SOME EARLY INSTANCES OF CONFIDENCE STATEMENTS

By E. L. Lehmann

Statistical Laboratory, University of California, Berkeley,
ONR 5 Technical Report to the Office of Naval Research
prepared under Contract Nonr-222(43) (NR 042-036)

September, 1958

This paper may be reproduced in whole or in part
for any purpose of the United States Government.

SOME EARLY INSTANCES OF CONFIDENCE STATEMENTS

By E. L. Lehmann¹University of California, Berkeley

A general theory for obtaining limits within which an unknown parameter can be asserted to lie with known probability was developed by Fisher and Neyman in the 1930's.² The basic device through which such limits are obtained consists in "solving" a probability statement which involves an unknown parameter θ , for this parameter. The resulting statement concerning θ is then equivalent to the original statement and thus holds with the same known probability.

The above inversion procedure seems to have been employed for the first time by Laplace. In his *Théorie Analytique des Probabilités* (p. 287),³ he applies it to the approximate equation

$$P\left\{\left|\frac{i}{n} - p\right| \leq T \sqrt{2p(1-p)/n}\right\} = \frac{2}{\sqrt{\pi}} \int_0^T e^{-t^2} dt \quad (1)$$

+ a correction term of order $n^{-1/2}$

where i/n denotes the frequency of success in n binomial trials with probability p of success. Replacing $p(1-p)$ by $i(n-i)/n^2$ and solving for p , he thus obtains the approximate equation

$$P\left\{-\frac{T \sqrt{2i(n-i)}}{n \sqrt{n}} \leq p \leq \frac{T \sqrt{2i(n-i)}}{n \sqrt{n}}\right\} = \frac{2}{\sqrt{\pi}} \int_0^T e^{-t^2} dt. \quad (2)$$

This appears to be the first instance of what, at least formally, can be considered a confidence statement.

Since the passage is short, it may be worth giving it

verbatim; however, in the formulae I shall suppress all terms of order less than that of the leading term; in particular, I have replaced Laplace's expression x , which is equal to np plus a term less than 1, by np :

Si l'on connaît le nombre de fois que sur n coups l'événement a est arrivé, la formule (o)⁵ donnera la probabilité que sa facilité p , supposée inconnue, sera comprise dans les limites données. En effet si l'on nomme i ce nombre de fois, on aura, par ce qui précède, la probabilité que la différence $\frac{i}{n} - p$ sera comprise dans les limites $\pm T\sqrt{2p(1-p)}/n\sqrt{n}$; par conséquent, on aura la probabilité que p sera compris dans les limites

$$\frac{i}{n} \pm \frac{T\sqrt{2p(1-p)}}{n\sqrt{n}}.$$

La fonction $T\sqrt{2p(1-p)}/n\sqrt{n}$ étant de l'ordre $1/\sqrt{n}$, on peut, en négligeant les quantités de l'ordre $1/n$, y substituer i au lieu de np et $n-i$ au lieu de $n(1-p)$; les limites précédentes deviennent ainsi, en négligeant les termes de l'ordre $1/n$,

$$\frac{i}{n} \pm \frac{T\sqrt{2i(n-i)}}{n\sqrt{n}}$$

et la probabilité que la facilité de l'événement a est contenue dans ces limites est égale à

$$\frac{2}{\sqrt{\pi}} \int dt e^{-t^2}.$$

On voit ainsi que, à mesure que les événements se multiplient, l'intervalle des limites se resserre de plus en plus, et la probabilité que la valeur de p tombe dans ces limites approche de plus en plus de l'unité ou de la certitude. C'est ainsi que les événements, en se développant, font connaître leurs probabilités respectives.

Laplace does not comment directly on the interpretation of the probability statement (2). However, from the way this and related results are applied it seems clear that both he and his successors thought of p as the variable which with the stated probability falls in the fixed interval $\pm Ti(n-i)/n\sqrt{n}$. This is indicated also by Laplace's alternative derivation of essentially the same formula (pp. 377/378). In this second derivation he uses a Bayesian argument, assuming a uniform a priori distribution for p , and he clearly considers the results obtained by these two methods as completely equivalent.

