A SHORT METHOD AND TABLES FOR THE CALCULATION OF THE AVERAGE AND STANDARD DEVIATION OF LOGARITHMIC DISTRIBUTIONS*

By

THOMAS N. JENKINS
New York University

In fitting various types of curves to reaction-time data,[1] the writer was impressed with the enormous amount of labor and boredom involved in the calculation of the constants of logarithmic distributions. Besides the constant use of a set of logarithm tables, it requires the tedium of squaring large numbers on a machine to compute the second moments of the distributions. In order to eliminate some of the labor involved in such a process, a short method was devised for the computation of the average and the standard deviation of logarithmic distributions.

The short method described in this paper was originally developed to facilitate the work of fitting logarithmic normal curves to a large number of reaction-time distributions, but dispersions approximating this type seem to be sufficiently common in economics and biology to warrant a more general use of short methods in the computation of the constants of such distributions. In the field of economics, logarithmic curves have been fitted with success to distributions of income and prices, and probably could be applied equally well to distributions of capital. Many skewed distributions can also be found in the fields of biology and psychology. Kapteyn fitted a logarithmic curve to a distribution of the minimum weights necessary to produce a sensation of pres

*A portion of the work involved in this paper was carried out during the writer's tenure as a National Research Fellow.

[1] Cf. Facilitation and Inhibition. *Arch. Psychol.* No. 86, 56 p.

sure.[2] Kapteyn attempts to show that logically the normal curve is the exception and skew curves the rule. For example, if "the diameters of certain ripe berries" are distributed in a normal curve, their *volumes* will be distributed in an asymmetrical curve; in other words, volume increase will be dependent upon size, so that volume changes are greater for large berries than for small ones. That skew curves are the rule can be shown analytically. Suppose certain quantities z are distributed normally, and *any* other quantities x are expressed as functions of z, thus,

$$z = f(x)$$

Then,

(1) $$dz = f'(x)dx$$

If the frequency curve for the z's is,

(2) $$y = \frac{N}{\sigma\sqrt{2\pi}} e^{-\frac{(z-M)^2}{2\sigma^2}}$$

then the frequency curve for the x's is,

(3) $$y \frac{N}{\sigma\sqrt{2\pi}} f'(x) e^{-\frac{[f(x)-M]^2}{2\sigma^2}}$$

It will be seen at once that the x's cannot be distributed normally provided x is a non-linear function of z. If we let $z = \log x$, then $dz = \frac{dx}{x}$ and equation (2) becomes

(4) $$y = \frac{N}{\sigma\sqrt{2\pi}} \frac{1}{x} e^{-\frac{(\log x - M)^2}{2\sigma^2}}$$

This is the logarithmic curve of distribution, the theory of which has been treated by several writers, one of the first and most important papers on this subject being that of McAllister.[3] The study

[2] J. C. Kapteyn, "Skew frequency curves in Biology and Statistics". Groningen. p. 42-43, 1903.

[3] The Law of the Geometric Mean. *Proc. Roy. Soc.* 29:367. (1879).

of the properties of the logarithmic curve of error was undertaken by McAllister at the suggestion of Galton,[4] who saw the possibility of applying it to psychological and social phenomena.

Dispersions approximating this type are illustrated by distributions which are definitely limited at the zero point, but a more definite presumption in favor of the logarithmic curve is indicated when the *real* origin, determined *a priori* or deduced from empirical considerations, does not correspond with the origin on the value scale.

Reaction-time distributions are good examples of dispersions where a displacement of the origin is indicated by empirical considerations. A little reflection will show that there must be a physiological limit for the speed of reaction. It takes a certain minimum time for the neuro-muscular machine to do its work. The time it takes for the machine to do its work constitutes an undisturbed region within which no deviations ever occur. Reaction-time dispersions approximate the logarithmic more closely than the normal curve of error. Investigations in the field of learning often give distributions which have origins other than the zero of the scale which can be determined *a priori* . . . .that is. the real origin follows *inevitably* from the conditions of the experiment.. If the norm of mastery for learning a maze is two perfect trials out of three, then the criterion is such that an animal to learn a maze must make at least two perfect runs. In other words, the experimenter's criterion is such that no deviations could *possibly* occur under two trials.

