The Annals of Probability
1989, Vol. 17, No. 1, 9–25

TWO MOMENTS SUFFICE FOR POISSON APPROXIMATIONS: THE CHEN–STEIN METHOD

BY R. ARRATIA,[1] L. GOLDSTEIN AND L. GORDON[2]

University of Southern California

Convergence to the Poisson distribution, for the number of occurrences of dependent events, can often be established by computing only first and second moments, but not higher ones. This remarkable result is due to Chen (1975). The method also provides an upper bound on the total variation distance to the Poisson distribution, and succeeds in cases where third and higher moments blow up. This paper presents Chen's results in a form that is easy to use and gives a multivariable extension, which gives an upper bound on the total variation distance between a sequence of dependent indicator functions and a Poisson process with the same intensity. A corollary of this is an upper bound on the total variation distance between a sequence of dependent indicator variables and the process having the same marginals but independent coordinates.

1. Introduction. Convergence to the Poisson distribution, for the number of occurrences of dependent events, can often be established by computing only first and second moments, but not higher ones. This remarkable result is due to Chen (1975). The method also provides an upper bound on the total variation distance to the Poisson distribution and succeeds in cases where third and higher moments blow up. This paper presents Chen's results in a form that is easy to use and gives a multivariable extension, which gives an upper bound on the total variation distance between a sequence of dependent indicator functions and a Poisson process with the same intensity. A corollary of this is an upper bound on the total variation distance between a sequence of dependent indicator variables and the process having the same marginals but independent coordinates.

The surprisingly wide applicability of Poisson approximations is very nicely described in notes on the "Poisson clumping heuristic" by Aldous (1987). Chen's method works directly in situations involving "clumps" of occurrences provided that each clump $\{Y_{\alpha(1)}, Y_{\alpha(2)}, \dots\}$ can be identified with a single index X_α. Such identification is used in Section 3, in Examples 3–5, which all involve the extremes of a stochastic process. The distribution of extremes is analyzed, as in Watson (1954), via the random number W of exceedances of a test value, so that the quality of the approximation of $P(W = 0)$ by e^{-EW} is given special attention in Theorem 1.

Chen's method is the adaptation to the Poisson distribution of Stein's differential method for the normal distribution, presented in Stein (1971). Both methods are discussed in a recent monograph by Stein (1986a).

Received July 1987; revised October 1987.
[1]Supported by NIH Grant GM-36230 and NSF Grant DMS-86-01986.
[2]Supported by NIH Grant GM-36230 and a grant from the System Development Foundation.
AMS 1980 *subject classifications*. 60F05, 60F17.
Key words and phrases. Poisson approximation, Poisson process, invariance principle, coupling, method of moments, inclusion–exclusion.

Chen's method is applied to some random graph problems in Barbour (1982) and in Bollobás (1985) and to some statistical problems in Barbour and Eagleson (1983). There are many situations where a law of large numbers is proved by the first and second moments method—see Erdös and Rényi (1960) or Bollobás (1985). In many of these situations, Chen's method could be used to get a Poisson limit. Better bounds on the Poisson convergence for independent trials are given in Barbour and Hall (1984) and Barbour (1987a, b). See also Barbour and Eagleson (1984), Barbour and Holst (1987) and Barbour and Jensen (1987). More references are given in Example 2 of Section 3.

This paper is organized as follows. Notation and the statements of our two theorems form Section 2. Theorem 1 is essentially contained in Chen (1975), and Theorem 2, which is a process version, is new. Theorem 3 is an easy corollary of Theorem 2 and gives a way of decoupling dependent events. Section 3 gives examples of applications. Section 4 defines and gives bounds on operators used in Chen's proof. Section 5 proves Theorem 1—all the ingredients of this proof are in Chen (1975) and Barbour and Eagleson (1983); but our presentation here is needed to prepare the way for the proof of Theorem 2, which is Section 6.

2. Notation and statement of results. Let I be an arbitrary index set, and for $\alpha \in I$, let X_α be a Bernoulli random variable with $p_\alpha \equiv P(X_\alpha = 1) = 1 - P(X_\alpha = 0) > 0$. Let

$$W \equiv \sum_{\alpha \in I} X_\alpha \quad \text{and} \quad \lambda \equiv EW = \sum_{\alpha \in I} p_\alpha. \tag{1}$$

We assume that $\lambda \in (0, \infty)$.

For each $\alpha \in I$, suppose we have chosen $B_\alpha \subset I$ with $\alpha \in B_\alpha$. We think of B_α as a "neighborhood of dependence" for α, such that X_α is independent or nearly independent of all of the X_β for β not in B_α. Define

$$b_1 \equiv \sum_{\alpha \in I} \sum_{\beta \in B_\alpha} p_\alpha p_\beta,$$

$$b_2 \equiv \sum_{\alpha \in I} \sum_{\alpha \neq \beta \in B_\alpha} p_{\alpha\beta}, \quad \text{where } p_{\alpha\beta} \equiv E(X_\alpha X_\beta),$$

$$b_3' \equiv \sum_{\alpha \in I} s_\alpha', \quad \text{and} \quad b_3 \equiv \sum_{\alpha \in I} s_\alpha,$$

where

$$s_\alpha' \equiv E\left| E\left\{ X_\alpha - p_\alpha \,\middle|\, \sum_{\beta \in I - B_\alpha} X_\beta \right\} \right|$$

$$\leq s_\alpha \equiv E\left| E\left\{ X_\alpha - p_\alpha \middle| \sigma(X_\beta \colon \beta \in I - B_\alpha) \right\} \right|.$$

Loosely speaking, our results are that when b_1, b_2 and b_3 are all small, then:

1. The total number W of events is approximately Poisson (Theorem 1).
2. The locations of the dependent events approximately form a Poisson process (Theorem 2).

3. The dependent events are almost indistinguishable from a collection of independent events having the same marginal probabilities (Theorem 3).

Loosely, b_1 measures the neighborhood size, b_2 measures the expected number of neighbors of a given occurrence and b_3' or b_3 measures the dependence between an event and the number of occurrences outside its neighborhood.

Let Z denote a Poisson random variable with mean λ, so that for $k = 0, 1, 2, \dots$, $P(Z = k) = e^{-\lambda}(\lambda^k/k!)$.

Let f, h: $Z^+ \to R$, where $Z^+ = \{0, 1, 2, \dots\}$, and write $\|h\| \equiv \sup_{k \geq 0}|h(k)|$. We denote the total variation distance between the distributions of W and Z by

$$\begin{aligned}\|\mathscr{L}(W) - \mathscr{L}(Z)\| &\equiv \sup_{\|h\|=1} |Eh(W) - Eh(Z)| \\ &= 2 \sup_{A \subset Z^+} |P(W \in A) - P(Z \in A)|.\end{aligned}$$

We observe that convergence in distribution is equivalent to convergence under the Prohorov metric, which coincides with half of the total variation distance on the set of probability measures supported on the integers.

