Statistical Science
1990, Vol. 5, No. 4, 403–434

Poisson Approximation and the Chen–Stein Method

Richard Arratia, Larry Goldstein and Louis Gordon

Abstract. The Chen–Stein method of Poisson approximation is a powerful tool for computing an error bound when approximating probabilities using the Poisson distribution. In many cases, this bound may be given in terms of first and second moments alone. We present a background of the method and state some fundamental Poisson approximation theorems. The body of this paper is an illustration, through varied examples, of the wide applicability and utility of the Chen–Stein method. These examples include birthday coincidences, head runs in coin tosses, random graphs, maxima of normal variates and random permutations and mappings. We conclude with an application to molecular biology. The variety of examples presented here does not exhaust the range of possible applications of the Chen–Stein method.

Key words and phrases: Poisson approximation, invariance principle, Stein's method.

1. INTRODUCTION

The central limit theorem has enjoyed a long and much deserved celebrated history. Overshadowed but perhaps of no less importance are theorems involving rare events and Poisson distributions. In generalizing the central limit theorem, one examines the consequences of relaxing the assumption that the summands are independent and identically distributed. In the same spirit, one may follow this path in the simplest possible Poisson limit theorem.

THEOREM 0. *Let* $X_{1,n}, \cdots, X_{n,n}$ *be independent indicator random variables with*

$$P(X_{i,n} = 1) = p_{i,n}.$$

Let $\lambda_n = \sum_{i=1}^n p_{i,n}$, *and* $W_n = \sum_{i=1}^n X_{i,n}$. *If* $n \to \infty$, $\max_{1 \le i \le n} p_{i,n} \to 0$, *and* $\lambda_n \to \lambda > 0$, *then* W_n *converges in distribution to* Z, *a Poisson random variable with mean* λ.

In what follows, we will write $Z \sim \mathcal{P}(\lambda)$ to mean that Z has a Poisson distribution with mean λ, that is, $P(Z = k) = e^{-\lambda}\lambda^k/k!$ for $k = 0, 1, \ldots$.

Focusing on occurrences of events, that is, on indicator random variables, the generalization of the above limit theorem to the case of other distributions is ruled out. However, generalization in the direction toward dependence is quite fruitful, as many important and interesting questions may be phrased in terms of sums of possibly dependent indicator random variables. In fact, our goal in this paper is to illustrate the broad range of problems that may be successfully attacked by a powerful Poisson approximation method due to Stein (1972) and Chen (1975). In Section 2, we present a review of this technique.

In Section 3, we present three Poisson approximation theorems based on the Chen–Stein method. In Section 4, these theorems are applied to a wide collection of examples that reduce to questions about sums of possibly dependent indicator random variables. The intuition that such a sum has a Poisson limit, and that the finite sum may therefore be approximated by a Poisson random variable, is essentially the same here as it is for the simplest theorem above. There are a large number of events, each of which has small probability of occurring. If the dependence between events is somehow confined, then the sum W should behave as in the case of no dependence. In addition, not only is W close to Poisson, but the entire process of indicators is close to a Poisson process. In practice, however, using a Poisson approximation to compute probabilities involving the indicators is not sufficient. One also needs to know what error is made in using the approximation. That the Chen–Stein method supplies an upper bound on this error is its main utility.

When the dependence structure is local, finding the Chen–Stein bounds involves the same effort as computing first and second moments of the total

Richard Arratia, Larry Goldstein and Louis Gordon are members of the Mathematics Department at the University of Southern California.

number of occurrences. In some of the examples below, we show that the rate achieved by the Chen–Stein method is sharp for the distance between the dependent indicator process and the approximating Poisson process.

The six subsections of Section 4 may be read independently of each other. Each is an example of using the Chen–Stein method to establish a Poisson approximation. In Section 4.1, we determine bounds on probabilities for the general birthday coincidence problem. In Section 4.2, we study the distribution of the length of the longest run of heads in a sequence of independent coin tosses. In Section 4.3, we consider the distribution of the number of cycles in a random graph. Next, in Section 4.4, we discuss the problem of approximating the distribution of the maxima of sequences of normal variates. Section 4.5 brings the Chen–Stein method to bear on the problem of permutations with restricted positions; the last example, Section 4.6, considers cycles in random permutations and mappings.

Our interest in Poisson approximation arose from problems in molecular biology and the statistical analysis of DNA. An example of the Poisson approximation method applied to this area is the subject of Section 5.

2. THE CHEN–STEIN METHOD

In 1972, Charles Stein published "*A bound on the error in the normal approximation to a distribution of a sum of dependent random variables.*" The goal of this work was to show convergence in distribution to the normal and produce an associated Berry–Esseen type theorem for sums of dependent random variables. The technique used was novel.

Stein's technique was free of Fourier methods and relied instead on the elementary differential equation

$$f'(x) - xf(x) = h(x) - Nh. \tag{1}$$

In equation (1) above, h is a function that is used to test convergence in distribution and $Nh = E[h(Z)]$, where Z is standard normal. The connection between this equation and the normal distribution is the following characterization. For W an arbitrary random variable and

$$(Lf)(x) = f'(x) - xf(x),$$

$E(Lf)(W) = 0$ for all differentiable functions f such that $E|Zf(Z)| < \infty$, if and only if W itself has a standard normal distribution. It now seems plausible that if $E(Lf)(W)$ is small for many functions f, then the distribution of W is close to that of Z. If W happens to be a normalized sum of an appropriate collection of random variables, then an argument involving a Taylor expansion about the sum W with a given term left out shows that the above expectation is indeed small if f is sufficiently smooth. The argument may be completed by demonstrating that smoothness properties assumed on h translate into the required smoothness properties on f through the differential equation (1). Stein's method has been applied with much success in the area of normal approximation (See, for example, Erickson, 1974; Chen, 1978; Chen and Ho, 1978; Bolthausen, 1984; Barbour and Hall, 1984b; Barbour and Eagleson, 1985; Stein, 1986; Baldi and Rinott, 1989; Baldi, Rinott and Stein, 1989; and Barbour, 1990).

There are other techniques that prove the central limit theorem without involving Fourier methods (for example, Breiman's 1968 treatment of the proof of Lindeberg, or Rosenblatt's 1974 treatment of a proof of Petrovsky and Kolmogorov). Stein's technique, however, is unique in that one may determine the bound on the error made in the approximation, a property of paramount importance in the examples to follow in Section 4.

Equation (1) above appears in other connections involving the normal distribution. Defining $h_0(x) = 1$, and $h_{n+1} = Lh_n$ for $n = 0, 1, \cdots$, one generates the Hermite polynomials, that complete orthogonal system of polynomials on $\mathbf{R}$ with measure $e^{-x^2/2}dx$. One may use a multidimensional version of equation (1) to recover and generalize Stein's (1956) remarkable result on the inadmissability of the normal mean in three or more dimensions (Hudson, 1978), or to study other questions arising in the estimation of the mean of a multivariate normal (Stein, 1981). Lastly we mention that Lf' is the generator of the Ornstein–Uhlenbeck process, which has a normal stationary distribution.

In 1975, Chen applied Stein's ideas in the Poisson setting. Corresponding to the differential equation in the normal case above, one has an analogous difference equation in the Poisson case. With Z now $\mathcal{P}(\lambda)$, if we define

$$(Lf)(x) = \lambda f(x+1) - xf(x), \tag{2}$$

then $E(Lf)(W) = 0$ for all f such that $E|Zf(Z)| < \infty$, if and only if $W \sim \mathcal{P}(\lambda)$. For W a sum of many Bernoulli random variables, each with small expectation, an argument involving leaving a given term out of the sum demonstrates that $E[(Lf)(W)]$ is small and so W is approximately Poisson. Again, one requires that properties of the "test function" h translate into the desired properties of f through the difference equation

$$\lambda f(x+1) - xf(x) = h(x) - \mathcal{P}_\lambda h; \tag{3}$$

here, $\mathcal{P}_\lambda h = E[h(Z)]$, where $Z \sim \mathcal{P}(\lambda)$. It is in this way that bounds may be obtained on the

distance between the distribution of such a sum and the Poisson. Chen's work has resulted in advances in the theory of Poisson approximation and has helped to develop and improve upon a body of interesting applications and examples. (For theoretical developments, see Barbour and Eagleson, 1983, 1984; Barbour and Hall, 1984a; Barbour, 1987; Arratia, Goldstein and Gordon, 1989; Barbour, Holst and Janson, 1988b. For applications and examples, see Barbour, 1982; Bollobás, 1985; Holst, 1986; Janson, 1986; Stein, 1986; Barbour, Holst and Janson, 1988; Heckman, 1988; Barbour and Holst, 1989; and Holst and Janson, 1990.)

3. POISSON APPROXIMATION THEOREMS

In this section, we will state three Poisson approximation theorems, each giving bounds in terms of the total variation distance between two distributions.

Here is the definition of total variation distance. Write $\mathcal{L}(Y)$ for the law or distribution of Y. For a real valued function h defined on the support of Y_0 and Y_1, let

$$\|h\| = \sup_k |h(k)|.$$

Define the total variation distance between Y_0 and Y_1, a real number between 0 and 2, by

$$\|\mathcal{L}(Y_0) - \mathcal{L}(Y_1)\| = \sup_{\|h\|=1} |E[h(Y_0)] - E[h(Y_1)]|.$$

Equivalently, one may write

$$\begin{aligned}\|\mathcal{L}(Y_0) - \mathcal{L}(Y_1)\| &= 2 \sup_A |P(Y_0 \in A) - P(Y_1 \in A)|. \\ &= 2 \min P(Y_0 \neq Y_1).\end{aligned}$$

In the last equality, the minimum is taken over all realizations of Y_0 and Y_1 on the same probability space.

The total variation distance has the following statistical interpretation. Consider the following two hypotheses on the distribution of the random variable Y:

$$H_0: \mathcal{L}(Y) = \mathcal{L}(Y_0)$$

versus

$$H_1: \mathcal{L}(Y) = \mathcal{L}(Y_1).$$

If we adopt the test with critical region C rejecting the null hypothesis when $Y \in C$ and accepting otherwise, then for any C that satisfies the natural condition $P(Y_1 \in C) \geq P(Y_0 \in C)$, the sum of the type I and type II error probabilities $\alpha_C + \beta_C$ is

$$\begin{aligned}&P(Y_0 \in C) + P(Y_1 \notin C) \\ &\quad = 1 - |P(Y_1 \in C) - P(Y_0 \in C)|.\end{aligned}$$

Hence,

$$\inf_C (\alpha_C + \beta_C) = 1 - \tfrac{1}{2}\|\mathcal{L}(Y_1) - \mathcal{L}(Y_0)\|.$$

All examples and theorems that follow will be set in the following framework. There is a finite or countable index set I. For each $\alpha \in I$, let X_α be a Bernoulli random variable with $p_\alpha = P(X_\alpha = 1) > 0$. Let

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

We assume $\lambda \in (0, \infty)$. Z will denote a Poisson random variable with the same mean as W. For each $\alpha \in I$, suppose we have chosen $B_\alpha \subset I$ with $\alpha \in B_\alpha$. We think of the set B_α as a neighborhood of α consisting of the set of indices β such that X_α and X_β are dependent.

Define

$$b_1 = \sum_{\alpha \in I} \sum_{\beta \in B_\alpha} p_\alpha p_\beta, \tag{4}$$

$$b_2 = \sum_{\alpha \in I} \sum_{\alpha \neq \beta \in B_\alpha} p_{\alpha\beta}, \quad \text{where } p_{\alpha\beta} = E[X_\alpha X_\beta], \tag{5}$$

and

$$b_3 = \sum_{\alpha \in I} E\,|\,E\{X_\alpha - p_\alpha \,|\, \sigma(X_\beta : \beta \notin B_\alpha)\}\,|. \tag{6}$$

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

Computing b_1 and b_2 usually involves the same work as computing the first and second moments of W. In applications where X_α is independent of the collection $\{X_\beta : \beta \notin B_\alpha\}$, the term $b_3 = 0$. When $b_3 = 0$, $b_2 - b_1 = E(W^2) - E(Z^2)$. Thus when $b_3 = 0$ and b_1 is small, the upper bounds on total variation distance given in the theorems below are comparable to the discrepancy between the second moment of W and that of the Poisson.

Together with error bounds, our results are that when b_1, b_2, and b_3 are all small, then

1. Theorem 1. The total number W of events is approximately Poisson.
2. Theorem 2. The locations of the dependent events approximately form a Poisson process.
3. Theorem 3. The dependent events are almost indistinguishable from a collection of independent events having the same marginal probabilities.

The following theorems are proved in Arratia, Goldstein and Gordon (1989).

THEOREM 1. *Let $W = \sum_{\alpha \in I} X_\alpha$ be the number of occurrences of dependent events, and let Z be a Poisson random variable with $EZ = EW = \lambda < \infty$. Then*

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

and

$$\begin{aligned} &| P(W = 0) - e^{-\lambda} | \\ &\quad \le (b_1 + b_2 + b_3)(1 - e^{-\lambda})/\lambda \\ &\quad < (1 \wedge \lambda^{-1})(b_1 + b_2 + b_3). \end{aligned}$$

The next theorem is a process version of the above theorem.

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*

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

Theorem 3 compares the dependent Bernoulli process $\mathbf{X}$ with an independent Bernoulli process $\mathbf{X}'$. Since $\sum_\alpha p_\alpha^2 \le b_1$, Theorem 3 implies that if the Chen–Stein method succeeds with b_1, b_2 and b_3 all small, then in the sense of total variation distance the dependent $\mathbf{X}$ process is close to being independent.

