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.

near the boundary. If a bound is sought for the Prohorov distance

$$\inf[\varepsilon > 0: \mathbb{P}\{S_n \in B\} \le \varepsilon + \mathbb{P}\{T_n \in B^\varepsilon\} \text{ for all } B]$$

it is the width of the transition region that is important. (Here B^ε denotes the set of points no further than ε from B: essentially B augmented by an ε strip around its boundary.) In either case there is a trade off between the size of the transition region and smoothness of f.

There are several ways to construct the smooth f. One can attempt a direct construction, as with Lindeberg's f. This requires ingenuity and, as Le Cam mentions, some sort of Lipschitz condition on a second derivative of a norm. One can also get a smooth approximation to the indicator function of B by convoluting it with a smooth distribution. Or one can combine these approaches, by applying some convolution smoothing to deterministic approximations that go only part of the way toward rounding off the rough edges of B. For example, Yurinskii (1977) got bounds on Prohorov distances, for strange norms on $\mathbb{R}_k$, by such means. One can even get rate results of Berry-Esseen type by convolution smoothing—the so-called method of compositions.

Roughly speaking, the method of compositions takes advantage of sources of smoothing untapped by Lindeberg's argument. Write W_{k+1} for the sum $\sum_{j>k} Y_j$ of Gaussian increments (so Le Cam's R_k is a sum of S_{k-1} and the Gaussian W_{k+1}). Write $g_k(t)$ for the smooth function $\mathbb{P}f(t + W_{k+1})$. When k is small, g_k is very smooth, even if f is discontinuous like the indicator of $(-\infty, x]$. To capture the effect of the increments X_k and Y_k carry out a Taylor expansion of g_k.

$$\begin{aligned}\mathbb{P}f(R_k + X_k) &= \mathbb{P}g_k(S_{k-1} + X_k) \\ &= \mathbb{P}g_k(S_{k-1}) + \tfrac{1}{2}\mathbb{P}X_k^2\mathbb{P}g_k''(S_{k-1}) \\ &\quad + \tfrac{1}{2}\mathbb{P}X_k^2[g_k''(S_{k-1}^*) - g_k''(S_{k-1})]\end{aligned}$$

and similarly for Y_k. For small k the Lipschitz constant for g_k'' will be smaller than the Lipschitz constant for f''.

A more subtle source of smoothing is the S_{k-1} itself. It should behave something like T_{k-1}; to some degree

$$\mathbb{P}g_k(S_{k-1} + t) \approx \mathbb{P}g_k(T_{k-1} + t).$$

For large values of k, the T_{k-1} provides extra smoothing for g_k. The combined effect of this T_{k-1} and the W_{k+1} is almost that of convolution of f with a $N(0, 1)$ Gaussian. Of course, the last approximation is practically the same assertion as $\mathbb{P}f(S_n) \approx \mathbb{P}f(T_n)$, except that it involves a smaller sample size. There is a glimmer of hope here for an inductive argument. If f is an indicator function of an interval there are slight complications for $k \approx n$. To overcome these one must first apply some convolution smoothing to f. For the details, as well as much more about the method of compositions, see Sazonov (1981).

Lindeberg's argument still has something to offer.

ADDITIONAL REFERENCES

BILLINGSLEY, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

BREIMAN, L. (1968). *Probability.* Addison-Wesley, Reading, Mass.

SAZONOV, V. V. (1981). Normal approximations—some recent advances. *Springer Lecture Notes in Mathematics* **879**.

YURINSKII, V. V. (1977). On the error of the Gaussian approximation for convolutions. *Theory Prob. Appl.* **22** 236–247.

Rejoinder

L. Le Cam

Many thanks are due to my colleagues for their constructive comments and criticisms, but particular thanks are due to Professor Doob for his wonderful explanation of why so many mathematical papers are unreadable! Doob also accuses me of writing a "history of (nonrigorous) early research in probability, of probability texts written by mathematicians ignorant of the subject . . .". This is partly true, but I believe that the roles of Bertrand, Poincaré, and Borel in that kind of history are particularly regrettable.

For the need to use "convolutions" instead of "sums of random variables" mentioned by Professors Doob and Trotter, one can only agree. Yet one can argue that those who went ahead and used such concepts before the publication of Kolmogorov's booklet were not as nonrigorous as it might seem. Most mathematicians would probably agree that it is legal to deal with certain objects called random variables without defining them provided that one sets down clearly what are the rules for handling them. Lévy, among others, was probably unclear when stating such rules. However, his attitude toward measure theory as a basis for probability was more complex than what Professor Doob implies. I think it was more in the