approximation fails, we still have an approximation—compound Poisson approximation, provided $(1 \wedge \lambda_1^{-1})e^{\lambda} \sum_{\alpha \in I} p_\alpha \xi_\alpha$ is small. A consequence of this is that we can avoid declumping in applications, as we shall see below.

Consider the example in Section 4.2.1 of Arratia, Goldstein and Gordon with $X_\alpha = C_\alpha C_{\alpha+1} \cdots C_{\alpha+t-1}$. Then the above dependence assumption is satisfied with $A_\alpha = \{1 \vee (\alpha - t + 1), \cdots, \alpha + 2t - 2\}$ and $B_\alpha = \{1 \vee (\alpha - 2t + 2), \cdots, \alpha + 3t - 3\}$. By (10), we obtain

$$\| \mathcal{L}(U) - \mathcal{L}(Z) \| \leq 4e^{\lambda}(1 \wedge ((p + nq)qp^t)^{-1})n(5t - 4)p^{2t},$$

where $q = 1 - p$. Note that $W = U$. In order that the distribution of $Z$ be determined, we need to compute $\lambda_i$ for all $i$. It can be shown that

$$\lambda_i = i^{-1} \sum_{\alpha=1}^{n} p^t P(V_{t-1} + V'_{\alpha \wedge t-1} = i - 1)$$

where $V_0 \equiv 0$, $V_m$ and $V'_m$ are geometric $(p)$ truncated at $m$, and $V_{t-1}$ and $V'_{\alpha \wedge t-1}$ are independent. We can either proceed to compute each $\lambda_i$ explicitly to determine $\mathcal{L}(Z)$ or approximate $\mathcal{L}(Z)$ by $\mathcal{L}(Z^*)$ to obtain the following result:

$$\| \mathcal{L}(U) - \mathcal{L}(Z^*) \| \leq 4e^{\lambda}(1 \wedge ((p + nq)qp^t)^{-1})n(5t - 4)p^{2t} + 4(2n - t)p^{2t} + 4q^{-1}p^{t+1}, \tag{11}$$

where $Z^*$ has the compound Poisson distribution $\exp[\lambda^*(\mu^* - \delta_0)]$ with $\lambda^* = nqp^t$ and $\mu^*(\{i\}) = qp^{i-1}$, $i = 1, 2, \cdots$ ("one plus a geometric $(p)$"). In approximating $\mathcal{L}(Z)$ by $\mathcal{L}(Z^*)$ we need not calculate $\lambda_i$ explicitly. For bounded $\lambda^*$, the order of the error bound in (11) is the same as that obtained by Arratia, Goldstein and Gordon. Note that $q\lambda^* \leq \lambda_1 \leq \lambda \leq np^t = q^{-1}\lambda^*$. Hence the result (11) not only provides an approximation for $\mathcal{L}(U)$ but also can be used to obtain the asymptotic distribution of $R_n$, the length of the longest run of heads beginning in the first $n$ tosses of a coin, since $\{R_n < t\} = \{U = 0\}$ and $P(Z^* = 0) = e^{-\lambda^*}$.

In the same way, (11) can also be applied to the biological example in Section 5 of the article by Arratia, Goldstein and Gordon to obtain an approximation result for $\mathcal{L}(\sum_{\alpha \in I} I(S_\alpha \geq s))$ and the asymptotic distribution of $M_n(t_n)$, the largest number of matches witnessed by any comparison of length $t_n$ substrings of two strands of DNA.

# Rejoinder

**Richard Arratia, Larry Goldstein and Louis Gordon**

At least one of us used to speak of the methods we have presented here as the philosopher's stone. None of us make such extravagant claims any longer; the discussants have put their collective fingers on a number of reasons why.

The method as we have presented it works best for dealing with local dependence, corresponding to situations in which $b_1$ is small and $b_3 = 0$. In these situations, $b_2$ is small and our approximations are useful if and only if second moments are well behaved. Steele gives an intriguing example having weak long-range dependence that is much harder to deal with. In Steele's problem, even if second moments were well controlled, there would still be difficulties due to the nonlocal dependence captured by $b_3$. Here is another such related example.