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*

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

Direct elementary computation shows that if X is Bernoulli and Y is Poisson, with $EX = EY = p \in [0, 1]$, then the total variation distance

$$\begin{aligned} &\| \mathcal{L}(X) - \mathcal{L}(Y) \| \\ &\quad = | 1 - p - e^{-p} | + | p - pe^{-p} | + | 0 - P(Y > 1) | \\ &\quad \le 2p^2. \end{aligned}$$

Thus X and Y can be coupled, i.e., constructed on a single probability space, so that $P(X \ne Y) = \frac{1}{2} \| \mathcal{L}(X) - \mathcal{L}(Y) \| \le p^2$. For the Poisson process $\mathbf{Y}$ of Theorem 2, and the independent events process $\mathbf{X}'$ of Theorem 3 above, coupling each coordinate shows that

$$\| \mathcal{L}(\mathbf{Y}) - \mathcal{L}(\mathbf{X}') \| \le 2P(\mathbf{Y} \ne \mathbf{X}') \le 2 \sum_\alpha p_\alpha^2.$$

Thus, Theorem 3 above is an elementary corollary of Theorem 2, using the triangle inequality

$$\begin{aligned} &\| \mathcal{L}(\mathbf{X}) - \mathcal{L}(\mathbf{X}') \| \\ &\quad \le \| \mathcal{L}(\mathbf{X}) - \mathcal{L}(\mathbf{Y}) \| + \| \mathcal{L}(\mathbf{Y}) - \mathcal{L}(\mathbf{X}') \|. \end{aligned}$$

Since $\sum p_\alpha^2$ is small in typical applications, Theorem 2 is "almost" equivalent to Theorem 3. More precisely, the weakening of Theorem 2, in which the bound is increased by $4 \sum p_\alpha^2$, is an elementary corollary of Theorem 3, using the triangle inequality.

3.1 Compound Poisson Process Limits

The Chen–Stein method is useful for situations in which occurrences happen in clumps and the distribution of number of clumps is approximately Poisson. In many situations, the distribution of the number of occurrences is approximately a compound Poisson distribution and the dependent process itself is close to a mosaic process in which locations are put down according to a spatial Poisson process and then at each location a type is assigned in some independent and identically distributed way. (See Aldous, 1989, or Hall, 1988.)

This situation can be handled by the Chen–Stein method. All that needs to be done is to enlarge the index set so that it keeps track of the *types* as well as the locations of the clumps. In these situations, Theorems 2 and 3 are a tool for showing that a dependent process is close to a mosaic process.

Here is a general overview; we will show how these considerations apply to the example of long head runs in Section 4.2. Start with a successful setup for the Chen–Stein method: an index set I, events X_α for $\alpha \in I$ and neighborhoods $B(\alpha)$ for $\alpha \in I$ such that b_1, b_2 and b_3 can be shown to be small. Suppose that each event $\{X_\alpha = 1\}$ can also be associated with a "type" chosen from some countable set T. Our new, enlarged index set will be $I^* \equiv I \times T$, and for $(\alpha, i) \in I^*$,

$$X_{\alpha,i} \equiv X_\alpha \mathbf{1}(\text{the occurrence at } \alpha \text{ is of type } i),$$

so that for each $\alpha \in I$, there is a partition:

$$X_\alpha = \sum_{i \in T} X_{\alpha,i}. \tag{7}$$

The new neighborhoods $B(\alpha, i)$ will be based on the old neighborhoods:

$$B(\alpha, i) \equiv B(\alpha) \times T = \{(\beta, j) \in I^* : \beta \in B(\alpha), j \in T\}. \tag{8}$$

The new value b_1^* is equal to the old value b_1:

$$\begin{aligned} b_1^* &= \sum_{\alpha \in I} \sum_{i,j \in T} (EX_{\alpha,i})(EX_{\alpha,j}) \\ &\quad + \sum_{\alpha \neq \beta \in B(\alpha)} \sum_{i,j \in T} (EX_{\alpha,i})(EX_{\beta,j}) \\ &= \sum_{\alpha \in I} \left(\sum_{i \in T} EX_{\alpha,i} \right)^2 \\ &\quad + \sum_{\alpha \neq \beta \in B(\alpha)} \left(\sum_{i \in T} EX_{\alpha,i} \right) \left(\sum_{j \in T} EX_{\beta,j} \right) \\ &= \sum_{\alpha \in I} (EX_\alpha)^2 + \sum_{\alpha \neq \beta \in B(\alpha)} (EX_\alpha)(EX_\beta) = b_1. \end{aligned}$$

Similarly, thanks to the partition structure (7) and the neighborhood structure (8), the value of b_2 is unchanged:

$$b_2^* = b_2.$$

In general $b_3^* \geq b_3$, but in many examples the neighborhoods B_α capture all of the dependence and it is easily verified that $b_3^* = b_3 = 0$.

Because of the partition structure (7), and because the Poisson process $\mathbf{Y}$ may be similarly constructed from the Poisson process $\mathbf{Y}^*$ by setting $Y_\alpha = \sum_i Y_{\alpha,i}$ for each α, the total variation distance for the processes involved in Theorem 2 cannot decrease:

$$\| \mathscr{L}(\mathbf{X}) - \mathscr{L}(\mathbf{Y}) \| \leq \| \mathscr{L}(\mathbf{X}^*) - \mathscr{L}(\mathbf{Y}^*) \|.$$

Here, $\mathbf{X}^* = (X_{\alpha,i})_{\alpha \in I, i \in T}$ is the dependent events process, with values in $\{0, 1\}^{I \times T}$, and $\mathbf{Y}^* = (Y_{\alpha,i})_{\alpha \in I, i \in T}$ is the Poisson process, with values in $\{0, 1, 2, \cdots\}^{I \times T}$, having independent components and the same intensity as $\mathbf{X}^*$.

4. APPLICATIONS

We demonstrate the utility of Poisson approximation by applying the above results to six examples, all of which reduce to questions about the number of occurrences of possibly dependent events.

4.1 The Birthday Problem

We first learned about Chen (1975a) from a lecture on the birthday problem and its variants by Persi Diaconis, who also suggested references on the birthday problem: Diaconis and Mosteller (1989), Janson (1986), Holst (1986) and Stein (1987), which gives proofs of more general results using similar techniques.

In the usual formulation of the birthday problem, we assume that birthdays of n individuals are independent over the d days in a year and compute the probability that at least two share the same birthday, that is, that there is at least one two-way coincidence. In the special case where birthdays are uniform, there is a simple exact formula. Letting W denote the number of birthday coincidences, that is, the number of pairs of people that share a birthday, we have the probability of no coincidence given by

$$P(W = 0) = \prod_{i=1}^{n-1} \left(1 - \frac{i}{d}\right).$$

If one were now interested in computing the probability of, say, exactly m two-way coincidences, or the probability of at least three people sharing the same birthday, or the probability that there are two people born within a week of each other, or probabilities under a nonuniform birthday distribution, then the counting arguments to arrive at exact formulas become much less tractable. However, extremely good approximate answers are quite easy to obtain using the Poisson approximation and one may use Theorem 1 to give an upper bound on the error.

Let us begin by considering the general birthday problem of a k-way coincidence when birthdays are uniform. Let $\{1, 2, \cdots, n\}$ denote a group of n people, and let the index set $I \equiv \{\alpha \subset \{1, 2, \cdots, n\} : |\alpha| = k\}$. For example, in the classical case $k = 2$ and I is the set of all pairs of people among whom a two-way coincidence could occur. Let X_α be the indicator of the event that the people indexed by α share the same birthday. The total number of coincidences is now given as the sum of dependent indicator random variables, $W = \sum_{\alpha \in I} X_\alpha$.

Because W is the sum of many Bernoulli random variables, each with small success probability $p_\alpha = d^{1-k}$, it seems reasonable to approximate W as a Poisson random variable Z with mean $\lambda = EW$. Easily then $\lambda = \binom{n}{k} d^{1-k}$ and the probability of no birthday coincidence is approximately

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

For the classical case of a birthday coincidence in a year of $d = 365$ days, it is widely known that $n = 23$ is the least number of people required to make such a coincidence more likely than not; amusingly, $\lambda = \binom{23}{2}/365$ is equal to $\ln(2)$ to 4 digits.

The probability of coincidence is approximated conservatively by the Poisson distribution in this case;

$$P(W = 0) \doteq 0.492 < 0.499998 \doteq \exp(-\lambda) = P(Z = 0).$$

The approximation is always conservative when birthdays are uniform;

$$P(W = 0) = \prod_{i=1}^{n-1}\left(1 - \frac{i}{d}\right) < \exp\left\{-\frac{1}{d}\sum_{i=1}^{n-1} i\right\} = e^{-\lambda} = P(Z = 0).$$

In addition, the probability of coincidence is minimized when birthdays are uniform (see, for example, Olkin and Marshall, 1979), making the Poisson approximation, assuming uniformity, conservative no matter what the true underlying distribution may be.

Poisson approximation using λ computed from the true distribution is not necessarily conservative when birthdays are nonuniform. One class of examples may be constructed by considering a distribution where one day has probability ε and all other days divide the remaining probability uniformly, each with mass $(1 - \varepsilon)/(d - 1)$. In particular, for $d = 7$ days, $n = 5$ individuals and $\varepsilon = \frac{2}{3}$, we have

$$\lambda = \binom{5}{2}\left(\left(\frac{2}{3}\right)^2 + 6\left(\frac{1}{18}\right)^2\right) = 4.63$$

and

$$P(W = 0) = 0.0118 > 0.0098 = \exp\{-EW\} = P(Z = 0).$$

We may bound the error in making the Poisson approximation with the help of Theorem 1 in Section 3. Recall that B_α is a "neighborhood of dependence" for the random variable X_α. Note that, if $\alpha \cap \beta = \varnothing$, then X_α and X_β are independent. This suggests that we should take the set

$$B_\alpha = \{\beta \in I : \alpha \cap \beta \neq \varnothing\}$$

as our set of dependence. With this choice

$$E\,|\,E\{X_\alpha - p_\alpha \,|\, \sigma(X_\beta : \beta \notin B_\alpha)\}\,| = 0$$

by independence; hence $b_3 = 0$.

Since all p_α are identical, we calculate

$$b_1 = |I|\,|B_\alpha|\,p_\alpha^2 = \binom{n}{k}\left\{\binom{n}{k} - \binom{n-k}{k}\right\}d^{2-2k}.$$

Specializing now to the case $k = 2$, we may use that X_α and X_β are pairwise independent, and that therefore $p_{\alpha\beta} = p_\alpha p_\beta$. Hence,

$$b_2 = |I|(|B_\alpha| - 1)p_{\alpha\beta} = \frac{b_1(|B_\alpha| - 1)}{|B_\alpha|}.$$

Putting the above together, one finds the following bound for the error in approximating $P(W = 0)$ by $e^{-\lambda}$ in the case $k = 2$:

$$|P(W = 0) - e^{-\lambda}| \le (b_1 + b_2)\frac{1 - e^{-\lambda}}{\lambda} = \frac{1}{d^2}\binom{n}{2}(4n - 7)\frac{1 - e^{-\lambda}}{\lambda}.$$

Although it is more difficult to exactly calculate the probability of a triple birthday coincidence, one may apply Poisson approximation with about the same ease as for the classical case. Suppose that we wish to compute the probability that in a group of 50, three or more share a birthday. We have then that $\lambda = \binom{n}{3}/d^2$ and the approximation $P(W = 0) \doteq e^{-\lambda}$; hence, in a group of 50, the probability that there is at least one triple coincidence is about $1 - e^{-\lambda} = 1 - 0.863 = 0.137$.

To determine a bound on the error, one may calculate

$$b_1 = |I|\,|B_\alpha|\,p_\alpha^2 = \binom{n}{3}\left\{\binom{n}{3} - \binom{n-3}{3}\right\}d^{-4},$$

and, for a given α, breaking up $B_\alpha - \{\alpha\}$ into those β such that $|\beta \cap \alpha| = 1$ and those for which $|\beta \cap \alpha| = 2$, we see

$$b_2 = |I|\left\{3\binom{n-3}{2}d^{-4} + 3(n-3)d^{-3}\right\}.$$

This shows the approximation above has an error of no more than

$$(b_1 + b_2)(1 - e^{-\lambda})/\lambda = 0.0597,$$

so that

$$0.803 \le P(W = 0) \le 0.923$$

Without too much difficulty, one can write down the exact formula for the probability of no triple coincidence. In order for there to be no triple coincidence, the d days of the year must be partitioned into h days where there are no birthdays, i days on which a single individual was born, and j days where exactly two individuals share a birthday. A factor of $n!/2^j$ is needed to count the number of arrangements of n persons into such a configuration of $i + j$ days. Hence,

$$P(W = 0) = d^{-n}\sum_{i+2j=n}\binom{d}{h, i, j}\frac{n!}{2^j}.$$

For $n = 50$ and $d = 365$ we have that $P(W = 0) = 0.8736$, for an actual error of $0.8736 - 0.8632 = 0.0104 < 0.0597$, the Chen–Stein bound on the error.

For general k, in the case where birthdays are *uniform*, it is possible to consider a slight improvement

on the choice on B_α. For $\alpha, \beta \in I$, knowing only the birthday of one member of α, say $\min(\alpha)$, does not change the probability that $X_\alpha X_\beta = 1$. Hence one could take

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

which is strictly smaller than the old choice of B_α; one still has $b_3 = 0$. Working through the calculations, one finds only a slight improvement in the error bound. For example, for the case of the triple coincidence with $n = 50$ and $d = 365$, the bound improves from 0.0597 to 0.0582. The change is slight because the main contribution to the upper bound comes from the part of b_2 where $\alpha \cap \beta = 2$; this contribution remains unchanged using the smaller B_α.

In the general case of computing k-way coincidences when birthdays are uniform, the Chen–Stein method gives the best possible rate of convergence of the total variation to zero. Take $n, d \to \infty$ in such a way that $\lambda/1$ stays bounded away from zero and ∞, which we denote by $\lambda \asymp 1$. This condition implies that $n^k \asymp d^{k-1}$ and hence that $b_1 = \lambda^2 |B_\alpha|/|I| \asymp n^{-1}$. The order of the Chen–Stein bound here is the same as the order of b_2,

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

The dominant contribution to b_2 comes from pairs (α, β) with $\alpha \cap \beta = j = k - 1$, and b_2 is of the order $n^{1+k}d^{-k} \asymp n/d \asymp n^{-1/(k-1)}$. Thus the Chen–Stein method yields that the total variation distance decays at a rate no slower than $O(n^{-1/(k-1)})$.