Theorem 1. *Let W be the number of occurrences of dependent events, and let Z be a Poisson random variable with $EZ = EW = \lambda$. Then*

$$\begin{aligned}\|\mathscr{L}(W) - \mathscr{L}(Z)\| &\leq 2\left[(b_1 + b_2)\frac{1 - e^{-\lambda}}{\lambda} + b_3'(1 \wedge 1.4\lambda^{-1/2})\right] \\ &\leq 2(b_1 + b_2 + b_3)\end{aligned}$$

and

$$|P(W = 0) - e^{-\lambda}| \leq (b_1 + b_2 + b_3')(1 - e^{-\lambda})/\lambda < (1 \wedge \lambda^{-1})(b_1 + b_2 + b_3).$$

Theorem 2. *For $\alpha \in I$, let Y_α be a random variable whose distribution is Poisson with mean p_α, with the Y_α mutually independent. The total variation distance between the dependent Bernoulli process $\mathbf{X} \equiv (X_\alpha)_{\alpha \in I}$, and the Poisson process $\mathbf{Y}$ on I with intensity $p_{(\cdot)}$, $\mathbf{Y} \equiv (Y_\alpha)_{\alpha \in I}$, satisfies*

$$\|\mathscr{L}(\mathbf{X}) - \mathscr{L}(\mathbf{Y})\| \leq 2(2b_1 + 2b_2 + b_3).$$

This follows easily from the following finite-dimensional bound: Let the index set I be partitioned into disjoint nonempty subsets $I(1), \dots, I(d)$, and let

$$W_j \equiv \sum_{\alpha \in I(j)} X_\alpha, \qquad Z_j \equiv \sum_{\alpha \in I(j)} Y_\alpha \quad \textit{and} \quad \lambda_j \equiv EW_j = EZ_j, \tag{2}$$

so that $W = W_1 + \cdots + W_d$ and $Z = Z_1 + \cdots + Z_d$. The total variation distance between the joint distribution of $(W_1, \dots, W_d)$ and $(Z_1, \dots, Z_d)$ satisfies

$$\begin{aligned}&\|\mathscr{L}((W_1, \dots, W_d)) - \mathscr{L}((Z_1, \dots, Z_d))\| \\ &\qquad \leq 2\big(1 \wedge 1.4(\min \lambda_i)^{-1/2}\big)(2b_1 + 2b_2 + b_3).\end{aligned}$$

The corollary below follows by applying Theorem 2 to compare each of the Bernoulli processes $\mathbf{X}$ and $\mathbf{X}'$, whose support is $\{0,1\}^I$, with the Poisson process $\mathbf{Y}$, whose support is $\{0,1,2,\dots\}^I$. For Theorem 2 applied to $\mathbf{X}'$, we use $B_\alpha = \{\alpha\}$, so that $b_1 = \Sigma p_\alpha^2$, $b_2 = b_3 = 0$.

THEOREM 3. *For $\alpha \in I$, let X_α' have the same distribution as X_α, with the X_α' mutually independent. The total variation distance between the dependent Bernoulli process $\mathbf{X} \equiv (X_\alpha)_{\alpha \in I}$, and the independent Bernoulli process $\mathbf{X}' \equiv (X_\alpha')_{\alpha \in I}$ having the same marginals, satisfies*

$$\|\mathscr{L}(\mathbf{X}) - \mathscr{L}(\mathbf{X}')\| \le 2(2b_1 + 2b_2 + b_3) + 4\sum p_\alpha^2.$$

We observe that the total variation distance $\|\mathscr{L}(\mathbf{X}) - \mathscr{L}(\mathbf{X}')\|$ can be interpreted as twice the minimum value of $P(\mathbf{X} \neq \mathbf{X}')$ over all realizations of both processes on the same probability space. By bounding the distance between $\mathbf{X}'$ and $\mathbf{Y}$ coordinatewise, we see that the $4\Sigma p_\alpha^2$ in Theorem 3 could be improved to $2\Sigma p_\alpha^2$.

OPEN PROBLEM. It is natural to ask, in cases in which Σp_α^2 is *not* small, so that the Poisson approximation is not useful, what comparison can be made between a dependent Bernoulli process $\mathbf{X}$ and the independent Bernoulli process $\mathbf{X}'$ having the same marginals?

In many applications, the appropriate choice of B_α makes X_α independent of $\sigma(X_\beta\colon \beta \in I - B_\alpha)$, so that $b_3 = 0$, and this can be verified without performing any calculations. In these situations, calculating b_1 and b_2 is essentially equivalent to computing the first and second moments of W—both tasks involve only the quantities p_α and $p_{\alpha\beta}$. The first sentence in this paper refers to these situations. In fact, when X_α is independent of $\sigma(X_\beta\colon \beta \in I - B_\alpha)$, our upper bound on $\|\mathscr{L}(\mathbf{X}) - \mathscr{L}(\mathbf{Y})\|$ is $4(b_1 + b_2)$, and $b_2 - b_1 = E(W^2) - \lambda - \lambda^2 = E(W^2) - E(Z^2)$, so our upper bound is small if and only if both b_1 is small, and the discrepancy in the second moment of W relative to the Poisson is small. In most applications, the quantities p_α and $|B_\alpha|$ are constant as α varies over I, so that $b_1 = \lambda^2|B_\alpha|/|I|$, hence for fixed λ, b_1 is small if and only if the neighborhood B_α is small relative to the entire index set.

There are situations involving long-range dependence in which the Chen–Stein method is applicable, with $b_3 > 0$. For example, to analyze the Mood test, which is based on the length of the longest head run in $m + n$ tosses of a coin *given* that there are exactly n heads, the number of head runs of a test length t can be approximated along the lines of Example 3, but upper bounds on b_3, as opposed to asymptotic upper bounds, are quite messy to derive. Chen (1975) discusses an example with a "ϕ-mixing" condition on the X_α, so that a bound on b_3 is available by hypothesis.

Of course, two moments of W alone cannot determine the distribution of W, which is an arbitrary nonnegative integer valued random variable. But it is not so naive to ask whether or not a Poisson approximation could be established in terms of just the quantities p_α and $p_{\alpha\beta}$. For example, consider

$$b_0 \equiv \sum_{\alpha \in I} p_\alpha^2 + \sum_{\alpha \neq \beta \in I} \left|\operatorname{cov}(X_\alpha, X_\beta)\right|. \tag{3}$$

In those of our applications in Section 3 in which λ stays bounded, b_0 is small and a Poisson approximation is valid. Is there a family of examples in which b_0 becomes arbitrarily small, whereas W stays bounded away from the Poisson? The reader is urged to try to resolve this question, before turning to our answer at the end of this paper, just before the references.

3. Examples of applications. The following five examples are all discussed only at the level of Theorem 1. The last three involve the maximum of a stationary random sequence or random field. Each example may be viewed as a sequence of problems of increasing size, in which the number W of occurrences has a Poisson limit. A bound on the rate of convergence is obtained as a bonus.

Theorem 2 gives "spatial" information about the locations of occurrences. It may be more convenient to use Theorem 2 to show that the locations of the occurrences converge to a spatial Poisson process in the usual sense, by taking appropriate rescalings and partitions of the index set. The approximation of a discrete intensity measure by its continuous limit then introduces another error (use your favorite metric on the space of nonnegative measures) on top of the approximation error controlled by Theorem 2.

In Example 5, for instance, when $\lambda_0 \in (0, \infty)$ and $m, n, t \to \infty$ so that $\lambda(m, n, t) \to \lambda_0$, the random measure $\sum_{\alpha=(i,j)\in I:\ X_\alpha = 1} \delta_{(i/m,\, i/n)}$, where $\delta_{(x,y)}$ denotes unit mass at the point (x, y), converges in distribution to the Poisson process on $[0,1]^2$ with constant intensity λ_0 times Lebesgue measure. This example was our original motivation for proving Theorem 2; the Poisson process limit was established for the special case $\alpha = 1$ and $\log(m)/\log(n) \to 1$ by the method of moments in Arratia, Gordon and Waterman (1986).

In Examples 2–4, the first half of Example 1 and some cases of Example 5, the Poisson convergence could also have been established by the method of moments. In the context of Poisson convergence, using the method of moments is equivalent to using Laplace transforms or using inclusion–exclusion; see, for example, Watson (1954) or Arratia, Gordon and Waterman (1986). Example 5, which arises naturally in trying to assess the significance of matchings between DNA sequences, has cases in which $E(W^3) \to \infty$, whereas Chen's method proves that W converges in distribution to a Poisson limit.

