FREQUENCY LAWS SHOWING STABILITY WITH REFERENCE TO THE GEOMETRIC MEAN AND OTHER MEANS*

BY E. L. DODD

1. *Stability with Reference to the Arithmetic Mean.* The purpose of this paper is to extend to various means the conception of stability as developed by Lévy† with reference to the arithmetic mean. The best known functions showing this stability, as mentioned by Lévy, are the Gaussian or normal probability function and the Cauchy function $(a/\pi)(a^2+x^2)^{-1}$. For this stability, the probability function or *distribution* of the arithmetic mean must have essentially the same form as that of individual variates. Indeed, in the Cauchy case, the distributions are identical. But in general, the distribution of the arithmetic mean contains the parameter n.

Lévy takes $\phi(t)$ as a characteristic function, and writes

(1) $\psi(t) = \log \phi(t) = - (c_0 + c_1 j) \, | \, t \, |^\alpha,$

where

(2) $c_0 > 0, \, | \, c_1 \, | \leq | \, c_0 \tan \pi\alpha/2 \, | \, , 0 < \alpha \leq 2, j = (- 1)^{1/2} \operatorname{sgn} t.$

If, now, X_i is an individual variate, and if Prob signifies *probability that*, then

(3)   Prob $\{ x < X_i < x + dx \} = f(x)dx, \quad (i = 1, 2, \cdots, n),$

where, if we use i in an *exponent* to signify $(- 1)^{1/2}$, we write

(4) $f(x) = f(x, c_0, c_1, \alpha) = \dfrac{1}{2\pi} \displaystyle\int_{-\infty}^{\infty} e^{-ixt}\phi(t)dt.$

Using F for the cumulative function, we find

(5)   Prob $\{ X_i < u \} = F(u, c_0, c_1, \alpha) = \displaystyle\int_{-\infty}^{u} f(x)dx.$

If now S is the sum of n independent variates X_i, the logarithm of the characteristic function of S is $n\psi(t)$; and hence its probability function or *density* takes the form $f(x, nc_0, nc_1, \alpha)$. For the arithmetic mean, A, it follows that $nf(nx, nc_0, nc_1, \alpha)$ is the density of distribution. Hence

(6) $\text{Prob } \{A < u\} = F(nu, nc_0, nc_1, \alpha)$.

2. *Extension of Stability.* To define* a mean M of n independent variates Y_i, I shall suppose here that

$$x = \xi(y)$$

is a *continuous increasing* function of y. Its inverse $y = \theta(x)$ will then exist, and we may set

(7) $$M = \theta \left[\frac{1}{n} \ \sum_{1}^{n} \xi(Y_i) \right].$$

I shall now regard the cumulative probability or frequency function F as *stable* with reference to the general mean M, if from

(8) $\text{Prob } \{Y_i < v\} = F[\xi(v), c_0, c_1, \alpha]$,

we can conclude that

(9) $\text{Prob } \{M < v\} = F[n\xi(v), nc_0, nc_1, \alpha]$.

A more general definition of stability might be framed, but the foregoing is adequate for this paper.

To show that (8) and (9) are satisfied when M is defined by (7), we need only set

(10) $X_i = \xi(Y_i), \quad u = \xi(v); \quad Y_i = \theta(X_i), \quad v = \theta(u)$.

Then, from (5) and (6), noting that ξ is an *increasing* function, we have

(11) $\text{Prob } \{Y_i < v\} = \text{Prob } \{X_i < u\} = F[u, c_0, c_1, \alpha]$;

(12) $\text{Prob } \{M < v\} = \text{Prob } \{A < u\} = F[nu, nc_0, nc_1, \alpha]$.

---

* I defined a somewhat more general mean in my paper *Functions of measurements under general laws of error*, Skandinavisk Aktuarietidskrift, vol. 5 (1922), p. 141.

3. *Particular Means.* The geometric mean G of n *positive* variates Y_i is given by

$$(13) \quad G = (Y_1 \cdot Y_2 \cdot \cdots \cdot Y_n)^{1/n}, \quad \log G = \frac{1}{n} \sum_1^n \log Y_i;$$

and thus (8) and (9) apply with $M = G$, $\xi(v) = \log v$, $v > 0$. We note in this case that $u = \xi(v) = \log v$ permits u to take on all real values, although v is restricted to positive values.

The cube root of the average cube of variates Y_i is a special case, with $p = 3$, of

$$(14) \quad M = |B|^{1/p} \operatorname{sgn} B, \quad B = \frac{1}{n} \sum_1^n |Y_i|^p \operatorname{sgn} Y_i, \quad p > 0,$$

where $\operatorname{sgn} Y_i = -1, 0, +1$, according as Y_i is negative, zero, or positive. Here, in (14), Y_i is *not* restricted to *positive* values. The case of $p = 2$ is a *signed root-mean-square*. For the general case (14), we take $\xi(v) = |v|^p \operatorname{sgn} v$, to obtain (8), (9).