A lower bound on the total variation distance in the case of a k-way coincidence can be given by considering the event E that $k + 1$ individuals share a birthday, that is, that there exist α, β of size k with $|\alpha \cap \beta| = k - 1$ such that $X_\alpha X_\beta = 1$. The actual probability $P(E)$ can be bounded from below by the first two terms of the inclusion–exclusion formula; the first term is dominant and of the order $\binom{n}{k+1}d^{-k} \asymp n/d \asymp n^{-1/(k-1)}$. Letting E' be the same event for the independent process, we have

$$\begin{aligned} P(E') &\leq \textstyle\sum_{|\alpha|=k, |\beta|=k, |\alpha\cap\beta|=k-1} EX'_\alpha X'_\beta \\ &= O(\tbinom{n}{k+1} d^{2-2k}) = o(n^{-1/(k-1)}). \end{aligned}$$

Thus, the order of the total variation distance is at least as large as $|P(E) - P(E')| \asymp P(E) \asymp n^{-1/(k-1)}$. Hence, the Chen–Stein method yields the correct order of the rate of decay of the total variation distance to zero.

4.2 The Length of the Longest Head Run

Consider many independent throws of a coin of success probability p, $0 < p < 1$. No matter what p, there will be some stretches where the coin comes up heads every time. To begin the analysis of the distribution of R_n, the length of the longest of these head runs, first note that for a test length t, appropriately chosen, one sees a head run of length t begin at a given position α only with small probability. As the number of positions where such a run could occur is large, a Poisson approximation should be valid.

However, one must first adjust for the fact that runs of heads occur in "clumps"; that is, if there is a run of heads of length t beginning at position α, then with probability p there will also be a run of heads of length t beginning at position $\alpha + 1$, with probability p^2 a run of heads of length t beginning at position $\alpha + 2$ and so forth. By counting only the first such run, the runs now counted are no longer clumped and, indeed, their number is Poisson in the limit. This is an example, with average clump size $1 + p + p^2, \cdots$, of the "Poisson clumping heuristic" as described by Aldous (1989). By using the fact that having no runs of length t is equivalent to having the longest head run shorter than t one may approximate the distribution function of the length of the longest run of heads.

Let then $C_1, C_2, \cdots$ be independent Bernoulli random variables with success probability p, and let R_n be the length of the longest run of heads *beginning* in the first n tosses. Set the index set to be $I = \{1, 2, \cdots, n\}$; the elements of the index set will denote locations where long head runs may begin. A head run of length t or more begins at position α if and only if the indicator random variable

$$Y_\alpha = \prod_{i=\alpha}^{\alpha+t-1} C_i$$

takes the value one. To declump, that is, in order to count only the first head run in a clump, we take $X_1 = Y_1$ and

$$X_\alpha = (1 - C_{\alpha-1})Y_\alpha, \quad \alpha = 2, 3, \cdots, n.$$

For $\alpha = 2, 3, \cdots, n$, X_α will be one if and only if a run of t or more heads begins at position α, preceded by a tail. If we had ignored clumping and simply taken $Y_\alpha = X_\alpha$, we would have b_2 not tending to zero, and, in fact, a Poisson approximation would not be valid.

Write now the total number of clumps of runs of length t or more as the sum of dependent indicator random variables

$$W = \sum_{\alpha \in I} X_\alpha.$$

The Poisson approximation heuristic says we should be able to approximate the distribution of W by a Poisson random variable with mean

$$\lambda = \lambda_n(t) = EW = p^t\{(n - 1)(1 - p) + 1\}.$$

In particular then, since we have as events

$$\{R_n < t\} = \{W = 0\},$$

the distribution function of R_n may be approximated as

$$P(R_n < t) = P(W = 0) \cong e^{-\lambda_n(t)}.$$

The test length is dictated by requiring λ to be bounded away from 0 and ∞; this is equivalent to the condition that $t - \log_{1/p}(n(1-p))$ is bounded. In fact, for integer t, with c defined by

$$t = \log_{1/p}((n-1)(1-p) + 1) + c,$$

the above approximation predicts that

$$P(R_n < t) \cong e^{-\lambda_n(t)} = \exp(-p^c),$$

that is, that $R_n - \log_{1/p}((n-1)(1-p)+1)$ has an asymptotic extreme value distribution. This is almost so; the limiting distribution is complicated by the fact that R_n can assume only integer values. However, this fact does not complicate the approximation itself.

For example, with $n = 2047$ and a fair coin with $p = \frac{1}{2}$, we look for runs of length $\log_{1/p}((n-1) \cdot (1-p) + 1) = \log_2(2046 \cdot \frac{1}{2} + 1) = 10$. Would a run of length, say $t = 14$ be unusual? By using the Poisson approximation, we see that $P(R_{2047} \geq 14) = 1 - P(R_{2047} < 10 + 4)$ may be approximated by $1 - \exp(-(\frac{1}{2})^4) = 0.06059$.

To assess the accuracy of the above Poisson approximation, we apply Theorem 1. Define $B_\alpha = \{\beta \in I: |\alpha - \beta| \leq t\}$ for all α. Since X_α is independent of $\sigma\{X_\beta: \beta \notin B_\alpha\}$, we have $b_3 = 0$. Furthermore, if $1 \leq |\alpha - \beta| \leq t$, we cannot have that both X_α and X_β are 1, since we have insisted that a run begin with a tail; therefore $p_{\alpha\beta} = 0$ for $\beta \in B_\alpha$, $\beta \neq \alpha$, hence $b_2 = 0$.

In order to calculate $b_1 = \sum_\alpha \sum_{\beta \in B_\alpha} p_\alpha p_\beta$, we break up the sum over $\beta \in B_\alpha$ into two parts, depending on whether or not p_1 appears. This yields the bound

$$b_1 < \lambda^2(2t+1)/n + 2\lambda p^t. \tag{9}$$

Theorem 1 now reveals that the Poisson approximation is quite accurate for the example considered above; the probability computed is correct to within $b_1 < 6.297 \times 10^{-5}$, so that

$$0.060527 \leq P(R_{2047} \geq 14) \leq 0.0606453.$$

4.2.1 Compound Poisson process limits and long head runs

What follows is a concrete illustration of the discussion in Section 3.1. Specifically, we show how the problem of long head runs may be treated to obtain a compound Poisson limit for the random variable

$$U \equiv \sum_{\alpha \in I} C_\alpha C_{\alpha+1} \cdots C_{\alpha+t-1},$$

which counts the number of locations among the first n at which a head run of length at least t begins. As we have noted, these locations tend to occur in clumps; W counts the number of clumps and is approximately Poisson in distribution. The size of each clump, minus one, is the length by which the associated head run exceeds t and is distributed as a geometric random variable with parameter p. The clump sizes are mutually independent of each other and approximately independent of the total number W of clumps, so the distribution of U is approximately Poisson compounded by geometric. Furthermore, the clump sizes are approximately independent of the locations of the clumps, so that we have approximately a mosaic process. The Chen–Stein method gives us total variation bounds to make all of this precise.

As described in Section 3.1, we will enlarge the index set from I to $I^* = I \times T$ in order to keep track of the *types* of clumps; the values of b_1^*, b_2^* and b_3^* are given by (4), (5) and (6) using I^* and the new neighborhoods B_α^*, defined below. Here we take $T = \{0, 1, \cdots, t\}$ as the set of possible types of clumps. Any run of exactly $t + i$ heads for $0 \leq i < t$ corresponds to a clump of size i; a run of $2t$ or more heads corresponds to a clump of type $i = t$. The interpretation is that each of the X_α runs of heads of length at least t starting at α can independently be assigned a type i, corresponding to a run of exactly $t + \min(i, t)$ heads. (For the purpose of proving convergence of U to a compound Poisson limit, the upper bound t could be replaced by anything tending to infinity as n grows.) For $\alpha \in I$, $i \in T$ let

$$X_{\alpha,i} \equiv (1 - \mathbf{1}\{\alpha > 1\}C_{\alpha-1})C_\alpha C_{\alpha+1} \cdots C_{\alpha+t+i-1}(1 - \mathbf{1}\{i < t\}C_{\alpha+t+i}),$$

so that for all $\alpha \in I$, $X_\alpha = \sum_{i \in T} X_{\alpha,i}$. Using the notation introduced in Section 3.1, in order to have $b_3^* = 0$, we expand the neighborhoods by a factor of two: Let $B_\alpha^* \equiv \{\beta \in I: |\alpha - \beta| \leq 2t\}$, which yields $b_1^* < 2b_1$, where an upper bound on b_1 is given in (9). We have $b_2^* < b_1^*$. Applied to the setup with index set $I \times T$, Theorem 2 yields the result

$$\tfrac{1}{2}\| \mathscr{L}(\mathbf{X}^*) - \mathscr{L}(\mathbf{Y}^*) \| \leq 2b_1^* + 2b_2^* + b_3^* < 8b_1. \tag{10}$$

In the example with $n = 2047$, $t = 14$, $p = \frac{1}{2}$ that we treated above, this upper bound is $8 \times 6.297 \times 10^{-5}$. The Poisson process $\mathbf{Y}^* \equiv (Y_{\alpha,i})_{\alpha \in I, i \in T}$ may be viewed as a refinement of the Poisson process $\mathbf{Y} \equiv (Y_\alpha)_{\alpha \in I}$, with $Y_\alpha = \sum_{i \in T} Y_{\alpha,i}$ for each $\alpha \in I$. The distribution of the type i of each clump is exactly geometric (p), truncated at height t.

To show that U is approximately compound Poisson, consider any set $A \subset \{0, 1, 2, \cdots\}$ and let

h be the functional

$$h(\mathbf{X}^*) = \mathbf{1}_A\left\{\sum_{\alpha,i}(i+1)X_{\alpha,i}\right\},$$

$$h(\mathbf{Y}^*) = \mathbf{1}_A\left\{\sum_{\alpha,i}(i+1)Y_{\alpha,i}\right\}.$$

First, observe that U equals $h(\mathbf{X}^*)$, except possibly on the event that there is a head run of more than $2t$, so that $\frac{1}{2}\|\mathscr{L}(U) - \mathscr{L}(h(\mathbf{X}^*))\| \le np^{2t}$. Second, observe that the distribution of $h(\mathbf{Y}^*)$ is exactly the compound of Poisson $(\lambda_n(t))$ by the distribution "one plus a geometric (p), truncated at t." If we write G for the compound Poisson $(\lambda_n(t))$ by the distribution "one plus a geometric (p)," (without truncation), then the net result is $\frac{1}{2}\|\mathscr{L}(U) - \mathscr{L}(G)\| \le 8b_1 + np^{2t} + \lambda p^t$, where the last error term, λp^t, bounds the truncation error.

4.3 Cycles in Random Graphs

In this section, we apply the Chen–Stein method to a problem treated in Takács (1988), who considers the total number of cycles in a random graph with n vertices, in which any pair of vertices, independently of all other pairs, is connected by an edge with probability $p = \rho/n \in (0, 1)$. The main result of Takács is that for fixed $\rho \in (0, 1)$, as $n \to \infty$, the limit distribution of the number of cycles is Poisson with parameter

$$a(\rho) = \frac{1}{2}\ln\left(\frac{1}{1-\rho}\right) - \frac{\rho}{2} - \frac{\rho^2}{4}.$$

The Chen–Stein method gives another proof of the Poisson convergence and also yields an upper bound on the distance to the Poisson. With no additional work, our Theorems 2 and 3 give approximations to the entire *process* of indicators of cycles. Theorem 3 confirms and quantifies the notion that the cycles occur jointly with a distribution that is close to that of mutual independence. The Poisson process in Theorem 2, indexed by individual cycles, is a refinement of a Poisson process, indexed by lengths, which approximates the process which counts the number of cycles of each length. This latter approximation supplies answers to questions such as: what is the probability that there are more cycles of even length than of odd length, what is the distribution of the number of different lengths represented by cycles, what is the distribution of the sum of the lengths of all cycles, and so on. Similar questions are considered by Wilf (1983) for the case of random permutations.

For $j \ge 3$, let I_j be the set of potential cycles of length j, where the vertex set is $\{1, 2, \cdots, n\}$. We have $|I_j| = (n)_j/(2j)$, where $(n)_j \equiv \binom{n}{j} j!$ counts the number of ways to select j distinct vertices in order, and the factor $1/(2j)$ appears since, for $j > 2$, a permutation of j vertices corresponds to a choice of a cycle in I_j together with a choice of any of two orientations and j starting points. For $\alpha \in I_j$, let X_α be the indicator of the event that the cycle α is a subgraph of our random graph, so that $EX_\alpha = p^j = \rho^j/n^j$. Let λ_j be the expected number of cycles of length j:

$$\lambda_j \equiv E\sum_{\alpha\in I(j)} X_\alpha = |I_j|EX_\alpha = \frac{\rho^j}{2j}\frac{(n)_j}{n^j},$$

with λ_j increasing up to $\rho^j/2j$ as $n \to \infty$.

Let $I \equiv \bigcup_{3\le j\le n} I_j$ be the set of all possible cycles, so that $W \equiv \sum_{\alpha\in I} I_\alpha$ is the total number of cycles, with

$$EW = \sum_{3\le j\le n}\lambda_j$$

$$= \sum_{3\le j\le n}\frac{\rho^j}{2j}\frac{(n)_j}{n^j} \to \frac{1}{2}\sum_{3\le j}\frac{\rho^j}{j} = a(\rho).$$

We define the neighborhoods B_α for $\alpha \in I$ by

$$B_\alpha \equiv \{\beta \in I\colon \alpha \text{ and } \beta \text{ have at least one edge in common}\},$$

so that $b_3 = 0$. Observe that here, α and β are neighbors if and only if X_α and X_β are strictly positively correlated, since

$$(11)\quad E(X_\alpha X_\beta) = p^{-(\#\text{edges common to }\alpha\text{ and }\beta)}EX_\alpha EX_\beta.$$

The result of using the Chen–Stein method in the above setup is summarized by the following theorem, whose proof we give at the end of this section.