All our examples have cases, such as the second half of Example 1, in which both $\lambda \to \infty$, and the total variation distance to the Poisson distribution tends to 0. In these cases, the Poisson distribution may be approximated by the normal, so that Chen's method is an easy way of proving a central limit theorem.

In summary, Chen's method of establishing a Poisson limit, compared with the method of moments or inclusion–exclusion,

1. is easier to use;
2. gives a rate of convergence; and
3. may work even when moments higher than the second blow up.

EXAMPLE 1 (A random graph problem). This problem comes from Rinott. On the cube $\{0,1\}^n$, assume that each of the $n2^{n-1}$ edges is assigned a random direction by tossing a fair coin and consider W, the number of vertices at which all n edges point inward. Here, I is the set of all 2^n vertices, X_α is the indicator that vertex α has all of its edges directed inward, $p_\alpha = 2^{-n}$ and $\lambda = 1$. We take $B_\alpha \equiv \{\beta\colon\ |\alpha - \beta| = 1\}$, so $b_2 = b_3 = 0$ and $b_1 = |I|\,|B_\alpha|p_\alpha^2 = \lambda|B_\alpha|p_\alpha = n2^{-n}$. Thus $\|\mathscr{L}(W) - \mathscr{L}(Z)\| \le 2b_1 = 2n2^{-n}$.

There are many other tractable variants of this problem. We give an example in which $\lambda \to \infty$ at the same time that Chen's method works, so that the Poisson approximation may be further approximated by a normal. With the same cube and random edges, let $W \equiv W(k,n)$ be the number of vertices at which exactly k edges point outward, so the special cases $k = 0$ was handled previously. Let I be the set of all 2^n vertices and X_α be the indicator that vertex α has exactly k of its edges directed inward. We have

$$p_\alpha = 2^{-n}\binom{n}{k} \quad \text{and} \quad \lambda = \binom{n}{k}.$$

Let $B_\alpha \equiv \{\beta\colon |\alpha - \beta| = 1\}$, so $b_3 = 0$ and

$$b_1 = np_\alpha\lambda = n2^{-n}\binom{n}{k}^2.$$

For $|\alpha - \beta| = 1$, by conditioning on the direction of the edge between α and β, we see that

$$p_{\alpha\beta} = 2^{2-2n}\binom{n-1}{k}\binom{n-1}{k-1} \le p_\alpha^2,$$

so $b_2 < b_1$. Using $\lambda \ge 1$, Theorem 1 gives

$$\|\mathscr{L}(W) - \mathscr{L}(Z)\| \le 2(b_1 + b_2)/\lambda < 4b_1/\lambda = 4n2^{-n}\binom{n}{k}.$$

Notice that there are cases in which $b_1 \to \infty$, whereas $b_1/\lambda \to 0$ and the Poisson convergence is established. Notice also how easily Chen's method has yielded a central limit theorem: For k, n with $0 < k < n$ and $n2^{-n}\binom{n}{k} \to 0$,

$$\left\{W(k,n) - \binom{n}{k}\right\} \Big/ \sqrt{\binom{n}{k}}$$

converges in distribution to the standard normal.

EXAMPLE 2 (The birthday problem). We first learned about Chen (1975) from a lecture on the birthday problem and its variants by Diaconis, who also suggested references on the birthday problem: Diaconis and Mosteller (1988), Janson (1986) and Stein (1986b), which gives proofs of more general results using similar techniques.

Suppose n balls (people) are uniformly and independently distributed into d boxes (days of the year), and we want to approximate the probability that at least one box receives k or more balls, for fixed $k = 2, 3, \dots$. Only in the "classical" case $k = 2$ is there a simple exact formula, $P(W = 0) =$

$d^{-n}d!/(d-n)!$, but the Chen–Stein method is robust and easily establishes a simple approximation in many variants of the classical birthday problem.

Let $I \equiv \{\alpha \subset \{1,2,\dots,n\}: |\alpha| = k\}$ and let X_α be the indicator of the event that the balls indexed by α all go into the same box. Then $\forall\ \alpha,\ p_\alpha = d^{1-k}$, $\lambda = \binom{n}{k}d^{1-k}$ and P(no box gets k or more balls) $= P(W=0)$ is approximated by

$$P(Z=0) = \exp(-\lambda) = \exp\left\{-\binom{n}{k}d^{1-k}\right\}.$$

We take $B_\alpha \equiv \{\beta \in I: \alpha \cap \beta \neq \varnothing\}$, hence $b_3 = 0$. Since

$$|B_\alpha| = \binom{n}{k} - \binom{n-k}{k},$$

we have $b_1 = p_\alpha^2|I|\,|B_\alpha| = \lambda^2|B_\alpha|/|I| < \lambda^2k^2n^{-1}$, with asymptotic equality as $n \to \infty$.

In the classical case $k = 2$ we have $\forall\ \alpha \neq \beta,\ p_{\alpha\beta} = p_\alpha p_\beta$, which is a nice natural example of pairwise but not mutual independence. Now

$$b_2 = |I|(|B_\alpha| - 1)p_{\alpha\beta} = b_1(|B_\alpha| - 1)/|B_\alpha| < b_1 < 4\lambda^2/n,$$

so that

$$\|\mathscr{L}(W) - \mathscr{L}(Z)\| \le \{16\lambda^2(1-e^{-\lambda})/\lambda\}n^{-1}.$$

Direct comparison of $P(W=0) = d^{-n}d!/(d-n)!$ and $P(Z=0) = e^{-\lambda}$ shows that if $n, d \to \infty$ in such a way that $\lambda = n(n-1)/2d$ is bounded away from 0 and ∞, then

$$\|\mathscr{L}(W) - \mathscr{L}(Z)\| > 2|P(W=0) - P(Z=0)| > Cn^{-1},$$

for some nonzero constant. Hence the bound on total variation distance, given easily by Chen's method, is sharp apart from a constant factor.

In the general case $k \ge 2$ we have

$$b_2 = \sum_{j=1}^{k-1}\binom{n}{k}\binom{k}{j}\binom{n-k}{k-j}d^{1+j-2k},$$

where the jth term is the contribution to b_2 from pairs (α,β) with $|\alpha \cap \beta| = j$ and $p_{\alpha\beta} = d^{1+j-2k}$. With d/n is large, the dominant contribution to b_2 comes from the pairs (α,β) with $|\alpha \cap \beta| = j = k-1$. Now take $n, d \to \infty$ in such a way that the ratio $\lambda/1$ is bounded away from 0 and ∞, which we denote by $\lambda \asymp 1$. Then $n^k \asymp d^{k-1}$, $b_1 \asymp |B_\alpha|/|I| \asymp n^{-1}$ and $b_2 \asymp n^{1+k}d^{-k} \asymp n/d \asymp n^{-1/(k-1)}$. Thus for $k \ge 3$ we have $\|\mathscr{L}(W) - \mathscr{L}(Z)\| = O(n^{-1/(k-1)})$, with b_2 making the main contribution.

The bounds above can be improved by a factor of $k^2/(k-1)^2$ if we change the natural definition of B_α to the following less natural definition, in which we make a canonical choice of some element of α and allow $\beta \in B_\alpha$ even if β overlaps α at this one element:

$$B_\alpha \equiv \{\beta \in I: (\alpha - \min(\alpha)) \cap \beta \neq \varnothing\},$$

so that

$$|B_\alpha| = \binom{n}{k-1} - \binom{n-(k-1)}{k-1}.$$

Since the distribution of the balls is *uniform*, we still have $b_3 = 0$.

