The Annals of Probability
1993, Vol. 21, No. 4, 2269–2279

BOOK REVIEW

D. ALDOUS, *Probability Approximations via the Poisson Clumping Heuristic*. Springer, Berlin, 1989.

A. D. BARBOUR, L. HOLST AND S. JANSON, *Poisson Approximation*. Oxford University Press, 1992.

REVIEW BY RICHARD ARRATIA AND SIMON TAVARÉ

University of Southern California

It might be that the two fundamental distributions in probability are the normal and the Poisson. Making these into processes with independent increments yields Brownian motions, Brownian sheets, ... and Poisson processes. Nevertheless, graduate training in probability tends to ignore Poisson processes—and none of the almost 200 entries in the Probability and Statistics category of the 1991 Mathematics Subject Classification explicitly mentions Poisson, either. Brownian motion is a fascinating object in its own right, even without considering its relation to martingales, diffusions, stochastic differential equations, and so on. In contrast, it is only the comparison of Poisson processes to various dependent processes that makes for rewarding study. Two recent books, *Probability Approximations via the Poisson Clumping Heuristic* (PCH) and *Poisson Approximation* (PA), combine to reveal the enormous depth and complexity of applications of the Poisson distribution and Poisson processes. These books are complementary. In the language of Breiman (1968), PCH addresses the left hand, intuition, and PA addresses the right hand, technique.

Poisson approximations for a given probability model involve three ingredients. The first, most easily overlooked, is the identification of suitable things to count, so that the random number W of occurrences is approximately Poisson. The second is an evaluation of, or approximation for, the Poisson parameter $\lambda = \mathbb{E}W$. The third ingredient is an analysis of the dependence structure to show that W is close to $\mathrm{Po}(\lambda)$, the Poisson distribution with parameter λ. Informally, PCH discusses the first two ingredients, and PA is concerned with the third.

Overview of PCH. Section VI.7 of Feller (1968) is titled "Observations fitting the Poisson distribution" and its footnote states:

> The Poisson distribution has become known as the law of small numbers or of rare events. These are misnomers which proved detrimental to the realization of the fundamental role of the Poisson distribution. The following examples will show how misleading the two names are.

Received February 1993

The five examples are (a) radioactive disintegrations, (b) flying-bomb hits on London, (c) chromosome interchanges in cells, (d) connections to wrong number, and (e) bacteria and blood counts. Contrast this to the opening paragraph of PCH, which includes the following problems.

> If you place a large number of points randomly in the unit square, what is the distribution of the radius of the largest circle containing no points?... Why do Brownian motion sample paths have local maxima but not points of increase, and how nearly do they have points of increase?... If an imaginary particle performs a simple random walk on the vertices of a high-dimensional cube, how long does it take to visit every vertex? If a particle moves under the influence of a potential field and random perturbations of velocity, how long does it take to escape from a deep potential well?... If you take a large i.i.d. sample from a 2-dimensional rotationally-invariant distribution, what is the maximum over all half-spaces of the deviation between the empirical and true distributions?

All of these problems may be posed as questions about the distribution of the number of occurrences of some rare events and may be answered using a Poisson approximation. In many cases, including Feller's examples, the rare events are close to independent, so that the count of occurrences is approximately Poisson. In many other situations, including most of Aldous's examples, the rare events are far from mutually independent, tending rather to occur in clumps. It is the *clumps*, suitably defined, that are approximately independent of each other, so that the count of clumps is approximately Poisson. The great service rendered by PCH is to reveal the true breadth and scope of this idea. Focussing on the clumps is the crucial first step in applying a Poisson approximation in these situations. Every probabilist should at least browse PCH to absorb this point of view.