The question is inspired by the important problem of analyzing the expected, as opposed to worst-case, behavior of the simplex method. See Borgwardt (1987) for an exposition. Specifically, one is led to study the number of edges or vertices in the convex hull of $n$ independent and identically distributed points in, say, $\mathbf{R}^2$. For a line segment joining two of the observed points to be an edge of the convex hull, all of the other points must lie on one of the half-planes determined by these points. The usual heuristic applies. There are a large number of pairs of points to serve as a potential edge, and the probability that a given pair is actually an edge in the convex hull is small. Hence, the total number of edges in the convex hull should be approximately Poisson. As with Steele's example, first moments are tractable. Unfortunately, second moments and nonlocal dependence are again a problem. If an edge is indeed in the convex hull, one of its endpoints is also on a second edge of the convex hull, this is reflected in the second moment and $b_2$. There is also some additional nonlocal dependence which is part of $b_3$. This type of behavior reinforcess the issues raised by Steele's example.

In discussing Section 3.1, Barbour gives an example involving a Bernoulli variable with $p_{1,n} = \frac{1}{2}$. This example shows that no negative power of $\lambda$ can be

brought into our Theorem 2, which compares the given Bernoulli process with an equal intensity Poisson process. In this example, the obstacle is the large atom of mass ½, rather than the dependence structure. Theorem 3 compares a given dependent Bernoulli process to an independent Bernoulli process, so the presence of atoms is no obstacle. To see that Theorem 3, like Theorem 2, admits no improvement for large $\lambda$, consider the following example. There are $n$ independent sites, each either empty with probability $1 - p$, or else occupied by a red or else a white ball, with probability $p/2$ each. Taking the obvious index size $2n$ and neighborhoods of size 2 we have $\lambda = np$, $b_1 = np^2$, $b_2 = b_3 = 0$, $\sum p_\alpha^2 = b_1/2$. For the process of $2n$ locally dependent events $\mathbf{X}$ compared with the corresponding process of $2n$ independent events $\mathbf{X}'$, Theorem 3 gives the upper bound $\frac{1}{2}\| \mathscr{L}(\mathbf{X}) - \mathscr{L}(\mathbf{X}') \| \leq 2b_1 + \sum p_\alpha^2 = \frac{5}{2}np^2$. Direct calculation yields $\frac{1}{2}\| \mathscr{L}(\mathbf{X}) - \mathscr{L}(\mathbf{X}') \| \geq P$ (some site in $\mathbf{X}'$ has both a red and a white ball) $= 1 - (1 - p^2/4)^n \approx 1 - \exp(-np^2/4)$. Taking $np^2$ fixed and $\lambda = np \to \infty$ shows that the bound from Theorem 3 cannot be strengthened by any function of $\lambda$ which approaches zero as $\lambda$ approaches infinity.

At the end of his discussion of cycles in random permutations, Barbour gives a beautiful coupling proof that a total variation distance is at most $4f(n)/n$. We observe that this result is stronger than anything which could be proved using the Chen–Stein method. The utility of the Chen–Stein method in this case was to focus on the possibility of giving bounds on the total variation distance from a complicated dependent process to a simpler Poisson process.

Waterman is interested in best matching segments from two long sequences, allowing mismatches, insertions and deletions. The dependence structure is local: an event involving two short segments is correlated with another such event only if there is some overlap in the segments. Hence when second moments from this dependence are well behaved, the Chen–Stein method yields distributional results almost as easily as the much cruder method of first and second moments yields strong laws. However, it is so far impossible to get useful estimates on second moments when deletions are allowed, i.e., when $\delta < \infty$. Nevertheless, the Chen–Stein method gives the useful insight that perhaps the only missing ingredient to a distributional theory for matching allowing deletions is an upper bound on certain correlations.

Chen raises many interesting new ideas in Poisson and Normal approximation. We comment briefly on two. We are pleased to see that there has been so much progress on large deviation estimates for Poisson approximation. In many cases, be it in the birthday problem or hypothesis testing, the results of applying the method yields a small error bound and an even smaller probability. In such a case, it is not possible to know how different the true probability is from zero. Hence, being able to deal with relative errors and large deviations is a significant advantage.

Chen's generalization of the Poisson difference equation to compound distributions is extremely clever. One would expect that dealing with the target compounded distribution through its own characterizing equation may give better estimates than what we derive through our process theorem. This is so because the process theorem may carry too much information extraneous to the problem. In particular, the process theorem contains information on the location of clumps, not just their quantity and types. In Section 4.2.1, in order to get our estimates of distance of $U$ to compound Poisson, we consider a functional $h$ on processes. As the theorem gives a like bound for arbitrary functionals with $\| h \| \leq 1$, the bound may suffer in a specific case. That some intentional loss of information may lead to better bounds was also noted by Barbour in his comments.

