THE DISTRIBUTION OF THE LOGARITHM OF SURVIVAL TIMES WHEN THE TRUE LAW IS EXPONENTIAL

By J. O. IRWIN

(With 2 Figures in the Text)

In a recent paper Withell (1942) has shown that for a wide range of micro-organisms and disinfectants or poisons the logarithms of the survival times are approximately normally distributed. Even when the number of survivors is adequately represented by an exponential function of the time (e^{-kt}) say, the former hypothesis still gives approximately correct results. This suggests that in many cases the data are not good enough to distinguish between the two hypotheses (1) of a constant force of mortality k and (2) of a normal distribution of the logarithms of survival times. It is worth while, therefore, to examine the form of the distribution of the logarithms of survival times when the exponential law is true, and to see how nearly normal it is. We shall show that except for the position of the mean, this distribution is independent of k.

Let the number of deaths between t and $t+dt$ be

$$df = ke^{-kt}\,dt, \tag{1}$$

and let

$$\log_e t = x,$$

then

$$df = ke^{x-ke^x}\,dx \tag{2}$$

gives the distribution of x.

The mode of this distribution is given by $e^x = 1/k$. Transferring the origin to the mode by writing $x = X - \log k$ we reach

$$df = e^{X-e^X}\,dX, \tag{3}$$

which is independent of k.

Except for the sign of X this distribution is identical in form with one of the limiting forms for the distribution of the greatest value in a sample when the sample size becomes large. Its properties have been discussed by Fisher & Tippett (1928) and Gumbel (1934, 1937). The mean value of X is $-\gamma$, where γ is Euler's constant (0·57722...), its variance is $\frac{1}{6}\pi^2$, the standard deviation $\pi/\sqrt{6} = 1{\cdot}28225$. The median is $\log_e(\log_e 2) = -0{\cdot}36651$. These properties may be proved as follows:

For the mean we have

$$\mu_1' = \int_{-\infty}^{\infty} Xe^{X-e^X}\,dX = \int_0^{\infty} (\log u)\,e^{-u}\,du, \tag{4}$$

while

$$\mu_2' = \int_{-\infty}^{\infty} X^2e^{X-e^X}\,dX = \int_0^{\infty} (\log u)^2\,e^{-u}\,du, \tag{5}$$

https://doi.org/10.1017/S0022172400035518 Published online by Cambridge University Press

on putting $X=\log u$. Now

$$\Gamma(z)=\int_0^\infty e^{-u}u^{z-1}\,dz$$

and

$$\frac{d}{dz}\Gamma(z)=\int_0^\infty \log u e^{-u}u^{z-1}\,dz, \tag{6}$$

$$\frac{d^2}{dz^2}\Gamma(z)=\int_0^\infty (\log u)^2 e^{-u}u^{z-1}\,dz. \tag{7}$$

But (Whittaker & Watson, 1915) it is known that

$$\frac{d}{dz}\log\Gamma(z)=-\gamma-\frac{1}{z}+z\sum_{n=1}^{\infty}\frac{1}{n(z+n)}, \tag{8}$$

$$\frac{d^2}{dz^2}\log\Gamma(z)=\sum_{n=0}^{\infty}\frac{1}{(z+n)^2}, \tag{9}$$

and since

$$\frac{d}{dz}\Gamma(z)=\Gamma(z)\frac{d}{dz}\{\log\Gamma(z)\}, \tag{10}$$

$$\frac{d^2}{dz^2}\Gamma(z)=\Gamma(z)\frac{d^2}{dz^2}\{\log\Gamma(z)\}+\Gamma(z)\left[\frac{d}{dz}\{\log\Gamma(z)\}\right]^2 \tag{11}$$

putting $z=1$ we find from (8) and (9)

$$\left[\frac{d}{dz}\log\Gamma(z)\right]_{z=1}=-\gamma,\quad \left[\frac{d}{dz}\Gamma(z)\right]_{z=1}=-\gamma, \tag{12}$$

$$\left[\frac{d^2}{dz^2}\log\Gamma(z)\right]_{z=1}=\frac{\pi^2}{6},\quad \left[\frac{d^2}{dz^2}\Gamma(z)\right]_{z=1}=\gamma^2+\frac{\pi^2}{6}. \tag{13}$$

Hence $\mu=-\gamma$ and $\sigma^2=\frac{1}{6}\pi^2$.