PROPOSITION 4. *For $\alpha \in I$, let X_α be the indicator that α occurs as a cycle in a random graph on n points, with $p = \rho/n$, let Y_α be Poisson, and let X'_α be Bernoulli, with $EY_\alpha = EX'_\alpha = EX_\alpha$, with all of the Y_α and X'_α mutually independent. Write $\mathbf{X} = (X_\alpha)_{\alpha\in I}$, $\mathbf{Y} \equiv (Y_\alpha)_{\alpha\in I}$, and $\mathbf{X}' \equiv (X'_\alpha)_{\alpha\in I}$ for the process of indicators of cycles, a Poisson process of the same intensity and an independent events process, respectively. For each positive $\delta < 1$, uniformly in $0 < \rho < \delta$, as $n \to \infty$, the total variation distances between the distributions of the processes $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{X}'$ are all $O(n^{-1})$. Furthermore, the rate n^{-1} is sharp in the sense that* $\liminf n\|\mathscr{L}(\mathbf{X}) - \mathscr{L}(\mathbf{Y})\|$, $\liminf n\|\mathscr{L}(\mathbf{X}) - \mathscr{L}(\mathbf{X}')\| > 0$.

The lower bound at the end of Proposition 4 arises by considering the probability that our random graph contains any pairs of triangles sharing a common edge. The expected number of such pairs is $(\rho/n)^5(n)_4/4 \sim n^{-1}\rho^5/4$, and the second term of the inclusion–exclusion series is $O(n^{-2})$, so this gives the asymp-

totic probability of at least one such pair. Thus $Eg(\mathbf{X}) \sim n^{-1}\rho^5/4$, where g is the functional with $g(\mathbf{X}) = \mathbf{1}(1 \le \sum_{\alpha,\beta\in I(3),\beta\in B(\alpha)} X_\alpha X_\beta)$. For the Poisson and independent events process we have

$$Eg(\mathbf{Y}),\ Eg(\mathbf{X}') \le \sum_{\alpha,\beta\in I(3),\beta\in B(\alpha)} EX_\alpha EX_\beta = O(n^{-2}).$$

Thus

$$\begin{aligned}\liminf n\|\mathscr{L}(\mathbf{X}) - \mathscr{L}(\mathbf{X}')\| &\ge \lim n(Eg(\mathbf{X}) - Eg(\mathbf{X}')) \\ &= \rho^5/4 > 0.\end{aligned}$$

Notice also that $b_2 \ge$ E(# pairs of neighboring triangles) so that b_2 decays no faster than $O(n^{-1})$.

For $j \ge 3$, let $W_j \equiv \sum_{\alpha\in I(j)} X_\alpha$, $Z_j \equiv \sum_{\alpha\in I(j)} Y_\alpha$, so that W_j is the number of cycles of length j, and Z_j is a Poisson random variable having the same mean as W_j. For $j > n$, W_j is identically zero. Thus $\mathbf{W} \equiv (W_j)_{j\ge 3}$ is the process that counts the number of cycles of each length, and $\mathbf{Z} \equiv (Z_j)_{j\ge 3}$ is a Poisson process with independent components. Since there is a functional $h(\cdot)$ such that $\mathbf{W} = h(\mathbf{X})$, $\mathbf{Z} = h(\mathbf{Y})$, we have as a corollary to Proposition 4 that

$$\begin{aligned}&\|\mathscr{L}(\mathbf{W}) - \mathscr{L}(\mathbf{Z})\| \\ &\quad \le \|\mathscr{L}(\mathbf{X}) - \mathscr{L}(\mathbf{Y})\| = O(1/n)\end{aligned} \tag{12}$$

uniformly in $\rho \le \delta < 1$. The convergence of the finite dimensional distributions of $\mathbf{W}$ to their independent Poisson process limit is given in Bollobás (1985 page 79).

We now show how the Poisson process $\mathbf{Z}$ supplies an answer to the question: what is the probability that there are more cycles of even length than of odd length. Consider the functionals (one for each value of n) defined by $f(c_3, c_4, \cdots, c_n) \equiv \mathbf{1}(c_4 + c_6 + \cdots > c_3 + c_5 + \cdots)$, so that our question is: what is the value of $Ef(\mathbf{W})$. Our answer is: approximately $Ef(\mathbf{Z})$, with error at most $\frac{1}{2}\|\mathscr{L}(\mathbf{W}) - \mathscr{L}(\mathbf{Z})\|$, since the functional f takes values in [0, 1].

One may simplify the answer further at the expense of some additional error of approximation as follows. The Poisson parameter for Z_j, namely $\lambda_j = n^{-j}(n)_j\rho^j/(2j)$ (which is zero for $j > n$), can be replaced by its limit value, $\rho^j/(2j)$, for $j = 1, 2, \cdots$. The total variation error introduced by this approximation is of the same order as the increase in expectation, namely

$$\sum_{j\ge 3}\left(\frac{\rho^j}{2j} - \frac{n^{-j}(n)_j\rho^j}{2j}\right), \tag{13}$$

which for fixed ρ is $O(1/n)$, the same as the error controlled by the Chen–Stein method. Now $Z_{\text{odd}} \equiv Z_3 + Z_5 + \cdots$ and $Z_{\text{even}} \equiv Z_4 + Z_6 + \cdots$ are independent, Poisson random variables with means

$$\begin{aligned}a_{\text{odd}}(\rho) \equiv EZ_{\text{odd}} &= \frac{1}{2}\left(\frac{\rho^3}{3} + \frac{\rho^5}{5} + \cdots\right) \\ &= \frac{1}{4}\ln\left(\frac{1+\rho}{1-\rho}\right) - \frac{\rho^2}{4}, \\ a_{\text{even}}(\rho) \equiv EZ_{\text{even}} &= \frac{1}{2}\left(\frac{\rho^4}{4} + \frac{\rho^6}{6} + \cdots\right) \\ &= \frac{1}{4}\ln\left(\frac{1}{1-\rho^2}\right) - \frac{\rho}{2}.\end{aligned} \tag{14}$$

We have just proved that the probability that there are more cycles of even length than of odd length converges to $P(Z_{\text{even}} > Z_{\text{odd}})$, where Z_{even} and Z_{odd} are independent Poisson, with parameters $a_{\text{even}}(\rho)$ and $a_{\text{odd}}(\rho)$, respectively. Furthermore, the distance between the actual probability for the graph on n vertices and its limiting value is no greater than the sum of (12) and (13).

The exact expression for b_2 is complicated; below we give an upper bound. In the second line of the bound (15), $j \ge 3$ is the number of vertices in α, $k \ge 1$ is the number of shared segments common to α and β and $l \ge 0$ is the number of vertices in β which are not on the common segments. Formally, a shared segment of α and β is an unoriented, maximal sequence of edges that occur consecutively in both α and β. Each of the k common segments corresponds to a factor of ρ/n in $E(X_\alpha X_\beta)$ that is not matched by a choice of one of n vertices, which suggests that for ρ not too large, the main contribution to b_2 comes from the case $k = 1$, and $b_2 = O(1/n)$; we prove this below for $\rho < \frac{1}{2}$. Unfortunately, for ρ sufficiently close to 1, both b_2 and the second moment of W blow up exponentially fast as $n \to \infty$. In these cases we must resort to a truncation argument to prove Proposition 4.

$$\begin{aligned}b_2 &= \sum_{\alpha\in I}\sum_{\beta\in B(\alpha)\setminus\{\alpha\}} E(X_\alpha X_\beta) \\ &\le \sum_{3\le j\le n}\frac{(n)_j}{2j}\left(\frac{\rho}{n}\right)^j \sum_{k\ge 1}\left(\frac{\rho}{n}\right)^k 2\binom{j}{2k} \\ &\quad \times \sum_{l\ge 0}\binom{n-2k}{l}\left(\frac{\rho}{n}\right)^l (k+l-1)!\,2^{k-1}.\end{aligned} \tag{15}$$

Here are further details for explaining the upper bound (15). Consider for example,

$$\alpha = (1\ \ 2\ \ 3\ \ 4\ \ 5\ \ 6\ \ 7\ \ 8)$$

and

$$\beta = (1\ \ 2\ \ 3\ \ 8\ \ 9\ \ 5\ \ 4\ \ 6\ \ 7),$$

which has $j = 8$ vertices in the cycle α. There are $k = 3$ common segments without regard to order of transversal; these are (1 2 3), (4 5), (6 7). The three common segments share the six common endpoints $\{1, 3, 4, 5, 6, 7\}$. There are $l = 2$ vertices of β not on the common segments; they are $\{8, 9\}$. The factor $2\binom{j}{2k}$ is the number of ways to choose the k common segments from α, since k segments must have $2k$ distinct vertices as endpoints, and a set of $2k$ endpoints determines two sets of k segments. In our example above, the other choice of segments using the same six endpoints is (3 4), (5 6), (7 8 1). The factor $(k + l - 1)!2^{k-1}$ is the number of ways that the k common segments from α and the l additional vertices can be arranged into a cycle β, choosing orientations for each segment after the first. Not all of these arrangements correspond to a choice of β sharing the given k segments: in our example, if the segment (4 5) were given the opposite orientation, β would be changed into $\beta' = (1\,2\,3\,8\,9\,4\,5\,6\,7)$, with α and β' classified as sharing not $k = 3$ but rather $k = 2$ segments, namely (1 2 3) and (4 5 6 7). Another reason that our bound on b_2 is an overestimate is the factor $\binom{n-2k}{l}$ for choosing l additional points for β. If i is the actual number of points used in the k common segments, with $2k \le i \le j$, then the number of ways to choose additional points for β is $\binom{n-i}{l} \le \binom{n-2k}{l}$.

To see that b_2 and therefore $EW^2 \to \infty$ for ρ sufficiently close to 1, consider $g(n, k, m, \rho)$, the contribution to b_2 from pairs of cycles α, β, each of length $2k + m$, and sharing k common segments, each consisting of a single edge. Observe that $g(n, k, m, \rho) = \rho^{3k+2m} g(n, k, m, 1)$. Let

$$f(a, b) = \lim_{n\to\infty} n^{-1} \log g(n, \lfloor an \rfloor, \lfloor bn \rfloor, 1), \quad \text{where } a > 0, \quad b > 0, \quad 2a + b < 1.$$

From a calculation below,

$$f(a, b) = a \log 2 - (3a + 2b) + L(1 - 2a) - L(a) + 2L(a + b) - 2L(b) - 2L(1 - 2a - b),$$

where $L(x) = x \log(x)$. Numerical search gives us, for $a = .02$, $b = .1$, that $0 < f(a, b) = .00398 \cdots$. Thus for ρ sufficiently close to 1, $g(n, \lfloor .02n \rfloor, \lfloor .1n \rfloor, \rho)$ and hence b_2 and the second moment of W blow up exponentially as $n \to \infty$. To derive the formula above for $f(a, b)$, we start with an asymptotic formula for $g(n, k, m, \rho)$, with $n, k \to \infty$:

$$g(n, k, m, \rho) \sim \left(\frac{\rho}{n}\right)^{3k+2m} \frac{(n)_{2k}}{k!\,2^k} \left(\binom{n - 2k}{m} \times (k + m - 1)!\, 2^{k-1} \right)^2 \exp(-\mu).$$

In the above formula for g, the first factor is $\mathrm{E}(X_\alpha X_\beta)$. The second factor counts the number of ways to form k nonoverlapping edges to serve as common segments for α and β. The third factor counts the number of ways to pick m additional vertices for α, to order the k edges and m vertices, and to orient the k edges, and to do the same for β. The final factor, an exponential that is bounded away from zero, is the Poisson approximation for what fraction of the arrangements just counted actually have k common segments of length 2. For an upper bound on $\mu = \mu(n, k, m)$, we have $\mu \le 2$, which, in case both α and β had chosen the same m additional vertices, is the expected number of pairs of objects, either edges or vertices, adjacent in both α and β.

The pairs α, β counted in $g(n, k, m, \rho)$ form part, but not all, of the term in (15) indexed by k, $j = 2k + m$, $l = m$. As a check on the above computation of exponential growth, the values of $g(n, .02n, .1n, 1)\exp(\mu)$ for $n = 1000, 2000, \cdots, 7000, 8000$ are approximately .0001132, .001077, .0210, .5496, 16.90, 575.4, 21020.3 and 808488.

However, for small ρ, we have $b_2 = O(n^{-1})$, which we show at the end of this section. We note that

$$\sum (EX_\alpha)^2 = \sum_{j \ge 3} \frac{(n)_j}{2j} \left(\frac{\rho}{n}\right)^{2j} = O(n^{-3}).$$

Using (11) to bound the off-diagonal terms $EX_\alpha EX_\beta$ of b_1 as multiples of the corresponding terms $\mathrm{E}(X_\alpha X_\beta)$ of b_2 shows that

$$b_1 \le \sum_{\alpha \in I} (EX_\alpha)^2 + \frac{\rho}{n} b_2. \tag{16}$$

Thus, when ρ is small enough that $b_2 = O(n^{-1})$, we have $b_1 = O(n^{-2})$, and Proposition 1 follows directly from the Chen–Stein method given in Theorems 1–3. For ρ close to 1, however, $b_2 \to \infty$, and we must resort to the truncation argument given below.