A second theme of PCH is evaluation of the Poisson parameter for situations with substantial clumping. To find the intensity λ of clumps, divide the intensity of occurrences by the expected size of a clump:

$$\lambda = \frac{\text{rate of occurrence}}{\text{mean number of occurrences per clump}}. \tag{1}$$

The top of the fraction is usually easy to calculate; the art is in estimating the bottom. (The situations without clumping are those in which the denominator is close to 1.) As Aldous describes it, "This mean clump size can be estimated by approximating the underlying random process locally by a simpler, known process for which explicit calculations are possible." In pursuit of these explicit calculations, a remarkable amount of probability theory is covered, albeit at an intuitive level. Regardless of the connection to Poisson approximation, which might be viewed as just an excuse to discuss a host of interesting processes, the ideas surveyed while calculating clump sizes are fascinating in their own right.

We suspect that this is the aspect of PCH that led Williams (1991) to state "Of course, intuition is much more important than knowledge of measure theory, and you should take every opportunity to sharpen your intuition. There is no better whetstone for this than Aldous (1989), though it is a very demanding book."

A third theme of PCH is an ideal for good probabilistic research. In the introduction, one finds:

> The proper business of probabilists is calculating probabilities. ... A limit theorem is an assertion of the form: 'the error in a certain approximation tends to 0 as (say $N \to \infty$). Call such limit theorem naive if there is no explicit error bound ...

A colleague expanded this to finish "... and call people who prove such theorems ...". The alternative to proving naive limit theorems is to give approximations with bounds. The book PA shows how to do this.

The expository style of PCH is unconventional. Although the book is very entertaining, and would be a good companion on a desert island, it is also frustrating. Many ideas are left raw, and in spite of the author's notes on the literature, it is hard to tell what has been proved and what hasn't.

Examples of clumping. We believe that the simplest example of the Poisson clumping heuristic, with nontrivial clumping, is the analysis of the length of the longest head run in a sequence of coin tosses. Consider a p-coin tossed repeatedly, and ask what is the distribution of L, the length of the longest consecutive run of heads within the first n tosses, for large n. For simplicity, we will ignore boundary effects throughout this discussion. Pick a positive integer test value t for L. At each epoch, the probability of t consecutive heads is p^t, so the expected number of occurrences of blocks of t consecutive heads is np^t. However, these occurrences come in clumps. A clump here may be defined as a maximal consecutive run of occurrences. For example, a tail followed by $t + 2$ heads followed by a tail should be viewed as one clump, consisting of three occurrences. To compute the mean number of occurrences per clump, follow a given clump back to its leftmost end. Since the chance that the clump contains more than k occurrences is p^k, the expected number of occurrences in a clump is $1 + p + p^2 + \cdots$. Thus, using (1), the number W of clumps has expectation

$$\lambda = \frac{np^t}{1 + p + p^2 + \cdots} = n(1 - p)p^t,$$

and the Poisson clumping heuristic predicts that

$$\mathbb{P}(L < t) \approx \mathbb{P}(W = 0) \approx \exp(-\lambda) = \exp(-n(1 - p)p^t).$$

A treatment of this with error bounds, both with and without the declumping factor of $(1 - p)$, appears in Section 8.4 of PA.

The appropriate definition of a clump is subject to taste. Another reasonable choice would be to identify clumps with renewals, so that in testing for t consecutive heads, a tail followed by $3t + 1$ heads followed by a tail would be counted as three clumps rather than as one clump.

In the previous example it is easy to label the clumps of occurrences, in terms of the index of the first occurrence. The event E_i that there is a clump at i may be defined as the event that t consecutive heads occur at i, and not at $i - 1$, so that $\mathbb{P}(E_i) = \mathbb{P}(\text{tail followed by } t \text{ heads}) = (1 - p)p^t$, and $W = \sum_1^n \mathbf{1}(E_i)$. One could analyze long head runs without noticing the phenomenon of clumping, either by using the events E_i or via renewal theory, which provides exact formulae.