The first derivation of (2) was taken over by Poisson into his *Recherches sur la Probabilité des Jugements* (1837). It is interesting that Poisson utilizes the result to obtain a fiducial distribution for p (p. 211). He gives the one-sided limits for $q = 1 - p$ corresponding to (2). Treating the totality of these limits as a cumulative distribution function, he differentiates it to obtain a formal probability density for q . In quoting the passage, I have changed the notation to make it correspond to that of Laplace:

La quantité Q , déterminée par la première formule (18)
 [that is, given up to terms of first order by $(1/\sqrt{\pi}) \int e^{-t^2} dt$],

étant la probabilité que le nombre $n-i$ est inférieur à $nq - T\sqrt{2npq}$, elle est également la probabilité que la chance inconnue q de l'événement F , arrivé $n-i$ fois dans n épreuves, est supérieure à $(n-i)/n + T\sqrt{2pq/n}$, ou bien à $(n-i)/n + (T/n)\sqrt{2i(n-i)/n}$, en substituant dans le second terme de cette limite, à la place de p et q , leur valeurs approchées i/n et $(n-i)/n$. Si l'on met $T-dT$ au lieu de T dans cette formule, et que l'on conserve seulement les infiniment petits du premier ordre, $Q - \frac{dQ}{dT} dT$ sera donc aussi la probabilité que q surpasse

$$\frac{i}{n} + \frac{T}{n} \sqrt{\frac{2i(n-i)}{n}} - \sqrt{\frac{2i(n-i)}{n}} \frac{dT}{n};$$

par conséquent, $-dQ/dT$ exprimera la probabilité infiniment petite que l'on a précisément

$$q = \frac{i}{n} + \frac{T}{n} \sqrt{\frac{2i(n-i)}{n}}$$

pour toutes les valeurs de T positives et très petites par rapport à $\sqrt{n}$, comme le suppose l'expression de Q .

Poisson's work was the basis for certain other writers, particularly Gavarret (1840) and Lexis (1875), who quote his statement of (2) which is obtained by Laplace's first method. However, generally Laplace's second (Bayesian) derivation was preferred as being on a firmer foundation. In fact, the first derivation was attacked by De Morgan (1837) as illegitimate; an attempt was made by Monro (1874) to legitimize it by showing it to be equivalent to the Bayesian derivation. Todhunter discusses both derivations in his *History of the Theory of Probability*

(1865, pp. 554-558). His own attitude is indicated by his remark: "As Poisson in his latest discussion of the question adopted the inversion of James Bernoulli's theorem,⁶ we may perhaps infer that he considered the amount of assumption thus involved to be no greater than that which is required in the use of Bayes' theorem." Both derivations are also given, without comment regarding any differences between them, by Czuber in his report on probability theory to the Deutsche Mathematiker Vereinigung in 1899 (pp. 82-84) and again in the third edition of his textbook on the theory of probability (1914, pp. 203-206).

For the first time as far as I have been able to determine, the proper interpretation of (2) is given by E. B. Wilson in 1927.⁷ He describes two procedures as the usual methods for assigning limits to a binomial probability p , one based on Tchebycheff's inequality, the other equivalent to (2). He then criticises this approach in the following passage in which p_0 stands for the frequency previously denoted by i/n : "Strictly speaking, the usual statement of probable inference as given above is elliptical. Really, the chance that the true probability p lies outside a specified range is either 0 or 1; for p actually lies within that range or does not. It is the observed rate p_0 which has a greater or less chance of lying within a certain interval of the true rate p . If the observer has the hard luck to have observed a relatively rare event and to have based his inference thereon, he may be fairly wide of the mark."

On the basis of Poisson's work, an extension of (2) was given by Lexis (1875, p. 106) to the case of two binomial probabilities.