EXAMPLE 3 (The longest perfect head run). Let $0 < p < 1$ and $Z_0, Z_1, Z_2, \ldots$ be an i.i.d. sequence with $p = P(Z_i = 1) = 1 - P(Z_i = 0)$. Let R_n be the length of the longest consecutive run of heads, i.e., 1's, starting within the first n tosses. We observe that R_n is the maximum of n terms from a stationary sequence of dependent, geometrically distributed random variables. The asymptotic distribution of R_n is discussed in Guibas and Odlyzko (1980) and in Gordon, Schilling and Waterman (1986), where a variant problem allowing a fixed number k of tails is also handled.

Let $I \equiv \{1, 2, \ldots, n\}$ and fix a positive integer "test" value t. Let $X_1 \equiv Z_1 Z_2 \cdots Z_t$, and

$$\text{for } \alpha = 2 \text{ to } n, \quad X_\alpha \equiv (1 - Z_{\alpha-1}) Z_\alpha Z_{\alpha+1} \cdots Z_{\alpha+t-1}.$$

As events, $\{R_n < t\} = \{W = 0\}$. Notice that we are dealing directly with "boundary effects," so that the X_α are not stationary—X_1 is different from the other X_α's. In Example 4 we handle the boundary by a different method, which would also work in this example. Now

$$\lambda \equiv \lambda(n, t) \equiv EW = p^t\{(n-1)(1-p) + 1\}. \tag{4}$$

Let $B_\alpha \equiv \{\beta \in I\colon |\alpha - \beta| \le t\}$ for $\alpha = 1$ to n, so that $b_2 = b_3 = 0$ and $b_1 < \lambda^2(2t+1)/n + 2\lambda p^t$. The distribution of R_n is controlled by

$$|P(R_n < t) - e^{-\lambda(n,t)}| \le b_1(1 \wedge 1/\lambda). \tag{5}$$

Now λ stays bounded away from 0 and ∞ if and only if $t - \log_{1/p}(n)$ stays bounded, and in this case, $b_1 \to 0$ as $n \to \infty$ [in fact, $b_1 = O(\log(n)/n)$ and careful analysis of $P(W = 0)$ using inclusion–exclusion shows that $|P(W = 0) - e^{-\lambda}|n/\log(n)$ is bounded away from 0]. From (4) and (5) it follows that the family $\{R_n - \log_{1/p}(n(1-p))\}$ is tight. All of the limit distributions of this family may be described as those of the "integerized extreme value" random variables $Y_r \equiv \lfloor Y + r \rfloor - r$ for $r \in [0, 1]$, where $P(Y < c) = \exp(-p^c)$. Furthermore, from (4) and (5) one can see that $R_n - \log_{1/p}(n(1-p)) \to Y_r$ in distribution if and only if $n \to \infty$ along a subsequence such that, taken modulo 1, $\log_{1/p}(n(1-p)) \to r$.

EXAMPLE 4 (The Erdös–Rényi law, in distribution). Let $0 < p < a \le 1$ and $\ldots, Z_{-1}, Z_0, Z_1, Z_2, \ldots$ be an i.i.d. sequence with $p = P(Z_i = 1) = 1 - P(Z_i = 0)$. Let R_n be the length of the longest consecutive run, starting within the first n tosses, in which the fraction of heads is at least a. The previous example is the special case $a = 1$ of the current example. Erdös and Rényi (1970) prove that

$$R_n/\log(n) \to 1/H(a, p) \quad \text{almost surely,}$$

where

$$(6)\qquad H(a,p) = a\log(a/p) + (1-a)\log((1-a)/(1-p))$$

is the relative entropy of a and p, with $H(1,p) = \log(1/p)$. Using Chen's method, it is not hard to approximate the distribution of R_n. The approximation implies that, for $a \neq 1$, the family of random variables

$$\{R_n - (\log(n) - \tfrac{1}{2}\log\log(n))/H(a,p)\}$$

is tight. This is sketched in the following discussion, and done in detail in Arratia, Gordon and Waterman (1988).

For t a positive integer and $a \in [0,1]$, define indicators

$$Y_{\alpha,t} \equiv 1\left(at \le \sum_{k=0}^{t-1} Z_{\alpha+k}\right), \qquad Y'_\alpha \equiv \max\{Y_{\alpha,t}, Y_{\alpha,t+1},\dots,Y_{\alpha,2t}\},$$

$$X_{\alpha,t} \equiv Y_{\alpha,t}\prod_{j=1}^{t}(1 - Y_{\alpha-j,t}), \qquad X'_\alpha \equiv Y'_\alpha \prod_{j=1}^{2t}(1 - Y'_{\alpha-j}).$$

Let $I \equiv \{1,2,\dots,n\}$, and define

$$W \equiv \sum_{\alpha\in I} X_{\alpha,t}, \qquad W' \equiv \sum_{\alpha\in I} X'_\alpha.$$

Apart from "boundary effects," W is the number of places within the first n tosses at which a "quality a, length t" head run begins, and W' is the number of places within the first n tosses at which a "quality a, length t or greater" head run begins, so the event $\{R_n < t\}$ can be approximated by the event $\{W' = 0\}$. The error in this approximation can be controlled by observing that

$$\{R_n < t, W' \neq 0\} \cup \{R_n \ge t, W' = 0\} \subset \{Y'_1 + \cdots + Y'_{2t} > 0\},$$

so that

$$(7)\qquad |P(W'=0) - P(R_n < t)| \le 2tEY'_1 \le 2te^{-tH(a,p)}.$$

The easy bounds $EY_{\alpha,t} \le EY'_\alpha \le e^{-tH(a,p)}$ can be proved by Cramér's argument: Compute expectations with respect to the probability Q under which $\dots, Z_{-1}, Z_0, Z_1, Z_2, \dots$ are a-coins, and observe that on the event $\{Y'_\alpha = 1\}$, the Radon–Nikodym derivative satisfies $dP/dQ \le e^{-tH(a,p)}$.

(In the definition of $X_{\alpha,t}$ in terms of Y, or of X'_α in terms of Y', the upper bound t or $2t$ in the product over j could easily be replaced by anything that tends to infinity with t, but in the definition of Y'_α in terms of $Y_{\alpha,t}, Y_{\alpha,t+1},\dots,Y_{\alpha,2t}$, the upper bound $2t$ is minimal, since we wish to use the following argument: "When the fraction of heads in a window is at least a, the same must be true in the first or second half of the window; hence, if there is a quality a window of length t or greater, there must be one of length between t and $2t$ inclusive.")