The analysis of coverage of the plane by randomly centered discs gives an example of the clumping heuristic in which it does not seem feasible to label clumps by the index of a particular occurrence within the clump. Consider the unit square with a large number θ of points placed uniformly and independently, and ask what is the distribution of L, the radius of the largest circle containing no points. To simplify the analysis, pretend that the number of points is also randomized, to be Poisson with mean θ, so that the given set of points is a Poisson process with intensity θ. Ignore boundary effects, by treating the unit square as a torus. Choose a test value r for L, so that the event that $L \le r$ is exactly the event that circles of radius r, with centers at our random points, completely cover the square. The uncovered part of the unit square is a random set $\mathscr{S}$ containing all points x such that the disc of radius r centered at x contains no points of the original Poisson(θ) point process. This random set $\mathscr{S}$ consists of clumps, since given that a point x is in $\mathscr{S}$, points very nearby are also likely to be in $\mathscr{S}$. Roughly, the clumps are the connected components of $\mathscr{S}$. The expected area of $\mathscr{S}$ is easy to calculate, being exactly $\mathbb{P}(x \in \mathscr{S}) = \mathbb{P}(\text{the original random set misses a given disc of radius } r) = \exp(-\theta\pi r^2)$. An argument involving the Poisson line process, as in Solomon (1978), predicts that the expected size of a clump is approximately $1/(\pi r^2\theta^2)$. Thus in (1) the number W of clumps has expectation

$$\lambda \approx \pi\theta^2 r^2 \exp(-\theta\pi r^2),$$

and the Poisson clumping heuristic predicts that

$$(2) \qquad \mathbb{P}(L \le r) \approx \mathbb{P}(W = 0) \approx \exp(-\lambda) \approx \exp\bigl(-\pi\theta^2 r^2 \exp(-\theta\pi r^2)\bigr).$$

In (2), the first $\approx$ reflects errors due to ignoring boundary effects and replacing a deterministic number θ of points by a random number. The second $\approx$ appears because W is not Poisson, and the third appears because λ has only been approximated. The sense of approximation meant by $\approx$ is deliberately vague, but implicit is the claim that for large θ, and r chosen so that the expected number λ of clumps is moderate, the overall approximation for $\mathbb{P}(L \le r)$ is good, so that (2) does describe the approximate distribution of L.

Establishing Poisson approximations: a historical view. Convergence in distribution to the Poisson has often been proved by establishing the appropriate limit for Laplace transforms, or by the method of moments. Let

W_n be nonnegative random variables, and let Z be Po(λ). For Laplace transforms, recall that W_n converges in distribution to Z if and only if $\log \mathbb{E}\exp(-sW_n) \to \lambda(e^{-s} - 1)$ for $s \geq 0$. For the method of moments, recall that W_n converges in distribution to Z if for $k = 1, 2, \ldots$ the kth moment of W_n converges to the kth moment of Z. Establishing convergence of all moments is equivalent to showing that the kth falling factorial moment of W_n converges to λ^k for each k.

These methods establish "naive" limit theorems. To get a rate of convergence for $\mathbb{P}(W_n = j)$ to $e^{-\lambda}\lambda^j/j!$ as $n \to \infty$, for each fixed $j = 0, 1, 2, \ldots$, inclusion–exclusion is a natural tool [Watson (1954)] that provides both upper bounds and asymptotics in many examples. However, for the total variation distance from W_n to Z, which involves summing over $j = 0, 1, \ldots$, the bounds obtained using inclusion–exclusion are usually intractable.

Inclusion–exclusion for Poisson approximation is described under the heading of sieve methods by Alon and Spencer [(1992), Chapter 8.3] and Bollobás [(1985), Chapter 1.4]. For a random variable $W \equiv \sum_{\alpha \in \Gamma} \mathbf{1}(E_\alpha)$ that counts the number of events E_α that occur,

$$\mathbb{P}(W = 0) = \sum_{k \geq 0} (-1)^k \sum_{J \subset \Gamma, |J| = k} \mathbb{P}\Big(\bigcap_J E_\alpha \Big). \tag{3}$$

