JIA 71 (1943) 0284-0291

NOTES ON THE POISSON FREQUENCY DISTRIBUTION

BY G. J. LIDSTONE, LL.D., ETC.

IN his recent important paper [this Volume, pp. 5–47] Mr H. L. Seal has incidentally called attention to Poisson's Frequency Distribution, hitherto rather unaccountably neglected by actuaries, and has pointed out that for a class of cases in which the Normal Law is of little use Poisson's form gives a close and useful approximation to the binomial distribution. Poisson showed [*Recherches sur les Probabilités des Jugements* (1837), § 81, p. 205] that if q is the chance that an event will happen at a single trial, the following expression will closely represent the chance that it will happen θ times in E trials if q is very small and E is large: viz.

$$e^{-m}\, m^{\theta}/\theta! \equiv \psi(\theta), \text{ say.}$$

Examples are given in Seal's Table 3 (*a*), loc. cit. pp. 43–4. On examining this Table Dr A. C. Aitken, F.R.S., pointed out to the writer that Poisson's function $\psi(\theta)$ is the first term in a series representing the probability or frequency in question, viz.

$$\psi(\theta) + B_2\,\Delta^2\psi(\theta-2) + B_3\,\Delta^3\psi(\theta-3) + B_4\,\Delta^4\psi(\theta-4) + \ldots,$$

which is Charlier's Type B as applied to the case where the observations are discrete, not continuous [cf. Elderton's *Frequency Curves*..., 3rd ed., p. 131; Aitken's *Statistical Mathematics* (reviewed by Seal, *J.I.A.* Vol. LXXI, p. 172), pp. 63 and 66]. The differences of the form $\Delta^t\psi(\theta-t)$ are the backward differences of $\psi(\theta)$ with the signs of the odd differences changed. In forming these differences for the earliest values of θ it must be assumed that $\psi(-1)$, $\psi(-2)$, ... exist and are equal to zero. The derivation of the formula is given later: the values of the B's are found to be

$$B_2 = -\tfrac{1}{2}qm = -\tfrac{1}{2}q^2\,E,$$

$$B_3 = +\tfrac{1}{3}q^2 m = +\tfrac{1}{3}q^3\,E,$$

$$B_4 = +\tfrac{1}{8}(q^2 m^2 - 2q^3 m) = \tfrac{1}{8}q^2 m\,(m-2q).$$

....

It will be convenient to use the symbol $\psi_2(\theta)$ to represent the sum of the first two terms of the series up to the term in Δ^2; $\psi_3(\theta)$ for the sum up to the term in Δ^3; and so on.

Aitken remarks that by using in place of Poisson's $\psi(\theta)$ the second approximation

$$\psi_2(\theta)=\psi(\theta)-\tfrac{1}{2}qm\Delta^2\psi(\theta-2)$$

the accuracy of the approximation to the binomial distribution is greatly increased, and therefore the range of cases over which the approximation is sufficient is greatly extended. Aitken applied this formula, $\psi_2(\theta)$, to the first four examples in Seal's Table 3 (*a*), and the writer has applied it to the remaining two examples (p. 44) and also to the examples given by Elderton, *J.I.A.S.S.* Vol. 1, No. 2, p. 47. It is unnecessary to take up space by tabulating the results in detail, since they can be simply described.

(1) In the case of Seal's last example, $q=\cdot 1$, $E=100$, $m=10$, the high value of q would suggest that the formula would be unsuitable and we should hardly expect good results: but in fact the $\psi_2(\theta)$ values are correct to 3 places, the maximum error being 6 in the fourth place, and on the whole they are nearly as good as the Type III values.

(2) In all the other examples (Seal's and Elderton's) the $\psi_2(\theta)$ values are as good as or rather better than the less convenient Type III values, and are in fact nearly accurate in the 4th place with occasional deviations up to 2 in that place, probably due largely to forcing of the last figure. The close approximation—as in Type III but not in the case of the Normal Curve—extends to the extreme tails of the distribution, indeed such slight errors as exist are most marked near the mode, where they are relatively unimportant.

These results are very remarkable and suggest that the $\psi_2(\theta)$ approximation may apply over a much more extensive range than is generally associated with the Poisson distribution, so that it may help appreciably in covering the range over which the Normal Curve is unsuitable: especially when it is considered that for many purposes actual or even approximate accuracy in the *third* place is sufficient. It in any case appears that the Tables of $\psi(\theta)$ given for values of m up to 15 in Table LI of *Tables for Statisticians*..., Part I, could usefully be extended very considerably, and second differences be given. It does not seem practicable to investigate theoretically the range of sufficient convergence. It would therefore be very useful to make a series of numerical tests, with specimen values of q and m, in this way "testing to destruction" as the engineers say—i.e. until the approximation is seen to break down. Since B_3 and B_4 are of the order of $q^2 m$ and $q^2 m^2$ respectively, it might appear at first sight that the accuracy of the $\psi_2(\theta)$ approximation will diminish not only as q increases but also as m increases. It will, however, be found that (as pointed out to the writer by Mr Seal) for a fixed value of q an increase in m *improves* the approximation, because $\Delta^3\psi$ and $\Delta^4\psi$ are reduced in greater proportion than B_3 and B_4 are increased. For example, the writer finds that if $q = \cdot 01$, $E = 10{,}000$, $qE = m = 100$, the $\psi_2(\theta)$ values are virtually correct to 6 places, and the unadjusted $\psi(\theta)$ values to at least 4 places.