For both W and W', it is easy to establish a Poisson approximation using the Chen–Stein method. In this paragraph we handle the case W', together with its relation to R_n. Let $\lambda' \equiv EW'$. Let $B_\alpha \equiv \{\beta \in I\colon |\alpha - \beta| < 4t\}$ for $\alpha = 1$ to n, so that $b_3 = 0$. If $|\alpha - \beta| \le 2t$, then $E(X'_\alpha X'_\beta) = 0$, but if $2t < |\alpha - \beta| < 4t$, then we can only conclude that $E(X'_\alpha X'_\beta) \le (EX'_\alpha)EY'_\beta$, so that $b_2 < 4t\lambda' EY'_\alpha$ and $b_1 = (8t - 1)\lambda' EX'_\alpha$. Combining (7) with Theorem 1, we have

$$(8)\qquad \begin{aligned} |P(R_n < t) - e^{-EW'}| &\le 2te^{-tH(a,p)} + (b_1 + b_2)(1 \wedge 1/\lambda') \\ &\le 2te^{-tH(a,p)} + 4t\lambda'(2EX'_\alpha + EY'_\alpha)(1 \wedge 1/\lambda') \\ &\le 2te^{-tH(a,p)} + 12te^{-tH(a,p)} \\ &\le 14te^{-tH(a,p)} \to 0 \quad \text{as } t \to \infty. \end{aligned}$$

To calculate $p_\alpha \equiv EX_{\alpha,t}$ is much easier than to calculate EX'_α. We can obtain a simple expression for the former, but only lower and upper bounds for the latter. To calculate p_α, let $s \equiv \lceil at \rceil$ and consider the events

$$A = \left\{ s = \sum_0^{t-1} Z_k \right\}, \qquad B = \left\{ \sum_{k=1}^{m} (Z_{t-k} - Z_{-k}) > 0 \text{ for } m = 1, \dots, t \right\},$$

so that $X_{0,t} = 1(A)1(B)$ and $p_\alpha = P(A)P(B|A)$. An argument involving exchangeability and the ballot theorem shows that $P(B|A) \to (a - p)$ as $t \to \infty$ (and a large deviation argument shows that this convergence is exponentially fast in t). Thus we have the asymptotic relations, as $t \to \infty$, $p_\alpha \sim (a - p)\binom{t}{s} p^s (1 - p)^{t-s}$ and

$$(9)\qquad \lambda \equiv \lambda(n, t) \equiv EW = np_\alpha \sim (a - p)n\binom{t}{s} p^s (1 - p)^{t-s}.$$

In the case $a \ne 1$, Stirling's formula lets us express λ in terms of the relative entropy $H(\cdot, p)$: as $t \to \infty$,

$$(10)\qquad \lambda \sim (a - p)n\{2\pi a(1 - a)t\}^{-1/2} \exp\{-tH(s/t, p)\},$$

and good explicit lower and upper bounds can be given. Although $s \equiv \lceil at \rceil$, it is not possible to replace s/t by a in the argument to $H(\cdot, p)$, since the resultant change would be approximately a factor of $\exp\{-t(s/t - a)\partial H/\partial a(a, p)\}$, where $s \equiv \lceil at \rceil$.

Now for $a \ne 1$, $\lambda \equiv EW$ stays bounded away from 0 and infinity if and only if $t - \{\log(n) - \frac{1}{2}\log\log(n)\}/H(a, p)$ stays bounded. It is not hard to show that there is a constant $c_{a,p} < \infty$, independent of t and n, such that $\lambda' \equiv EW'$ satisfies $\lambda/c_{a,p} < \lambda' < c_{a,p}\lambda$. Write $R_n^* \equiv R_n - \{\log(n) - \frac{1}{2}\log\log(n)\}/H(a, p)$. From (10) and (8) it follows that the family $\{R_n^*\}$ is tight.

EXAMPLE 5 (The Erdös–Rényi law for matching two random sequences). This example is closely related to Examples 3 and 4; the special case $a = 1$, corresponding to perfect matching, has been discussed in Arratia and Waterman (1985a, b), and Arratia, Gordon and Waterman (1986). Let $\dots, A_{-1}, A_0, A_1, A_2, \dots$ and $\dots, B_{-1}, B_0, B_1, B_2, \dots$ be i.i.d. integer valued random "letters," say, with common nontrivial distribution μ. Let $p \equiv \sum_{l \in Z} \mu_l^2$, so that $\forall\ i, j$, $p = P(A_i = B_j)$. Let $a \in (p, 1]$. We are interested in the asymptotic distribution, for large m and n, of the length $M_{m,n}$ of the longest "quality a" matching consecutive segment common to the two sequences $A_1, \dots, A_m$ and $B_1, \dots, B_n$.

For t a positive integer and $a \in [0,1]$, define indicators

$$Z_{ij} \equiv 1(A_i = B_j), \qquad Y_{ij} \equiv 1\left(\text{for some } l \in [t, 2t],\ al \le \sum_{k=0}^{l-1} Z_{i+k,\, j+k}\right),$$

$$X_{ij} \equiv Y_{ij} \prod_{l=1}^{2t} (1 - Y_{i-l,\, j-l}).$$

Let $I \equiv \{1, 2, \dots, m\} \times \{1, 2, \dots, n\}$ and

$$W \equiv W(m, n, a, t) \equiv \sum_{\alpha \in I} X_\alpha.$$

Define $M^*_{m,n}$ by $M_{m,n} = \max\{t\colon W(m, n, a, t) > 0\}$. Apart from boundary effects, which can be controlled as in the previous example, $M^*_{m,n}$ agrees with the length of the longest quality a matching, $M_{m,n}$. We observe that $M^*_{m,n}$ is the maximum of an $m \times n$ rectangle of random variables from a two-parameter stationary sequence, where the same stationary sequence is used for all m, n. If a Poisson approximation can be established by the Chen–Stein method, the net result can be expressed as a comparison between the distribution of the length of the longest quality a matching for sequences of length m and n, and the distribution of the length of the longest quality a head run in a sequence of tosses of a p-coin of length mn: As $m, n \to \infty$,

$$\max_t \left| P(M_{m,n} < t) - P(R_{(mn)} < t) \right| \to 0,$$

with an explicit bound on the rate of convergence.

The essential novel feature of this problem, in contrast to Examples 3 and 4, is that distinct "coins" Z_{ij} and Z_{kl} are *strictly* positively correlated if $i = k$ or else $j = l$, provided that μ is *not* the uniform distribution on a finite set. Let

$$p_3 \equiv E(Z_{11} Z_{12}) = P(A_1 = B_1 = B_2) = \sum_{l \in Z} (\mu_l)^3.$$

By Jensen's inequality, $p_3 < p^{3/2}$. Define $\theta \equiv \theta(a, p, p_3)$ by

$$1 + \theta = \lim \log\bigl(E(Y_{00} Y_{0,4t})\bigr) / \log(EY_{00}),$$

where large deviation arguments show that the limit $\theta \in (0,1]$ exists, and furthermore $1 + \theta = \lim \log(E(X_{00} X_{0,4t})) / \log(EX_{00})$. For the case $a = 1$, it is obvious that $1 + \theta = \log(p_3)/\log(p)$, so that $\theta > \frac{1}{2}$. We will show in the next paragraph why $\theta > \frac{1}{2}$ is necessary and sufficient for Chen's method to succeed in establishing a Poisson convergence for W, as $m = n \to \infty$ and t grows appropriately. In Arratia, Gordon and Waterman (1988) it is shown that, perhaps surprisingly, there are nontrivial cases in which $\theta(a, p, p_3) < \frac{1}{2}$, so that a Poisson convergence cannot be established by this method. Nevertheless, the analog of the Erdös–Rényi law for matching two sequences of *equal* length, namely that

$$M_{n,n} / \log(n^2) \to 1/H(a, p) \quad \text{almost surely},$$

always holds; this is proved in Arratia and Waterman (1989).

