region must be the same whether it is expressed in terms of θ or ϕ, so that we should have for all values $\frac{\partial(\theta_1, \theta_2, \ldots, \theta_r)}{\partial(\phi_1, \phi_2, \ldots, \phi_r)} = 1$ a condition which is manifestly not satisfied by the general transformation.

In conclusion I should like to acknowledge the great kindness of Mr. F.J.M. Stratton, to whose criticism and encouragement the present form of this note is due.