INTERNAL AND EXTERNAL MEANS ARISING FROM THE SCALING OF FREQUENCY FUNCTIONS

By Edward L. Dodd

The scaling[1] of frequency functions has been discussed from the standpoint of maximum likelihood. But the likelihood criterion to be satisfied sometimes leads to a minimum likelihood; and sometimes to neither a maximum nor a minimum. Scaling will be studied in this paper with reference to the likelihood actually secured, and also with reference to the character of means obtained, whether internal or external.

Section 1. Introduction

It is well known that a scale obtained in a curve-fitting process is sometimes a mean. Thus, with the normal function

$$\frac{1}{a\sqrt{2\pi}}\, e^{-(x/a)^2/2}, \tag{1}$$

if the scale a is to be obtained from measurements, $x_1, x_2, \cdots, x_n$, we commonly accept the value

$$a = \left\{\frac{1}{n}\sum x_i^2\right\}^{1/2}; \tag{2}$$

that is, the root-mean square of the measurements. Here, the positive value of a is naturally taken. It is called the standard deviation, and thought of as an appropriate new unit of measure.

But even with the x's all negative, and the a taken positive, O. Chisini[2] considered it proper to regard a as a mean of the x's, albeit an *external* mean. From Chisini's viewpoint, this a whether regarded as positive or negative is primarily a solution of

$$x_1^2 + x_2^2 + \cdots + x_n^2 = a^2 + a^2 + \cdots + a^2. \tag{3}$$

In this sum of squares, the single number a may be *substituted* for each of the x's. Perhaps this kind of mean should be called a *substitutive* mean to distinguish it from the means of general analysis which are always internal.

---

[1] Fisher, R. A., "On the mathematical foundation of theoretical statistics," Philosophical Transactions of the Royal Society of London, Series A, Vol. 222, 309–368, (1921). See p. 338.

[2] Chisini, O., "Sul concetto di media," Periodico di matematico, Series 4, Vol. 9, 106–116, (1929).

The normal function is a particular case of a more general function:

$$\text{(4)} \qquad \text{Constant}\cdot a^{-1}\, e^{\phi(t)}, \qquad \phi(t) = -t^p/p, \qquad t = x/a.$$

The likelihood method to find the scale a for this function leads to power means, including the arithmetic mean, the root-mean-square, root-mean-cube, etc., for $p = 1, 2, 3$, etc.

The word *scale* will be used only for a positive number,—which then may be regarded as a unit of measurement.

For measurements, $x_1, x_2, \cdots, x_n$ Chisini regarded M as a mean, relative to a function G, provided

$$\text{(5)} \qquad G(x_1, x_2, \cdots, x_n) = G(M, M, \cdots, M).$$

If a solution of this equation is

$$\text{(6)} \qquad M = F(x_1, x_2, \cdots, x_n),$$

and c is a possible value for the x's, it follows at once that

$$\text{(7)} \qquad F(c, c, \cdots, c) = c,$$

or at least one value of this F is c. Conversely, if (7) is satisfied, it is but a change of notation to replace c in (7) by M, and to combine this with (6) to obtain

$$\text{(8)} \qquad F(x_1, x_2, \cdots, x_n) = F(M, M, \cdots, M).$$

Hence, this F which in (6) gives explicit form to the implicit M found in (5) may also be thought of as a mean-forming function, such as G in (5). Briefly, F is a particular G. Thus $F(x_1, x_2, \cdots x_n)$ is a mean of $x_1, x_2, \cdots, x_n$, if F is so constructed that (7) is satisfied when the arguments are all equal.

