Biometrika Trust

An Explanation of Deviations from Poisson's Law in Practice

Author(s): Student

Source: *Biometrika*, Vol. 12, No. 3/4 (Nov., 1919), pp. 211-215

Published by: [Biometrika Trust](#)

Stable URL: <http://www.jstor.org/stable/2331767>

Accessed: 22/06/2014 11:06

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at
<http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Biometrika Trust is collaborating with JSTOR to digitize, preserve and extend access to *Biometrika*.

<http://www.jstor.org>

AN EXPLANATION OF DEVIATIONS FROM POISSON'S LAW IN PRACTICE.

By "STUDENT."

In her paper on the Poisson Law of small numbers, *Biometrika*, x, p. 36 *et seq.* Miss Whitaker after a very interesting analysis of the various attempts which have been made to test Poisson's Law on actual statistics concludes that "A general interpretation based on a very simple conception seems needed for those demographic cases in which the law of small numbers appears far more often to correspond to a negative than to a positive binomial."

The following is an attempt to explore the general question of what effect various departures from the conditions which lead to Poisson's Law have on the resulting statistics, and especially which conditions lead to positive and which to negative binomials when the exponential might at first sight be expected.

Poisson's Law has been applied to the occurrence of different numbers of individuals in divisions of space or time: thus of yeast cells in squares of a haemocytometer, of deaths from the kick of a horse in Prussian Army Corps which may be taken as individuals occurring in divisions of space, or of suicides of children per year in Prussia which are individuals occurring in divisions of time. In such cases it has been asserted that if the chance of an individual being found in a given division is so small that when multiplied by the very large number of individuals the product is still a reasonably small number, then the frequency of divisions containing 0, 1, 2 ... r individuals will be given by the terms of the exponential

$$Ne^{-m} \left\{ 1 + m + \frac{m^2}{2} + \dots + \frac{m^r}{r} + \dots \right\},$$

where N is the number of divisions and m the mean number of individuals occurring in a division.

For the above to be true it is necessary

- (1) That the chance of falling in a division is the same for each individual.
- (2) That the chance of an individual falling in it is the same for each division.
- (3) That the fact that an individual has fallen in a division does not affect the chance of other individuals falling therein.

As to these three conditions (1) is seldom or never true. I propose to show that this is generally unimportant; unless the chances of some individuals falling in a particular division are relatively high the Poisson law holds; the tendency however is towards a positive binomial.

Next (2) is comparatively seldom true except in the case of artificial divisions. The result of this, as Pearson has shown, is that a negative binomial fits the results better than the exponential.

Lastly (3) is often untrue. It will be shown that if the presence of an individual makes another less likely to fall into a division the positive binomial, but if more likely, the negative binomial will fit the figures best.

We may start from the fact that if the chance of an event happening be q and of its not happening p , then the chances of its happening 0, 1, 2, etc. times in n trials are given by the terms of the expansion of $(p + q)^n$, viz.

$$p^n : np^{n-1}q : \frac{n(n-1)}{2} p^{n-2}q^2 : \text{etc.}$$

As the moment coefficients of this series about the zero end of the range are

$$\nu_1 = nq,$$

$$\nu_2 = npq + n^2q^2 \text{ whence } \mu_2 = npq,$$

the binomial is completely determined if we know ν_1 and μ_2 for

$$p = \frac{\mu_2}{\nu_1} q = 1 - p = 1 - \frac{\mu_2}{\nu_1} \text{ and } n = \frac{\nu_1}{q} = \frac{\nu_1^2}{\nu_1 - \mu_2},$$

and in particular the binomial is positive (i.e. n and q are positive) if $\frac{\mu_2}{\nu_1} < 1$ and negative if $\frac{\mu_2}{\nu_1} > 1$. In the particular case when $\frac{\mu_2}{\nu_1} = 1$ the binomial becomes the Poisson exponential.

It is therefore unnecessary to deal with higher moments than the second for the purpose in hand.

Let us first consider the result of each individual having a different chance of falling in a given division:—

Let the chances of n individuals falling in a given division be $q_1, q_2, q_3 \dots q_n$. The chances of their not doing so are therefore $(1 - q_1), (1 - q_2), (1 - q_3) \dots (1 - q_n)$, and the chances that 0, 1, 2 ... n of them will fall in that division are given by the various terms of the expansion of

$$\{(1 - q_1) + q_1\} \{(1 - q_2) + q_2\} \{(1 - q_3) + q_3\} (\dots) \{(1 - q_n) + q_n\},$$

i.e. by

$$\begin{aligned} & (1 - q_1)(1 - q_2) \dots (1 - q_n) + S \{q_1(1 - q_2) \dots (1 - q_n)\} \\ & + S \{q_1q_2(1 - q_3) \dots (1 - q_n)\} + \dots + S \{q_1q_2q_3 \dots q_r(1 - q_{r+1}) \dots (1 - q_n)\} + \dots \\ & + q_1q_2q_3 \dots q_n, \end{aligned}$$

the term $S \{q_1q_2q_3 \dots q_r(1 - q_{r+1}) \dots (1 - q_n)\}$ giving the chance that exactly r individuals will fall in the division.