This follows from the identity

$$\mathbf{1}(w = 0) = \sum_{k \geq 0} (-1)^k \binom{w}{k}, \qquad w \in \mathbb{Z}^+, \tag{4}$$

which can be applied to any nonnegative integer-valued random variable. The connection between (3) and (4) is that the falling factorial may be expanded as $(W)_k = \sum_* \mathbf{1}(E_{\alpha_1}) \cdots \mathbf{1}(E_{\alpha_k})$, where $\sum_*$ denotes a sum taken over ordered k-tuples $(\alpha_1, \ldots, \alpha_k)$ of distinct indices from Γ, so that

$$\binom{W}{k} = \sum_{J \subset \Gamma, |J| = k} \mathbf{1}(\cap_J E_\alpha).$$

The Bonferroni inequalities sharpen the inclusion–exclusion formula by noting that the partial sums provide upper and lower bounds. In terms of the falling factorial moments $\mathbb{E}(W)_k$ of an arbitrary $\mathbb{Z}^+$-valued random variable, we have upper and lower bounds on $\mathbb{P}(W = 0)$ given by $\sum_{k=0}^{r} (-1)^k \mathbb{E}(W)_k/k!$, with the direction of the bound depending only on the parity of r. Likewise, the partial sums $\sum_{k=0}^{r} (-\lambda)^k/k!$ provide upper and lower bounds for $\exp(-\lambda)$. Let $d_k \equiv \mathbb{E}(W)_k - \lambda^k$ be the discrepancy between the kth falling factorial moment of W and that of the Poisson distribution with parameter λ. Combining the upper and lower bounds, for r even we have

$$\sum_0^{r-1} (-1)^k \frac{d_k}{k!} - \frac{\lambda^r}{r!} \leq \mathbb{P}(W = 0) - e^{-\lambda} \leq \sum_0^{r} (-1)^k \frac{d_k}{k!} + \frac{\lambda^{r+1}}{(r+1)!}.$$

In summary, to show that $\mathbb{P}(W = 0)$ is close to $\exp(-\lambda)$ via inclusion–exclusion, it suffices to pick r large enough that $\lambda^r/r!$ is small, and then control all moments of order up to r.

For $j = 0, 1, 2, \ldots$, the difference between $\mathbb{P}(W = j)$ and $e^{-\lambda}\lambda^j/j!$ can be bounded by similar considerations applied to the general inclusion–exclusion identity:

$$\mathbf{1}(w = j) = \sum_{k \geq 0} (-1)^k \frac{(w)_{k+j}}{j!k!}, \qquad w \in \mathbb{Z}^+.$$

Stein's method. That a random variable Z has the distribution $\mathrm{Po}(\lambda)$ is characterized by the recursion

$$\lambda \mathbb{P}(Z = k - 1) = k\mathbb{P}(Z = k), \qquad k = 0, 1, 2, \ldots .$$

The basis of Stein's method [Stein (1972)], as applied to Poisson approximation by Chen (1975), is to extend this from indicators of singletons, $g(z) = \mathbf{1}(z = k)$, to arbitrary bounded $g\colon \mathbb{Z}^+ \to \mathbb{R}$. Thus Z has distribution $\mathrm{Po}(\lambda)$ if and only if for all bounded g

$$\mathbb{E}(\lambda g(Z + 1) - Zg(Z)) = 0. \tag{5}$$

Stein's insight is that if $\mathbb{E}(\lambda g(W + 1) - Wg(W))$ is small for enough functions g, then the random variable W is close in distribution to $\mathrm{Po}(\lambda)$.

The mechanism for carrying this out is to write (5) as $\mathbb{E} f(Z) = 0$, where f and g are related by *Stein's equation*,

$$f(k) = \lambda g(k + 1) - kg(k), \qquad k = 1, 2, \ldots . \tag{6}$$

Given $f = (f(0), f(1), \ldots)$ there is a unique $g = (g(1), g(2), \ldots)$ which satisfies (6), namely,

$$g(j) = \frac{\mathbb{E}(f(Z)\mathbf{1}(Z < j))}{\lambda \mathbb{P}(Z = j - 1)},$$

where henceforth we write Z for a random variable having the distribution $\mathrm{Po}(\lambda)$. For subsets $A \subset \mathbb{Z}^+$, let g_A denote the solution of Stein's equation using $f(\cdot) = \mathbf{1}_A(\cdot) - \mathbb{P}(Z \in A)$, so that

$$\mathbb{P}(W \in A) - \mathbb{P}(Z \in A) = \mathbb{E} f(W) = \mathbb{E}(\lambda g_A(W + 1) - Wg_A(W)).$$