Inasmuch as a frequency function $f(t)$ is non-negative, $\log_e f(t)$ is real,—say $\phi(t)$ plus constant. Following R. A. Fisher, it will be convenient to write

$$\text{(9)} \qquad f(t) = Ca^{-1}\, e^{\phi(t)}, \qquad C = \text{Constant}$$

With location m already determined, the x's will be thought of as measured. from m. And we set

$$\text{(10)} \qquad t = x/a, \qquad t_i = x_i/a, \qquad i = 1, 2, \cdots, n.$$

The "productive" probability—to yield $x_1, x_2, \cdots, x_n$—is then

$$\text{(11)} \qquad L = \Pi f(t_i) = C^n\, a^{-n}\, e^{\Sigma\phi(t_i)}.$$

This is proportional[3] to the "likelihood" of a. Also—it may be noted in passing—the productive probability is also proportional to the *a posteriori* probability, if a constant *a priori* probability is postulated. The likelihood will here be taken as $\Pi f(t_i)$ itself; and it will be designated by L,—in Fisher's

[3] Loc. Cit., Fisher, p. 310.

notation, $L = \log \Pi$. Of course, Π and $\log \Pi$ take maximum values simultaneously, if at all. From (11) it follows[4] that

$$(12) \qquad -a \cdot \partial \log L/\partial a = n + \Sigma t_i \phi'(t_i) = \Sigma\{t_i \phi'(t_i) + 1\}.$$

The equation

$$(13) \qquad \Sigma t_i \phi'(t_i) + n = 0 \qquad (i = 1, 2, \cdots, n)$$

will be called the likelihood condition, whether this leads to maximum likelihood, to minimum likelihood, or to neither. A second differention[5] leads to

$$(14) \qquad a^2 \cdot \partial^2 \log L/\partial a^2 = \Sigma t_i^2 \phi''(t_i) - n = \Sigma\{t_i^2 \phi''(t_i) - 1\}.$$

When negative, this indicates a maximum likelihood; when positive, a minimum likelihood for the a obtained from (13).

Preparatory to the theorems of the next section, just one more matter will be discussed. The unit for t is arbitrary; and it may be convenient to write, with $k \neq 0$,

$$(15) \qquad \phi(t) = \phi(ku) = \Phi(u), \qquad t = ku.$$

Then

$$(16) \qquad t\phi'(t) = u\Phi'(u).$$

Suppose, now, that a positive constant k can be found such that $k\phi'(k) = -1$. Then, with $t = ku$, as postulated,

$$(17) \qquad 1 \cdot \Phi'(1) = k\phi'(k) = -1.$$

Thus $\Phi'(1) = -1$,—or as it will now be written $\phi'(1) = -1$,—is no more restrictive than the condition that some positive k exists such that $k\phi'(k) = -1$.

SECTION 2. GENERAL THEOREMS CONCERNING THE SCALE AS A MEAN

THEOREM I

Given the frequency function

$$(18) \qquad f(t) = Ca^{-1} e^{\phi(t)}, \qquad t = x/a, \qquad t_i = x_i/a, \qquad C = \text{Constant}.$$

And suppose that

$$(19) \qquad \phi'(1) = -1.$$

Suppose, also, that for given $x_1, x_2, \cdots, x_n$, the likelihood condition (13), now written

$$(20) \qquad \textstyle\sum_1^n (x_i/a)\phi'(x_i/a) + n = 0,$$

[4] Loc. Cit., Fisher, p. 338.
[5] Loc. Cit., Fisher, p. 339.

has a positive solution.

$$a = F(x_1, x_2, \cdots, x_n). \tag{21}$$

Then this a, the scale, is a mean.

Proof. With each $x_i = 0$, (20) cannot be satisfied.

But if, with $c \neq 0$, we take each $x_i = c$, and at the same time set $a = c$, then, by (19), $\Sigma = -n$; and thus (20) which gives a implicitly is satisfied. The explicit a in (21) is therefore such a function F that (7) is satisfied. Hence, the scale a is a mean.