4. *Stability for Symmetrical Distributions.* If in (1) we take $c_1 = 0$, $c_0 = c$, then (4) becomes

$$(15) \quad f(x) = f(x, c, \alpha) = \frac{1}{\pi} \int_0^\infty e^{-ct^\alpha} \cos txdt.$$

In this case, $f(-x) = f(x)$, and the distribution is symmetrical, with mode at the origin. The further specialization $\alpha = 1$ or $\alpha = 2$, gives the Cauchy or the Gaussian distribution, respectively. In what follows we merely require, as before, $0 < \alpha \leq 2$, with $c > 0$.

We shall also introduce a total frequency function

$$(16) \quad T[u, c, \alpha] = \int_{-u}^u f(x)dx = 2 \int_0^u f(x)dx, \quad u \geq 0.$$

Then if $g(x, c, \alpha, n)$ is the frequency density for the sum of n independent variates X_i,

$$(17) \quad g(x, c, \alpha, n) = \frac{1}{\pi} \int_0^\infty e^{-nct^\alpha} \cos txdt = n^{-1/\alpha} f(n^{-1/\alpha}x, c, \alpha),$$

as may be seen by a change of variables $nt^\alpha = \tau^\alpha$. Then, passing to the arithmetic mean A, by changing x into nx, dx into ndx, we obtain

(18) $\qquad$ $\text{Prob } \{ \, |A| < u \} = T[n^{1-1/\alpha}u, c, \alpha].$

From the definition (16), it is seen that $T[u, c, \alpha]$ is an increasing function of u, a function, indeed, that approaches unity (certainty) when u approaches infinity. Thus, from (18), with $1 < \alpha \leqq 2$, and $n > 1$,

$$\text{Prob } \{ |A| < u \} > T[u, c, \alpha] = \text{Prob } \{ |X_i| < u \}.$$

Suppose, now, that when n measurements of a quantity have been made, we postulate a *true value* for the measured quantity. Sometimes this true value is actually ascertainable; for example, if ten balls have been drawn from an urn containing white and black balls, and w are found to be white, the ratio $w/10$ is a *measurement* of an ascertainable proportion of white balls in the urn. By translation, now, suppose the origin moved to the true value. Thus, if in the above urn 40% of the balls are white, $X = w/10 - 40$. The X's may now be called *errors*. The inequality obtained above,

$$\text{Prob } \{ |A| < u \} > \text{Prob } \{ |X_i| < u \},$$

now tells us that the arithmetic mean A is more likely to fall inside a fixed interval about the true value *zero* than is an individual measurement X_i; and in this sense, A is more *reliable* than X_i when $1 < \alpha \leqq 2$. On the other hand, A is less reliable than X_i when $0 < \alpha < 1$. These facts regarding the *arithmetic mean* are not new.

5. *The Geometric Mean, with Symmetric Distribution.* Let G be defined as in (13); but taking a positive number a, suppose that

(19) $\qquad$ $X_i = \log (Y_i/a), \qquad (i = 1, 2, \cdots, n),$

with X_i subject to (15). We have

(20) $\qquad$ $\log (G/a) = \dfrac{1}{n} \sum \log (Y_i/a),$

where $\log (G/a) = A$ in (18). It follows that

(21) $\quad$ $\text{Prob } \{ \, |\log (G/a)| < u \} = T[n^{1-1/\alpha}u, c, \alpha], \; u \geqq 0.$

Then, with $u = \log v$ as before, but now with $v \geq 1$, since $u \geq 0$,

$$(22) \qquad \text{Prob} \left\{ \frac{a}{v} < G < av \right\} = T[n^{1-1/\alpha} \log v, c, \alpha].$$

If we further postulate that a is the *true value*, G is less *reliable* than an individual measurement Y_i, if $0 < \alpha < 1$, but more reliable, if $1 < \alpha \leq 2$. Moreover, with $1 < \alpha \leq 2$, if $\epsilon > 0$ is arbitrarily small, and $v = 1 + \epsilon$, we may take n large enough so that it is asymptotically certain that G will be confined to an arbitrarily small interval about the true value a.

6. *Another Illustration.* As an added requirement, let us suppose that in the definition of M in (7) the function ξ is *odd*:

$$(23) \qquad \xi(-Y_i) = -\xi(Y_i), \qquad \xi(-v) = -\xi(v);$$
$$\theta(-X_i) = -\theta(X_i), \qquad \theta(-u) = -\theta(u).$$

Then, from (10), the inequality $|X_i| < u$ becomes $|Y_i| < v$. Thus, if

$$(24) \qquad \text{Prob} \left\{ |Y_i| < v \right\} = T[\xi(v), c, \alpha], \quad v \geq 0,$$

it follows that

$$(25) \qquad \text{Prob} \left\{ |M| < v \right\} = T[n^{1-1/\alpha}\xi(v), c, \alpha].$$

Means of the form (14) are included here.

THE UNIVERSITY OF TEXAS