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

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

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 following vein. Between 1919 and 1926, Lévy convinced himself that one could define and handle measures on infinite dimensional cubes. He gives such a theory (Lévy, 1925) and repeats it (Lévy, 1937a) (with the added comment that one can do it on cubes of uncountable dimensionality, but that the result is utter chaos!). With the theory stated in his paper in 1925, Lévy could handle countable or countably generated families of random variables perfectly legitimately. He just did not feel it necessary to say repeatedly that the legitimacy could derive from his theory stated in Lévy (1925). That he did not use much the standard tools of measure theory (e.g., Radon-Nikodym) is another matter.

Professor Trotter takes me to task for saying that Gauss used a "circular" argument. Perhaps I should have said "semicircular" or would "elliptical" be better? (Gauss's work was about bodies moving in conic sections.) Be this as it may, it is certain that the argument was found very unconvincing by subsequent authors (including the loathsome Bertrand and Poincaré). It is also hard to apply to Gauss the wonderful excuse described by Professor Doob: He did not rush into print but claimed on July 30, 1806 that he had used the method since 1794. There are some indications, however, that the justification for the method was invented only later, perhaps in 1807, and included in the 1809 Latin version of a work claimed to have been written in 1806. For this see Plackett ("Studies in the history of probability and statistics. XXIX. The discovery of the method of least squares." *Biometrika* **59** 239–251, 1972).

At the same time Gauss claimed that Laplace's method of minimizing sums of absolute deviations instead of their squares "is not admissible on the basis of the calculus of probability and leads to contradictions." Thus, even if Gauss did not intend to give a rigorous proof, he must have believed the argument solid enough to warrant an attack on Laplace.

I shall say no more on Gauss, according to a principle stated by Chinese friends in connection with Mao Zedong, "Why pick on a dead man?"

Professor Pollard's advertising for Lindeberg's method is well taken and very nice. I am sorry that I did not mention the books by Pat Billingsley and Leo Breiman. I had in mind more elementary texts, suitable for not so advanced undergraduates. Apologies to all concerned.