After first testing the hypothesis that the two probabilities p and p' are equal,⁸ he wishes to estimate the difference between these probabilities. With

$$F_u = \frac{2}{\sqrt{\pi}} \int_0^u e^{-t^2} dt,$$

he obtains the approximate probability $F_3 = 99998$ that the difference between the frequency of successes i/n in the first and i'/n' in the second sequence of trials falls in the interval

$$p - p' \pm 3 \sqrt{\frac{2i(n-i)}{n^3} + \frac{2i'(n'-i')}{n'^3}}$$

and hence that $p-p'$ falls in the interval

$$\frac{i}{n} - \frac{i'}{n'} \pm 3 \sqrt{\frac{2i(n-i)}{n^3} + \frac{2i'(n'-i')}{n'^3}}.$$

There is another line of development, which utilizes the inversion of probability statements to obtain special cases of formally correct confidence intervals. In the theory of errors of observations, probable errors are frequently used by 19th century authors to obtain intervals within which the true value of a parameter lies with probability $1/2$, that is confidence intervals with confidence coefficient $1/2$. This application is made by Gauss in the first publication (1816) on the subject, in a paper concerned with the "determination of the accuracy of observations." It is clear that he suggests the value $1/2$ purely as a convention; at the beginning of the paper, he gives the limits which a normally distributed error will not exceed with probability $1/2$, but also those corresponding to the probabilities $1/10$ and $1/100$.

In a similar context, an even more explicit statement is given by Fourier (1826). Having previously discussed the concept of probable error, and having defined the quantities $A = (a+b+\dots+n)/m$, $B = (a^2+b^2+\dots+n^2)/m$, $g = \sqrt{2(B-A^2)/m}$, where $a, b, \dots, n$ are m observations of the "true" quantity H , he writes (Section XI):

"Pour compléter cette discussion, il faut déterminer quelle probabilité il y a que la quantité cherchée H est comprise entre des limites proposées $A + D$ et $A - D$. A est le resultat moyen qu'on a trouve, H est la valeur fixe que donnerait un nombre infini d'observations, et D est une quantité proposée que l'on ajoute a la valeur A ou que l'on retranche. La Table suivante fait connaître la probabilité P d'une erreur positive ou negative plus grande que D , et cette quantite D est le produit de g par un facteur proposee ∂

∂	P
0.47708	1/2
1.38591	1/20
1.98495	1/200
2.46130	1/2000
2.86783	1/20000

Chacun des nombres de la colonne P fait connaître quelle probabilité il y a que la valeur exacte H , qui est l'objet de la recherche, est comprise entre les limites $A+g\partial$ et $A-g\partial$"¹⁰

Although most later authors restrict their attention to the conventional value $\frac{1}{2}$, they sometimes mention in passing that the methods apply equally well to other values. Thus for example,

Jevons in his Principles of Science writes (2nd edition, 1883, p. 387):

"The probable error is taken by mathematicians to mean the limits within which it is as likely as not that the truth will fall. ... Any other limits might have been selected at will. We might calculate the limits within which it was one hundred or one thousand to one that the truth would fall; but there is a convention to take the even odds one to one, as the quantity of probability of which the limits are to be estimated."

The connection between the idea of probable error and statements of the type (2) is recognized by Lexis who advocates the use of a higher probability of covering the true value thereby coming closer to the modern point of view. In his discussion of (2) he writes (1875, pp. 99/100):⁹

...so ist mit der Wahrscheinlichkeit F_u die Unbekannte p höchstens um $\pm T/\sqrt{n} \sqrt{2i(n-i)}$ von i/n verschieden, oder es besteht mit der bezeichneten Wahrscheinlichkeit die Grenzgleichung

$$p = \frac{i}{n} \left(\pm \frac{T}{\sqrt{n}} \sqrt{2i(n-i)} \right)$$

wenn wir durch die einseitige Klammer andeuten, dass p irgend einen Wert in den angegebenen Grenzen haben kann. Setzt man z.B. $u = .476936$, so ist $F_u = 1/2$, und man kann also 1 gegen 1 wetten dass die Abweichung der empirischen Grösse i/n von p höchstens

$$\pm \frac{.476936}{\sqrt{n}} \sqrt{2i(n-i)}$$

beträgt. Diese letztere Grösse wuerde die sogenannte

"wahrscheinliche" Abweichung des Beobachtungsergebnisses darstellen. Für die Zwecke der Statistik dürfte es jedoch geeigneter sein, für u eine so grosse Zahl zu nehmen, dass F_u der Einheit sehr nahe kommt, also eine Wahrscheinlichkeit ausdrückt, die in der Praxis der Gewissheit gleich geachtet werden kann.¹⁰ Es genügt, wie oben, u gleich 3 zu setzen, und man erhält alsdann die Wahrscheinlichkeit $F_3 = .999978$ für die Grenzgleichung:

$$p = \frac{1}{n} \left(\pm \frac{3}{\sqrt{n}} \sqrt{2i(n-1)} \right).$$

Die mit dieser Wahrscheinlichkeit zulaessige positive und negative Abweichung koennte man als die "aeusserste" Abweichung bezeichnen."