Fix $\varepsilon \le 1$ and consider only cycles α of length $|\alpha|$ up to εn. Formally, consider the truncated indicators of cycles: for $\alpha \in I$,

$$X_\alpha^\varepsilon \equiv \mathbf{1}(|\alpha| \le \varepsilon n) X_\alpha,$$

which form the process $\mathbf{X}^\varepsilon \equiv (X_\alpha^\varepsilon)_{\alpha \in I}$. We have

$$\tfrac{1}{2} \| \mathscr{L}(\mathbf{X}^\varepsilon) - \mathscr{L}(\mathbf{X}) \| \le P(\mathbf{X}^\varepsilon \ne \mathbf{X}) \le \sum_{j > \varepsilon n} \rho^j (n)_j n^{-j} < \frac{\rho^{\varepsilon n}}{1 - \rho},$$

so that the approximation error in replacing $\mathbf{X}$ by $\mathbf{X}^\varepsilon$ is exponentially small as $n \to \infty$. The same holds for truncation of the Poisson process $\mathbf{Y}$ and the independent events process $\mathbf{X}'$. The bound (16) applies also to the truncated process, so that Proposition 1

for the truncated process will be proved if we can show that for the truncated process, $b_2 = O(n^{-1})$. The original version of Proposition 1 then follows by using the triangle inequality to compare the original with the truncated processes.

Using the Chen–Stein method with the same neighborhoods as before, we have the following upper bounds, corresponding to (15) for the truncated process.

$$
\begin{aligned}
b_2(\varepsilon) &= \sum_{\alpha\in I}\sum_{\beta\in B(\alpha)\setminus\{\alpha\}} \mathrm{E}(X_\alpha^\varepsilon X_\beta^\varepsilon) \\
&\le \sum_{j=3}^{en} \frac{(n)_j}{2j}\left(\frac{\rho}{n}\right)^j \sum_{k\ge 1}\left(\frac{\rho}{n}\right)^k 2\binom{j}{2k} \\
&\quad \times \sum_{l\ge 0}\binom{n-2k}{l}\left(\frac{\rho}{n}\right)^l (k+l-1)!\,2^{k-1} \\
&\le \sum_{j=3}^{en} \frac{\rho^j}{2j}\sum_{k\ge 1}\left(\frac{2\rho}{n}\right)^k\binom{j}{2k}\sum_{l\ge 0}\rho^l(k+l-1)_{k-1} \\
&= \sum_{j=3}^{en} \frac{\rho^j}{2j}\sum_{k\ge 1}\left(\frac{2\rho}{n}\right)^k\binom{j}{2k}\frac{(k-1)!}{(1-\rho)^k} \\
&< \sum_{j=3}^{en} \frac{\rho^j}{2j}\sum_{k=1}^{j/2}\binom{j}{2k}\left(\frac{2\rho k/n}{1-\rho}\right)^k \\
&< \sum_{j=3}^{en}\sum_{k\ge 1}\frac{\rho^j}{2j}\binom{j}{2k}\left(\frac{\rho j/n}{1-\rho}\right)^k \\
&< \sum_{j=3}^{en}\frac{j\rho^j}{2}\frac{\rho j/n}{1-\rho}\left(1+\sqrt{\frac{\rho j/n}{1-\rho}}\right)^j
\end{aligned}
\tag{17}
$$

To get the second equality, use the identity $\sum_{l\ge 0} x^l(k+l-1)_{k-1} = (k-1)!(1-x)^{-k}$. To get the next line, just replace $(k-1)!$ by k^k. To get the next line, we use $2k \le j$. For the final line, use the inequality $\sum_{k\ge 1}\binom{j}{2k}x^{2k} \le x^2 j^2(1+x)^j$ for $x > 0$.

For $0 < \rho < \delta < 1$, the final line of (17), which has $j/n \le \varepsilon$, shows that

$$
b_2(\varepsilon) < n^{-1}\frac{\delta\varepsilon}{1-\delta}\sum_{j\ge 3}\frac{j}{2}\left(\delta+\delta\sqrt{\frac{\delta\varepsilon}{1-\delta}}\right)^j.
$$

Given $\delta < 1$, we can find $\varepsilon > 0$ so small that $\delta + \delta\sqrt{\delta\varepsilon/(1-\delta)} < 1$. For such a choice of ε, we thus have $b_2(\varepsilon) < C(\delta)/n$, uniformly in $0 < \rho < \delta < 1$. This completes the proof of Proposition 1. In particular, we observe that $\varepsilon = 1$ works if $\delta < \frac{1}{2}$, so that for $\rho < \frac{1}{2}$, $b_1 + b_2 = O(n^{-1})$ and the Chen–Stein method works directly with no truncation.

4.4 Maxima of Independent and Dependent Normal Variates

The power of the Chen–Stein method is well illustrated by the classical problem of determining the distribution of the maximum of normal variates. An extensive treatment of this topic also using Stein's method appears in Holst and Janson (1990). See also Barbour, Holst, and Janson (1988b) for a treatment of the m-dependent case very similar to ours. Consider a sequence of independent standard normal variates $\{Z_1, Z_2, \cdots\}$. Let $M_n = M_{1,n} = \max_{\alpha\le n} Z_\alpha$.

Hall (1980) analyzes the distribution of M_n. He concludes that the usual approximation as scaled extreme value is too slowly convergent to be satisfactory for practically occurring sample sizes, and he suggests alternative approximations for it and for $M_{k,n}$, the kth largest of the first n observations. His derivation involves a careful asymptotic analysis. A number of similar results may be obtained using the Chen–Stein method in the independent case.

Choose a test value t. We wish to approximate $P\{M_n \le t\}$. Let $X_\alpha = \mathbf{1}\{Z_\alpha > t\}$, so that $EX_\alpha = p(t) = 1 - \Phi(t)$. If the test value is to be sufficiently large to be of interest, we may expect $p(t)$ to be rather small, so that the Poisson approximation should apply. With $I = \{1, 2, \cdots, n\}$, we have $W = \sum_{\alpha\in I} X_\alpha$, $EW = \lambda_n(t) = np(t)$; we are led to believe that $P\{M_n \le t\} \cong e^{-\lambda_n(t)}$.

Bounds on the quality of the approximations are given by Theorem 1. Using independence, we choose the neighborhood of dependence $B(\alpha) = \{\alpha\}$ and find

$$
\begin{aligned}
b_1 &= np^2(t) = \lambda_n^2(t)/n \\
b_2 &= 0 \\
b_3 &= 0.
\end{aligned}
$$

We may conclude immediately that since $\{M_n \le t\} = \{W = 0\}$

$$
\begin{aligned}
e^{-\lambda_n(t)} - \lambda_n^2(t)/n \le P\{M_n \le t\} \\
\le e^{-\lambda_n(t)} + \lambda_n^2(t)/n.
\end{aligned}
\tag{18}
$$

Hall's approximations essentially involve writing

$$
\begin{aligned}
P\{M_n \le t\} &= (\Phi(t))^n \\
&= (1-(1-\Phi(t)))^n \cong \exp(-\lambda_n(t)),
\end{aligned}
\tag{19}
$$

approximating the upper tail of the normal by the asymptotic expansion (26.2.12) of Abramowitz and Stegun (1964) and providing usably simple explicit bounds for the error of approximation. Note that the last term of (19) can be interpreted as the Poisson probability whose error of approximation is bounded by (18).

In Table 1, we give various lower and upper bounds for $P\{M_n \le t\}$. Compared are the bounds given in Hall (1980) with the bounds (18), and with modified bounds given by replacing the normal distribution with the bounds (20) for Mill's ratio.

Following Hall (1980), write:

$$Q_1 = Q_{1n}(t) = \exp\left(-n\frac{\phi(t)}{t}\left(1 - \frac{1}{t^2} + \frac{3}{t^4} + n\frac{\phi(t)}{2(n-1)}\right)\right),$$

$$Q_{2n}(t) = \exp\left(-n\frac{\phi(t)}{t}\left(1 - \frac{1}{t^2}\right)\right),$$

$$Q_{3n}(t) = \exp\left(-n\frac{\phi(t)}{t}\left(1 - \frac{1}{t^2} + \frac{3}{t^4} - \frac{15}{t^6}\right)\right),$$

all valid when $2\pi t^2 e^{t^2} > n$. Hall shows the function Q_{1n} is a lower bound on the distribution of M_n, and that Q_{2n} and Q_{3n} are upper bounds.

In the following analysis we make repeated use of the inequalities:

$$(20)\qquad \frac{2}{t + \sqrt{t^2+4}} < \frac{1-\Phi(t)}{\phi(t)} < \frac{4}{3t + \sqrt{t^2+8}},$$

for $t \geq 0$. The lower bound is due to Birnbaum (1942); the upper bound is proved in Sampford (1953). Both bounds can be obtained as corollaries to Karlin (1982), in which total positivity is used to prove the monotonicity of the variance of certain families of truncated distributions, including the left-truncated normal family.

Write $\bar{\lambda}_n(t) = 4n\phi(t)/(3t + \sqrt{8+t^2})$ and $\underline{\lambda}_n(t) = 2n\phi(t)/(t + \sqrt{4+t^2})$. Observe that $\underline{\lambda}_n(t) < \lambda_n(t) < \bar{\lambda}_n(t)$, following directly from (20). We present in Table 1 a comparison of the lower bounds $Q_{1n}(t)$, the Chen–Stein bound $L_{0n}(t)$ from (18) and $L_{1n}(t)$ obtained by substituting the upper bound $\bar{\lambda}_n(t)$ in (18). The upper bounds U_{0n} and U_{1n} are similarly obtained, save that both $\bar{\lambda}_n(t)$ and $\underline{\lambda}_n(t)$ need to be used in obtaining U_{1n} from U_{0n}. All errors are reported as percentages of the actual upper tail of the exact distribution of the maximum.

Note that the Chen–Stein bounds L_0 and U_0 using the normal distribution function are by far the tightest of all the bounds displayed over the range of values tabulated. If one wishes to approximate the upper tail of the normal distribution using Mill's ratio, then the bounds L_1 and U_1 are next preferred, save when the test value t exceeds 4, in which case Q_3, which uses an asymptotic expansion to sixth order, is preferred to U_1. As Hall notes, taking more terms in an asymptotic expansion is not always desirable. Compare the errors for Q_2 and Q_3 when $n = 10$.

An appealing feature of the Poisson approximation is its versatility. Since $\{M_{k,n} \leq t\} = \{W < k\}$, from the

TABLE 1

Percent relative errors for bounds on the distribution of the maximum of independent normal variates

n	t	$\Phi^n(t)$	$\frac{Q_{1n}-\Phi^n}{1-\Phi^n}$	$\frac{L_{1n}-\Phi^n}{1-\Phi^n}$	$\frac{L_{0n}-\Phi^n}{1-\Phi^n}$	$\frac{U_{0n}-\Phi^n}{1-\Phi^n}$	$\frac{U_{1n}-\Phi^n}{1-\Phi^n}$	$\frac{Q_{2n}-\Phi^n}{1-\Phi^n}$	$\frac{Q_{3n}-\Phi^n}{1-\Phi^n}$
						%			
10	1.6	.5692	−24.26	−5.40	−4.89	9.05	11.14	20.02	73.76
	2.0	.7944	−10.30	−1.78	−1.50	3.53	5.05	10.84	15.90
	2.4	.9210	−4.33	−.62	−.46	1.24	2.29	6.08	4.78
	2.8	.9747	−1.94	−.22	−.13	.38	1.10	3.55	1.62
	4.0	.9997	−.28	−.02	−.00	.00	.25	.96	.11
50	2.2	.4966	−4.76	−1.60	−1.44	2.40	3.35	6.14	6.50
	2.6	.7917	−2.58	−.42	−.31	.73	1.52	4.22	2.50
	3.0	.9347	−1.30	−.14	−.07	.20	.78	2.70	.96
	3.4	.9833	−.67	−.06	−.02	.05	.46	1.74	.38
	4.5	.9998	−.14	−.01	−.00	.00	.17	.62	.05
100	2.4	.4391	−2.82	−1.04	−.93	1.46	2.17	4.18	3.27
	2.8	.7743	−1.71	−.26	−.18	.40	1.04	3.17	1.45
	3.2	.9336	−.91	−.09	−.04	.10	.58	2.12	.59
	3.6	.9842	−.49	−.04	−.01	.02	.37	1.41	.25
	4.5	.9997	−.14	−.01	−.00	.00	.17	.62	.05
500	3.0	.5090	−.94	−.19	−.14	.23	.65	1.97	.69
	3.4	.8449	−.62	−.06	−.02	.05	.43	1.62	.36
	3.8	.9645	−.36	−.03	−.00	.01	.30	1.14	.16
	4.2	.9933	−.21	−.02	−.00	.00	.21	.80	.08
1000	3.2	.5029	−.65	−.11	−.07	.12	.46	1.54	.43
	3.6	.8529	−.46	−.04	−.01	.02	.34	1.31	.23
	4.0	.9688	−.27	−.02	−.00	.00	.25	.95	.11
	4.4	.9946	−.16	−.01	−.00	.00	.18	.67	.05

identical calculations, the Chen–Stein method yields with no more work

$$\left| \sum_{j=0}^{k-1} \frac{\lambda_n^j(t)}{j!} e^{-\lambda_n(t)} - P\{M_{k,n} \le t\} \right| \le \frac{\lambda_n^2(t)}{n}.$$

Similar bounds are not explicitly available in Hall (1980) and would require substantially more work.

Our interest in the Chen–Stein Poisson approximation arose from our study of maxima of weakly dependent random sequences. Our initial tool was the Bonferroni inequalities, whose effective use we learned from the seminal paper of Watson (1954). Watson's method implicitly requires that one compute all moments of the sum of indicators $\sum X_\alpha$. Hence the use of Watson's method is equivalent to proving convergence in distribution to the distribution hoped to be determined by limits of moments of the counting process. Watson illustrates the utility of his method by evaluating the limiting distribution of the maximum of a stationary k-dependent sequence of jointly normal variates.

