

- BREIMAN, L. (1968). *Probability*. Addison-Wesley, Reading, Mass.
- CHEN, L. H. Y. (1975a). Poisson approximation for dependent trials. *Ann. Probab.* **3** 534–545.
- CHEN, L. H. Y. (1975b). An approximation theorem for sums of certain randomly selected indicators. *Z. Wahrsch. Verw. Gebiete* **33** 69–74.
- CHEN, L. H. Y. (1978). Two central limit problems for dependent random variables. *Z. Wahrsch. Verw. Gebiete* **43** 223–243.
- CHEN, L. H. Y. and HO, S. T. (1978). An L_p bound for the remainder in a combinatorial central limit theorem. *Ann. Probab.* **6** 231–249.
- DIACONIS, P. and MOSTELLER, F. (1989). Methods for studying coincidences. *J. Amer. Statist. Assoc.* **84** 853–861.
- ERICKSON, R. V. (1974). L_1 bounds for asymptotic normality of m -dependent sums using Stein's technique. *Ann. Probab.* **2** 522–529.
- FICKETT, J. W. and BURKS, C. (1988). Development of a database for nucleotide sequences. In *Mathematical Methods for DNA Sequences* (M. S. Waterman, ed.) 1–44. CRC Press, Boca Raton, Fla.
- HALL, P. (1980). Estimating probabilities for normal extremes. *Adv. in Appl. Probab.* **12** 491–500.
- HALL, P. (1988). *Introduction to the Theory of Coverage Processes*. Wiley, New York.
- HECKMAN, N. (1988). Bump hunting in regression analysis. Preprint.
- HOLST, L. (1986). On birthday, collectors', occupancy and other classical urn problems. *Internat. Statist. Rev.* **54** 15–27.
- HOLST, L. and JANSON, S. (1990). Poisson approximation using the Stein–Chen method and coupling: Number of exceedances of Gaussian random variables. *Ann. Probab.* **18** 713–723.
- HUDSON, H. M. (1978). A natural identity for exponential families with applications in multiparameter estimation. *Ann. Statist.* **6** 473–484.
- JANSON, S. (1986). Birthday problems, randomly colored graphs, and Poisson limits of dissociated variables. Tech. report 1986 **16**. Dept. Math., Uppsala Univ.
- KARLIN, S. (1982). Some results on optimal partitioning of variance and monotonicity with truncation level. In *Statistics and Probability: Essays in Honor of C. R. Rao* (P. Kallianpur, R. Krishnaiah and J. K. Ghosh, eds.) 375–382. North-Holland, Amsterdam.
- KARLIN, S., GHANDOUR, G., OST, F., TAVARÉ, S. and KORN, L. J. (1983). New approaches for computer analysis of nucleic acid sequences. *Proc. Nat. Acad. Sci. U.S.A.* **80** 5660–5664.
- KARLIN, S. and OST, F. (1987). Counts of long aligned word matches among random letter sequences. *Adv. in Appl. Probab.* **19** 293–351.
- LEADBETTER, M. R., LINDGREN, G. and ROOTZÉN, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer, New York.
- MOLER, C., ULLMAN, M., LITTLE, J. and BANGERT, S. (1987). *Pro-MATLAB User's Manual*. The Math Works, Sherborn, Mass.
- OHYAMA, K., FUKUZAWA, H. and KOHCHI, T. ET AL. (1986). Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. *Nature* **322** 572–574.
- OLKIN, I. and MARSHALL, A. (1979). *Inequalities: Theory of Majorization and Its Applications*. Academic, New York.
- RIORDAN, J. (1978). *An Introduction to Combinatorial Analysis*. Princeton Univ. Press.
- ROOTZÉN, H. (1983). The rate of convergence of extremes of stationary normal sequences. *Adv. in Appl. Probab.* **15** 54–80.
- ROSENBLATT, M. (1974). *Random Processes*. Springer, New York.
- SAMPFORD, M. R. (1953). Some inequalities on Mill's ratio and related functions. *Ann. Math. Statist.* **24** 130–132.
- SERFLING, R. J. (1975). A general Poisson approximation theorem. *Ann. Probab.* **3** 726–731.
- STEIN, C. M. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *Proc. Third Berkeley Symp. Math. Statist. Probab.* **1** 197–206. Univ. California Press, Berkeley, Calif.
- STEIN, C. M. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **2** 583–602. Univ. California Press, Berkeley, Calif.
- STEIN, C. M. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9** 1135–1151.
- STEIN, C. M. (1986). *Approximate Computations of Expectations*. IMS, Hayward, Calif.
- STEIN, C. M. (1987). The number of monochromatic edges in a graph with randomly colored vertices. Unpublished manuscript.
- TAKÁCS, L. (1988). On the limit distribution of the number of cycles in a random graph. *J. Appl. Prob.* **26** 359–376.
- TAVARÉ, S. and GIDDINGS, B. W. (1989). Some statistical aspects of the primary structure of nucleotide sequences. In *Mathematical Methods for DNA Sequences* (M. S. Waterman, ed.) 117–132. CRC Press, Boca Raton, Fla.
- WATSON, G. S. (1954). Extreme values in samples from m -dependent stationary stochastic sequences. *Ann. Math. Statist.* **25** 798–800.
- WILF, H. S. (1983). Three problems in combinatorial asymptotics. *J. Combin. Theory Ser. A* **35** 199–207.

