Statistics & Probability Letters 1 (1982) 33–35
North-Holland Publishing Company

July 1982

Poisson on the Poisson Distribution

Stephen M. Stigler
Department of Statistics, University of Chicago, Chicago, IL 60637, U.S.A.

Received January 1982; revised version received May 1982

Abstract. A translation of the totality of Poisson's own 1837 discussion of the Poisson distribution is presented, and its relation to earlier work of De Moivre is briefly noted.

Keyword. Poisson distribution.

Simeon Denis Poisson (1781–1840) wrote widely on mathematics, mechanics, physics and probability, but he is best known today through the discrete probability distribution which bears his name. Curiously, the Poisson distribution appears but once in all of Poisson's works, and then on but a single page (Haight, 1967, p. 113). This reference is on p. 206 of Poisson's 1837 book, *Recherches sur la Probabilité des Jugements en Matière Criminelle et en Matière Civile Précédées des Règles Générales du Calcul des Probabilités*. The importance of the distribution today and the rarity of Poisson's book would seem to justify presenting a translation of that passage, together with sufficient other material to describe the context. The distribution appears in Chapter 3 as limit to the binomial. Actually, as will be seen below, Poisson derived the distribution directly as an approximation to the *negative* binomial cumulative distribution. There is no indication that he sensed the wide applicability of the distribution; rather, it was one of several approximations and received no special comment. Poisson may thus be cited as exemplifying both an aphorism of Whitehead ("*Everything of importance has been said before by somebody who did not discover it*" (Merton, 1968, p. 1)) and the Law of Eponymy ("*No scientific discovery is named after its original discoverer*" (Stigler, 1980)). Paradoxically, there are even stronger grounds for linking Poisson to the Cauchy distribution (Stigler, 1974).

In the passage to follow, E and F are complementary events, $p = \mathbf{P}(E)$, $q = \mathbf{P}(F)$, $p + q = 1$, and $\mu = m + n$ is the total number of trials.

Section 73, pp. 189–190 [1]. "I now return to the case where the chances p and q of the two events E and F are constant, and I shall consider the probability that in a number μ or $m+n$ of trials, E will happen at least m times and F at most n times. This probability will be the sum of the first m terms in the development of $(p+q)^\mu$, ordered in increasing powers of q; thus designating it by P we will have (Section 15) [2]

$$P = p^\mu + \mu p^{\mu-1} q + \frac{\mu(\mu-1)}{1 \cdot 2} p^{\mu-2} q^2 + \cdots + \frac{\mu(\mu-1)\cdots(\mu-n+1)}{1\cdot 2\cdot 3\cdots n} p^{\mu-n} q^n. \quad (8)$$

In this form it is difficult to transform the probability into an integral to which we can apply the method of Section 67, when m and n are very large numbers. [3] We thus seek an alternative expression

[1] All footnotes are supplied by the translator.

[2] Poisson had presented this same expression for the cumulative binomial distribution earlier in his book, in Section 15, on p. 54.

[3] In Section 67 (pp. 173–175) Poisson had first replaced $1\cdot 2\cdot 3\cdots n$ by the complete gamma integral; then, supposing n large, approximated the integrand by a multiple of a normal density, and by expanding the coefficient of this density in a series he proceeded to derive Stirling's formula.

for P that is better suited to this aim.

We may also describe the composite event we are concerned with as that where F does not happen more than n times in the μ trials. I shall call this event G. It can be seen to occur in any of the following $n+1$ cases:

Case 1. The first m trials all result in the event E; because then there only remain $\mu-m$ or n trials, which cannot produce F more than n times. The probability of this first case is p^m.

Case 2. The first $m+1$ trials produce E m times and F once, without F occurring last, or this second case would reduce to the first. It is evident that the following $n-1$ trials cannot produce F more than $n-1$ times, and thus this event cannot happen more than n times in the totality of the trials. The probability of the occurrence of E m times and F once in a specified position is $p^m q$, and since the specified position may be any of the first m, it follows that the probability of the second case favorable to G is $mp^m q$.