We are grateful to the discussants for pointing out the limitations of the method. The great variety of concepts the discussants raise reinforces our belief that the scope of ideas one is led to think about while attempting Poisson approximation is broad.

## ADDITIONAL REFERENCES


ARRATIA, R., MORRIS, P. and WATERMAN, M. S. (1988). Stochastic scrabble: A law of large numbers for sequence matching with scores. *J. Appl. Probab.* **25** 106–119.

BARBOUR, A. D. (1988). Stein's method and Poisson process convergence. *J. Appl. Probab.* **25A** 175–184.

BARBOUR, A. D. and BROWN, T. C. (1990). The Stein–Chen method, point processes and compensators. Preprint.

BARBOUR, A. D., CHEN, L. H. Y. and LOH, W. L. (1990). Compound Poisson approximation for nonnegative random variables using Stein's method. In preparation.

BARBOUR, A. D., HOLST, L. and JANSON, S. (1991). *Poisson approximation.* In preparation.

BAXTER, G. (1961). A combinatorial lemma for complex numbers. *Ann. Math. Statist.* **32** 901–904.

BORGWARDT, K. H. (1987). *The Simplex Method.* Springer, New York.

BOROVKOV, A. A. and UTEV, S. A. (1984). On an inequality and a related characterization of the normal distribution. *Theory Probab. Appl.* **28** 219–228.

BRASCAMP, H. J. and LIEB, E. H. (1976). On extensions of the Brunn–Minkowski and Prékopa–Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation. *J. Funct. Anal.* **22** 366–389.

CHEN, L. H. Y. (1975c). An approximation theorem for convolutions of probability measures. *Ann. Probab.* **3** 992–999.

CHEN, L. H. Y. and CHOI, K. P. (1990). Some asymptotic and large deviation results in Poisson approximation. In preparation.

CHEN, L. H. Y. and LOU, J. H. (1987). Characterization of probability distributions by Poincaré-type inequalities. *Ann. Inst. H. Poincaré Sect B. (N.S.)* **23** 91–110.

CHEN, L. H. Y. and LOU, J. H. (1989). A characterization of probability measures which admit Poincaré inequalities. Preprint.

CHERNOFF, H. (1981). A note on an inequality involving the normal distribution. *Ann. Probab.* **9** 533–535.

CHVÀTAL, V. and SANKOFF, D. (1975). Longest common subsequences of two random sequences. *J. Appl. Prob.* **12** 306–315.

DEKEN, J. (1979). Some limit results for longest common subsequences. *Discrete Mathematics* **26** 17–31.

ERDÖS, P. and RENYI, A. (1970). On a new law of large numbers. *J. Anal. Math.* **22** 103–111.

KARLIN, S. and ALTSCHUL, S. F. (1990). Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. *Proc. Natl. Acad. Sci. U.S.A.* **87** 2264–2268.

KAC, M. (1954). Toeplitz matrices, transition kernels and a related problem in probability theory. *Duke Math. J.* **21** 501–509.

MICHEL, R. (1988). An improved error bound for the compound Poisson approximation of a nearly homogeneous portfolio. *ASTIN Bull.* **17** 165–169.

SMITH, R. L. (1988). Extreme value theory for dependent sequences via the Stein–Chen method of Poisson approximation. *Stochastic Proc. Appl.* **30** 317–327.

SMITH, T. F., BURKS, C. and WATERMAN, M. S. (1985). The statistical distribution of nucleic acid similarities. *Nucleic Acids Res.* **13** 645–656.

SMITH, T. F. and WATERMAN, M. S. (1981). Identification of common molecular subsequences. *J. Molecular Biology* **147** 195–197.

SNYDER, T. and STEELE, J. M. (1990). Convex hulls of random walks. Technical report, Dept. Statist., Univ. Pennsylvania.

SPITZER, F. and WIDOM, H. (1961). The circumference of a convex polygon. *Proc. Amer. Math. Soc.* **12** 506–509.

STEIN, C. (1990). A way of auxiliary randomization. In *Probability Theory: Proc. Singapore Probab. Conf., Singapore, 8–16 June 1989.* To appear.

WATERMAN, M. S. and EGGERT, M. (1987). A new algorithm for best subsequence alignments with application to tRNA-rRNA comparisons. *J. Molecular Biology* **197** 723–728.

WATERMAN, M. S., GORDON, L. and ARRATIA, R. (1987). Phase transitions in sequence matches and nucleic acid structure. *Proc. Natl. Acad. Sci. U.S.A.* **84** 1239–1243.