Studies in the History of Probability and Statistics. XXV. On the history of some statistical laws of distribution

By O. B. SHEYNIN

Institute for the History of Natural Science and Technology, Academy of Sciences, Moscow

SUMMARY

The history of the χ^2 distribution and the contributions by Herschel, Maxwell and Boltzmann are discussed. The use of the beta distribution by Bayes is also described.

1. The use of the χ^2 distribution in physical theory

The deduction of the χ^2 distribution is due to Abbe (1863); see also Sheynin (1966). While considering the accuracy of archery, Herschel (1869) almost arrived at this distribution for two degrees of freedom, a fact mentioned to me in 1967 by Professor W. Kruskal in a private communication. For three degrees of freedom similar results were arrived at by Maxwell (1860) somewhat prior to Herschel. Boltzmann arrived at this distribution at first (1878) for two and three degrees of freedom and then (1881) in the general case.

Thus, these works should find their place in the history of the χ^2 distribution beside those of Laplace, Bienaymé and Ellis (Lancaster, 1966), and, of course, beside those of Abbe (Helmert, 1876); see also David (1957) and Pearson (1900).

Considering the probability of an arrow deviating by z from the centre of a circular target Herschel (1869, pp. 506–7) starts from an obvious formula

$$\operatorname{pr}(z|\phi) \propto e^{-kz^2}\,dz,$$

where ϕ is the polar angle of the point of hit, the polar coordinate system being (z, ϕ), and k is to be estimated from observations. Then, since the circumference of a circle is proportional to z,

$$\operatorname{pr}(z \leqslant z_0) \propto \int_0^{z_0} z\, e^{-kz^2}\,dz. \tag{1}$$

Herschel could have arrived at the χ^2 distribution for two degrees of freedom had he considered instead of (1) the probability that $z^2 \leqslant z_0$.

But the introduction of z^2 does not seem sufficiently natural. More natural, however, would have been a similar procedure of Maxwell (1860, p. 381), this being his classical paper where he arrived at the normal distribution of velocities of molecules

$$f(x) = \frac{1}{\alpha\sqrt{\pi}}\, e^{-x^2/\alpha^2} \quad (-\infty < x < +\infty). \tag{2}$$

It is not our intention to discuss his deduction of (2); what we notice is that Maxwell, starting from (2) and giving no intermediate explanation, observes that

$$\frac{4N}{\alpha^3\sqrt{\pi}}\, v^2\, e^{-v^2/\alpha^2}\,dv \tag{3}$$

molecules out of N, the whole number of them, have space velocities belonging to the segment $[v, v+dv]$. Reconstruction of (3) is easy enough: starting from (2), we have a similar distribution for the simultaneous probability of $\{\xi < x, \eta < y, \zeta < z\}$ and

$$\operatorname{pr}(\text{velocity} \in [v, v+dv]) = \int_0^{2\pi} d\phi \int_0^{\pi} \sin\Theta\, d\Theta \int_v^{v+dv} z^2\, e^{-z^2/\alpha^2}\,dz,$$

this being equivalent to (3). Here ξ, η and ζ are the components of velocity in each of the three dimensions.

Introducing probability proper, i.e. eliminating N, and studying the square of the velocity rather than the velocity itself, Maxwell could have found $\operatorname{pr}(\xi^2+\eta^2+\zeta^2 \in [v, v+dv])$ which would have been the χ^2 distribution for three degrees of freedom. This transformation would have been quite natural, for instance, in studying the kinetic energy (K.E.) of molecules. Maxwell, however, restricted himself to remarking that the mean value of v^2 is equal to $\frac{3}{2}\alpha^2$.

On the other hand, Boltzmann (1878, p. 252) studied just this distribution of the K.E. and arrived at the χ^2 distribution for two, and then (p. 257), for three degrees of freedom. Only in the first instance did he mention his proceeding from the normal law

$$\frac{k}{\pi}e^{-k(u^2+v^2)}\,du\,dv \tag{4}$$

and explain that the integration had been made with respect to the polar angle.