It follows that the total variation distance from the distribution of W to $\mathrm{Po}(\lambda)$ is given by

$$d_{TV}(W, Z) \equiv \sup_{A \subset \mathbb{Z}^+} |\mathbb{P}(W \in A) - \mathbb{P}(Z \in A)| \tag{7}$$

$$= \sup_A |\mathbb{E}(\lambda g_A(W + 1) - Wg_A(W))|. \tag{8}$$

We will return to a discussion of (8) in the context of PA. As a word of caution, we note that some authors define total variation distance as $d_{TV}(\mu, \nu) \equiv \sup_{\|h\| = 1} |\int h\, d\mu - \int h\, d\nu|$, which has values twice as big as the d_{TV} given by (7).

Overview of PA. PA is devoted to providing bounds for Poisson approximations. The book features both general theory for Poisson random variables and processes, and concrete examples. The theory includes upper and lower bounds, often with explicit universal constants. The examples are given as chapters on the following: random permutations; random graphs; occupancy and urn models; spacings; and exceedances and extremes. The expository style of PA is rigorous but unforgiving. It satisfies an ideal for research papers in this journal: "Every sentence should challenge the reader."

The main theme of PA is the development of upper bounds for the distance to the Poisson distribution using Stein's method and coupling techniques. Among the choices of metric are the total variation distance and the Wasserstein distance. Later sections of this review discuss some of the details of this. For now, we quote two theorems that illustrate the typical style of the results of PA. Theorem 1.C compares the mixed Poisson distribution $\mathrm{Po}(\Lambda)$, governed by a nonnegative random variable Λ [i.e., $\mathbb{P}(W = k|\Lambda = x) = \exp(-x)x^k/k!$], and the Poisson distribution with parameter λ. The proof using Stein's method is elegant. The result is a pair of upper bounds on the total variation distance:

(i) $d_{TV}(\mathrm{Po}(\Lambda), \mathrm{Po}(\lambda)) \le \min(1, \lambda^{-1/2})\mathbb{E}|\Lambda - \lambda|$

(ii) $d_{TV}(\mathrm{Po}(\Lambda), \mathrm{Po}(\lambda)) \le \lambda^{-1}(1 - \exp(-\lambda))\mathrm{Var}\,\Lambda, \quad \text{if } \mathbb{E}\Lambda = \lambda.$

The factor $\lambda^{-1}(1 - \exp(-\lambda))$ is less than $\min(1, \lambda^{-1})$.

The second example is Theorem 8.H, which looks at rare sets in a Markov chain. Suppose X is a stationary ergodic Markov chain on $\mathbb{Z}$, with j-step transition function $P^{(j)}$ and stationary measure μ. If W is the number of visits to the set A in the first n steps, so that $\lambda = \mathbb{E}W = n\mu(A)$, then

$$(9)\quad d_{TV}(W, \mathrm{Po}(\lambda)) \le (1 - e^{-\lambda})\left(\mu(A) + \frac{2}{\mu(A)} \sum_{r,s \in A} \mu_r \sum_{j \ge 1} |P^{(j)}_{rs} - \mu_s|\right).$$

Section B12 of PCH discusses rare sets in Markov chains and gives a heuristic for computing the expected size of a clump of visits to that set. We note that the bound in (9) provides a useful approximation only in the situation in which the expected clump size is close to 1.