The sum of the above series is clearly unity so that the 1st and 2nd moment coefficients about the zero end of the series are given by two series of which the r th terms are

$$rS\{q_1q_2 - q_r(1 - q_{r+1}) \dots (1 - q_n)\} \text{ and } r^2S\{q_1q_2 \dots q_r(1 - q_{r+1}) \dots (1 - q_n)\}$$

respectively.

These series may be summed by rearranging them in the ascending order of the q products thus:

$$\begin{aligned} S\{q_1(1 - q_2)(1 - q_3) \dots (1 - q_n)\} &= S(q_1) - 2S(q_1q_2) + \dots + (-1)^{r-1}r.S(q_1q_2 \dots q_r) + \dots \\ 2S\{q_1q_2(1 - q_3)(1 - q_4) \dots (1 - q_n)\} &= 2S(q_1q_2) + \dots + (-1)^{r-2}r(r-1)S(q_1q_2 \dots q_r) + \dots \\ &\dots\dots\dots \\ tS\{q_1q_2 \dots q_t(1 - q_{t+1}) \dots (1 - q_n)\} &= tS(q_1q_2 \dots q_t) + \dots \end{aligned}$$

$$+ (-1)^{r-t} \frac{r \cdot \overline{r-1}}{\overline{t-1} \overline{r-t}} S(q_1q_2 \dots q_r) + \dots$$

$$rS\{q_1q_2 \dots q_r(1 - q_{r+1}) \dots (1 - q_n)\} = \dots\dots\dots r \cdot S(q_1q_2 \dots q_r) + \dots$$

Adding these we get on the left ν_1 and on the right $S(q_1) +$ a number of terms of the form $r(1 - 1)^{r-1}S(q_1q_2 \dots q_r)$ which accordingly vanish and we get

$$\nu_1 = S(q_1).$$

In a similar manner it can be shown that

$$\nu_2 = S(q_1) + 2S(q_1q_2),$$

and other moment coefficients about zero can be found in the same way, but we are not here concerned with them*.

If $\bar{q}$, $\bar{q}^2$ are the mean values of q and q^2 , obviously

$$\nu_1 = S(q_1) = n\bar{q} \dots\dots\dots (1),$$

and

$$\begin{aligned} \nu_2 = S(q_1) + 2S(q_1q_2) &= S(q_1) + \{S(q_1)\}^2 - S(q_1^2) \\ &= n\bar{q} + n^2\bar{q}^2 - n\bar{q}^2 \dots\dots\dots (2), \end{aligned}$$

$$= n\bar{q} + n^2\bar{q}^2 - n\bar{q}^2 - n\sigma_q^2 \dots\dots\dots (3),$$

$$\begin{aligned} \therefore \mu_2 &= n\bar{q} - n\bar{q}^2 - n\sigma_q^2 \\ &= n\bar{q} \left(1 - \bar{q} - \frac{\sigma_q^2}{\bar{q}}\right) \dots\dots\dots (4). \end{aligned}$$

* The moment coefficients are :

$$\mu_2 = n\bar{p}\bar{q} - n_q\mu_2,$$

$$\mu_3 = n\bar{p}\bar{q}(\bar{p} - \bar{q}) - 3n(\bar{p} - \bar{q})_q\mu_2 + 2n_q\mu_3,$$

$$\mu_4 = n\bar{p}\bar{q}\{1 + 3(n-2)\bar{p}\bar{q}\} - n\{7 + 6(n-6)\bar{p}\bar{q}\}_q\mu_2 + 12n(\bar{p} - \bar{q})_q\mu_3 - 6n_q\mu_4 + 3n^2_q\mu_2^2,$$

where ${}_q\mu_2$ etc. are the moment coefficients of the q distribution and $\bar{p} = 1 - \bar{q}$.

If now the distribution of chances is to be represented by the binomial $(P + Q)^N$. Then

$$Q = 1 - \frac{\mu_2}{\nu_1} = 1 - \frac{n\bar{q}(1 - \bar{q} - \sigma_q^2/\bar{q})}{n\bar{q}} \\ = \bar{q} + \frac{\sigma_q^2}{\bar{q}} \dots\dots\dots(5).$$

Since the original q 's are the chances of events happening they are always positive so that the above expression must be positive and the binomial positive.