THEOREM II

Given the frequency function

$$f(t) = Ca^{-1} e^{\phi(t)}, \qquad t = x/a, \qquad t_i = x_i/a, \quad C = \text{Constant}. \tag{18}$$

Suppose that

$$\phi'(1) = -1, \tag{19}$$

and that

$$| t\phi'(t) | < 1 \quad \text{if} \quad | t | < 1. \tag{22}$$

Moreover, suppose that the likelihood condition (20) for measurements $x_1, x_2, \cdots, x_n$, has a positive solution a. Then

$$a \leqq \text{Maximum } | x_i |. \tag{23}$$

Or, suppose that, in place of (22), we have

$$| t\phi'(t) | > 1 \quad \text{if} \quad | t | > 1; \tag{24}$$

and that $t\phi'(t)$ keeps the same sign, if $| t | > 1$. Then

$$\text{Minimum } | x_i | \leqq a. \tag{25}$$

Proof. Suppose, if possible, that $a > \text{Max } | x_i |$. Then each $| x_i/a | < 1$, and by (22), $| (x_i/a)\phi'(x_i/a) | < 1$. Then (20) is not satisfied, since $| \Sigma | < n$. Thus the hypothesis is contradicted.

Now (25) is satisfied at once if any $x_i = 0$. But suppose, on the other hand, that $\text{Min } | x_i | > 0$; and, if possible, that $a < \text{Min } | x_i |$. Then, by (24) et seq., since $| x_i/a | > 1$, it follows that $| \Sigma | > n$. And thus (20) is again contradicted.

THEOREM III

Given the frequency function

$$f(t) = Ca^{-1} e^{\phi(t)}, \qquad t = x/a, \qquad t_i = x_i/a, \quad C = \text{Constant}; \tag{18}$$

and set $\psi(t) = t\phi'(t) + 1$. Suppose that

$$\lim_{t \to 0} \psi(t) = \alpha, \qquad \lim_{| t | \to \infty} \psi(t) = \beta, \qquad \alpha\beta < 0. \tag{26}$$

And suppose that $\psi(t)$ is continuous when $t \neq 0$.

Then, for any set of real numbers, $x_1, x_2, \cdots, x_n$, of which none is zero, there exists a positive number a, as scale, such that the likelihood condition

$$\sum_1^n (x_i/a)\,\phi'(x_i/a) + n = 0 \tag{20}$$

is satisfied.

The conclusion is also valid, if in place of the limit β, there is postulated

$$\lim_{t\to -b+0} \psi(t) = -\alpha \mid \infty \mid = \lim_{t\to c-0} \psi(t), \tag{27}$$

where $b > 0$, $c > 0$, and $\psi(t)$ is continuous for $-b < t < 0$ and for $0 < t < c$. That is, the new limits are to be infinite with sign opposite to that of α.

Proof. The limits for $t \to 0$ and for $\mid t \mid \to \infty$ are the same as the limits for $a \to \infty$ and $a \to 0+$,—noting that $t = x/a$, $x \neq 0$. Thus $\Sigma\psi(t_i)$ changes sign as a goes from $0+$ to ∞. Hence, since $\psi(t)$ is continuous, (20) is satisfied for some positive a.

For the proof of the second part of the theorem, suppose that $x_n > 0$ and that x_n is the greatest x_i. Then with $a > x_n/c$, but approaching x_n/c, $\psi(x_n/a)$ becomes infinite with sign opposite to that of α. Furthermore, in $\Sigma\psi(x_i/a)$, the positive x's $< x_n$ have a negligible effect; and thus $\lim \Sigma\psi(x_i/a)$, as $a \to (x_n/c) + 0$, is infinite with sign opposite to that of α, when this sum Σ is taken for the positive x's. Likewise, if $x_1 < 0$, and is the least x_i, $\lim \Sigma\psi(x_i/a)$, as $a \to (-x_1/b) + 0$, is infinite with sign opposite to that of α, when this sum is taken for the negative x's. If, now, the measurements happen to be all positive, we think of a as approaching $x_n/c + 0$; and the continuity condition leads to an a which makes $\Sigma\psi(x_i/a) = 0$. Likewise, if the measurements happen to be all negative, we use $-x_1/b + 0$. If both positive and negative x's appear, we use the greater of the two ratios $-x_1/b$ and x_n/c.