Here is the corresponding computation using the Chen–Stein method for the case of a 1-dependent moving average of normal variates. Let $Y_\alpha = (Z_\alpha + \theta Z_{\alpha+1})/\sqrt{1+\theta^2}$ be a stationary sequence of normal variates with mean 0, unit variance, and common lag-1 autocorrelation $\rho = \theta/(1+\theta^2)$. Let $M_n^* = \max_{\alpha \le n} Y_\alpha$. Again choose test value t. Form $X_\alpha = \mathbf{1}\{Y_\alpha > t\}$ so that

$$P\{M_n^* \le t\} = P\left\{ \sum_{\alpha=1}^{n} X_\alpha = 0 \right\}.$$

Choose neighborhoods of dependence $B_\alpha = \{\alpha - 1, \alpha, \alpha + 1\} \cap \{1, \cdots, n\}$. Let $p(t) = 1 - \Phi(t)$ and $\lambda_n(t) = np(t)$ be as before. We then have for positive t

$$b_1 < \frac{3\lambda_n^2(t)}{n}$$

$$b_2 < 2C(\rho)\lambda_n(t)\left(\frac{1}{u^2+1}\right)^{\rho/(1+\rho)}\left(\frac{\lambda_n(t)}{n}\right)^{(1-\rho)/(1+\rho)}$$

$$b_3 = 0,$$

where $u = t\sqrt{2/(1+\rho)}$, and

$$C(\rho) = \sqrt{2\pi}^{(1-\rho)/(1+\rho)} \sqrt{\frac{2(1+\rho)}{\pi(1-\rho)}}. \tag{21}$$

The bound on b_1 is immediate from the definition of B_α. The bound on b_2 is a consequence of the elementary inequality of Lemma 1—stated and proved at the end of the section. The term b_3 is zero because dependence is local. Hence by Theorem 1

$$\left| P\left\{ \max_{\alpha \le n} Y_\alpha \le t \right\} - e^{-\lambda_n(t)} \right|$$

$$< (1 - e^{-\lambda_n(t)})\left(\frac{3\lambda_n(t)}{n} + 2C(\rho)\left(\frac{1}{u^2+1}\right)^{\rho/(1+\rho)} \times \left(\frac{\lambda_n(t)}{n}\right)^{(1-\rho)/(1+\rho)}\right). \tag{22}$$

The bounds are useful when $\lambda_n(t)/n$ is close to 0, and so are Poisson approximations to the distributions of other extreme order statistics.

Although bounds for rates of convergence are implicit in Watson's (1954) use of the Bonferroni inequalities, they are certainly more conveniently available in the Chen–Stein formulation.

An exhaustive treatment of rates of convergence for stationary Gaussian time series is given in Rootzén (1983). There, rates of convergence are established and bounds with explicit constants are given in substantial generality. Connections are made with Poisson approximation using coupling methods due to Serfling (1975). The chief tool is a technical lemma relating the distribution of dependent and independent Gaussian variates.

The bounds obtained with the Chen–Stein method are frequently quite good. This is true in our simple example above, in which, when t grows like $\sqrt{2\ln(n)}$, the rates of convergence of the bounds given above are exactly those of the bounds obtained by Rootzén (1983), shown there to be of best possible order. In this case, the coefficient of the leading term in (22) is about 1.92, compared to Rootzén's 4.47. The computation sketched above carries over with obvious modification for finite moving averages.

Finally, we end the section with the promised lemma:

LEMMA 1. *Let Y_1, Y_2 be jointly standard normal with covariance ρ. For $t > 0$, write $u = t\sqrt{2/(1+\rho)}$. Then*

$$P\{\min\{Y_1, Y_2\} > t\}$$

$$< \sqrt{\frac{2(1+\rho)}{\pi(1-\rho)}}\,(\phi(u) - u(1 - \Phi(u))) \tag{23}$$

$$< C(\rho)\left(\frac{1}{u^2+1}\right)^{\rho/(1+\rho)} (1 - \Phi(t))^{2/(1+\rho)}, \tag{24}$$

where $C(\rho)$ is defined in (21).

PROOF. Note that $Y_1 + Y_2$ and $Y_1 - Y_2$ are uncorrelated, and that the event $\{\min\{Y_1, Y_2\} > t\} =$

$\{|Y_1 - Y_2|/2 < (Y_1 + Y_2)/2 - t\}$. Hence

$$
\begin{aligned}
&P\{\min\{Y_1, Y_2\} > t\} \\
&\quad = 2P\{0 < Y_1 - Y_2 \\
&\qquad\qquad < Y_1 + Y_2 - 2t \mid Y_1 + Y_2 > 2t\} \\
&\qquad \times P\{Y_1 + Y_2 > 2t\} \\
&\quad < 2\sqrt{\frac{1}{4\pi(1-\rho)}}\, \mathrm{E}\{Y_1 \\
&\qquad\qquad + Y_2 - 2t \mid Y_1 + Y_2 > 2t\} \\
&\qquad \times (1 - \Phi(u)) \\
&\quad = \sqrt{\frac{1}{\pi(1-\rho)}}\, \sqrt{2(1+\rho)}\left(\frac{\phi(u)}{1-\Phi(u)} - u\right) \\
&\qquad \times (1 - \Phi(u)),
\end{aligned}
$$

proving the first inequality. To prove the second, use (20) repeatedly:

$$
\begin{aligned}
&\phi(u) - u(1 - \Phi(u)) \\
&\quad < \left[1 - \frac{2u}{\sqrt{u^2+4} + u}\right]\phi(u) \\
&\quad = \left[\frac{\sqrt{u^2+4} - u}{\sqrt{u^2+4} + u}\right]\phi(u) \\
&\quad = \sqrt{2\pi}^{(1-\rho)/(1+\rho)}\left[\frac{2}{\sqrt{u^2+4} + u}\right]^2 \phi(t)^{2/(1+\rho)} \\
&\quad = \sqrt{2\pi}^{(1-\rho)/(1+\rho)}\left[\frac{2}{\sqrt{u^2+4} + u}\right]^{2-2/(1+\rho)} \\
&\qquad \times \left[\frac{2}{\sqrt{u^2+4} + u}\phi(t)\right]^{2/(1+\rho)} \\
&\quad < \sqrt{2\pi}^{(1-\rho)/(1+\rho)}\left[\frac{2}{\sqrt{u^2+4} + u}\right]^{2\rho/(1+\rho)} \\
&\qquad \times [1 - \Phi(t)]^{2/(1+\rho)} \\
&\quad < \sqrt{2\pi}^{(1-\rho)/(1+\rho)}\left[\frac{1}{u^2+1}\right]^{\rho/(1+\rho)} \\
&\qquad \times [1 - \Phi(t)]^{2/(1+\rho)}.
\end{aligned}
$$

□

4.5 Permutations with Restricted Positions

Consider a probability model in which all $n!$ permutations π on $\{1, 2, \cdots, n\}$ are equally likely. For $i = 1, 2, \cdots, n$, *let* $F_i \subset \{1, 2, \cdots, n\}$ be given, to be thought of as the set of restricted positions for element i. The random variable

$$W \equiv W(\pi) \equiv \sum_{1 \le i \le n} \mathbf{1}(\pi_i \in F_i)$$

counts the number of restricted positions taken by a random permutation. Its expectation is

$$\lambda \equiv EW = \frac{1}{n}\sum |F_i|.$$

Our goal is to understand the relation between the number of permutations with no elements in restricted positions, and the Poisson approximation, $n!e^{-\lambda}$. More generally, we are concerned with how well the distribution of W matches the Poisson distribution with parameter λ, and how close in distribution are the family of dependent events $(\{\pi_i \in F_i\})_{1 \le i \le n}$ and a family of independent events of the same individual probabilities.

The problem of permutations with restricted positions is also treated in Chen (1975b) and Barbour and Holst (1989), which contains many references. These papers also start with Stein's method as embodied by Equations (2) and (3). The next step, as presented clearly in Barbour and Holst (1989), is to look, for each of the events being counted, for a good coupling between the total number W of events, and a random variable equal in distribution to the number of events, minus one, conditioned on the occurrence of the one selected event. That treatment of Stein's method allows the user more freedom of choice than what we are presenting in this paper as the "Chen–Stein method." For the benchmark example of permutations with restricted positions, the Chen–Stein method as presented here is both harder to use and gets a worse bound on the Poisson approximation for W. In detail, apart from constant factors, the bounds in the other two papers, and our term b_1, are equivalent, but we also have a term b_3, which is greater than b_1 by a factor which is of the order of $\log n$.

Overall then, for the class of problems as described in Example 1.3, our bound shows that the number of restricted positions taken by a uniformly selected random permutation is approximately Poisson, with a bound on the error decreasing at rate $\log n/n$. However, for no additional work, the Chen–Stein method yields information about the entire process of occurrences, via Theorems 2 and 3.

EXAMPLE 1.1. Derangements. Let $F_i = \{i\}$ for $i = 1$ to n. Then W is the number of fixed points of a random permutation, $\{W = 0\}$ is the set of derangements of n objects and $\lambda = 1$. This example is exceptional and misleading, in that the error in the Poisson approximation is superexponentially small as $n \to \infty$.

EXAMPLE 1.2. The ménage problem. Let $F_i = \{i, i+1\}$ for $i = 1$ to $n-1$ and $F_n = \{n, 1\}$. Here $\lambda = 2$, and careful use of inclusion–exclusion shows that the Poisson approximation satisfies $|P(W=0) - e^{-2}| \sim cn^{-1}$, in contrast to the superexponential decay of Example 1.1.

EXAMPLE 1.3. Derangements, ménage problems, etc. Let $F_i = \{i, i+1, \cdots, i+d-1\}$ for $i = 1$ to n, with the additions taken modulo n. Here $\lambda = d$, and the two examples above are the special cases $d = 1, 2$ of this more general example. In even greater generality, following Riordan (1978), we may consider $\{W = 0\}$ to be the set of permutations discordant with d given permutations $\sigma_1, \sigma_2, \cdots, \sigma_d$, by taking $F_i = \{\sigma_1(i), \cdots, \sigma_d(i)\}$ for $i = 1$ to n. We will analyze this example using the Chen–Stein method below to get $b_1 \le d(2d-1)/n$, $b_2 = 0$ and $b_3 = O(\log n/n)$.

EXAMPLE 2. Let $F_i = \{i, n\}$ for $i = 1$ to $n-1$ and $F_n = \{n, 1\}$. Here $\lambda = 2$, and the Poisson approximation is not at all valid, since $P(W=0) = 0$.

The "natural" way to use the Chen–Stein method would be to take $I = \{1, \cdots, n\}$ and $X_i \equiv \mathbf{1}(\pi_i \in F_i)$ for $i \in I$. The neighborhood of dependence in this setup would then be $B_i \equiv \{j \in I: F_i \cap F_j \neq \varnothing\}$.

Instead, we take an approach which gives symmetric treatment to the domain and range of the random permutation. Thus, we let

$$I = \{\alpha = (i, j): j \in F_i\}$$

and

$$\text{for } \alpha = (i, j) \in I, \quad X_\alpha \equiv \mathbf{1}(\pi_i = j),$$

so that I may be thought of as the set of restricted edges in the bipartite graph $K_{n,n}$ of possible edges between n men and n women, with $EX_\alpha = 1/n$ for all $\alpha \in I$, and $|I| = \lambda n$. Our choice of the neighborhood of dependence of α is the set of edges sharing an endpoint with α:

$$\text{for } \alpha = (i, j),$$

$$B_\alpha \equiv \{\beta = (i', j') \in I: i = i' \text{ or } j = j'\}.$$

The first two components of the Chen–Stein bound then are

$$b_1 = \sum_{\alpha \in I} \sum_{\beta \in B_\alpha} EX_\alpha EX_\beta = n^{-2} \sum_{\alpha \in I} |B_\alpha|, \quad b_2 = 0.$$

In Example 1.3 we have $|B_\alpha| \le 2d-1$, so $b_1 \le d(2d-1)/n$. If we have used the "natural" setup described in the first paragraph of this section, then b_1 would be increased by a factor of d in this example. Instead of having $b_2 = 0$, in the "natural" setup we would use the easily established bound $E(X_\alpha X_\beta) \le n/(n-1)(EX_\alpha EX_\beta)$, so that $b_2 < n/(n-1)b_1$. For b_3, the technique we use below leads to the same upper bound on b_3 with either setup.

Random permutations embody long-range, global dependence, so any choice of neighborhoods will have $b_3 > 0$, other than the choice $B_\alpha \equiv I$, which gives the useless large value $b_1 = \lambda^2$. It is possible but difficult to give a useful upper bound on b_3; we carry this out in the next three lemmas. The first lemma states the bound on b_3 and relies on the next two lemmas. The thing most worth observing in the lemmas below is the technique for getting a handle on b_3, displayed by the equality in (26): since each term of b_3 is the expectation of the absolute value of the conditional expectation of something with mean zero, each term can be expressed as twice the expectation of the positive part (or the negative part) of that conditional expectation.