For the median $$\int_{-\infty}^{X} e^{X-e^X}\,dX=(1-e^{-e^X})=\tfrac{1}{2},$$

or

$$X=\log_e(\log_e 2). \tag{14}$$

In comparing the normal distribution of log survival time with the true distribution, the normal distribution has been centred at the median M, as is usual in work of this kind. Table 1 gives for values of X from $-5\cdot5$ to $+2\cdot0$, the standardized deviate $(X-M)/\sigma$, where $M=-0\cdot36651$, $\sigma=\pi/\sqrt{(6)}$, and the corresponding cumulative percentage of survivors calculated from the normal approximation and the true distribution. It gives also the normal equivalent deviations corresponding to the true percentage of survivors. In Fig. 1 the standardized deviates for given X are compared with the normal equivalent deviations. In Fig. 2 the distribution of deaths at equal intervals of X are compared, on the two hypotheses. From Table 1 or Fig. 1 it may be seen that the normal approximation almost always slightly underestimates the percentage of survivors (between $X=1\cdot0$ and $X=-1\cdot5$ the reverse is true); but it would clearly require very good data to distinguish between the two hypotheses. We should, in many cases, be doubtful with actual observations

https://doi.org/10.1017/S0022172400035518 Published online by Cambridge University Press

whether the deviations from the straight line were not simply sampling fluctuations. If we were fortunate enough to obtain a curve always concave to the base, the systematic nature of the deviations might lead us to suspect a departure from normality, but this would hardly happen in practice owing to errors of sampling.

With an initial count of 300 organisms, in fact, we could not discriminate between the two hypotheses for values of X less than 0·5, that is, for survival

Table 1. *Cumulative percentage of survivors*

Survival time Mean S.T.	X	$-(X-M)/\sigma$	% survivors: Normal approximation	% survivors: Exact distribution	Normal equivalent deviation for exact distribution
0·00408	−5·5	+4·003	100·00	99·59	2·645
0·00675	−5·0	+3·613	99·99	99·33	2·472
0·0111	−4·5	+3·223	99·94	98·90	2·289
0·0183	−4·0	+2·833	99·77	98·18	2·094
0·0302	−3·5	+2·443	99·27	97·02	1·884
0·0498	−3·0	+2·053	98·00	95·14	1·659
0·0821	−2·5	+1·664	95·19	92·12	1·413
0·135	−2·0	+1·274	89·86	87·34	1·143
0·223	−1·5	+0·884	81·16	80·00	0·842
0·368	−1·0	+0·494	68·93	69·22	0·502
0·607	−0·5	+0·104	54·15	54·52	0·114
1·000	0	−0·286	38·75	36·79	−0·338
1·649	0·5	−0·676	24·96	19·23	−0·869
2·718	1·0	−1·066	14·33	6·60	−1·506
4·481	1·5	−1·455	7·28	1·13	−2·279
7·389	2·0	−1·845	3·25	0·062	−3·2
12·18	2·5	−2·235	1·27	0·001	−4·3

$X = \log_e \{(\text{survival time})/(\text{mean survival time})\}$.
$M = (\text{median value of } X) = \log_e (\log_e 2)$.
$\sigma^2 = (\text{variance of } X) = \frac{1}{6}\pi^2$.

times less than 1·65 times the mean survival time. We may reasonably suppose the sampling distribution of individual counts to be Poisson. With a true initial count of N organisms and a true later count of Np organisms, the standard error of the estimate of p is $\sqrt{(p\,(1+p)/N)}$. If $N=300$, $p=95$ %, the standard error is 7·9 %; with $p=19·23$ % it is 2·8 %, and 24·96 % is at just about the 5 % level of significance. Beyond this point an examination of the appropriate Poisson distributions leaves no doubt that the difference between the two hypotheses could usually be detected.

If the normal curve had been centred at the mean instead of the median and standardized deviates from the mean had been calculated, the straight line in Fig. 1 would have been shifted downwards, parallel to itself, so as to pass through the point ($X=-0·577$, $y=0$). This would have given a worse fit at the centre, but a better fit in the tails. In addition to being less logical (since it seems reasonable to measure both sets of deviations from the same point), this procedure would have made any discrepancy between the two hypotheses harder to detect.

Withell has estimated the value of the standard deviation of the logarithm of the survival time (Gaddum's λ or our σ) (a) for sixteen observed curves

https://doi.org/10.1017/S0022172400035518 Published online by Cambridge University Press