SECTION 3. SOME FAIRLY REGULAR FREQUENCY FUNCTIONS

To illustrate the foregoing theorems in a somewhat general manner, consider the measurements, $x_1, x_2, \cdots, x_n$, and with $t = x/a$, $t_i = x_i/a$, set up the function:

$$f(t) = Ca^{-1} \mid kt \mid^p (1 + k^2t^2)^{-q} e^{-r \mid kt \mid^s}, \tag{28}$$

where, as before, C is a suitably chosen constant.

Suppose also that

$$p > -1, \qquad q \geqq 0, \qquad r \geqq 0, \qquad s \geqq 0; \tag{29}$$

and that either

$$r > 0,\ s > 0 \qquad \text{or} \qquad r = 0,\ 2q > p + 1. \tag{30}$$

Then with $\phi(t) = \log f(t)$, it follows that, when $t \neq 0$,

$$t\phi'(t) + 1 = (p + 1) - rsk^s \mid t \mid^s - 2qk^2t^2(1 + k^2t^2)^{-1}. \tag{31}$$

Now the condition $1 \cdot \phi'(1) = -1$ would be satisfied if $\Psi(k) = 0$, where

$$\Psi(k) = rsk^{s+2} + rsk^s + (2q - p - 1)k^2 - (p + 1). \tag{32}$$

But, under the conditions (29) and (30) $\Psi(0) < 0$, and $\Psi(\infty) > 0$. Hence, there is a positive k for which $\Psi(k) = 0$. Then if k be assigned this value, (19) is satisfied; and by Theorem I, any scale a that the likelihood condition (20) may lead to is a mean. But, by Theorem III a scale a will actually exist —indeed, for any positive k that may be used in (29); since the limit of $t\phi'(t) + 1$ is positive as $t \to 0$, and is negative as $|t| \to \infty$.

Moreover, if in (29), the further condition $-1 < p \leqq 0$ is introduced, (22) is satisfied. And, thus, $a \leqq$ Maximum $|x_i|$. Also, $|t\phi'(t)|$ increases with $|t|$. Hence, by (24) et seq., Minimum $|x_i| \leqq a$.

If in (28), we set $q = 0$, $s = 1$, $r > 0$, and confine our attention to positive x and t, there is obtained the Pearson Type III. Reference to (32) shows that $\Psi(k) = 0$ if $k = (p + 1)/r$. With this substitution,

$$f(t) = C^1 a^{-1} t^p e^{-(p+1)t}, \qquad C' = \text{Constant}. \tag{33}$$

Since $\phi'(1) = -1$, any solution of the likelihood condition is a mean. Here, with $t > 0$, $t\phi'(t) = p - (p + 1)t$, and $t^2\phi''(t) - 1 = -(p + 1)$. From (14) we see that, with $p + 1 > 0$, any mean obtained corresponds to maximum likelihood and the single maximum found is actually the largest value. Moreover, with the measurements, $x_1, x_2, \cdots, x_n$, all positive, a scale a will exist,—as noted in the general case (28).

In passing, it may be noted that Type III appears[6] rather naturally in a form giving $\phi'(1) = -1$ at once, without any transformation. Here, then, a scale is a mean.

Given the Pearson Type I in the form

$$f(t) = Ca^{-1}(b + kt)^p(c - kt)^q, \quad t = x/a, \quad b > 0, \quad c > 0, \quad |pq| > 0. \tag{34}$$

If $p + q + 1 > 0$, it is possible to find a positive k so that with $\phi = \log f$, $\phi'(1) = -1$. In this case, any scale found by the likelihood condition is a mean. With k thus chosen, $f(t)$ has essentially the same form as it would have if $k = 1$. Hence for convenience, let us simply set $k = 1$ in the above equation. Then for $-b < t < c$,

$$\psi(t) = t\phi'(t) + 1 = 1 + pt(b + t)^{-1} - qt(c - t)^{-1}.$$

Suppose now that $p > 0$ and $q > 0$. Then Theorem III may be applied; since $\lim \psi(t) = 1$, as $t \to 0$; but $\lim \psi(t) \to -\infty$, as $t \to -b + 0$, or as $t \to c - 0$.