To assess the adequacy of upper bounds, one tries to find lower bounds of comparable order. Chapter 3 provides tools for this, using both Stein's method and various ad hoc techniques. To give the flavor of these bounds, we quote the simplest example, where the variance is less than the mean. Let δ be the total variation distance between W and $\mathrm{Po}(\lambda)$, where $\lambda = \mathbb{E}W$. The authors observe that the upper bound given by Stein's method is often of order ε, where

$$\varepsilon \equiv \min(1, \mathbb{E}W)\left|\frac{\mathrm{Var}\,W}{\mathbb{E}W} - 1\right|.$$

Theorem 3.A(b) provides a lower bound on the total variation distance δ, for

the case Var $W < \lambda = \mathbb{E}W$, in the form

$$\delta \geq c\varepsilon\left(1 + \log\left(\frac{1}{\varepsilon}\right)\right)^{-1}\left(1 \wedge (\lambda \vee 1)\left(1 + \log\left(\frac{1}{\varepsilon}\right)\right)^{-1}\right),$$

for some universal constant c. Even for the reader who is daunted by such details, there is an important lesson to be learned: When the variance is below and bounded away from the mean, there is no possibility of Poisson convergence.

The first nine chapters of PA deal with one-dimensional Poisson approximations, and Chapter 10, which is the most technical part of the book, extends these ideas to process approximations. The enormous scope of this project may be appreciated by noting an analogy: The one-dimensional central limit theorem extends to Donsker's invariance principle, martingale central limit theorems, and so on. Viewing $W = \sum_{\alpha \in \Gamma} I_\alpha$ as one particular functional, the sum, applied to a process $(I_\alpha, \alpha \in \Gamma)$, it is natural to consider arbitrary functionals, that is, to consider approximations to the distribution of the entire dependent process of indicators. PA develops results using coupling and a wide variety of metrics. The simplest form of process approximation, using total variation distance and the "local method," may be found in Arratia, Goldstein and Gordon (1989, 1990). An appealing feature of giving total variation upper bounds on the process is that exactly the same bounds apply to arbitrary functionals.

More on Stein's method. Techniques for expanding and bounding (8) occupy most of PA. The straightforward ingredients are the bounds, valid uniformly in A,

$$\|g_A\| \equiv \sup_{k \geq 1} |g_A(k)| \leq \min(1, \lambda^{-1/2}) \tag{10}$$

and

$$\|\Delta g_A\| \equiv \sup_{k \geq 1} |g_A(k+1) - g_A(k)| \leq \frac{1 - e^{-\lambda}}{\lambda} \leq \min(1, \lambda^{-1}). \tag{11}$$

[In earlier papers on Poisson approximation using Stein's method, upper bounds such as $\min(1, 1.4\lambda^{-1/2})$ appear, corresponding to earlier versions of the bound in (10); it should be a matter of taste whether to quote these verbatim with the original bound or simply to insert the improved bound expressed by (10).]

We express W as a sum of indicators of dependent events, indexed by an arbitrary set Γ, so that $W = \sum_{\alpha \in \Gamma} I_\alpha$. There are many choices of how to proceed in expanding and bounding (8). Henceforth we will write $p_\alpha = \mathbb{P}(I_\alpha = 1) = 1 - \mathbb{P}(I_\alpha = 0)$ and assume that $\lambda = \sum p_\alpha < \infty$, so that $\lambda = \mathbb{E}W$.

An easy method, referred to in PA as the local method [proposed originally by Chen (1975)], is to assign "neighborhoods of strong dependence" $B_\alpha \subset \Gamma$, with $\alpha \in B_\alpha$. (PA uses the notation $\Gamma_\alpha^s \equiv B_\alpha - \{\alpha\}$.) Using the notation of Arratia, Goldstein and Gordon (1989), the total variation distance from W to $\mathrm{Po}(\lambda)$ can be bounded above in terms of three quantities, b_1, b_2, b_3', which

must all be small for a useful approximation. First,

$$b_1 = \sum_{\alpha\in\Gamma} p_\alpha \sum_{\beta\in B_\alpha} p_\beta$$

reflects the size of the neighborhoods. Second,

$$b_2 = \sum_{\alpha\in\Gamma} \sum_{\alpha\neq\beta\in B_\alpha} \mathbb{E}(I_\alpha I_\beta) = \sum_{\alpha\in\Gamma} p_\alpha \sum_{\alpha\neq\beta\in B_\alpha} \mathbb{E}(I_\beta | I_\alpha = 1)$$