Case 3. The first $m+2$ trials produce E m times and F twice, without F occurring last, or this third case would reduce to one of the first two cases. The probability of the occurrence of E m times and F twice, in specified positions, is $p^m q^2$. Taking two-by-two the first $m+1$ positions for F, we have $\frac{1}{2}m(m+1)$ different combinations; the probability of the third case favorable to G is therefore $\frac{1}{2}m(m+1)p^m q^2$.

Continuing in this manner, we will finally arrive at the $(n+1)$st case, in which the μ trials produce E m times and F n times, without F occurring in the last position, or this case would reduce to one of the preceding cases; its probability is

$$\frac{m(m+1)(m+2)\cdots(m+n-1)}{1\cdot 2\cdot 3\cdots n}p^m q^n.$$

These $n+1$ cases being distinct from one another, and representing all the different ways in which the event G can occur, the complete probability will be the sum of their respective probabilities (Section 10)[4] and we have

$$P=p^m\left[1+mq+\frac{m(m+1)}{1\cdot 2}q^2+\frac{m(m+1)(m+2)}{1\cdot 2\cdot 3}q^3+\cdots+\frac{m(m+1)(m+2)\cdots(m+n-1)}{1\cdot 2\cdot 3\cdots n}q^n\right]. \quad (9)$$

This expression[5] agrees with formula (8), but has the advantage that it can be easily transformed to definite integrals whose numerical values can be calculated by the method of Section 67, to a better approximation when m and n are large numbers."

Section 81, pp. 205–207. "In the preceding calculation, we have excluded (Section 78)[6] the case where one of the two chances p and q is very small; in consequence it remains to consider this case.

I suppose that q is a very small fraction, or that the event F has a very weak probability. In a very large number μ of trials, the ratio n/μ of the number of times F happens to the number μ will also be a very small fraction. Putting $\mu-n$ in place of m in the formula (9), setting $q\mu=\omega$, $q=\omega/\mu$, and then neglecting the fraction n/μ, the quantity contained within the parentheses in this formula becomes

$$1+\omega+\frac{\omega^2}{1\cdot 2}+\frac{\omega^3}{1\cdot 2\cdot 3}+\cdots+\frac{\omega^n}{1\cdot 2\cdot 3\cdots n}.$$

At the same time, we will have

$$p=1-\omega/\mu,$$
$$p^m=(1-\omega/\mu)^{\mu}(1-\omega/\mu)^{-n}.$$

Here we can replace the first factor by the exponential $e^{-\omega}$, and the second by unity. Consequently, from formula (9) we will have, very nearly,

[4] Poisson had stated the law of total probability in Section 10, on p. 44.

[5] Thus Poisson has reexpressed the upper tail cumulative binomial probability (8) as a lower tail cumulative negative binomial probability (9). Actually, the form (9) is more thoroughly exploited in the omitted Sections 74–80 than in the derivation of the Poisson distribution we present.

[6] Sections 74–80 (pp. 190–205) have been concerned with asymptotic approximations to the probability P, including in Sections 78–80 a derivation of a normal approximation for the case where neither p nor q is 'very small'.

$$P = \left(1 + \omega + \frac{\omega^2}{1 \cdot 2} + \frac{\omega^3}{1 \cdot 2 \cdot 3} + \cdots + \frac{\omega^n}{1 \cdot 2 \cdot 3 \cdots n}\right) e^{-\omega}$$

for the probability that an event, whose chance on each trial is the very small fraction ω/μ, will not happen more than n times in a large number μ of trials.