For $\alpha = (i, j) \in I$, let $B_\alpha = \{\beta = (k, l) \in I\colon\ |i - k| \wedge |j - l| < 4t\}$ and $C(\alpha) = \{\beta = (k, l) \in I\colon |i - k| \vee |j - l| < 4t\}$, so that B_α is the union of two perpendicular strips and $C(\alpha)$ is the square where the strips intersect. We have $b_3 = 0$ and $b_1 = \lambda^2|B_\alpha|/|I| < \lambda^2 8t(m + n)/(mn)$. Assume that $m, n, t \to \infty$ with both λ and $\log(m)/\log(n)$ bounded away from 0 and ∞. The central contribution to b_2, $b_2^* \equiv \sum_{\beta \in C(\alpha)} p_{\alpha\beta}$, converges to 0 in all cases. There are asymptotically $mn(m + n)$ pairs (α, β), which are identical in one coordinate and differ by at least $4t$ in the other coordinate. These pairs have $p_{\alpha\beta} = E(X_{00}X_{0,4t}) \approx (p_\alpha)^{1+\theta}$, where we write $x \approx y$ to mean $\log(x) \sim \log(y)$. A little more work shows that the remaining contributions to b_2 do not increase the exponential growth rate, so $b_2 - b_2^* \approx mn(m + n)(p_\alpha)^{1+\theta}$. Suppose further now that $\log(n)/\log(mn) \to \rho \in [\frac{1}{2}, 1)$. Then

$$b_2 - b_2^* \approx mn(m + n)(p_\alpha)^{1+\theta} \approx (mn)^{1+\rho}(p_\alpha)^{1+\theta} = (mn)^{\rho-\theta}\lambda^{1+\theta} \approx (mn)^{\rho-\theta}.$$

Thus if $\theta < \frac{1}{2}$, then $b_2 \to \infty$ and Chen's method fails (in fact $EW^2 \to \infty$) for all values of ρ, including the important special case with $m = n$ and $\rho = \frac{1}{2}$. In cases where $\theta > \frac{1}{2}$, we have that $b_2 \to 0$ and Chen's method yields a Poisson limit, uniformly in m, n, t such that λ stays in a compact subset of $(0, \infty)$ and $\log(n)/\log(mn)$ stays in a compact subset of $(1 - \theta, \theta)$. This latter condition on the relative growth of m and n is the same as would be required to have EW^2 stay bounded. For any nonuniform distribution μ for the i.i.d. letters, strictly stronger conditions would be required to have the higher moments of W stay bounded. Thus, for example with $a = 1$, we have $EW^3 \to \infty$ if $\rho \equiv \lim \log(n)/\log(mn) \in (\phi/2, \theta)$, where $\phi \geq 1$ is defined by $p_4 \equiv \Sigma(\mu_l)^4 = p^{1+\phi}$ and $\phi < 2\theta$ iff μ is nonuniform. In these cases, Chen's method establishes the Poisson convergence, together with a bound on the rate of convergence, whereas the method of moments blows up.

4. Bounds on the inverse operator.

Recall that Z denotes a Poisson random variable with mean λ. Define linear operators S and T, which depend on the parameter λ, by

$$(Tf)(w) \equiv wf(w) - \lambda f(w + 1), \quad \text{for } w \geq 0$$

and

$$(Sh)(w + 1) \equiv -\lambda^{-1}P(Z = w)^{-1}E(h(Z); Z \leq w), \quad \text{for } w \geq 0.$$

To be definite, we let $(Sh)(0) = 0$, but this is an arbitrary choice; the value $Sh(0)$ is never used. Note that S is inverse to T in that $\forall\ h$, $T(Sh) = h$, and that "$\forall\ f$ bounded, $E[(Tf)(Z)] = 0$" precisely characterizes the distribution of Z as Poisson with mean λ. We write Δf for the function defined by $(\Delta f)(w) = f(w + 1) - f(w)$. The proof of the following lemma is in the appendix to Barbour and Eagleson (1983).

LEMMA 1. *Suppose that* $\forall\ w \geq 0$, $h(w) \in [0, 1]$ *and* $f = S(h(\cdot) - Eh(Z))$. *Then*

$$\|\Delta f\| \leq (1 - e^{-\lambda})/\lambda \quad \textit{and} \quad \|f\| \leq 1 \wedge 1.4\lambda^{-1/2}.$$

Furthermore, if $h(w) = 1(w = 0) - e^{-\lambda}$, *then* $\|f\| = (1 - e^{-\lambda})/\lambda$.

The starting point for obtaining these bounds is the observation that if $Eh(Z) = 0$, then

$$(Sh)(w+1) = -\lambda^{-1}P(Z = w)^{-1}\operatorname{cov}(h(Z), 1(Z \le w)).$$

For fixed $k \ge 0$, if $h(w) = 1(w \le k) - P(Z \le k)$, then

$$\operatorname{cov}(h(Z), 1(Z \le w)) = P(Z \le k \wedge w) - P(Z \le k)P(Z \le w).$$

Since $d/d\lambda P(Z \le j) = -P(Z = j)$, we have $P(Z \le j) = 1 - \int_0^\lambda e^{-v}v^j/j!\,dv = \int_\lambda^\infty e^{-v}v^j/j!\,dv$. Combining these ingredients, for the special case $k = 0$, we have $(1 - e^{-\lambda})/\lambda = -f(1) > -f(2) > \cdots > 0$, which proves the last part of this lemma.

5. Proof in the one-variable case. In this section we give the proof of Theorem 1. Let h be given with $\|h\| = 1$. Let $\bar{h}(\cdot) \equiv h(\cdot) - Eh(Z)$ and $f \equiv S\bar{h}$, so that $Tf = \bar{h}$, and $E[Tf(W)] = E[h(W) - h(Z)]$. The series of equalities and one inequality below show that

$$|E\{h(W) - h(Z)\}| \le (b_1 + b_2)\|\Delta f\| + b_3'\|f\|. \tag{11}$$

Combining (11) with the bounds on $\|f\|$ and $\|\Delta f\|$ from Lemma 1 completes the proof; a factor of 2 is introduced in handling the positive and negative parts of h separately.

Write $V_\alpha \equiv \sum_{\beta \in I - B_\alpha} X_\beta$ and $W_\alpha \equiv W - X_\alpha$, so that $V_\alpha \le W_\alpha \le W$, $X_\alpha f(W) = X_\alpha f(W_\alpha + 1)$ and $f(W_\alpha + 1) - f(W + 1) = X_\alpha[f(W_\alpha + 1) - f(W_\alpha + 2)]$. We compute

$$\begin{aligned}
E\{h(W) - h(Z)\}& \\
&= E\{Wf(W) - \lambda f(W+1)\} \\
&= \sum_{\alpha \in I} E\{X_\alpha f(W) - p_\alpha f(W+1)\} \\
&= \sum_{\alpha \in I} E\{p_\alpha f(W_\alpha + 1) - p_\alpha f(W+1)\} \\
&\quad + \sum_{\alpha \in I} E\{X_\alpha f(W_\alpha + 1) - p_\alpha f(W_\alpha + 1)\} \\
&= \sum_{\alpha \in I} E\{p_\alpha X_\alpha[f(W_\alpha + 1) - f(W_\alpha + 2)]\} \qquad (12)\\
&\quad + \sum_{\alpha \in I} E\{(X_\alpha - p_\alpha)[f(W_\alpha + 1) - f(V_\alpha + 1)]\} \\
&\quad + \sum_{\alpha \in I} E\{(X_\alpha - p_\alpha) f(V_\alpha + 1)\} \\
&\le \|\Delta f\| \sum_{\alpha \in I} p_\alpha^2 + \|\Delta f\|\Big(\sum_{\alpha \in I} \sum_{\alpha \ne \beta \in B_\alpha} p_{\alpha\beta} + p_\alpha p_\beta\Big) + b_3'\|f\| \\
&= (b_1 + b_2)\|\Delta f\| + b_3'\|f\|.
\end{aligned}$$

The upper bound on (12) is justified as follows. The first sum is less than or equal to $\|\Delta f\|\sum_{\alpha\in I}p_\alpha^2$. For the second sum, each term can be written as a telescoping sum of $|B_\alpha| - 1$ terms, each with one more summand X_β being left out. Each such term in the telescoping sum is of the form

$$\begin{aligned} E\{(X_\alpha - p_\alpha)(f(U + X_\beta) - f(U))\} &= E\{(X_\alpha - p_\alpha)X_\beta(f(U+1) - f(U))\} \\ &= E\{X_\alpha X_\beta \Delta f(U)\} - E\{p_\alpha X_\beta \Delta f(U)\} \\ &\le \|\Delta f\|(p_{\alpha\beta} + p_\alpha p_\beta). \end{aligned}$$