More interesting is that Boltzmann (1881, p. 576), also with no explanations given, writes down the distribution of the K.E. of molecules in the general case of a dimensions, i.e.

$$\mathrm{pr}\,(\tfrac{1}{2}mv^2 \in [\chi, \chi+d\chi]) = \frac{h^{\frac{1}{2}a}}{\Gamma(\frac{1}{2}a)}e^{-h\chi}\chi^{\frac{1}{2}a-1}\,d\chi. \tag{5}$$

The meaning of h, first introduced in 1878, is not given in either article but it should be equal to $2k/m$, where k is the same as in (4) and m, as in (5), is the mass of the molecule.

In deducing (5), Boltzmann may have used the discontinuity, Dirichlet, factor as both Abbe and Helmert did. But then, he would possibly have said so. And at least two other methods for the calculation of the a dimensional integral of the exponential function were open to him.

The lack of references to Abbe and Helmert testify to the fact that Boltzmann did not know their works.

None of the nineteenth-century authors mentioned thought of compiling a table of the χ^2 distribution. The mode of using statistical considerations in physics proper could be seen in Maxwell (1860) who was evidently satisfied with deducing several distributions and giving a corresponding overall numerical estimate of the behaviour of molecules.

In error theory, the criteria for estimating systematic influences, bias, in observations developed by Abbe and Helmert were nonparametric and the χ^2 distribution itself was not used.

2. The beta distribution

This distribution is due to Bayes (1764). The beta function had been repeatedly considered by several scholars prior to Bayes, but it was his probabilistic problem that led to the calculation of

$$\frac{\int_b^c x^p(1-x)^q\,dx}{\int_0^1 x^p(1-x)^q\,dx} = I_c(p+1, q+1) - I_b(p+1, q+1), \tag{6}$$

where $0 \leqslant b < c \leqslant 1$, p and q are large positive integers and I is the symbol for the incomplete beta function. Of course, Bayes used neither these terms nor the symbols.

For the denominator of (6) he easily deduced

$$\mathrm{B}(p+1, q+1) = \frac{1}{(q+1)\binom{p+q+1}{p}}$$

and, for the estimation of the whole fraction (6), used (Bayes, 1765) a supplementary curve

$$y = c\left(1 - \frac{n^2x^2}{pq}\right)^{\frac{1}{2}n} \quad (n = p+q).$$

A more detailed description, including an improvement of Bayes's estimate due to Price (see Sheynin, 1969) and, of course, including the related work of Laplace (*Théor. anal. prob.*, chap. 6 of book 2) would be of interest for the history of mathematical calculus. We, however, shall restrict ourselves to probabilistic problems proper.

The integral limit theorem in De Moivre could be written as

$$\mathrm{pr}\left[-z \leqslant \frac{\mu/n - p}{\{(pq)/n\}^{\frac{1}{2}}} \leqslant z\right] \to \frac{2}{(2\pi)^{\frac{1}{2}}}\int_0^z e^{-\frac{1}{2}x^2}dx, \tag{7}$$

where pq/n is the dispersion of μ/n, the relative frequency, the notation being that in standard use (which is not the case with Bayes).

Neither Bayes nor Price used this limit theorem while the latter observed, moreover (in the covering letter to Bayes, 1764) that it is not 'rigorously exact' for the case of a finite n. Price also noticed that the problem being solved by Bayes is the converse to that of De Moivre's. We shall now study this point.