In using the short method for finding the first and second moments of a logarithmic distribution, the computer must still resort to a table, but in this case it is only necessary to use a single page table instead of an extensive logarithm table. Furthermore, the labor of squaring the logarithms is eliminated. The short method can best be explained by following the process through an

[4] The Geometric Mean in Vital and Social Statistics. *Proc. Roy. Soc.* 29:365. (1879).

actual example.[5] This is illustrated in Table II on a distribution of reaction-times. Beginning at 70 (the real origin of the distribution is assumed to be at 70) the step-intervals are numbered from zero to the end of the distribution. Under the log x column of Table I the value for each step is found and multiplied by the frequency for each step. This operation gives the values shown in the $F \log x$ column of Table II. The sum of these values divided by N (number of cases) gives the correction C. The average ($\log G_l$) for the logarithmic distribution is finally found by adding a factor K to the correction C. The constant K depends upon the length of the step-interval. In this distribution, the length of the step-interval is ten. Looking under column K of Table I, we find that the value of K for a step-interval of ten units is equal to .69897. The geometric mean (G) of the distribution is found by adding 70 to (G_l).[6]

The process of finding the second moment and standard deviation (σ_g) is similar to that for finding the first moment and the average. In one respect it is simpler: no correction has to be added for the length of step. The $F \log^2 x$ column is obtained by multiplying the value for each step in the $\log^2 x$ column of Table I by its appropriate frequency. The sum of these divided by N (number of cases) gives the crude unit moment. The square of the correction C is then subtracted to give the corrected unit moment around the average. The square root of the corrected unit moment around the average gives the standard deviation (σ_g), and the antilog of σ_g gives the standard deviation ratio (σ_r).

The formula for finding the average of a logarithmic distribution is,

$$\log G_l = \frac{\Sigma F \log x}{N} + K = C + K$$

[5] For those interested, the proof of the formulae for getting the average and standard deviation is given in an appendix at the end of this paper.

[6] One would expect the geometric mean to be different if the origin were taken at a point other than 70. An origin at 70 was assumed because it results in an extremely good fit to the distribution. In this case, 70 would correspond to the physiological limit below which no deviations ever occur.

TABLE II

Time	F	Step	$F \log x$	$F \log^2 x$
70	1	1		
80	3	2	1.431363	.682934
90	14	3	9.785580	6.839826
100	40	4	33.803921	28.567627
10	55	5	52.483338	50.081832
20	60	6	62.483561	65.069923
30	52	7	57.925054	64.525229
40	46	8	54.100197	63.626769
50	33	9	40.604814	49.962150
60	28	10	35.805100	45.785901
70	21	11	27.766605	36.713541
80	14	12	19.064189	25.960237
90	9	13	12.581460	17.588126
200	7	14	10.019546	14.341615
10	7	15	10.236785	14.970255
20	4	16	5.965446	8.896638
30	3	17	4.555541	6.917653
40	1	18	1.544068	2.384146
50				
60	1	20	1.591064	2.531486
70				
80	1	22	1.633468	2.668219
	400		400) 443.381100	400) 508.114107
				1.270285
		$C =$	1.108452	$C^2 =$ 1.228667
		$K =$	.698970	
				$\sigma_0^2 =$ .041618
		log $G_1 =$	1.807422	
				$\sigma_9 =$ .204004
For origin at 70		$G_1 =$	65.8	
		$G =$	135.8	$\sigma_r =$ 1.59