expresses the positive correlations captured within the neighborhoods. One can view b_2 as the average number of occurrences among neighbors, given the occurrence of an event. Note that both b_1 and b_2 become smaller if the neighborhoods B_α are chosen smaller. The third term is

$$b_3' = \sum_{\alpha\in\Gamma} s_\alpha', \quad \text{where } s_\alpha' = \mathbb{E}\left|\mathbb{E}\left\{I_\alpha - p_\alpha \middle| \sum_{\beta\in\Gamma - B_\alpha} I_\beta\right\}\right|,$$

which penalizes the choice of too small a neighborhood. The resulting bound is

$$d_{TV}(W, Z) \leq (b_1 + b_2)\frac{1 - e^{-\lambda}}{\lambda} + b_3' \min(1, \lambda^{-1/2}), \tag{12}$$

where the factors that depend on λ come from (10) and (11). The natural choice of neighborhoods in many examples is the smallest one that makes each I_α independent of $\sigma(I_\beta\colon \beta \in \Gamma - B_\alpha)$, so that $b_3' = 0$.

An alternative to the local method for bounding and expanding (8) is the "coupling method," which is the main focus of PA. A simple version of this is expressed by Theorem 2.A in the form

$$d_{TV}(W, Z) \leq \frac{1 - e^{-\lambda}}{\lambda} \sum_{\alpha\in\Gamma} p_\alpha \mathbb{E}|U_\alpha - V_\alpha|, \tag{13}$$

where (U_α, V_α) is any coupling with $U_\alpha =_d W$ and $V_\alpha =_d (W - 1 | I_\alpha = 1)$. Much of the fun and art of this method is in constructing couplings.

Comparing the local method and the coupling method. We can compare the bounds achievable by the local method and the coupling method, as follows. Consider the local method used with the smallest possible neighborhoods, that is, $B_\alpha = \{\alpha\}$, so that $b_2 = 0$. The term b_1 simplifies to $b_1 = \Sigma_\alpha p_\alpha^2$. The mutual dependence of the I_α is measured entirely by b_3'. Write $W_\alpha = W - I_\alpha$, so that $b_3' = \Sigma_\alpha s_\alpha'$, with

$$\begin{aligned}
s_\alpha' &= \mathbb{E}|\mathbb{E}(I_\alpha - p_\alpha | W_\alpha)| \\
&= \sum_k \mathbb{P}(W_\alpha = k)|\mathbb{P}(I_\alpha = 1 | W_\alpha = k) - p_\alpha| \\
&= \sum_k |\mathbb{P}(I_\alpha = 1, W_\alpha = k) - p_\alpha \mathbb{P}(W_\alpha = k)| \\
&= \sum_k p_\alpha |\mathbb{P}(W_\alpha = k | I_\alpha = 1) - \mathbb{P}(W_\alpha = k)| \\
&= 2p_\alpha d_{TV}(W_\alpha, (W_\alpha | I_\alpha = 1)).
\end{aligned}$$

In the notation of the coupling method, $V_\alpha =_{\mathrm{d}} (W - 1 | I_\alpha = 1) =_{\mathrm{d}} (W_\alpha | I_\alpha = 1)$, so that

$$b'_3 = 2 \sum_\alpha p_\alpha d_{TV}(W_\alpha, V_\alpha).$$

The net result of the local method, using $B_\alpha = \{\alpha\}$, is

$$(14) \quad d_{TV}(W, Z) \le \frac{1 - e^{-\lambda}}{\lambda} \sum_\alpha p_\alpha^2 + 2 \min(1, \lambda^{-1/2}) \sum_\alpha p_\alpha d_{TV}(W_\alpha, V_\alpha).$$

