Normal approximation for stochastic geometry and allocations

Mathew D. Penrose

University of Bath, UK

Progress in Stein's Method, Singapore, January 2009

Let $B_1, \ldots, B_n$ be interpenetrating unit balls, independently uniformly randomly scattered in a cube C_n of volume n in d-space (with periodic boundary conditions). Define the variables

$$V_n := \text{Volume}\left(\cup_{i=1}^{n} B_i\right),$$

$$S_n := \sum_{i=1}^{n} \mathbf{1}\left\{B_i \cap \left(\cup_{j \neq i} B_j\right) = \emptyset\right\}.$$

Thus V_n is the total volume covered by the balls, and S_n is the number of isolated balls.

S_n is also the number of singletons in a certain *geometric graph.*

See [Hall (1988); Penrose (2003); Stoyan et al. (1995)].

Let θ denote the volume of the unit ball in $\mathbf{R^d}$. As $n \to \infty$,

$$\mathbf{E}V_n \sim n(1 - e^{-\theta}); \qquad \mathbf{E}S_n \sim ne^{-2^d\theta}.$$

One can also show that there are constants c_1, c_2 such that

$$\mathrm{Var}(V_n) \sim c_1 n; \qquad \mathrm{Var}(S_n) \sim c_2 n,$$

and to give formulae for c_1 and c_2. It is not so clear from the formulae for c_1 and c_2 that their values are non-zero for all d and all choices of radius (our choice of unit radius was arbitrary).

Let Φ be the standard normal distribution function. For random X with positive finite standard deviation $\mathrm{SD}(X)$, set

$$D(X) := \sup_{t \in \mathbf{R}} \left\{ \left| P\left[\frac{X - \mathbf{E}X}{\mathrm{SD}(X)} \leq t \right] - \Phi(t) \right| \right\}$$

(i.e., the Kolmogorov dist. between $\mathcal{L}(X)$ and the normal)

As $n \to \infty$ (Moran (1973); P. and Yukich (2001)). we have CLTs:

$$D(V_n) \to 0 \tag{1}$$

$$D(S_n) \to 0 \tag{2}$$

What about the rate of normal approximation in (1) and (2)?

In recent work with Larry Goldstein, we provide explicit Berry-Esséen type error bounds which show that as $n \to \infty$ we have

$$D(V_n) = O(n^{-1/2}); \tag{3}$$

$$D(S_n) = O(n^{-1/2}). \tag{4}$$

Since S_n is integer valued, it is not hard to show that there is a lower bound of the same order for S_n, i.e., that we can change the right hand side of (4) to $\Theta(n^{-1/2})$. The same ought to be true in the case of (3) but we do not have a proof of this.

Chatterjee (2008) obtains similar bounds to (3) and (4) for the Kantorovich-Wasserstein (rather than the Kolmogorov) distance between probability distributions.

In the Poissonized case with a Poisson point process of unit intensity on C_n, rather than exactly n points as considered here, the Kolmogorov distance bounds corresponding to (3) and (4) were already known (see P. and Rosoman (2008)).

It is not clear that the proof of error bounds in the Poissonized setting, using spatial independence properties of the Poisson process, is of any use in deriving (3) and (4).

Our proof uses the idea of *size biasing*. For a nonnegative random variable Y with distribution F and finite mean μ, the size biased distribution of Y is defined to be the distribution $\tilde{F}$ with

$$d\tilde{F}(x) = x dF(x)/\mu, \quad x \geq 0.$$

We prove (3) and (4) using a result of Goldstein (2005) which says, loosely speaking, that if one can closely couple a random variable Y to another variable with the size-biased distribution of Y, then one may be able to obtain a good bound on $D(Y)$.