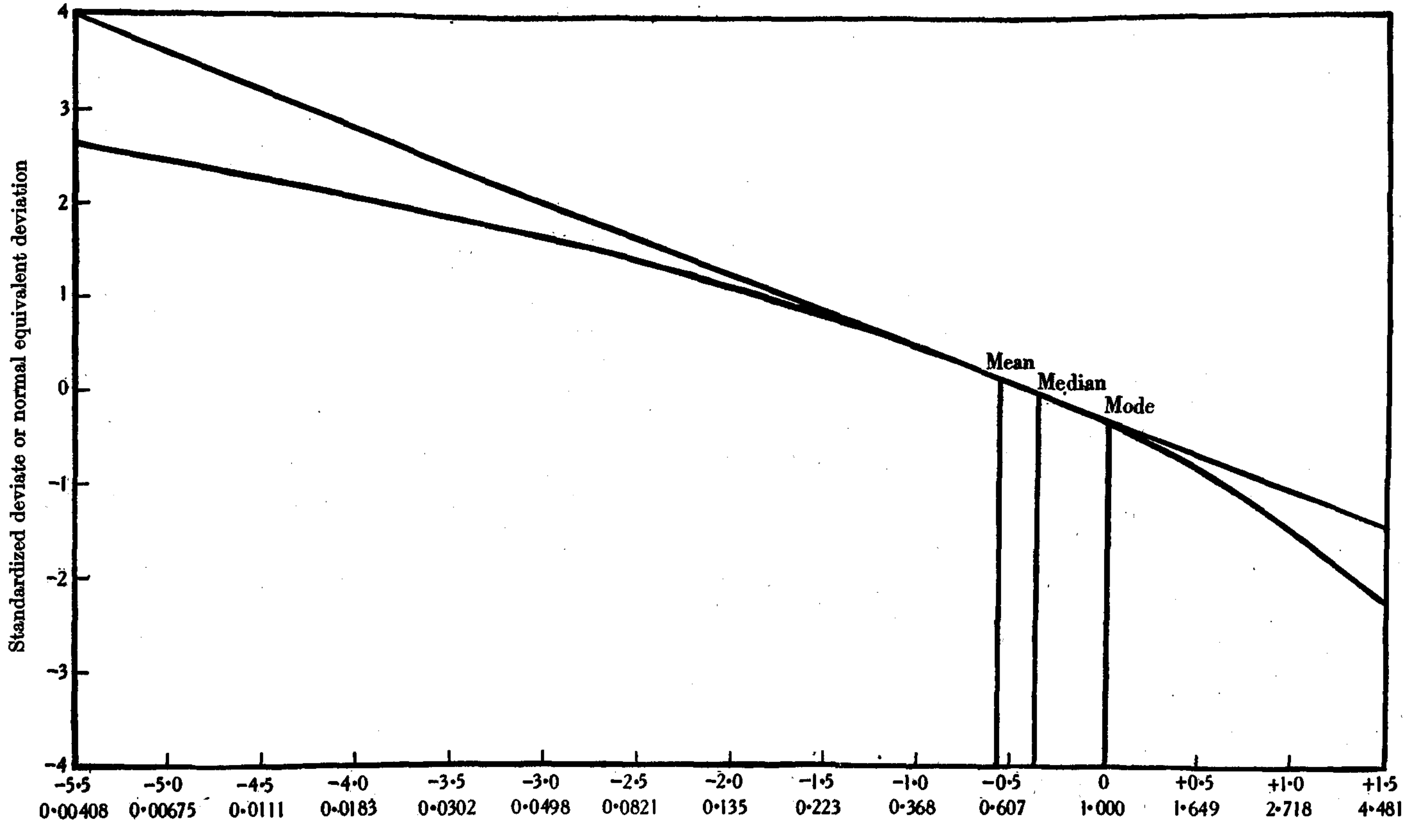

Fig. 1. Approximation to normality of the distribution of log survival time, when the true law is exponential.

First horizontal scale $= \log_e$ (survival time) $- \log_e$ (mean survival time).
Second horizontal scale = ratio of survival time to mean survival time.
Straight line = standardized deviate $(X - M)/\sigma_X$ ($X = \log_e$ survival time, M = median value of X).
Curve = normal equivalent deviation corresponding to true percentage of survivors.
Positions of mean, median and mode of $\log_e$ survival time are shown.

https://doi.org/10.1017/S0022172400035518 Published online by Cambridge University Press