If, notwithstanding the opinion of Price (and, obviously, Bayes), the Bayesian formula is to be considered in the limiting case, $n \to \infty$, it would give

$$\mathrm{pr}\left\{-z \leqslant \frac{\bar{p}-a}{(pq/n^{\frac{3}{2}})^{\frac{1}{2}}} \leqslant z\right\} \to \frac{2}{(2\pi)^{\frac{1}{2}}} \int_0^z e^{-\frac{1}{2}x^2}\,dx, \tag{8}$$

where $a = p/n$ and $\bar{p}$ is the estimate of p, the classical probability in the binomial scheme. In a somewhat different form (8) is to be found in a commentary by Timerding to the German translation of Bayes (1908). Lacking in this commentary, however, is the understanding that, as a first approximation,

$$a = E(\bar{p}), \quad pq/n^{\frac{3}{2}} = \mathrm{var}\,(\bar{p}).$$

Neither Bayes nor Price possessed any notion about dispersion but, what is remarkable, they both obviously understood the practical worthlessness of using (7) for the converse problem of deducing p from the relative frequency of the event studied.

It is generally known that the so-called Bayes's formula

$$\mathrm{pr}\,(A_i|B) = \frac{\mathrm{pr}\,(B|A_i)\,\mathrm{pr}\,(A_i)}{\sum\limits_{j=1}^{n} \mathrm{pr}\,(B|A_j)\,\mathrm{pr}\,(A_j)} \tag{9}$$

is not to be found in Bayes. Expressed in words only, (9) is contained in Laplace (*Essai philosophique*) and it is possibly Cournot (1843, § 88, p. 158) to whom this incorrect expression (Bayes's formula) is due.

Acknowledgement is due to Professor H. O. Lancaster for reprints of his own paper (1966) and that of David (1957) and to Professor W. Kruskal for permitting us to mention his own discovery of Herschel's connexion with the χ^2.

References

Abbe, E. (1863). Über die Gesetzmässigkeit der Vertheilung der Fehler bei Beobachtungsreihen. *Ges. Abh.* **2**, 55–81.

Bayes, T. (1764). An essay towards solving a problem in the doctrine of chances. Reprinted in *Biometrika*, 1958, **45**, 296–315.

Bayes, T. (1765). A demonstration of the second rule in the essay, etc. *Phil. Trans. R. Soc.* **54**, 296–325. [This includes a covering note and subsequent remarks by Price.]

Bayes, T. (1908). *Versuch zur Lösung eines Problems der Wahrscheinlichkeitsrechnung*. Herausgeber, H. E. Timerding. Ostwald Klassiker No. 169. Leipzig: Engelmann.

Boltzmann, L. (1878). Weitere Bemerkungen über einige Probleme der mechanischen Wärmetheorie. *Wiss. Abh.* **2**, 250–88.

Boltzmann, L. (1881). Über einige das Wärmegleichgewicht betreffende Sätze. *Wiss. Abh.* **2**, 572–81.

Cournot, A. A. (1843). *Exposition de la théorie des chances et des probabilités*. Paris: Hachette.

David, H. A. (1957). Some notes on the statistical papers of F. R. Helmert. *Bull. Stat. Soc. N.S.W.* **19**, 25–8.

Helmert, F. R. (1876). Über die Wahrscheinlichkeit der Potenzsummen der Beobachtungsfehler etc. *Z. Math. und Phys.* **21**, 192–218.

Herschel, J. F. W. (1869). *Familiar Lectures on Scientific Subjects*, pp. 496–507. New York: Routledge.

Lancaster, H. O. (1966). Forerunners of the Pearson χ^2. *Aust. J. Statist.* **8**, 117–26.

Maxwell, J. C. (1860). Illustrations of the dynamical theory of gases. *Scient. Papers*, vol. **1**, 377–410. Paris, 1927: Libr. scient. Hermann.

Pearson, K. (1900). On a criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Phil. Mag.* **50**, 157–75.

Sheynin, O. B. (1966). Origin of the theory of errors. *Nature, Lond.* **211**, 1003–4.

Sheynin, O. B. (1969). On the work of T. Bayes in the theory of probability (Russian). *Proc. 12th Conf. Inst. Hist. Nat. Sci. and Technology, Hist. of Math. and Mech. Sect.* 40–57. Moscow: Viniti publ.

[*Received August* 1970. *Revised October* 1970]

Some key words: History of normal, chi-squared and beta distributions; History of kinetic theory and Bayes's formula.