Next consider the Wasserstein distance $d_W(X, Y)$, defined as the minimum of $\mathbb{E}|X - Y|$ over all couplings. For integer-valued random variables, $d_{TV} \le d_W$, since

$$\begin{aligned} d_{TV}(X, Y) &= \min_{\text{couplings}} \mathbb{P}(X \ne Y) = \min_{\text{couplings}} \mathbb{E}\mathbf{1}(X \ne Y) \\ &\le \min_{\text{couplings}} \mathbb{E}|X - Y| \equiv d_W(X, Y). \end{aligned}$$

The minimal value of the expression $\mathbb{E}|U_\alpha - V_\alpha|$ in (13) is precisely $d_W(W, V_\alpha)$. The bound from the coupling method in (13) can therefore be rewritten as

$$(15) \qquad d_{TV}(W, Z) \le \frac{1 - e^{-\lambda}}{\lambda} \sum_\alpha p_\alpha d_W(W, V_\alpha).$$

One of the differences between (14) and (15) is the appearance of W_α in place of W. To bound the effect of this change, observe that $W = W_\alpha + I_\alpha \ge W_\alpha$, so that stochastic monotonicity implies $d_W(W, W_\alpha) = p_\alpha$. Hence, by the triangle inequality,

$$|d_W(W, V_\alpha) - d_W(W_\alpha, V_\alpha)| \le p_\alpha.$$

Thus, to within $\pm \lambda^{-1}(1 - e^{-\lambda}) \sum_\alpha p_\alpha^2$, which is usually negligible, the local method has the upper bound

$$(16) \qquad 2 \min(1, \lambda^{-1/2}) \sum_\alpha p_\alpha d_{TV}(W_\alpha, V_\alpha)$$

and the coupling method has the upper bound

$$(17) \qquad \frac{1 - e^{-\lambda}}{\lambda} \sum_\alpha p_\alpha d_W(W_\alpha, V_\alpha).$$

Roughly speaking, the coupling method increases $d_{TV}(W_\alpha, V_\alpha)$ to $d_W(W_\alpha, V_\alpha)$, and saves a factor $\max(2, 2\sqrt{\lambda})$ in the coefficient of b'_3.

Discussion. The books reviewed here provide a coherent overview of areas of research that are sure to flourish for many years. Each could make a rewarding basis for a semester or more of graduate-level study. PA does an admirable job of presenting a cohesive toolkit for assessing Poisson approximations, and sets the standard for results and proofs in this area. These methods,

together with the idea of clumping from PCH, deserve to be a part of any applied probabilist's arsenal.

Two frontiers for further research stand out. One is process approximation. The other is the development of new techniques to analyze the myriad examples in PCH where clumping plays a crucial role. The techniques in PA seem applicable only in cases where the clumps are readily identifiable.

An ideal for the style of such research is expressed in the postscript of PCH:

> A mathematical area develops best when it faces hard concrete problems which are not in the 'domain of attraction' of existing proof techniques. An area develops worst along the 'lines of least resistance' in which existing results are slightly generalized or abstracted. I hope this book will discourage theoreticians from the pursuit of minor variations of the known and the formalization of the heuristically obvious, and encourage instead the pursuit of the unknown and the unobvious.

REFERENCES

ALON, N. and SPENCER, J. H. (1992). *The Probabilistic Method*. Wiley, 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., GOLDSTEIN, L. and GORDON, L. (1990). Poisson approximation and the Chen–Stein method. *Statist. Sci.* **5** 403–434.

BOLLOBÁS, B. (1985). *Random Graphs*. Academic, London.

BREIMAN, L. (1968). *Probability*. Addison-Wesley, Reading, MA.

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

FELLER, W. (1968). *An Introduction to Probability Theory and its Applications* **1**, 3d ed. Wiley, New York.

SOLOMON, H. (1978). *Geometric Probability*. SIAM, Philadelphia.

STEIN, C. (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.

STEIN, C. (1986). *Approximate Computation of Expectations*. IMS, Hayward, CA.

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

WILLIAMS, D. (1991). *Probability with Martingales*. Cambridge Univ. Press.

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