LEMMA 2.

$$b_3 \le \min_{1<k<n} \left(\frac{2\lambda k}{n-k} + 2n\lambda 2^{-k} e^{\lambda e} \right)$$

$$\sim 2\lambda \frac{(2\log_2(n) + \lambda e/\ln 2)}{n} \quad \text{if } \lambda = o(n).$$

PROOF. Fix $\alpha \in I$, let $V = \sum_{\beta \notin B_\alpha} X_\beta$, and for $J \subset I - B_\alpha$ define the event

$$E_J \equiv \{X_\beta = 1 \ \forall \beta \in J, \quad X_\beta = 0 \ \forall \beta \in I - B_\alpha - J\}, \tag{25}$$

so that on the event E_J we have $V = |J|$. There are $n\lambda$ contributions to b_3 of the form

$$\begin{aligned} s_\alpha &\equiv E\left| E\left(X_\alpha - \frac{1}{n} \middle| \sigma(X_\beta: \beta \notin B_\alpha)\right)\right| \\ &= \sum_J \left| E\left(X_\alpha - \frac{1}{n} \middle| E_J\right)\right| P(E_J) \\ &= \sum_J 2\left(E\left(X_\alpha - \frac{1}{n} \middle| E_J\right)\right)^+ P(E_J) \qquad (26) \\ &\le 2 \sum_J \left(\frac{1}{n - |J|} - \frac{1}{n}\right)^+ P(E_J) \\ &\qquad \text{(using Lemma 3)} \\ &= 2 \sum_{0 \le j < n} \left(\frac{1}{n-j} - \frac{1}{n}\right) P(V = j) \\ &\le 2\left(\frac{1}{n-k} - \frac{1}{n}\right) + 2P(V \ge k). \end{aligned}$$

To bound the last term, we use the upper bound based on Lemma 4: since $V \le W$, $P(V \ge k) \le P(W \ge k) \le 2^{-k} E 2^W \le 2^{-k} e^{\lambda e}$. Multiplying by $|I| = n\lambda$, for every

positive integer k, we have the upper bound

$$b_3 \le 2\lambda k/(n-k) + 2n\lambda 2^{-k}e^{\lambda e}.$$

If $\lambda = o(n)$, then taking $k = \lceil 2\log_2(n) + \lambda e/\ln 2 \rceil$ makes the second term negligible and demonstrates the asymptotic claim. □

LEMMA 3. *For the event E_J defined in* (25),

$$E(X_\alpha \mid E_J) \le 1/(n - |J|). \tag{27}$$

PROOF. To describe the conditioning event E_J, let $k = |J|$, and relabel the men and women so that $\alpha = (1, 1)$ and $J = \{(n-k+1, n-k+1), \cdots, (n, n)\}$. The conditioning event E_J can then be viewed as a collection of matchings between a set of $n-k$ men and $n-k$ women, and the conditions of the form $X_\beta = 0$ forbid certain matchings, which do not involve man 1 or woman 1 since $\beta \notin B_\alpha$. This event E_J can be partitioned into $n-k$ subsets according to the mate chosen by man 1, $U_j \equiv \{\pi \in E_J : \pi_1 = j\}$, so that the conditional probability above is the ratio $|U_1|/|E_J|$. Thus, it suffices to show that, for $j = 1$ to $n-k$, we have $|U_1| \le |U_j|$.

We prove that $|U_1| \le |U_j|$ for $j = 2$ to $n-k$ by presenting a one-to-one map f which maps U_1 into U_j, namely composition with the appropriate transposition:

$$f(\pi) = (1 \quad j) \circ \pi.$$

Informally, f is the map that has women 1 and j swap their mates. We observe that when $\pi \in U_1$, then $f(\pi) \in U_j$, because the two new matchings created involve man 1 or woman 1, and hence E_J places no restriction on the use of these edges. We have inequality in Lemma 3 because f may not be onto, for if $\sigma \in U_j$ with $\sigma_i = 1$, then $\pi = f^{-1}(\sigma)$ has $\pi_1 = 1$ and $\pi_i = j$, but if $\beta = (i, j) \in I$ is a restricted edge, then $X_\beta(\pi) = 1$ so that $\pi \notin E_J$, hence $\pi \notin U_1$. □

LEMMA 4.

$$E2^W \le e^{\lambda e}.$$

PROOF. Observe that, for any $J \subset I$ with $|J| = k$, $E\prod_{\alpha\in J} X_\alpha$ is either zero, in case any of the edges in J intersect, or else is $(n-k)!/n! = 1/(n)_k$. Recall that $|I| = n\lambda$. Thus

$$\begin{aligned} E2^W &= E\prod_{\alpha\in I}(1 + X_\alpha) \\ &= \sum_{J\subset I} E\prod_{\alpha\in J} X_\alpha \\ &\le \sum_{0\le k\le n} \binom{n\lambda}{k}\frac{1}{(n)_k}. \end{aligned}$$

For the kth term of the last sum we have

$$\binom{n\lambda}{k}\frac{1}{(n)_k} \le \frac{(n\lambda)^k}{k!(n/e)^k} = \frac{(\lambda e)^k}{k!}.$$ □

4.6 Cycles in Permutations and Random Mappings

As in Section 4.5, we again consider a probability model in which all $n!$ permutations π on $\{1, 2, \cdots, n\}$ are equally likely. Our goal is to understand to what extent cycles of a random permutation occur approximately independently and whether a Poisson approximation holds for the numbers of cycles of various lengths. For any fixed $j \ge 1$, it is easy to show by inclusion–exclusion that the number W_j of cycles of length j converges to Z_j, a Poisson random variable, with $EZ_j = 1/j$.

The first interesting phenomenon illustrated by this example is that the Chen–Stein method and the idea of computing the total variation distance to a process with independent coordinates let us compute a "critical boundary" for Poisson approximation. Consider $\mathbf{Z} \equiv (Z_1, Z_2, \cdots)$, the Poisson process with $EZ_j = 1/j$ and independent coordinates. It can be shown by inclusion–exclusion that finite dimensional distributions of the cycle counting process converge to those of $\mathbf{Z}$, and, since $\sum j W_j = n$, it is easy to see that the full process counting cycles, $\mathbf{W} \equiv (W_1, \cdots, W_n)$, is not close, in total variation, to the first n coordinates of the Poisson process $\mathbf{Z}$. We will consider jointly all cycles of length 1 (i.e., fixed points), 2, 3, $\cdots$, $f(n)$, where, for example, $f(n)$ grows like $\sqrt{n}$ or $n/\log n$. It turns out that a Poisson approximation for $\{W_1, \cdots, W_{f(n)}\}$ is good as long as $f(n) = o(n)$.

The second phenomenon illustrated is that a Poisson approximation for the process $\{W_1, \cdots, W_{f(n)}\}$ may be valid even when the Chen–Stein method fails. This occurs here in all cases where $f(n)/\sqrt{n} \to \infty$ and $f(n)/n \to 0$. In these cases, the process of indicators is *not* approximately independent, the total variation distances in Theorems 2 and 3 tend to 2 and, hence, $(b_1 + b_2 + b_3)$ cannot tend to zero and Theorem 1 cannot yield a successful approximation for W. This reflects what is usually a virtue of the Chen–Stein method; when the method does certify a successful approximation for the number W of occurrences via $(b_1 + b_2 + b_3) \to 0$, the method would also imply, through Theorems 2 and 3, that the process of indicators is approximately independent.

4.6.1 Independence among the short cycles

The natural way to establish a Poisson approximation for $(W_1, \cdots, W_{f(n)})$ is to use an index set I consisting of all cycles of length at most $f(n)$. For $j \ge 1$, let I_j be the set of cyclic permutations α of exactly j elements of $\{1, 2, \cdots, n\}$, so that $|I_j| = (n)_j/j$. For $\alpha \in I_j$, let X_α be the indicator of the event

that α is a cycle of the random permutation π, so that $W_j \equiv \sum_{\alpha \in I(j)} X_\alpha$ is the number of cycles of length j. Now $EX_\alpha = 1/(n)_j$, and the expected number of cycles of length j is therefore

$$EW_j = E \sum_{\alpha \in I(j)} X_\alpha = \frac{|I_j|}{(n)_j} = \frac{1}{j}, \quad \text{if } j \le n.$$

In a slight abuse of notation, we speak of the intersection of two cycles rather than the intersection of the corresponding sets of elements involved. Two cycles α, β will be considered neighbors if their intersection is nonempty. Note that distinct neighbors cannot both be cycles of the same permutation, so that

$$\alpha \neq \beta \in B_\alpha \quad \text{implies } X_\alpha X_\beta \equiv 0,$$

hence $b_2 = 0$.

Fix a function f mapping the positive integers to positive integers, and let the index set $I \equiv I_f$ consist of all "short" cycles, i.e., cycles of length at most $f(n)$: $I \equiv \bigcup_{j \le f(n)} I_j$.

This paragraph and the next discuss situations in which the Chen–Stein method implies a Poisson approximation for the number of cycles of length at most $f(n)$. We claim that

$$b_1 \to 0 \quad \text{if and only if} \quad \frac{(f(n))^2}{n} \to 0$$

(28) and furthermore

$$b_1 \le \frac{f(n)^2}{n}.$$

We prove here only the second half of the above claim. Note that for each j, the expected number of elements in cycles of length j is $jEW_j = 1$, so the expected number of elements in cycles in our index set I is $f(n)$. This implies that the expected number of cycles in I containing any fixed element $i \in \{1, 2, \cdots, n\}$ is $f(n)/n$. Thus

$$\begin{aligned} b_1 &= \sum_{\alpha,\beta \in I} \mathbf{1}(\alpha \cap \beta \neq \emptyset) EX_\alpha EX_\beta \\ &\le \sum_{\alpha,\beta \in I} |\alpha \cap \beta| EX_\alpha EX_\beta \\ &= \sum_i \left(\sum_{\alpha \in I} \mathbf{1}(i \in \alpha) EX_\alpha \right)^2 \\ &= n\left(\frac{f(n)}{n}\right)^2 = \frac{f(n)^2}{n}. \end{aligned}$$

If $f(n) < n^{1/2-\varepsilon}$ for some $\varepsilon > 0$, then techniques similar to those in the last section show that $b_3 \to 0$. Furthermore, if b_1, b_2 and b_3 tend to zero, then Theorem 3 implies that the short cycles (i.e., those of length at most $f(n)$) occur with a joint distribution that is close to mutual independence: $\| \mathscr{L}(\mathbf{X}) - \mathscr{L}(\mathbf{X}') \| \to 0$. (Here, $\mathbf{X}$ and the independent process $\mathbf{X}'$ are indexed by $\alpha \in I \equiv I_f$.)

Next we consider cases of f for which the Chen–Stein method cannot possibly succeed. If $f(n)^2/n$ is bounded away from zero [respectively, goes to infinity], then b_1 is bounded away from zero [goes to infinity]. Consider $T \equiv \sum_{\alpha \in I} \sum_{\alpha \neq \beta \in B(\alpha)} X'_\alpha X'_\beta$, which is the number of pairs of distinct neighbors in $\mathbf{X}'$, and recall that the number of pairs of distinct neighbors in $\mathbf{X}$ is identically zero, so that the total variation distance between $\mathbf{X}$ and $\mathbf{X}'$ is at least $2P(T > 0)$. Now b_1, apart from its diagonal terms, is ET: $b_1 - \sum_{\alpha \in I} (EX_\alpha)^2 = \sum_{\alpha \in I} \sum_{\alpha \neq \beta \in B(\alpha)} EX_\alpha EX_\beta = \sum_{\alpha \in I} \sum_{\alpha \neq \beta \in B(\alpha)} X'_\alpha X'_\beta$, $= ET$. If $f(n)^2/n$ is bounded away from zero [goes to infinity], then b_1 and ET are bounded away from zero [go to infinity], and $P(T > 0)$ and $\frac{1}{2} \| \mathscr{L}(\mathbf{X}) - \mathscr{L}(\mathbf{X}') \|$ are bounded away from zero [tend to 1]. The implication that $ET > 0$ can be shown by the method of first and second moments or another application of the Chen–Stein method. Even though $(W_1, \cdots, W_{f(n)})$ is approximately Poisson for all $f = o(n)$, this cannot possibly be established by the Chen–Stein method for f with $f^2(n)/n \to \infty$, since in this case the process of indicators is not close to being independent.

Instead of random permutations, consider random mappings of $\{1, \cdots, n\}$ into $\{1, \cdots, n\}$, with all n^n mappings equally likely. Using exactly the same setup as above, we have $b_2 = b_3 = 0$, so that b_1 tells the entire story. For a cycle α of length j, the value of EX_α changes from $1/(n)_j$ to n^{-j}, but the statements in (28) and all the qualitative relations above involving $f(n)^2/n$, b_1, $P(T > 0)$ and $\| \mathscr{L}(\mathbf{X}) - \mathscr{L}(\mathbf{X}') \|$ remain the same as in the case of permutations.

4.6.2 Independence among the numbers of cycles of short lengths

A necessary and sufficient condition for the first $f(n)$ coordinates of the cycle counting process to be close, in total variation, to the first $f(n)$ coordinates of its limiting Poisson process, is that $f(n)/n \to 0$:

$$|(W_1, \cdots, W_{f(n)}) - (Z_1, \cdots, Z_{f(n)})| \to 0 \quad \text{if and only if } f(n) = o(n). \tag{29}$$

This result, from Arratia and Tavaré (1990), has a surprisingly easy direct proof, starting from Cauchy's formula for the number of permutations with a given cycle structure.

What the last section shows is that the Chen–Stein method applied to cycles of length at most $f(n)$ cannot establish (29) beyond $f(n) \cong \sqrt{n}$, because the events indexed by cycles carry too much information to be

approximately independent. The cycle events process $\mathbf{X}$ of the previous section not only detects the number of cycles of each length up to $f(n)$, it also carries information about which elements of $\{1, 2, \cdots, n\}$ are related by being in the same short cycle.

It is possible to get around this problem, by assigning to each cycle of a random permutation π exactly one of its elements as its "marker." In more detail, let $I = \{1, \cdots, n\} \times \{1, \cdots, f(n)\}$, and for $(i, j) \in I$,

$$X_{ij} = \mathbf{1}\{i \text{ "marks" a cycle in } \pi \text{ of length } j\}.$$

With this setup, W_j, the number of cycles of length j, equals $\sum_{1 \le i \le n} X_{ij}$. Two elements $\alpha = (i, j)$, $\beta = (k, l)$ are taken to be neighbors if $i = k$, so $b_2 = 0$. The computation of b_1 shows that we should be careful in picking a notion of "marking" a cycle: if a cycle is marked by its smallest element, then for each j, EX_{ij} is a messy decreasing function of i. If instead we take an independent auxiliary random permutation to serve as a ranking and mark a cycle by its element of smallest rank, then $EX_{ij} = 1/(nj)$ and

$$b_1 = n\left(\frac{\sum_{j \le f(n)} 1}{nj}\right)^2 \sim \frac{(\log f(n))^2}{n} \to 0,$$

even for $f(n) = n$.

