dently not meant to be taken completely seriously, but I think he goes too far when he accuses Gauss of circular reasoning. I find no foundation for that in what Gauss actually wrote. From the assumption that the arithmetic mean of repeated observations of a single quantity is the maximum likelihood estimate, he derives the normal distribution, and from that the more general principle of least squares for getting maximum likelihood estimates from observations on a number of related quantities. He asserts that the latter principle should be considered to be just as valid as the principle of the arithmetic mean; he does *not* close the circle by claiming that this in turn justifies the principle of the arithmetic mean. (My use of the term "maximum likelihood" is anachronistic; I am not claiming that Gauss had an exact equivalent of the modern notion clearly in mind.)

Gauss argument involves some tacit assumptions, and one can debate the extent to which he was aware of them; the argument is loose by modern standards, but it is not absurd. The extent to which he claimed to be giving a rigorous proof is not clear to me. (Neither Gauss (1809) nor his later extensive exposition of the method of least squares, which appeared in 1821 and 1823 with a supplement in 1826 (Volume 4 of his collected works, pp. 1–93) is at all a treatise in pure mathematics.) He clearly considered the method of least squares to be of great practical value. He was also interested in justifying it philosophically, but he explicitly stated that it cannot be singled out as the only reasonable method on purely theoretical grounds (see Section 186 of Gauss (1809) and Section 6 of the work on least squares mentioned above.)

Professor Le Cam's dissatisfaction with the performance of the hypothetical referee of Gauss (1809) brings to mind one more conversation with Feller, when he expressed some preference for the old days before the present refereeing system. An author who submitted inferior work for publication then ran this risk of damage to his reputation if it appeared. Of course the old system only worked well when the research community was smaller, the volume of publication was much less, and many papers actually got read by at least a few experts, not just counted by deans evaluating candidates for promotion. Gauss' reputation does not seem to have suffered from either the original publication in 1809 or its reprinting over a century and a half later.

I also wonder whether the influence of Gauss' work among nonmathematicians (astronomers, physicists, surveyors) may have played some role (along with the law of eponymy) in attaching his name to the distribution that bears it.

Comment

J. L. Doob

Le Cam's interesting account can be described with only slight exaggeration as a history of (nonrigorous) early research in probability, of probability texts written by mathematicians ignorant of the subject, and finally of frequently clumsy research published before the writers had digested their own work or consulted that of others. Is such the history of all probability research? Of all mathematical research? Should trade secrets be disclosed?

The following quotations are relevant. Lévy, who plays an important role in Le Cam's account, once remarked to me that reading other mathematicians' research gave him actual physical pain. A well known nonprobabilist mathematician remarked to me that the first time a mathematician publishes a research result the treatment is likely to be both wrong and unreadable; the second time the treatment will be correct but obscure; finally, a third treatment may be both correct and clear.

As a partial explanation of the second remark, and of the sometimes unseemly haste to publish, it must be acknowledged that no matter how much a mathematician admires his own work, the writing of it may finally make it so loathesome to his sight that he hastens to send it off for premature publication.

Influential on the nature and speed of probability research is the fact that probabilists, until about the last 30 years, have labored under the psychological disadvantage that their field was not considered a mathematical discipline by their colleagues, who for one thing did not understand why standard mathematical nomenclature was insufficient, why old concepts had to be rechristened "random variables" and "expectations." Moreover, probability books were full of nonmathematical concepts: dice, gambling houses, Peter, and Paul.