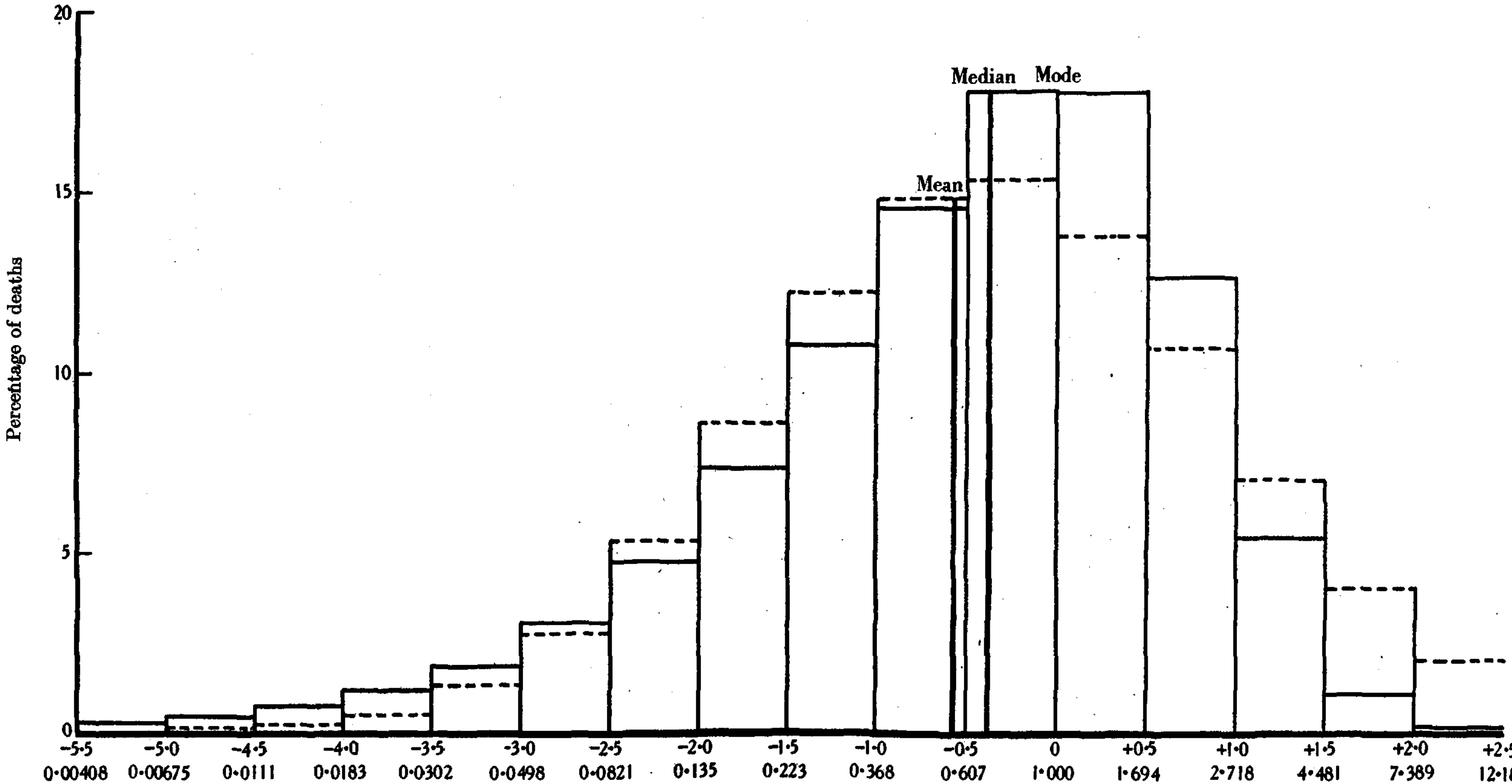

Fig. 2. Approximation to normality of the distribution of log survival time, when the true law is exponential.

First horizontal scale $= \log_e$ (survival time) $- \log_e$ (mean survival time).
Second horizontal scale = ratio of survival time to mean survival time.
Continuous histogram gives true frequencies of death for exact distribution of ($X = \log_e$ survival time).
Dotted histogram gives the frequencies from the normal approximation.
Normal curve is centred at the median of the true distribution.

https://doi.org/10.1017/S0022172400035518 Published online by Cambridge University Press

which he classes as exponential and (*b*) for six curves which he describes as 'exponential with lag'. In terms of logarithms to the base 10, the average value of λ for the former is 0·492, for the latter 0·317, while $\pi/\sqrt{(6)}\log_e 10$ is equal to 0·557. Withell estimated his standard deviations from half the difference of the deviates corresponding to 16 and 84 % of survivors, which is only correct for a normal distribution. More exactly these values are 15·87 and 84·13 %, and when the exponential law is true values of the deviates from the mode of the log-time curve corresponding to these percentages are found by equating them in turn to $(1-e^{-e^x})$. This gives $\lambda=0{\cdot}504$ against Withell's observed values of 0·492, a close enough agreement.

REFERENCES

Fisher, R. A. & Tippett, L. H. C. (1928). *Proc. Camb. Phil. Soc.* **24**, 180.
Gumbel, E. J. (1934). *Ann. Inst. Poincaré*, **5**, 115.
Gumbel, E. J. (1937). *La Durée Extrême de la Vie Humaine.* Paris: Hermann et Cie.
Whittaker, E. T. & Watson, G. N. (1915). *Modern Analysis*, 2nd ed. Camb. Univ. Press.
Withell, E. R. (1942). *J. Hyg., Camb.*, **42**, 124.

(*MS. received for publication* 3. iii. 42.—Ed.)