It is the achievement of Laplace and Gauss, and of their followers throughout the 19th century, to have recognized the importance of obtaining probability limits for unknown parameters. They also developed a method for determining such limits, which is free of a priori assumptions regarding these parameters: The inversion of direct probability statements based on large sample approximations. However, it appears to have been understood only in the present century that in the resulting statements the parameter is constant and the probability refers to the variation of the limits.

REFERENCES

- Czuber, Emanuel (1899). Jahresbericht der Deutschen Mathematiker vereinigung, 7, No. 2.
- Czuber, Emanuel (1914). Wahrscheinlichkeitsrechnung. Dritte Auflage. Leipzig.
- Fourier, J. B. (1826). Recherches statistiques sur la ville de Paris et le departement de la Seine, 3.
- Gauss, C. F. (1816). Zeitschrift fuer Astronomie und verwandte Wissenschaften, 1, 185. (Reprinted in Gauss: Werke, 4, 109-119, Goettingen, 1873).
- De Morgan, A. (1837). Cambridge Philos. Trans., 6, 422-
- Hotelling, Harold (1931). Ann. Math. Stat., 2, 360.
- Jevons, W. Stanley (1873). The principles of science. MacMillan, London.
- Laplace, P. S. (1812). Théorie analytique des probabilités. Paris. (The 3rd edition of 1820 is reprinted as Vol. 7 of Laplace's collected works.)
- Lexis, W. (1875). Einleitung in die Theorie der Bevoelkerungsstatistik. Strassburg.
- Monro, C. J. (1874). Proc. London Math. Soc., 5, 74.
- Neyman, J. (1952). Lectures and conferences on mathematical statistics and probability. 2nd edition. Grad. School U.S. Dept. of Agriculture, Washington.
- Poisson, S. D. (1837). Recherches sur la probabilité des jugements. Paris.
- Todhunter, I. (1865). A history of the mathematical theory of probability. MacMillan. Cambridge.
- Wilson, E. B. (1927). J. Am. Stat. Assoc., 27, 209.
- Working, Halbrook and Hotelling, Harold (1929). J. Am. Stat. Assoc. Suppl. 24, 73.

Translation of the Lexis passage

...then with probability F_u , the unknown quantity p differs from i/n by at most $\pm T/\sqrt{n}\sqrt{2i(n-i)}$, or there holds with this probability the relation $p = i/n (\pm T/\sqrt{n}\sqrt{2i(n-i)})$ if we indicate by a one-sided parenthesis that p can take on any value within the stated limits. If for example $u = .476936$ then $F_u = 1/2$ and one can bet 1 against 1 that the deviation of the observed quantity i/n differs from p by at most $\pm .476936 \sqrt{2i(n-i)/n}$. This latter quantity would represent the so-called "probable" deviation of the observed frequency. For the purposes of statistics it would seem to be more appropriate to take for u a number so large that F_u comes very close to unity, and thus represents a probability, which for all practical purposes can be considered equal to certainty. It is sufficient to put, as above, u equal to 3, and one then obtains the probability $F_3 = .999973$ for the relation: $P = i/n (\pm 3\sqrt{2i(u-i)/n})$. The positive and negative deviation which is permitted with this probability could be called the "extreme" deviation.

FOOTNOTES

¹This paper was prepared with the partial support of the Office of Naval Research (Nonr-222-43). This paper in whole or in part may be reproduced for any purpose of the United States Government.

²Special cases were treated by Working and Hotelling (1929) and by Hotelling (1931).

³The page references here are to the 3rd edition of 1820.

⁴The event whose probability is p .

⁵Essentially our formula (1).

⁶That is, the first derivation.

⁷Attention was recently drawn to this paper by J. Neyman (1952, p. 222).

⁸I have here changed the notation so that it agrees with that used above.

⁹The notation has again been changed to agree with that used earlier.

¹⁰A similar definition is given by Fourier who attaches to the quantity βg the name "la plus grande erreur."