If now we introduce the Poisson condition that $\bar{q}$ though positive is negligibly small (5) becomes in general zero, for σ_q is usually of the same order as $\bar{q}$, and in that case Poisson's law holds in spite of the inequality of the original q 's. If however $\frac{\sigma_q^2}{\bar{q}}$ is appreciably greater than zero (as in the extreme case

$$q_1 = \frac{1}{2} q_2 = q_3 = \dots = q_n = 0 \text{ when } \frac{\sigma_q^2}{\bar{q}} = \frac{n-1}{2n} = \frac{1}{2} \Big),$$

the distribution of chances is to be represented by a positive binomial.

Next we have to consider the effect of disregarding condition (2), namely that the chance of an individual falling into it must be the same for each division.

Let us suppose then that the q 's are all different for each division so that $n\bar{q}$ is also different.

Then writing m for $n\bar{q}$ and $\bar{m}$, $\bar{m}^2$, $\overline{nq^2}$ for the means of m , m^2 and nq^2 taken over all the divisions.

We get from (1) $\nu_1 = \bar{m} \dots\dots\dots(6),$

from (2) $\nu_2 = \bar{m} + \bar{m}^2 - \overline{nq^2}$
 $= \bar{m} + \bar{m}^2 + \sigma_m^2 - \overline{nq^2} \dots\dots\dots(7),$

$\therefore \mu_2 = \bar{m} + \sigma_m^2 - \overline{nq^2}^* \dots\dots\dots(8).$

As before if $(P + Q)^N$ is the best fitting binomial,

$$Q = 1 - \frac{\mu_2}{\nu_1} = \frac{\overline{nq^2} - \sigma_m^2}{m}.$$

Hence if $\sigma_m^2 > \overline{nq^2}$, which if there is any appreciable variation in m is probable, since as explained above nq^2 is generally negligible, a negative binomial will be found to fit better than the exponential.

Clearly condition (2) is usually not fulfilled in the vital and demographic statistics; divisions either of space or time are generally governed by different

* If we suppose that q does not vary with the individual but that $nq (=m)$ varies with the division, the moment-coefficients of the m distribution being written ${}_m\mu$, then the moment-coefficients of the resulting distribution of divisions are as follows :

$$\mu_2 = \bar{m} + {}_m\mu_2, \\ \mu_3 = \bar{m} + 3{}_m\mu_2 + {}_m\mu_3, \\ \mu_4 = \bar{m} + 3\bar{m}^2 + (7 + 6\bar{m}){}_m\mu_2 + 6{}_m\mu_3 + {}_m\mu_4.$$

environments which will vary the chances of an individual falling into them, and so we may expect that as a rule negative binomials will occur in place of the exponential.

Finally, suppose that the presence of an individual in a division influences the chance of other individuals falling in that division.

Clearly it may do so either by way of increasing the chance or diminishing it.

If the chance be increased it is clear that we shall get for the same mean number of individuals per division a larger number of divisions containing high numbers of individuals and a larger number of zero divisions. In other words, for the same mean we shall get a larger Standard Deviation, so that μ_2/ν_1 will be greater than 1 and a negative binomial will fit better than the exponential. On the other hand, if the chance of other individuals is decreased by the presence of one already in a division μ_2/ν_1 will become less than unity and the best fitting binomial will be positive. The first of these two cases includes linking or clumping of events or bacteria, the second such a thing as the counting of large cells on a haemocytometer whose divisions are comparable in size with them.

We have now shown that a population which might be expected at first sight to follow Poisson's law

(1) Will do so if the only deviation from the ideal conditions is that the chances of different individuals falling into the same division are not equal, as long as these chances are all small.

(2) If in addition to this the chances of some individuals are large a positive binomial will fit the results better than the exponential.

(3) If the different divisions have different chances of containing individuals, as is usual, a negative binomial will fit the results better than the exponential, except in so far as (2) may interfere.

(4) If the presence of one individual in a division increases the chance of other individuals falling into that division, a negative binomial will fit best, but if it decreases the chance a positive binomial.

Generally speaking (3) is the operating deviation from Poisson's conditions and accordingly most statistics give negative binomials.

Finally I should like to point out that the object of my original paper (*Biometrika*, Vol. v) was to give the user of the haemocytometer a guide to the error which he may expect from its use, and that the net result was that the probable error of his count was $\cdot 6745 \sqrt{N}$ where N was the total number counted* and that if N be a reasonably large number tables of the probability integral may be used, otherwise the exponential (or better still go on counting). This result is not affected by slight deviations from the Poisson law, any more than slight deviations from the normal law affect our use of the probability integral tables.

* *Biometrika*, Vol. v, p. 355. The probable error of mean is $\cdot 6745 \sqrt{m/M}$ where m is the mean and M the number of unit areas counted. If in this we put $M=1$, then $m=N$ and the total count is $N \pm \cdot 6745 \sqrt{N}$ as above.