Comment

J. Michael Steele

This beautiful exposition leaves little room for quibbles. Still, if forced to raise some issue, I suspect my

J. Michael Steele is Professor of Statistics, Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6302, and Editor of The Annals of Applied Probability.

best shot is to point out that, despite its power, the Chen–Stein method is not omnipotent. In fact, there are simple problems where one might suspect that a Poisson law lurks below the surface, yet the hooks provided by the Chen–Stein method leave us without a catch.

Consider a simple random walk $S_n = X_1 + X_2 + \dots + X_n$ in \mathbf{R}^2 where the X_i are iid. To make life as

simple as possible, we further assume that the X_i have a density, and we take $S_0 = 0$ to have a clear convention. The convex hull H_n of S_0, S_1, \dots, S_n is a natural object that turns out to be intriguing both for its probability theory and geometry. A priori one might expect results on H_n to be difficult and incomplete, but—at least as far as first moments go—the theory is surprisingly easy and precise.

The first contributions to the theory of H_n are due to Spitzer and Widom (1961). Their seminal observation was to show that an ancient result of Cauchy could be combined with a purely combinatorial result of Kac (1954) to obtain an exact formula for the expectation of the length L_n of the boundary of H_n :

$$EL_n = 2 \sum_{k=1}^n E|S_k|/k.$$

From this beautiful formula one can obtain considerable information about EL_n , and, in particular, one can use it to show that $EL_n \sim c\sqrt{n}$ provided $EX_i^2 < \infty$ and $EX_i = 0$. This observation does not yet put us in the territory of the Poisson law—that comes later—but it does give the first suggestion of a counting law to be discovered.

The second paper to treat the geometry of H_n is due to Baxter (1961). This work shows, among other results, that the number N_n of sides of H_n has a remarkably simple expectation. In fact, it is just twice the n th harmonic number, i.e.,

$$EN_n = 2 \sum_{k=1}^n \frac{1}{k}.$$

Comment

A. D. Barbour

The Chen–Stein method has added a new dimension to the techniques available for justifying Poisson approximations. In fields such a random graph theory (Bollobás, 1985, Chapter 4), extreme value theory (Smith, 1988; Holst and Janson, 1990) and spatial statistics (Barbour and Eagleson, 1984), where Poisson approximation plays an important role, the Chen–Stein method has already proved to be the best general approach, and its potential has by no means been exhausted. Its strengths are that it makes many sorts

In Snyder and Steele (1990), a common generalization of these results is given. If we let $e_i, i = 1, 2, \dots$ denote the lengths of the faces of H_n and if f is any function, then provided both sides make sense we have the identity

$$E \sum_i f(e_i) = 2 \sum_{k=1}^n \frac{1}{k} Ef(|S_k|).$$

Naturally, this identity yields that of Spitzer and Widom by taking $f(x) = x$, and we can also get Baxter’s identity just by taking $f(x) = 1$.

Now, here is where it may pay to start looking for a Poisson law. If we let f_n denote the indicator of an interval $[a_n, b_n]$, then for any given distribution of the X_i it is not hard to determine a_n and b_n so that for each n the sum $G_n = \sum_i f_n(e_i)$ satisfies $EG_n = \lambda > 0$. It may be most natural to take $a_n = 0$ in order to focus on the small faces of H_n . The variable G_n is nothing more than a sum of a random number of dependent random variables, the $f_n(e_i)$. Further, these variables do not seem all *that* dependent. Thus, there is a serious possibility of a Poisson approximation to the distribution of G_n .

Still, in this problem the Poisson law seems a long way away. The first moments were obtained through somewhat slippery trickery, and second moments do not seem to be open to more of the same. The Poisson law is honestly in play, yet the Chen–Stein method has far to come to meet the challenge. Can sufficient information be found on the second moments of G_n to complete the Chen–Stein program?

of weak dependence easy to handle, it gives explicit estimates of the accuracy of approximation, and it continues to give good results even when the expectation λ is large. The preceding survey illustrates the first two of these aspects admirably, but it gives rather less weight to the third, to which the following comments are addressed. For details and much more about the Chen–Stein method, see the forthcoming book of Barbour, Holst and Janson (1991).

A remarkable feature of the Chen–Stein method is the form of the estimate of Theorem 1. When applied in the simplest setting, that of Theorem 0, it gives an error estimate no greater than $2 \min(1, \lambda^{-1}) \sum_{i=1}^n p_{i,n}^2$. Were only an estimate of the form $c \sum_{i=1}^n p_{i,n}^2$ required, for some real c , it could be obtained

A. D. Barbour is Professor of Mathematics, Angewandte Mathematik, Universität Zurich, Rämistrasse 74, Zürich CH-8001, Switzerland.