Finally, with either notion of marking, it should be the case that $b_3 \to 0$ if and only if $f(n)/n \to 0$. We believe this on intuitive grounds but we have not tried to give a detailed proof, since the direct proof of (29) is simple and yields a stronger result than the Chen–Stein method would yield with this setup.

5. A BIOLOGICAL EXAMPLE

Our continuing desire is to solve problems relevant to molecular biology. This desire was the original motivation for studying the Chen–Stein Poisson approximation. In this section, we present its application to a problem motivated by current data analytic techniques in molecular biology.

A strand of DNA can be represented as a long string of letters from the finite alphabet $\{a, c, g, t\}$. Currently, a large amount of laboratory effort is being expended in the determination and subsequent compilation of genetic information from various organisms. This information consists of listings of these long strings. These data are collected in international databases, GenBank in the United States and EMBL in Europe. Currently, release 62.0 of GenBank contains 37,183,950 letters of DNA, made up roughly of sequences of 1,000 letters each. Given two strings of n and m letters, information about their comparison is conceptually summarized as a matrix of $n \times m$ positions in which a match of letters in positions i and j is traditionally represented by a dot in position (i, j)—a practice leading to the biologists' term "dot matrix analysis."

A natural question arises from comparison of two or more such strings, when the scientist wants to know whether a comparison detects an unusual congruence shared among the strings. In our problem, congruence is measured by the number of letters that match between two subwords of these sequences. Although important biological questions involve the more general notions of insertion and deletion, we restrict our study to the simpler question of matching and nonmatching positions. Such statistical problems are naturally cast in the usual hypothesis-testing context in which we need to compute the tail probability (the biologists' p-value) for a seemingly unusual event.

The standard tool used to solve such problems has been a probabilistic use of the Bonferroni inequalities as pioneered in Watson (1954). See, for example, the moment calculations in Karlin and Ost (1987) and the discussion in Karlin, Ghandour, Ost, Tavaré and Korn (1983). Use of the Bonferroni inequalities requires computation of moments of arbitrarily large order; the task is always tedious and frequently technically demanding. The technical difficulties that can now be avoided are exemplified in Arratia, Gordon and Waterman (1986). The Chen–Stein method allows for an easier treatment of the same problem that leads to stronger results with no extra work.

In Arratia, Gordon and Waterman (1990), we study a more general version of the following problem.

Let $A_1, \cdots, A_n, \cdots$ and $B_1, \cdots, B_n, \cdots$ be independently chosen according to the same common distribution $\{\mu_l\}$ from a common alphabet $\{1, 2, \cdots, d\}$. Choose a test value t and compute

$$M_n(t) = \max_{1 \le i,j \le n-t+1} \sum_{k=0}^{t-1} \mathbf{1}\{A_{i+k} = B_{j+k}\}, \tag{30}$$

the largest number of matches witnessed by any comparison of length t substrings. What is the distribution of $M_n(t)$?

An asymptotic analysis is possible using the Chen–Stein method. Effectively, we rigorously use Aldous' (1989) Poisson clumping heuristic to obtain rates of convergence for the Erdös–Rényi strong law, with a two-dimensional index set. While a proof is beyond the scope of this paper, we can easily guess the ultimate result.

Write $\alpha = (i, j)$, and write $S_\alpha = \sum_{k=1}^{t} \mathbf{1}\{A_{k+i} = B_{k+j}\}$. Choose $s \le t$ and let $Y_\alpha = \mathbf{1}\{S_\alpha \ge s\}$. The special case $s = t$ corresponds to perfect matching, which is similar to the case of perfect head runs dealt with in Section 4.2. We have $P\{M_n(t) < s\} = P\{\sum Y_\alpha = 0\}$. Denote by $p = \sum \mu_l^2$ the probability of seeing a match between two arbitrarily selected letters. Each S_α is distributed as binomial(t, p), and there are

$(n-t+1)^2$ possible index pairs among two sequences, each of length n. There is only local dependence among the Y_α; if $\alpha = (i, j)$ and $\alpha' = (i', j')$, then Y_α and $Y_{\alpha'}$ are independent whenever $|i - i'|$ and $|j - j'|$ both exceed t. Hence the natural neighborhoods of dependence used in computing the Chen–Stein bounds will ensure that $b_3 = 0$. However, the intuition that $\sum_\alpha Y_\alpha$ admit a Poisson approximation is incorrect; as in the case of head runs in Section 4.2, there is a substantial clumping which would result in an excessively large value of b_2. In order to succeed, we must declump, performing the analysis instead on the modified indicators $X_{(i,j)} = Y_{(i,j)} \prod_{k=1}^{t} (1 - Y_{i-k,j-k})$. The indicators X have negative correlations, ensuring a small b_2. The price paid for negative correlation is an extended—but still finite—neighborhood of dependence, and a less obvious explicit representation for EX_α. The latter is obtained with the help of the ballot theorem. To make the method work at all, we also need to impose a weak technical condition on the alphabet probabilities $\{\mu_l\}$ to deal with dependence engendered by pairs α, α' where $i - i' \neq j - j'$. The net result is

$$P\{M_n(t_n) < s\} - \exp\left(-\left(\frac{s}{t_n} - p\right) n^2 P\{\text{binomial}(t_n, p) \geq s\}\right) \to 0, \tag{31}$$

as $n \to \infty$ whenever $t_n/\ln(n^2) \to c > 1/\ln(1/p)$ and $c < \infty$. Loosely interpreted, there are about n^2 blocks of length t requiring comparison. The ballot theorem and declumping reduces the effective number of comparisons to $(s/t_n - p)n^2$. Multiplying by the indicated binomial probability gives an "expectation." The Poisson probability of seeing zero events gives us a distribution function.

Finally we present in Figure 1 a test of the applicability of our results. This graph portrays data taken from a larger experiment whose results are reported in Arratia, Gordon and Waterman (1990). Calculations are done as if the sequences studied were generated independently with letters from the alphabet $\{a, c, g, t\}$ with probabilities $\mu = (.3544, .1430, .1451, .3575)$ and $p = \sum \mu_l^2 = .2949$. The left-hand glyph summarizes 200 simulations with $n = 512$ and $t = 21$ in which the assumptions concerning frequencies and independence are exactly obeyed. The random number generator used was that provided by the MATLAB package, in which all programming was done on a Sun 3/260 computer (see Moler, Ullman, Little and Bangert, 1987).

The probabilities $\{\mu_l\}$ are determined by the incidence of base pairs in the subject of the second glyph—the complete chloroplast genome of the liverwort *Marchantia polymorpha*, taken from the GenBank database. See Fickett and Burks (1988) for a description of the GenBank database. The complete genome of *Marchantia polymorpha* is there given as a sequence of 121,024 letters from the alphabet $\{a, c, g, t\}$. The genomic sequence was cut into 236 blocks of exactly 512 letters with the remaining letters ignored. The highest number of matches over all 21-segments was again computed for a simple random sample of 200 pairs from the population of all block pairs.

The glyphs themselves are hanging histograms, in which the empirical fraction hangs from the predicted density given by the Poisson approximation (31). The predicted density is shown above a bar, actual observed frequency is shown at the lower part of a bar. For example, the richest matching segments were predicted to have exactly 16 of 21 matches in exactly 57% of the 200 trials. In the simulation experiment displayed on the left, 40% of the simulations were found to have exactly 16 matches in the richest matching 21-segments. In the sample of 200 pairs from blocks of *Marchantia polymorpha*, 32% were found to have had 16 matches in the richest matching pair of 21-segments.

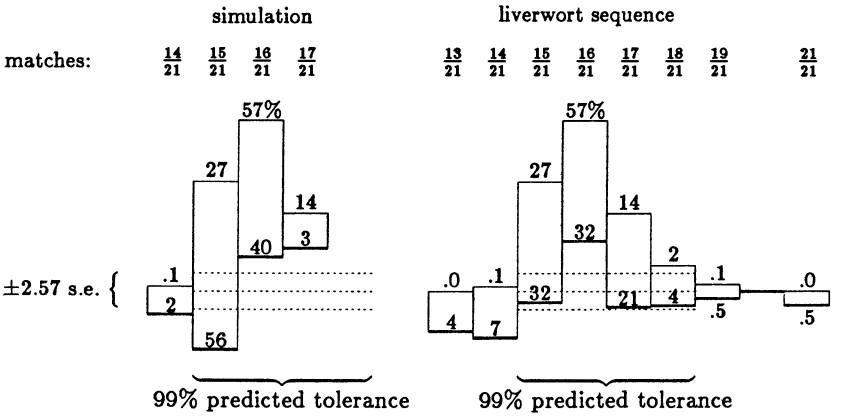

FIG. 1. *Maximum matches between 21-segments in 200 pairs of 512 letter sequences, simulated and biological. Empirical histogram hanging from predicted histogram after arcsine transformation.*

Relative frequencies on a base of 200 are plotted after arcsine transformation in order to stabilize variance. The horizontal dotted lines lie at heights 0 and ±2.56 standard errors, giving barwise 99% confidence limits. For example, in the rightmost glyph, 27% of 200 pairs were predicted to have had 15 matches in their richest matching 21-segments; 32% were observed. The lower end of the bar is plotted at relative height $\arcsin(\sqrt{.271}) - \arcsin(\sqrt{64/200})$, corresponding to $(.5475 - .6013)/.0354 = -1.52$ standard error units of size $1/(2\sqrt{200})$. Tolerance intervals based on (31) are given by the horizontal extent of the dotted lines. These cover horizontally at least 99% of the predicted distribution, from the lower .005 quantile to the upper .995 quantile. Were the approximate distribution exact, one would expect fewer than two simulated values to exceed the tolerance limits in either glyph. It is remarkable, given the known dependence of adjacent nucleotides, that predictions based on assumptions of iid generation of sequences should fit as well as they do. For an analysis of dependence among adjacent nucleotides, see Tavaré and Giddings (1989).

The empirical distribution is less concentrated than the simulated distribution, no doubt attributable to departures from distributional assumptions. For example, slow systematic variation in fraction of letters represented in a section of a genome could result in a more spread out distribution than the Poisson approximation would predict.

Interestingly, the outlying comparison—a perfect match of 21/21 with approximate p-value $< .00006$—has biological implications. The outlier corresponds to a perfect match beginning at nucleotide 26,665 and at nucleotide 67,475. Ohyama, Fukuzawa and Kohchi (1986) study the chloroplast gene organization using the complete DNA sequence. From Ohyama, the first segment is located in an intervening sequence within the gene coding a tRNA for lysine, while the second segment is located in an intervening sequence within open reading frame ORF203. Higher organisms have gene (protein encoding) DNA sequences interrupted by so-called intervening sequences which are removed from transcribed RNA by a mechanism known as splicing. There is as yet no consensus regarding the biological role of intervening sequences. Open reading frames indicate regions of DNA that could encode proteins, although it has not yet been determined whether the region's DNA is actually translated into proteins.

ACKNOWLEDGMENTS

We are grateful to Persi Diaconis who first brought the Chen–Stein method to our attention, to Michael Waterman, whose DNA sequence matching questions motivated our interest in Poisson approximation and to Samuel Karlin, whose questions about the locations of matching regions led us to look for process versions of Poisson approximation theorems using the Chen–Stein method. Supported by NSF Grant DMS-88-15106 and NIH Grant 5 R01 GM36230-04.

REFERENCES

ABRAMOWITZ, M. and STEGUN, I. A. (1964). *Handbook of Mathematical Functions*. U.S. Department of Commerce, Washington, D.C.

ALDOUS, D. (1989). *Probability Approximations via the Poisson Clumping Heuristic. Applied Mathematical Sciences* **77**. Springer, New York.

ARRATIA, R., GOLDSTEIN, L. and GORDON, L. (1989). Two moments suffice for Poisson approximations: The Chen–Stein method. *Ann. Probab.* **17** 9–25.

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. (1990). The Erdös–Rényi law in distribution, for coin tossing and sequence matching. *Ann. Statist.* **18** 539–570.

ARRATIA, R. and TAVARÉ, S. (1990). The cycle structure of random permutations. Preprint.

BALDI, P. and RINOTT, Y. (1989). On normal approximations of distributions in terms of dependency graphs. *Ann. Probab.* **17** 1646–1650.

BALDI, P., RINOTT, Y. and STEIN, C. (1989). A normal approximation for the number of local maxima of a random function on a graph. In *Probability, Statistics and Mathematics: Papers in Honor of Samuel Karlin* (T. W. Anderson, K. B. Athreya and D. L. Iglehart, eds.) 59–81. Academic, New York.

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

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

BARBOUR, A. D. (1990). Stein's method for diffusion approximations. *Probab. Theory Related Fields* **84** 297–322.

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 EAGLESON, G. K. (1985). Multiple comparisons and sums of dissociated random variables. *Adv. in Appl. Probab.* **17** 147–162.

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

BARBOUR, A. D. and HALL, P. (1984b). Reversing the Berry–Esseen inequality. *Proc. Amer. Math. Soc.* **90** 107–110.

BARBOUR, A. D. and HOLST, L. (1989). Some applications of the Stein–Chen method for proving Poisson convergence. *Adv. in Appl. Probab.* **21** 74–90.

BARBOUR, A. D., HOLST, L. and JANSON, S. (1988a). Poisson approximation in occupancy problems. Preprint.

BARBOUR, A. D., HOLST, L. and JANSON, S. (1988b). Poisson approximation with the Stein–Chen method and coupling. Preprint.

BIRNBAUM, Z. W. (1942). An inequality for Mill's ratio. *Ann. Math. Statist.* **13** 245–246.

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

BOLTHAUSEN, E. (1984). An estimate of the remainder in a combinatorial central limit theorem. *Z. Wahrsch. Verw. Gebiete* **66** 379–386.

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). Inadmissability 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 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 $S_n = X_1 + X_2 + \cdots + X_n$ in $\mathbf{R}^2$ where the X_i are iid. To make life as

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