TABLE I

Step	Log x	$(\text{Log } x)^2$	Step Interval	K
1	.00000 00000	.00000 00000	1	.30102 99956
2	.47712 12547	.22764 46917	2	.00000 00000
3	.69897 00043	.48855 90669	3	.17609 12590
4	.84509 80400	.71419 06972	4	.30102 99956
5	.95424 25094	.91057 87668	5	.39794 00086
6	1.04139 26851	1.08449 87247	6	.47712 12547
7	1.11394 33523	1.24086 97921	7	.54406 80443
8	1.17609 12590	1.38319 06496	8	.60205 99913
9	1.23044 89213	1.51400 45481	9	.64321 25137
10	1.27875 36009	1.63521 07719	10	.69897 00043
11	1.32221 92947	1.74826 38633	11	.74036 26894
12	1.36172 78360	1.85430 26993	12	.77815 12503
13	1.39794 00086	1.95423 62678	13	.81291 33566
14	1.43136 37641	2.04880 22253	14	.84509 80400
15	1.46239 79978	2.13860 79042	15	.87506 12633
16	1.49136 16938	2.22415 97018	16	.90308 99869
17	1.51851 39398	2.30588 45856	17	.92941 89257
18	1.54406 80443	2.38414 61255	18	.95424 25094
19	1.56820 17240	2.45925 66473	19	.97772 36052
20	1.59106 46070	2,53148 65837	20	1.00000 00000
21	1.61278 38567	2.60107 17684	21	1.02118 92990
22	1.63346 84555	2.66821 91953	22	1.04139 26851
23	1.65321 25137	2.73311 16157	23	1.06069 78403
24	1.67209 78579	2.79591 12465	24	1.07918 12460
25	1.69019 60800	2.85676 27889	25	1.09691 00130
26	1.70757 01760	2.91579 59062	26	1.11394 33523
27	1.72427 58696	2.97312 72744	27	1.13033 37684
28	1.74036 26894	3.02886 22909	28	1.14612 80356
29	1.75587 48556	3.08309 65087	29	1.16136 80022
30	1.77085 20116	3.13591 68471	30	1.17609 12590
31	1.78532 98350	3.18740 26197	31	1.19033 16981
32	1.79934 05494	3.23762 64129	32	1.20411 99826
33	1.81291 33566	3.28665 48386	33	1.21748 39442
34	1.82607 48027	3.33454 91850	34	1.23044 89213
35	1.83884 90907	3.38136 59785	35	1.24303 80486
36	1.85125 83487	3.42715 74737	36	1.25527 25051

37	1.86332 28601	3.47197 20810	37	1.26717 17284
38	1.87506 12633	3.51585 47414	38	1.27875 36009
39	1.88649 07251	3.55884 72561	39	1,29003 46113
40	1.89762 70912	3.60098 85775	40	1.30102 99956
41	1.90848 50188	3.64231 50672	41	1.31175 38610
42	1.91907 80923	3.68286 07246	42	1.32221 92947
43	1.92941 89257	3.72265 73909	43	1.33243 84599
44	1.93951 92526	3.76173 49312	44	1.34242 26808
45	1.94939 00066	3.80012 13980	45	1.35218 25181
46	1.95904 13923	3.83784 31768	46	1.36172 78360
47	1.96848 29485	3.87492 51187	47	1.37106 78622
48	1.97772 36052	3.91139 06589	48	1.38021 12417
49	1.98677 17342	3.94726 19240	49	1.38916 60843
50	1.99563 51945	3.98255 98299	50	1.39794 00086

and, $$G_l = \text{antilog}(C + K)$$

where G_l is the geometric mean measured from an origin which may be other than the zero of the scale. The geometric mean (G) measured from the zero of the value scale is,

$$G = G_l + (\text{displacement of the origin})$$

The formula for finding the standard deviation around the average is,

$$\sigma_g = \sqrt{\frac{\Sigma F \log^2 x}{N} - C^2}$$

and, $$\sigma_r = \text{antilog } \sigma_g$$

Summary of steps in the calculation of the average ($\log G_l$) by the short method:

1. Beginning at the origin, find the deviation of the mid-point of each step-interval from the origin in units of step-interval.
2. Using Table I, find the log x of each step-deviation and weight it by its appropriate F (frequency).

3. Find the *sum* of the $F\log^2 x$'s, and divide this sum by N (number of cases). This gives the *correction* C.
4. Using Table I, find the value of K corresponding to the number of units in the step-interval. Add the factor K to the correction C to get the average $(\log G_1)$.

Summary of steps in the calculation of the standard deviation (σ_g) around the average $(\log G_1)$ by the short method:

1. Using Table 1, find the $\log^2 x$ of each step-deviation and weight it by its appropriate frequency.
2. Find the *sum* of the $F \log^2 x$'s, and divide this sum by N.
3. Then subtract the square of the correction C to get the *second unit moment* around the average.
4. Extract the square root of the second unit moment to obtain the standard deviation (σ_g).