---

[6] Carver, H. C., Handbook of Mathematical Statistics, Chap. VII, see p. 105, Line 4, noting that $\phi' = y'/y$.

Hence a scale a satisfying the likelihood condition exists. Moreover, the likelihood is at a maximum; since, with $-b < t < c$,

$$t^2\phi''(t) - 1 = -pt^2(b + t)^{-2} - qt^2(c - t)^{-2} - 1 < 0.$$

This maximum is also the largest value for all values of a.

If the Pearson Type IV is given in the form

$$f(t) = Ca^{-1}(1 + k^2t^2)^{-p} e^{q \text{ arc tan } kt}, \qquad t = x/a \tag{35}$$

then if $p > 1/2$, it is possible to find a positive k which will make $\phi'(1) = -1$. In this case, any scale a is a mean. Moreover—for any $k \neq 0$—the limit of $t\phi'(t) + 1$ is 1 for $t \to 0$ and is $1 - 2p$ for $t \to \infty$. Hence, by Theorem III, if $p > 1/2$, as above, then a scale a exists satisfying the likelihood condition (20).

SECTION 4. FREQUENCY FUNCTIONS WITH CERTAIN PECULIARITIES

The theorems of section 2 give sufficient conditions, which in some cases may not be necessary. Nevertheless, by violating certain hypotheses, particular functions may be set up which exhibit various peculiarities.

For the Pearson Types, the differential equation is

$$\frac{y'(t)}{y(t)} = \phi'(t) = \frac{a_0 + a_1t}{b_0 + b_1^t + b_2t^2}, \qquad t = x/a. \tag{36}$$

The determination of a positive scale a by the Fisher likelihood process is impossible here, in case $a_0 = 0$, $a_1 > 0$, $b_0 + b_1t + b_2t^2 > 0$. For in this case $t\phi'(t) \geqq 0$; and thus (20) cannot be satisfied. The U-shaped Type II curves are in this class. Likewise, if $a_0 \neq 0$, $a_1 = 0$, and $b_0 + b_1t + b_2t^2 > 0$,—for example, with $b_2 > 0$, $b_1^2 < 4b_0b_2$,—and the measurements all happen to have the same sign as a_0, such scaling is impossible.

For the purpose of constructing peculiar functions we may take $c > 0$ and require that the measurements x_i be either $-c$ or c—with at least one $-c$ and at least one c—and that $\phi(t)$ be an even function. Then $\phi(-c) = \phi(c)$ and (11) becomes

$$L = [Ca^{-1} e^{\phi(c/a)}]^n. \tag{37}$$

The likelihood condition (13) reduces to

$$0 = \psi(t) = t\phi'(t) + 1 = (c/a)\phi'(c/a) + 1, \tag{38}$$

with the right member an even function of c/a. And from (14), a maximum likelihood is indicated when

$$(c/a)^2\phi''(c/a) - 1 < 0, \tag{39}$$

with the left member likewise an even function. A minimum likelihood is indicated if the left member is positive.

Let us apply this to the case where

(40) $\phi(t) = (-2/3) \log (1 - 3 \mid t \mid); \qquad t\phi'(t) = 2 \mid t \mid (1 - 3 \mid t \mid)^{-1}.$

The likelihood condition (38) is satisfied only when $t = \pm 1$. Also $\phi'(1) = -1$. Thus the only means are the internal means $\pm c$; and the only scale conformable to (38) is $a = c$. But this has minimum likelihood; since $1 \cdot \phi''(1) - 1 = \frac{1}{2} > 0$. For positive t, this function (40) is a Pearson Type.