[*Added in proof.*] The following formulae, contributed by Dr Aitken after reading the paper in proof, illustrate the foregoing remarks and add to their precision. It is easily shown that at the point $\theta = m$ we have (taking m integral)

$$\Delta^3\psi(m-3) = -\frac{2}{m^2}\psi(m),$$

$$\Delta^4\psi(m-4) = \left(\frac{3}{m^2} - \frac{6}{m^3}\right)\psi(m).$$

Now it is shown in a later section of the paper that

$$\psi(m) \doteq 1/\sqrt{2m\pi} \doteq \tfrac{2}{5}/\sqrt{m}.$$

Hence we find

$$B_3\,\Delta^3\psi(m-3) \doteq -\tfrac{4}{15}q^2/m^{\frac{3}{2}},$$

$$B_4\,\Delta^4\psi(m-4) \doteq \tfrac{3}{20}q^2/m^{\frac{1}{2}}, \text{ ignoring small terms.}$$

These results show that, at least near the mode, the Δ^3 and Δ^4 terms vary as q^2 but diminish as m increases: also the Δ^3 term is small compared with the Δ^4 term if m is at all considerable, as it practically always is in a mortality experience. It is not easy to determine the position and magnitude of the *maximum* numerical values of $\Delta^3\psi$ and $\Delta^4\psi$, but it seems safe to assume that they cannot greatly *exceed* the value of $\Delta^4\psi(m)$.* Hence the following table of

* For any function u_x we have the following very rough inequalities:

$$|\Delta^3 u_x| \ngtr 2\,|\Delta^2 u|, \quad \text{since } \Delta^3 u_x = \Delta^2 u_{x+1} - \Delta^2 u_x,$$

$$|\Delta^4 u_x| \ngtr 4\,|\Delta^2 u|, \quad \text{since } \Delta^4 u_x = \Delta^2 u_{x+2} - 2\Delta^2 u_{x+1} + \Delta^2 u_x,$$

where $\Delta^2 u$ is the greatest of the 2nd-differences involved.

$\frac{3}{20}q^2/m^{\frac{1}{2}}$ seems to supply a sufficient guide to the range of values over which the $\psi_2(\theta)$ approximation is adequate. The satisfactory result seems to be that it is so at all points of a mortality experience except the most extreme old ages.

Table: Values of $10^5 \times \frac{3}{20}q^2/m^{\frac{1}{2}}$

q	$m=1$	$m=4$	$m=9$	$m=100$
·01	2	1	1	·
·02	6	3	2	1
·03	14	7	5	1
·04	24	12	8	2
·05	38	19	13	4
·06	54	27	18	5
·07	75	37	25	7
·08	96	48	32	10
·09	122	61	41	12
·10	150	75	50	15

In view of what has been remarked above as to the relative magnitude of the coefficients B_3 and B_4 it would appear that if a higher degree of approximation is required $\psi_3(\theta)$ will not be very effective, and we should pass at one step from $\psi_2(\theta)$ to $\psi_4(\theta)$. This is quite practicable, though rather laborious if the distribution is a long one.

THE DERIVATION OF THE SERIES

Prof. Charles Jordan, in his *Statistique Mathématique* (1927), § 38, p. 98, gives the series in variant forms and with different proofs: he records the coefficients up to B_6. Aitken, in his *Statistical Mathematics*, p. 66, indicates how the series may be obtained by the powerful method of Generating Functions, discussed in the earlier part of his book [see also Soper's *Frequency Arrays*], but the proof is not given in full. It is learned from Aitken that B_n may be shown to be equal to the coefficient of α^n in the expansion of the exponential

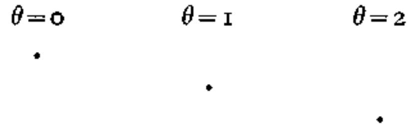
$$\exp\left[-\tfrac{1}{2}qm\alpha^2+\tfrac{1}{3}q^2m\alpha^3-\tfrac{1}{4}q^3m\alpha^4\ldots\right],$$

from which the coefficients may be easily derived, by direct expansion or by the use of Formula II, *J.I.A.* Vol. LI, p. 40.