In this case $n = 0$, this value of P reduces to $e^{-\omega}$; there is therefore the probability $e^{-\omega}$ that the event we are concerned with will not happen a single time in the μ trials, and consequently the probability $1 - e^{-\omega}$ that it will happen at least once, as we have already seen in Section 8.[7] When n is not a very small number, the value of P will differ very little from unity, as we may see by writing the preceding expression in the form

$$P = 1 - \frac{\omega^{n+1} e^{-\omega}}{1 \cdot 2 \cdot 3 \cdots (n+1)} \times \left(1 + \frac{\omega}{(n+2)} + \frac{\omega^2}{(n+2)(n+3)} + \cdots\right).$$

If we have, for example, $\omega = 1$, and we suppose that $n = 10$, the difference $1 - P$ will be nearly a hundred-millionth, and it is then nearly certain that an event whose very weak chance is $1/\mu$ on each trial will not happen more than 10 times in μ trials."

Poisson's derivation of his distribution, was foreshadowed by an analysis, De Moivre had presented over a century earlier. In a series of problems in his book, *The Doctrine of Chances*, De Moivre had sought (in Poisson's later notation) the value of μ for which the P of formula (8) or (9) was one-half, for various of n. For the case where $p/(1-p)$ was "*supposed infinite or pretty large in respect to unity*", he had expressed the solution, for example for the case $n = 3$, as given in terms of the root of the equation

$$z = \log 2 + \log\left(1 + z + \tfrac{1}{2}zz + \tfrac{1}{6}z^3\right)$$

where, in Poisson's notation, $z = \mu(1-p)/p = \omega$.

This equation can be seen to agree with

$$\tfrac{1}{2} = \left(1 + \omega + \tfrac{1}{2}\omega^2 + \tfrac{1}{6}\omega^3\right) e^{-\omega},$$

and some (e.g., Newbold, 1927, David, 1962, p. 168) have felt the distribution should be attributed to De Moivre. It must be admitted that Poisson added little to De Moivre's mathematical approximation, with which he was quite familiar, although one would have to stretch the point to claim the discrete distribution $e^{-\omega}\omega^n/n!$ is found in De Moivre. The relevant portion of De Moivre's work can be found as Problems 5–7 (pp. 14–21) of the first edition (1718), and in the recently reprinted second (1738) and third (1756) editions as Problems 3–5 (pp. 32–42 of the second edition, pp. 36–46 of the third edition). An account of the subsequent history of the distribution can be found in Chapter 9 of Haight (1967), which, however, overlooks an early work of Simon Newcomb's where the distribution is suggested as a fit to data for perhaps the first time (Newcomb, 1860).

[7] In Section 8 (pp. 40–41) Poisson had given the approximation $1-(1-p)^n = 1-e^{-np}$.

References

David, F.N. (1962), *Games, Gods and Gambling* (Griffin, London).

De Moivre, A. (1718, 1738, 1756), *The Doctrine of Chances*, The 2nd and 3rd editions have been recently reprinted, the latter in New York by Chelsea in 1967.

Haight, F.A. (1967), *Handbook of the Poisson Distribution* (Wiley, New York).

Merton, R.K. (1968), *Social Theory and Social Structure* (The Free Press, New York, enlarged ed.).

Newbold, E.M. (1927), Practical applications of the statistics of repeated events, particularly to industrial accidents, *J. Roy. Statist. Soc.* **90**, 487–547.

Newcomb, S. (1860), Notes on the theory of probabilities, *The Mathematical Monthly* **2**, 134–140.

Poisson, S. (1837), *Recherches sur la Probabilité des Jugements en Matière Criminelle et en Matière Civile Précédées des Règles Générales du Calcul des Probabilités* (Bachelier, Paris), a xerographic reprint is available from University Microfilms International, Ann Arbor, MI, U.S.A.

Stigler, S.M. (1974), Cauchy and the Witch of Agnesi: An historical note on the Cauchy distribution, *Biometrika* **61**, 375–380.

Stigler, S.M. (1980), Stigler's law of eponymy, *Trans. New York Acad. Sci. Ser. 2* **39**, 147–158.