Consider next a function of the form (28),—with $p = -1.25$, $q = -0.5$, however,—for which (31) becomes

(41) $t\phi'(t) + 1 = -1/4 - t^2/4 + t^2/(1 + t^2) = -(1 - t^2)^2/4(1 + t^2).$

whence $\phi'(1) = -1$, $\phi''(1) = +1$, $\phi'''(1) = -3$. Here the likelihood condition (38) has but a single absolute solution $\mid t \mid = 1$, leading to the single scale $a = c$, and to the two internal means, $\pm c$. But, in this case $1 \cdot \phi''(1) - 1 = 0$, so that $\partial^2 \log L/\partial a^2 = 0$. Moreover, for $t = 1$, $\partial^3 \log L/\partial a^3 = a^{-3} \neq 0$. Thus, the only scale obtained by the likelihood method (38)—viz., $a = c$— has a likelihood which is neither at a maximum nor at a minimum.

Another anomalous function is that given by

(42) $$\phi(t) = t^4 - 2.5t^2, \qquad t = \pm c/a.$$

The likelihood condition (38) leads to

$$\psi(t) = (1 - t^2)(1 - 4t^2) = 0.$$

The only solutions are $t = \pm 1$, giving internal means $\pm c$; and $t = \pm 1/2$, giving external means $\pm 2c$. And from (39) et seq., it can be shown that the internal mean and scale, $a = c$ has minimum likelihood, while the external mean and scale, $a = 2c$, has maximum likelihood.

But it will be noted that a maximum value for a vicinity does not always signify a largest value for the entire possible range. Indeed, for the function (42), $a = 2c$ has maximum likelihood without having the largest likelihood. To avoid such an anomaly, a necessary condition is that as $\mid t \mid \to \infty$, $\psi(t) \to -\infty$; as seen by taking the logarithm of L in (37), noting that as $a \to 0$, $(-\log a) \to +\infty$.

Finally avoiding the anomaly just mentioned, let us set up a frequency function, using the $\psi(t)$ in (38), and writing

$$\psi(t) = 1 + t\phi'(t) = (1 - 2t^2)(1 - t^2)(1 - 0.9t^2).$$

From this it follows readily that

(43) $$\phi(t) = K - 1.95t^2 + 1.175t^4 - 0.3t^6, \qquad K = \text{Constant}.$$

This, with $t_i = \pm c/a$, leads to an internal mean or scale $a = c$ with minimum likelihood, a nearby scale $a = c\sqrt{0.9}$ with maximum likelihood—differing indeed only slightly from the minimum just mentioned—and another scale $a = c\sqrt{2}$ having maximum likelihood, and this likelihood is indeed greater than that for any other positive value of a. The external mean $a = c\sqrt{2}$ in this case has the largest likelihood. This may be checked by the use of the logarithm of L as it appears in (37), in which the important part is $\phi(c/a) - \log_e a$.

In passing it may be noted that if $\psi(t)$ has the form $\psi(t) = (1 - t)H(t)$, with $H(1) \neq \infty$, and $t_i = x_i/a$; then any solution a of the likelihood condition $\psi(t) = 0$ is a mean,—by Theorem I.

SECTION 5. SUMMARY

When the R. A. Fisher likelihood method is used to find an "optimum" scale for frequency functions, it sometimes happens that this scale is a well known mean or at least is a *substitutive* mean—See Equation (5). Or a simple transformation (15) may often put the frequency function into such a form. Conditions are given under which a scale will be a mean. Under further conditions this mean will be internal—at least as regards absolute values. Finally, under certain conditions, a scale will exist.

But for certain functions not satisfying these conditions, anomalies appear. The scale given by the usual likelihood condition may be a scale with a minimum likelihood. Sometimes the likelihood will be at neither a maximum nor a minimum. In certain simple cases, no scale exists. Furthermore, it may happen that the scales which are internal means have minimum likelihood and those that are external means have maximum likelihood. Among Pearson Types are found both anomalous functions and functions which would be regarded as regular as regards maximum likelihood.

In this problem of scaling, likelihood is proportional to *a posteriori* probability with the *a priori* probability taken as constant.