Thus our upper bound for the second term is

$$\|\Delta f\| \sum_{\alpha\in I} \sum_{\alpha\neq\beta\in B_\alpha} (p_{\alpha\beta} + p_\alpha p_\beta),$$

which combined with the bound on the first term yields $(b_1 + b_2)\|\Delta f\|$. The third term is bounded by $b_3'\|f\|$. □

6. Proof in the multivariable case. In this section we give the proof of Theorem 2. The process bound is a consequence of the d-dimensional bound, as follows. Given $\varepsilon > 0$, since $\lambda < \infty$, there exist $d < \infty$ and $d - 1$ distinct indices $\alpha(1),\dots,\alpha(d-1)$ with $\sum_{1\le j<d} p_{\alpha(j)} > \lambda - \varepsilon$. Applying the d-dimensional bound with singletons $I(j) = \{\alpha(j)\}$ and $I(d) = I - \{\alpha(1),\dots,\alpha(d-1)\}$ and using Chebyshev's bound $P(W_d > 0) < \varepsilon$, $P(Z_d > 0) < \varepsilon$, we have

$$\|\mathscr{L}(\mathbf{X}) - \mathscr{L}(\mathbf{Y})\| \le 2(2b_1 + 2b_2 + b_3) + 2\varepsilon,$$

which establishes the first part of Theorem 2.

For $i = 1,\dots,d$, write $\mathbf{e}_i \in R^d$ for the unit vector with 1 as its ith coordinate. Write $\mathbf{j} \equiv (j_1,\dots,j_d)$ for an element of $(Z^+)^d$ and write $f, h, f_1,\dots,f_d$ for functions from $(Z^+)^d$ into R. Define linear operators S_i and T_i to be the analogs of S and T from Theorem 1, but acting only on the ith coordinate:

$$(T_i f)(\mathbf{j}) \equiv j_i f(\mathbf{j}) - \lambda_i f(\mathbf{j} + \mathbf{e}_i),$$

$$(S_i h)(\mathbf{j} + \mathbf{e}_i) \equiv -(\lambda_i P(Z_i = j_i))^{-1} E\{h(j_1,\dots,Z_i,\dots,j_d)1(Z_i \le j_i)\}.$$

As before, for $i = 1,\dots,d$, for all h, $T_i S_i h = h$.

Define a linear operator P_i that averages over the ith coordinate with respect to the Poisson distribution with parameter λ_i:

$$(P_i h)(\mathbf{j}) \equiv Eh(j_1,\dots,Z_i,\dots,j_d),$$

so that $P_i h$ is a function of the $d - 1$ coordinates other than the ith coordinate. Write Δ_i for the difference operator on the ith coordinate, so that $(\Delta_i f)(\mathbf{j}) \equiv f(\mathbf{j} + \mathbf{e}_i) - f(\mathbf{j})$. Fix h with $\|h\| \le 1$ and let

$$f_i \equiv S_i(h - P_i h),$$

so that $T_i f_i = h - P_i h$. By Lemma 1 (treating the $d - 1$ arguments other than the ith as fixed parameters and handling the positive and negative parts of h separately), $\|f_i\| < 2(1 \wedge 1.4\lambda_i^{-1/2})$, so $\|\Delta_k f_i\| < 4(1 \wedge 1.4\lambda_i^{-1/2})$.

An argument similar to the following will be used d times. At the ith step, each f_1 is replaced by f_i, and the first $i-1$ coordinates are replaced by $Z_1, \dots, Z_{i-1}$, in the vector argument to every function in the series of equalities (13)–(14). On the left side of (13), the ith coordinate is W_i in one term and Z_i in the other, so that the result is an upper bound on

$$|Eh(Z_1, \dots, Z_{i-1}, W_i, W_{i+1}, \dots, W_d) - Eh(Z_1, \dots, Z_{i-1}, Z_i, W_{i+1}, \dots, W_d)|.$$

The combination of these d results yields an upper bound on

$$|Eh(W_1, W_2, \dots, W_d) - Eh(Z_1, Z_2, \dots, Z_d)|,$$

establishing the theorem. Instead of presenting the general ith step, we write out the version for $i = 1$, and then briefly discuss the modifications needed for the steps with $i = 2, \dots, d$.

To begin the step for $i = 1$, for $\alpha \in I(1)$ let $\mathbf{W}_\alpha \equiv (W_1 - X_\alpha, W_2, \dots, W_d)$ and let $\mathbf{V}_\alpha$ be the random element of $(Z^+)^d$ whose kth coordinate is

$$(\mathbf{V}_\alpha)_k \equiv \sum_{\beta \in I(k) - B_\alpha} X_\beta,$$

so that $\mathbf{V}_\alpha \le \mathbf{W}_\alpha \le \mathbf{W}$, coordinatewise. We compute

$$\begin{aligned}
&Eh(W_1, W_2, \dots, W_d) - Eh(Z_1, W_2, \dots, W_d) \\
&\quad = E\{(h - P_1 h)(W_1, W_2, \dots, W_d)\} \\
&\quad = E\{(T_1 f_1)(W_1, W_2, \dots, W_d)\} \qquad (13)\\
&\quad = E\{W_1 f_1(\mathbf{W}) - \lambda_1 f_1(\mathbf{W} + \mathbf{e}_1)\} \\
&\quad = \sum_{\alpha \in I(1)} E\{X_\alpha f_1(\mathbf{W}) - p_\alpha f_1(\mathbf{W} + \mathbf{e}_1)\} \\
&\quad = \sum_{\alpha \in I(1)} E\{p_\alpha f_1(\mathbf{W}_\alpha + \mathbf{e}_1) - p_\alpha f_1(\mathbf{W} + \mathbf{e}_1)\} \\
&\qquad + \sum_{\alpha \in I(1)} E\{X_\alpha f_1(\mathbf{W}_\alpha + \mathbf{e}_1) - p_\alpha f_1(\mathbf{W}_\alpha + \mathbf{e}_1)\} \\
&\quad = \sum_{\alpha \in I(1)} E\{p_\alpha X_\alpha (f_1(\mathbf{W}_\alpha + \mathbf{e}_1) - f_1(\mathbf{W}_\alpha + 2\mathbf{e}_1))\} \\
&\qquad + \sum_{\alpha \in I(1)} E\{(X_\alpha - p_\alpha)[f_1(\mathbf{W}_\alpha + \mathbf{e}_1) - f_1(\mathbf{V}_\alpha + \mathbf{e}_1)]\} \qquad (14)\\
&\qquad + \sum_{\alpha \in I(1)} E\{(X_\alpha - p_\alpha) f_1(\mathbf{V}_\alpha + \mathbf{e}_1)\} \\
&\quad \le 2\| f_1\| \sum_{\alpha \in I(1)} p_\alpha^2 + 2\| f_i\| \left(\sum_{\alpha \in I(1)} \sum_{\alpha \ne \beta \in B_\alpha} p_{\alpha\beta} + p_\alpha p_\beta \right) + \| f_1\| \sum_{\alpha \in I(1)} s_\alpha,
\end{aligned}$$

which is less than or equal to the contribution to $2(1 \wedge 1.4(\min_j \lambda_j)^{-1/2})(2b_1 + 2b_2 + b_3)$ from terms with $\alpha \in I(1)$.