SIZE BIASING LEMMA

If $X_1, \ldots, X_n$ are exchangeable Bernoulli random variables, and $Y = X_1 + \cdots + X_n$, and Y' has the Y size biased distribution, then

$$\mathcal{L}(Y') = \mathcal{L}(Y|X_1 = 1)$$

so if I is an independent random index, with the discrete uniform distribution on $\{1, \ldots, n\}$ then

$$\mathcal{L}(Y') = \mathcal{L}(Y|X_I = 1)$$

APPROXIMATION LEMMA (Goldstein) Let $Y \geq 0$ be a random variable with mean μ and variance $\sigma^2 \in (0, \infty)$, and let Y^s be defined on the same space, with the Y-size biased distribution. If $|Y^s - Y| \leq B$ for some $B \leq \sigma^{3/2}/\sqrt{6\mu}$, then

$$D(Y) \leq \frac{0.4B}{\sigma} + \frac{\mu}{\sigma^2}\left(\frac{64B^2}{\sigma} + \frac{4B^3}{\sigma^2} + 23\Delta\right),$$

where

$$\Delta := \sqrt{\mathrm{Var}(\mathbf{E}(Y^s - Y|Y))}.$$

Let $Y := n - V_n = \sum_{i=1}^n X_i$ with $X_i := \mathbf{1}\{\text{ball } i \text{ NOT isolated}\}$

$$\mathcal{L}(Y^s) = \mathcal{L}(Y|X_1 = 1) = \mathcal{L}(Y|N \geq 1)$$

with

$$N = \sum_{j=2}^{n} \mathbf{1}\{B_j \cap B_1 \neq \emptyset\} \sim \mathrm{Bin}(n-1, \phi/n), \quad \phi := 2^d \pi$$

To get Y^s:

(1) Sample location of B_1 (uniform on C_n).

(2) Sample N' with $\mathcal{L}(N') = \mathcal{L}(N|N \geq 1)$

(3) Place N' balls B_i $(2 \leq i \leq n)$ overlapping B_1, and the other $n - N'$ not ovelapping B_1.

(4) Count the number of non-isolated balls.

A COUPLING LEMMA Let $m \in \mathbf{N}$ and $p \in (0,1)$. Suppose $N = \sum_{i=1}^{m} \xi_i$ where ξ_i are independent $\mathrm{Bin}(1,p)$. Define π_k by

$$\pi_k := \begin{cases} \frac{P[N>k|N>0]-P[N>k]}{P[N=k](1-(k/m))} & \text{if} \quad 0 \le k \le m-1 \\ 0 & \text{if} \quad k = m, \end{cases}$$

Then $0 \le \pi_k \le 1$ for all $k \in \{0, \ldots, m\}$.

If $\mathcal{B}$ is a further Bernoulli variable with $P(\mathcal{B} = 1|\xi_1, \ldots, \xi_m) = \pi_N$, and suppose I is an independent discrete uniform random variable over $\{1, 2, \ldots, m\}$. Set $M := N + (1 - B_I)\mathcal{B}$, i.e. let M be given by the same sum as N except that if $\mathcal{B} = 1$ the Ith term is set to 1. Then $\mathcal{L}(M) = \mathcal{L}(N|N > 0)$.

COUPLED CONSTRUCTION OF Y, Y^s

(1) Place the balls $B_1, \ldots, B_n$ (unif. random centres).

Let Y be the number of non-isolated balls.

(2) Let $I \sim U\{1, \ldots, n\}$. Let $N = \sum_{j:j \neq I} \mathbf{1}\{B_j \cap B_I \neq \emptyset\}$

(3) If $N = k$ then with probability $1 - \pi_k$, STOP.

(4) Randomly pick $J \neq I$. Move U_J to overlap U_I.

Let Y^s be the new number of non-isolated balls.

AN URN MODEL: Throw n balls uniformly at random into m urns.

Let Y be the number of isolated balls.

A similar proof shows that in the limit $n \to \infty$ with $m/n \to \alpha > 0$,

$$D(Y) = O(n^{-1/2})$$

More generally, if $P[\text{land in urn } j]$ is p_j , $1 \le j \le m.$ and $\mathbf{p}$ varies with n so $\max_j p_j$ bounded, then

$$D(Y) = \Theta(\mathrm{SD}(Y)^{-1}) = \Theta\left(\left(\sum_j n^2 p_j^2\right)^{-1}\right)$$