THE NATURE OF THE POISSON DISTRIBUTION NEAR THE MODE

Let $qE \equiv m = n + f$, where n is the greatest integer contained in m and f is zero or a positive fraction less than unity.

If $n = 0$, $\psi(\theta)$ evidently decreases steadily as θ increases, i.e. the maximum term is $\psi(0)$, and the curve is of reversed-*J* form, thus:



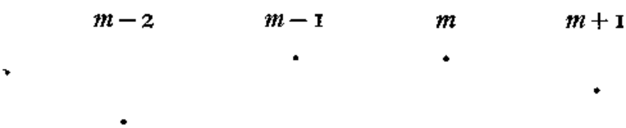
If $n > 0$, there are two cases:

(i) If $f = 0$, evidently $\psi(m-1) = \psi(m)$ and these are greater than any other terms. The successive terms $\psi(m-2)$, $\psi(m-1)$, $\psi(m)$, $\psi(m+1)$ are in the ratios of

$$\frac{m-1}{m} : 1 : 1 : \frac{m}{m+1}.$$

Hence $$\psi(m-2) : \psi(m+1) :: m^2 - 1 : m^2,$$

i.e. $\psi(m+1) > \psi(m-2)$, and the graph is like this:



(ii) If $f > 0$ the greatest term is $\psi(n)$. The terms $\psi(n-1)$, $\psi(n)$ and $\psi(n+1)$ are in the ratios

$$\frac{n}{n+f} : 1 : \frac{n+f}{n+1}.$$

Hence

$$\psi(n-1) : \psi(n+1) :: \frac{n}{n+f} : \frac{n+f}{n+1}$$

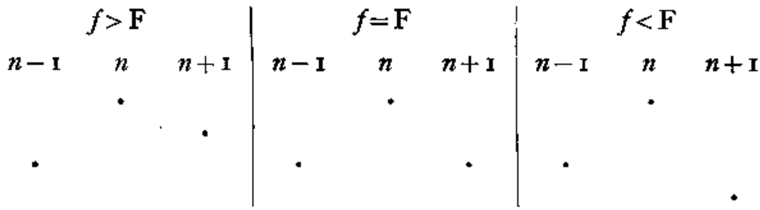
$$:: n^2 + n : n^2 + 2nf + f^2.$$

Now $$n^2 + n > = < n^2 + 2nf + f^2$$

as $$f^2 + 2nf < = > -n,$$

i.e. as $$f < = > [(\sqrt{(n^2+n)} - n) \div \tfrac{1}{2} - 1/8\,(n + \tfrac{1}{2})].$$

Practically, for any considerable value of n, $\psi(n+1)$ is greater or less than $\psi(n-1)$ as f is greater or less than $\frac{1}{2}$. Using F to represent $\frac{1}{2}-1/8(n+\frac{1}{2})$, the graph is like this:



In all cases the maximum value is at the point n or the two adjacent points $(n-1)$ and n. This maximum value is

$$e^{-(n+f)}\cdot\frac{(n+f)^n}{n!}=\frac{e^{-(n+f)}\cdot n^n\cdot(1+f/n)^n}{n!}$$
$$=\frac{e^{-(n+f)}\cdot n^n\cdot e^{f-\frac{1}{2}f^2/n\ldots}}{n!}.$$

Using Stirling's formula for $n!$ this reduces to

$$\frac{1}{\sqrt{2\pi n}}\left(1-\frac{1}{12n}-\frac{1}{2}\cdot\frac{f^2}{n}\ldots\right),$$

which is very approximately equal to

$$1/\sqrt{2\pi\left(n+\tfrac{1}{6}+f^2\right)}.$$

The approximation improves rapidly as n increases, but even for $n=2$ the approximate value is correct to 3 decimal places.

THE SIGN OF $\Delta^2\psi(\theta)$

We shall now find the points at which $\Delta^2\psi(\theta)$ changes sign. Consider the second central difference $\delta^2\psi(\theta)=\Delta^2\psi(\theta-1)$. This may be written as

$$\left[\frac{\theta}{m}-2+\frac{m}{\theta+1}\right]\psi(\theta)$$
$$=[\theta^2+\theta-2m\theta-2m+m^2]\,\psi(\theta)/m(\theta+1).$$

The roots of the coefficient in [ ] are

$$\theta=m-\tfrac{1}{2}\pm\sqrt{m+\tfrac{1}{4}}.$$

Thus this coefficient, and hence also $\delta^2\psi(\theta)=\Delta^2\psi(\theta-1)$, is zero if θ takes either of these values, negative if θ lies between them, and positive if θ falls before the smaller root or beyond the larger root.

All the above results may be tested by means of the values given in Table LI of *Tables for Statisticians...*, Part I.