As in the proof of Theorem 1, the first sum in (14) is bounded by $\|\Delta_1 f_1\| \Sigma_{\alpha \in I(1)} p_\alpha^2$, and the third sum is bounded by $\| f_1\| \Sigma_{\alpha \in I(1)} s_\alpha$, which is the contribution to $\|f_1\|b_3$ from terms $\alpha \in I(1)$. For the second sum, the term indexed by α can be written as a telescoping sum of terms indexed by $\beta \in B_\alpha - \{\alpha\}$. When the index β is in $I(k)$, the corresponding summand is of the form

$$\begin{aligned} &E\{(X_\alpha - p_\alpha)(f_1(\mathbf{U} + X_\beta \mathbf{e}_k) - f_1(\mathbf{U}))\} \\ &\quad = E\{(X_\alpha - p_\alpha)X_\beta(f_1(\mathbf{U} + \mathbf{e}_k) - f_1(\mathbf{U}))\} \\ &\quad = E\{X_\alpha X_\beta \Delta_k f_1(\mathbf{U})\} - E\{p_\alpha X_\beta \Delta_k f_1(\mathbf{U})\} \\ &\quad \le 2\|f_1\|(p_{\alpha\beta} + p_\alpha p_\beta). \end{aligned} \tag{15}$$

Thus our upper bound for the second term is

$$2\|f_1\| \sum_{\alpha \in I(1)} \sum_{\alpha \ne \beta \in B_\alpha} (p_{\alpha\beta} + p_\alpha p_\beta),$$

which combined with the bound on the first term yields the contribution to $(b_1 + b_2)2\|f_1\|$ from terms with $\alpha \in I(1)$. This ends the argument for the step with $i = 1$.

For the general ith step, the first $i - 1$ components of V_α are changed to $Z_1, \dots, Z_{i-1}$, and the telescoping sum for bounding the α term of the second sum involves only those β in $I(i) \cup \cdots \cup I(d)$. □

We note that the bound after equation (15) accounts for the difference between the bounds of Theorems 1 and 2—for $\Delta_k f_1$ the only available bound is $\|\Delta_k f_1\| \le 2\|f_1\|$.

Acknowledgments. The authors are very grateful to Persi Diaconis for having brought Chen's work to our attention and to Michael S. Waterman for many stimulating conversations.

In answer to the puzzle at formula (3) at the end of Section 2: "Be wise, symmetrize." Let W be any bounded, positive integer valued random variable having equal mean and variance; say $P(W \le m) = 1$ and $\lambda \equiv EW = \mathrm{var}(W)$. Fix W, and let $n \ge m$; we will consider the limit as n tends to ∞. Let $I \equiv \{1, \dots, n\}$ and let $\{X_\alpha\colon \alpha \in I\}$ be an exchangeable family of Bernoulli variables with $\sum_{\alpha \in I} X_\alpha = W$. Note that the family $\{X_\alpha\colon \alpha \in I\}$ depends on n, but its sum W does not. We have $p_\alpha = \lambda/n$ and

$$\lambda = \mathrm{var}(W) = \sum \mathrm{var}(X_\alpha) + \sum_{\alpha \ne \beta} \mathrm{cov}(X_\alpha, X_\beta) = \lambda - \lambda^2/n + \sum_{\alpha \ne \beta} \mathrm{cov}(X_\alpha, X_\beta).$$

By exchangeability, all of the covariance terms have the same sign, hence

$$\lambda^2/n = \sum_{\alpha \ne \beta} \mathrm{cov}(X_\alpha, X_\beta) = \sum_{\alpha \ne \beta} |\mathrm{cov}(X_\alpha, X_\beta)|$$

and $b_0 = 2\lambda^2/n$.

REFERENCES

ALDOUS, D. (1987). Probability approximations via the Poisson clumping heuristic. *Lecture Notes in Math.* Springer, Berlin. To appear.

ARRATIA, R. and WATERMAN, M. S. (1985a). An Erdös–Rényi law with shifts. *Adv. in Math.* **55** 13–23.

ARRATIA, R. and WATERMAN, M. S. (1985b). Critical phenomena in sequence matching. *Ann. Probab.* **13** 1236–1249.

ARRATIA, R. and WATERMAN, M. S. (1989). The Erdös–Rényi strong law for pattern matching with a given proportion of mismatches. *Ann. Probab.* To appear.

ARRATIA, R., GORDON, L. and WATERMAN, M. S. (1986). An extreme value theory for sequence matching. *Ann. Statist.* **14** 971–993.

ARRATIA, R., GORDON, L. and WATERMAN, M. S. (1988). The Erdös–Rényi law in distribution, for coin tossing and sequence matching. Unpublished manuscript.

BARBOUR, A. D. (1982). Poisson convergence and random graphs. *Math. Proc. Cambridge Philos. Soc.* **92** 349–359.

BARBOUR, A. D. (1987a). Asymptotic expansions in the Poisson limit theorem. *Ann. Probab.* **15** 748–766.

BARBOUR, A. D. (1987b). Stein's method and Poisson process convergence. Preprint.

BARBOUR, A. D. and EAGLESON, G. K. (1983). Poisson approximation for some statistics based on exchangeable trials. *Adv. in Appl. Probab.* **15** 585–600.

BARBOUR, A. D. and EAGLESON, G. K. (1984). Poisson convergence for dissociated statistics. *J. Roy. Statist. Soc. Ser. B* **46** 397–402.

BARBOUR, A. D. and HALL, P. (1984). On the rate of Poisson convergence. *Math. Proc. Cambridge Philos. Soc.* **95** 473–480.

BARBOUR, A. D. and HOLST, L. (1987). Some applications of the Stein–Chen method for proving Poisson convergence. Preprint.

BARBOUR, A. D. and JENSEN, J. L. (1987). Local and tail approximations near the Poisson limit. Preprint.

BOLLOBÁS, B. (1985). *Random Graphs.* Academic, New York.

CHEN, L. H. Y. (1975). Poisson approximation for dependent trials. *Ann. Probab.* **3** 534–545.

DIACONIS, P. and MOSTELLER, F. (1988). On coincidences. Unpublished manuscript.

ERDÖS, P. and RÉNYI, A. (1960). On the evolution of random graphs. *Publ. Math. Inst. Hungar. Acad. Sci.* **5** 17–61.

ERDÖS, P. and RÉNYI, A. (1970). On a new law of large numbers. *J. Anal. Math.* **22** 103–111. Reprinted (1976) in *Selected Papers of Alfred Rényi* **3** 1962–1970. Akademiai Kiado, Budapest.

GORDON, L., SCHILLING, M. F. and WATERMAN, M. S. (1986). An extreme value theory for long head runs. *Probab. Theory Related Fields* **72** 279–288.

GUIBAS, L. J. and ODLYZKO, A. M. (1980). Long repetitive patterns in random sequences. *Z. Wahrsch. verw. Gebiete* **53** 241–262.

JANSON, S. (1986). Birthday problems, randomly colored graphs, and Poisson limits of dissociated variables. Technical Report 1986: 16, Dept. Mathematics, Uppsala Univ.

STEIN, C. M. (1971). 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.* **3** 583–602. Univ. California Press.

STEIN, C. M. (1986a). *Approximate Computation of Expectations.* IMS, Hayward, Calif.

STEIN, C. M. (1986b). The number of monochromatic edges in a graph with randomly colored vertices. Unpublished manuscript.

WATSON, G. S. (1954). Extreme values in samples from m-dependent stationary stochastic sequences. *Ann. Math. Statist.* **25** 798–800.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF SOUTHERN CALIFORNIA
DRB 306, UNIVERSITY PARK
LOS ANGELES, CALIFORNIA 90089-1113