Even as late as the 1930s it was not quite obvious to some probabilists, and it was certainly a matter of

J. L. Doob is Emeritus Professor, Department of Mathematics, University of Illinois, Urbana, IL 61801.

doubt to most nonprobabilists, that probability could be treated as a rigorous mathematical discipline. In fact it is clear from their publications that many probabilists were uneasy in their research until their problems were rephrased in what was then nonprobabilistic language. For example, difference and differential equations for transition probabilities were suggested by sketchily described probability contexts, contexts then avoided as much as possible in the treatment and discussion of the equations. This uneasiness explains why it seemed more natural to Feller in 1935 than it does to Le Cam in 1985 to discuss convolutions of distribution functions rather than the corresponding sums of independent random variables. Feller had a superb background in classical analysis, and accordingly devised a heavily formal version of the central limit theorem, whereas Lévy produced a rather vague but correct in principle corresponding version. As always, Lévy exploited his unparalleled intuition to the despair of his readers, who found his work vague and obscure, although insightful and instructive when finally mastered. Lévy was one of the first probabilists to treat sample functions and sequences in depth, but never fully accepted measure theory as the mathematical basis of probability. For example, to him conditional expectations were a part of the essence of probability, needing no formal general definition.

Comment

David Pollard

Professor Le Cam deserves our thanks for a fine piece of scholarship. I hope that others will be inspired by his example to share with us their understanding of important ideas in probability and statistics.

I was particularly pleased to read the high praise in Section 3 for Lindeberg's proof of the central limit theorem. It is indeed surprising that the proof does not appear more often in standard texts (although Billingsley (1968) and Breiman (1968) should be added to the list of texts where it does appear), especially since the characteristic function approach is an effective source of confusion for beginners.

As Le Cam notes, the proof has even more to recommend it than its simplicity. It can be modified to give more information on the rate at which S_n converges in distribution to T_n, and it is easily extended beyond the case of distribution functions on the real line. I'll indicate briefly how this can be done.

Lindeberg's argument depends on not much more than Taylor's theorem to compare the expected value $\mathbb{P}f(S_n)$ of a smooth function of S_n with the corresponding expected value $\mathbb{P}f(T_n)$ for the sum of Gaussian increments. This translates into a bound on the difference $\Delta(x) = \mathbb{P}\{S_n \le x\} - \mathbb{P}\{T_n \le x\}$ between distribution functions when f is chosen as a smooth approximation to the indicator function of $(-\infty, x]$. The f used by Lindeberg was sandwiched between the indicator functions of $(-\infty, x]$ and $(-\infty, x + L]$, for a small L, and was piecewise cubic in $(x, x + L)$. The Lipschitz constraint on the second derivative (actually, Lindeberg put a bound on the third derivative) forces L to be of the order $A^{-1/3}$; a function with this degree of smoothness cannot negotiate the descent from 1 down to 0 in a shorter interval. Because this f fits between the two indicator functions,

$$\mathbb{P}\{S_n \le x\} \le |\mathbb{P}f(S_n) - \mathbb{P}f(T_n)| + \mathbb{P}\{T_n \le x + L\}.$$

As Le Cam shows, the first term on the righthand side is bounded by $A\beta$, with β a sum of third absolute moments; the second term exceeds $\mathbb{P}\{T_n \le x\}$ by the probability that T_n lies in $(x, x + L]$, that is, by a term of order L. An A of the order $\beta^{-3/4}$ balances these two contributions to the difference $\Delta(x)$ between distribution functions. A similar argument gives a similar-looking lower bound. Since the method works uniformly in x, this produces the bound of order $\beta^{1/4}$ that Le Cam quotes from Lindeberg.

The same idea works for subsets of other linear spaces. If B is such a subset, the challenge is to find a smooth approximation f to the indicator function of B: an f for which a Taylor expansion is possible; which takes values close to 1 well inside B, and values near 0 well outside B; and which makes the transition between these two levels as rapidly as possible near the boundary of B. If a bound on

$$\Delta(B) = \mathbb{P}\{S_n \in B\} - \mathbb{P}\{T_n \in B\}$$

is sought, attention must be paid to how much mass the distribution of T_n puts in the transition region

David Pollard is Professor of Statistics and Mathematics at Yale University. His mailing address is Department of Statistics, Yale University, Box 2179, Yale Station, New Haven, CT 06520.