THE MEDIAN

In the case of a discrete (non-continuous) distribution the "median" requires special definition. In order that it may have a perfectly definite value it will, for the present purpose, be defined as follows. If the distribution is y_x and

$$\tfrac{1}{2}=y_0+y_1+\ldots+y_t+\lambda y_{t+1} \quad (\lambda<1),$$

the median is $t+\lambda$. Thus $\begin{smallmatrix}\text{less}\\\text{more}\end{smallmatrix}$ than one-half of the total distribution will be represented by the sum of the first $\begin{smallmatrix}t\\t+1\end{smallmatrix}$ terms.

The eminent Indian mathematician Ramanujan enunciated, without proof, the remarkable result that, if m is integral,

$$\tfrac{1}{2}e^m=1+\frac{m}{1!}+\frac{m^2}{2!}+\ldots+\frac{m^{m-1}}{(m-1)!}+\lambda\frac{m^m}{m!},$$

or dividing by e^m

$$\tfrac{1}{2}=e^{-m}\left(1+\frac{m}{1!}+\frac{m^2}{2!}+\ldots+\frac{m^{m-1}}{(m-1)!}+\lambda\frac{m^m}{m!}\right),$$

where λ lies between $\frac{1}{3}$ and $\frac{1}{2}$. This shows at once that if in the Poisson distribution m is integral, the median as above defined is $(m-1+\lambda)$, i.e. falls between $(m-1)$ and m. It seems reasonable to assume that if m is not integral but (as before) is equal to $n+f$ the median will still be approximately $m-1+\lambda$ or $n+f-1+\lambda$: hence the median lies between $(n-1)$ and n or between n and $(n+1)$ as $f >$ or $<$ $(1-\lambda \fallingdotseq \frac{2}{3})$. These results are confirmed by Table LI, already referred to. As a rather extreme example, the median should be zero, i.e. $\psi(0)=\frac{1}{2}$, when m is about $\frac{2}{3}$, the true value being $\log_e 2 \fallingdotseq \cdot 693$.

Ramanujan's proposition was proved by Szegö [*J. Lond. Math. Soc.* Vol. III (1928), pp. 225–32] and by G. N. Watson [*Proc. Lond. Math. Soc.* 2nd Ser., Vol. XXIX (1929), pp. 293–308]. The latter also showed—subject to the truth of two unproved but reasonable hypotheses—that (as also enunciated by Ramanujan)

$$\lambda = \tfrac{1}{3} + \frac{4}{135\,(m+k)}, \quad k \text{ between } \tfrac{8}{45} \text{ and } \tfrac{2}{21},$$

so that λ tends rapidly to $\frac{1}{3}$ as m increases.

THE SUM OF t TERMS

Let $\Psi(t) = \sum_{\theta=0}^{t} \psi(\theta)$. Then

$$\Psi(t) = e^{-m}\left(1 + m + \frac{m^2}{2!} + \ldots + \frac{m^t}{t!}\right).$$

It is easily found that (all terms cancelling out except one)

$$\frac{d}{dm}\Psi(t) = -e^{-m} \,.\, m^t/t!$$

Hence, integrating from $m=0$ to $m=m$, and noting that at the lower limit $\Psi(\theta)=1$, we find

$$\Psi(t) = 1 - \int_0^m e^{-m} m^t \, dm/t! \qquad \ldots\ldots(\text{A})$$

$$= \int_m^\infty e^{-m} m^t \, dm/t! \equiv \int_m^\infty e^{-x} x^t \, dx/t! \qquad \ldots\ldots(\text{B})^*$$

In the notation of the well-known *Tables of the Incomplete Γ-Function*† the integral in (A) is $\mathrm{I}\,(m/\sqrt{t+1},\, t)$, so that those Tables afford the means of evaluating $\Psi(t)$, the chance that the event will happen not more than t times.

For the second approximation, $\psi_2(\theta)$, an easy adjustment is all that is required. We have

$$\sum_{\theta=0}^{t} \psi_2(\theta) = \sum_{\theta=0}^{t} [\psi(\theta) - \tfrac{1}{2} qm \, \Delta^2 \psi(\theta-2)]$$

$$= \Psi(t) - \tfrac{1}{2} qm \, \Delta^2 \Psi(\theta-2)$$

$$= \Psi(t) - \tfrac{1}{2} qm \, \Delta\psi(t-1),$$

since $\Delta\psi(-2)=0$.

* This result was given by Szegö, *loc. cit.*, and previously on p. viii of the Introduction to the Γ-*Function Tables*. [*For* $\mathrm{I}\,(p, x)$ *read* $\mathrm{I}\,(x/\sqrt{(p+1)},\, p)$.]

† See Review, *J.I.A.* Vol. LIII, p. 309. This unsigned review may be confidently attributed to the late Ralph Todhunter.