APPENDIX

Deduction of the Formulae for the Short Method

Let $\log G$ be the logarithmic mean, x, the length of step, $f_1 \ldots\ldots f_n$ the frequencies for successive steps, and $m_1, m_2 \ldots\ldots m_n$ the mid-points of the steps for origin at zero. Then the first mid-point, m_1, is at $x/2$ the second, m_2, is at $3x/2$ etc. For convenience, these items may be arranged in the form of a table.

Midpoint (M)	F	$F \log M$	$F \log^2 M$
m_1	f_1	$f_1 \log x/2$	$f_1 \log^2 x/2$
m_2	f_2	$f_2 \log 3x/2$	$f_2 \log$
.	.	.	
.	.	.	
.	.	.	
.	.	.	
m_n	f_n	$f_n \log \frac{2n-1}{2} x$	$f_n \log^2 \frac{2n-1}{2} x$

$$N \log G = \frac{\Sigma f_n \log \frac{2n-1}{2} x}{N}, \quad \sigma_g^2 = \frac{\Sigma f_n \log^2 \frac{2n-1}{2} x}{N} - \log^2 G$$

where G is the geometric mean, and σ_g is the standard deviation around $\log G$.

We have, therefore,

$$\log G = \frac{f_1 \log 1/2 + f_1 \log 3/2 + \cdots + f_n \log \left(\frac{2n-1}{2}\right) x}{N}$$

Stating the logarithm of each fraction as the sum or difference of the logarithms of its factors, we have,

$$\log G = \frac{f_1 (\log 1 - \log 2 + \log x) + \cdots + f_n [\log (2n-1) - \log 2 + \log x]}{N}$$

$$= \frac{f_1 \log 1 + f_2 \log 3 + \cdots + f_n \log (2n-1)}{N} + \log x - \log 2$$

$$= \frac{\Sigma [f_n \log (2n-1)]}{N} + \log x - \log 2$$

Since

$$\log x = \frac{\Sigma f \log x}{N}$$

and,

$$\log 2 = \frac{\Sigma f \log 2}{N}$$

Letting

$$K = \log x - \log 2$$

and,

$$C = \frac{\Sigma [f_n \log (2n-1)]}{N}$$

we finally have,

$$\log G = C + K$$

Where C is the correction, and K is the constant indicated in Table I.

For the second unit moment around $\log G$, we have,

$$\sigma_g^2 = \frac{\Sigma\left[f_n \log^2\left(\frac{2n-1}{2}\right)x\right]}{N} - \left(\frac{\Sigma\left[f_n \log\left(\frac{2n-1}{2}\right)x\right]}{N}\right)^2$$

$$= \frac{\Sigma\left\{f_n\left[\log^2\left(\frac{2n-1}{2}\right) + 2\log\left(\frac{2n-1}{2}\right)\log x + \log^2 x\right]\right\}}{N}$$

$$- \left\{\left(\frac{\Sigma f_n \log\left(\frac{2n-1}{2}\right)}{N}\right)^2 + \frac{2\Sigma\left[f_n \log\left(\frac{2n-1}{2}\right)\right]\log x}{N} + \log^2 x\right\}$$

Expanding again and collecting terms, we have,

$$\sigma_g^2 = \frac{\Sigma\left\{f_n\left[\log(2n-1)\right]^2\right\}}{N} - \left\{\frac{\Sigma\left[f_n \log(2n-1)\right]}{N}\right\}^2$$

$$= \frac{\Sigma\left\{f_n\left[\log(2n-1)\right]^2\right\}}{N} - C^2$$

In Table I, $x = (2n-1)$ so that,

$$\log x = \log(2n-1)$$

and $$(\log x)^2 = \left[\log(2n-1)\right]^2$$

For each step, 1, 2, 3 . . . . n, the corresponding values of x are 1, 3, 5 . . . . $2n-1$. Note that x as used in the table is not the same as x as used in the deduction of the formulae.

The figures in Table I are accurate to ten places of decimals.

The log x column consists simply of the logarithms of odd numbers from one to one hundred. K was computed by subtracting log 2 from the logarithm of each number indicated in the "step interval" column. The $(\log x)^2$ column was computed by squaring fifteen place logarithms with the aid of calculating machines. This had to be done by indirect methods through the use of the simple algebraic relationship, $(a+r)^2 = a^2 + 2ar + r^2$, where a is the first part of the number and is the remainder. The table was computed by two different persons and checked on two different calculating machines by each person.