DOI: 10.1214/ECP.v16-1605

*Elect. Comm. in Probab.* **16** (2011), 70–83

ELECTRONIC
COMMUNICATIONS
in PROBABILITY

# APPLICATIONS OF SIZE BIASED COUPLINGS FOR CONCENTRATION OF MEASURES

SUBHANKAR GHOSH
*Dept. of Mathematics, University of Southern California, 3620 South Vermont Avenue, Los Angeles,CA*
email: subhankar.ghosh@gmail.com

LARRY GOLDSTEIN
*Dept. of Mathematics, University of Southern California, 3620 South Vermont Avenue, Los Angeles,CA*
email: larry@usc.edu

*Submitted* September 22, 2010, *accepted in final form* December 20,2010

AMS 2000 Subject classification: Primary 60E15; Secondary 60C05,60D05.
Keywords: Large deviations, size biased couplings, Stein's method.

*Abstract*

Let $Y$ be a nonnegative random variable with mean $\mu$ and finite positive variance $\sigma^2$, and let $Y^s$, defined on the same space as $Y$, have the $Y$ size biased distribution, that is, the distribution characterized by

$$E[Yf(Y)] = \mu Ef(Y^s) \quad \text{for all functions } f \text{ for which these expectations exist.}$$

Under a variety of conditions on the coupling of $Y$ and $Y^s$, including combinations of boundedness and monotonicity, concentration of measure inequalities such as

$$P\left(\frac{Y-\mu}{\sigma} \geq t\right) \leq \exp\left(-\frac{t^2}{2(A+Bt)}\right) \quad \text{for all } t \geq 0$$

are shown to hold for some explicit $A$ and $B$ in [8]. Such concentration of measure results are applied to a number of new examples: the number of relatively ordered subsequences of a random permutation, sliding window statistics including the number of $m$-runs in a sequence of coin tosses, the number of local maxima of a random function on a lattice, the number of urns containing exactly one ball in an urn allocation model, and the volume covered by the union of $n$ balls placed uniformly over a volume $n$ subset of $\mathbb{R}^d$.

## 1 Introduction

Theorem 1.1, from [8], demonstrates that the existence of a bounded size bias coupling to a nonnegative variable $Y$ implies bounds for the degree of concentration of the distribution of $Y$. In this work we explore a spectrum of new consequences of Theorem 1.1.

The couplings required here which yield concentration of measure results for $Y$ are to a random variable having the size biased distribution of $Y$, denoted $Y^s$. Size biasing of a random variable is

essentially sampling it proportional to its size, and is a well known phenomenon in the literature of both probability and statistics; see, for example, the waiting time paradox in Feller [7], Section I.4, and the method of constructing unbiased ratio estimators in [16]. Size biased couplings are used in Stein's method for normal approximation (see, for instance, [11], [9] and [5]), and is a method which in some sense parallels the exchangeable pair technique. In fact, these two techniques are somewhat complementary, with size biasing useful for the approximation of distributions of nonnegative random variables such as counts, and the exchangeable pair for mean zero variates.
Recently, the objects of Stein's method have also proved successful in deriving concentration of measure inequalities, that is, deviation inequalities of the form $P(|Y-E(Y)| \geq t\sqrt{\text{Var}(Y)})$, where typically one seeks bounds that decay exponentially in $t$; for a guide to the literature on the concentration of measures, see [14] for a detailed overview. As far as the use of techniques related to Stein's method is concerned, Raič [22] obtained large deviation bounds for certain graph related statistics using the Stein equation (see [24]) along with the Cramér transform. Chatterjee [3] derived Gaussian and Poisson type tail bounds for Hoeffding's combinatorial CLT and the net magnetization in the Curie-Weiss model in statistical physics in [3] using the exchangeable pair of Stein's method (see [24]). Considering the complementary method, Ghosh and Goldstein [8] proved Theorem 1.1 which relies on the existence of bounded size bias couplings. Here we demonstrate the broad range of applicability of Theorem 1.1 by presenting a variety of examples.
First recall that for a given nonnegative random variable $Y$ with finite nonzero mean $\mu$, we say that $Y^s$ has the $Y$-size biased distribution if

$$E[Yf(Y)] = \mu E[f(Y^s)] \quad \text{for all functions } f \text{ for which these expectations exist.} \tag{1}$$

**Theorem 1.1.** *Let $Y$ be a nonnegative random variable with mean and variance $\mu$ and $\sigma^2$ respectively, both finite and positive. Suppose there exists a coupling of $Y$ to a variable $Y^s$ having the $Y$-size bias distribution which satisfies $|Y^s - Y| \leq C$ for some $C > 0$ with probability one. Let $A = C\mu/\sigma^2$.*

*If $Y^s \geq Y$ with probability one, then*

$$P\left(\frac{Y-\mu}{\sigma} \leq -t\right) \leq \exp\left(-\frac{t^2}{2A}\right) \quad \text{for all } t > 0. \tag{2}$$

*If the moment generating function $m(\theta) = E(e^{\theta Y})$ is finite at $\theta = 2/C$, then*

$$P\left(\frac{Y-\mu}{\sigma} \geq t\right) \leq \exp\left(-\frac{t^2}{2(A+Bt)}\right) \quad \text{for all } t > 0, \text{ where } B = C/2\sigma. \tag{3}$$

In typical examples the variable $Y$ is indexed by $n$, and the ones we consider have the property that the ratio $\mu/\sigma^2$ remains bounded as $n \to \infty$, and $C$ does not depend on $n$. In such cases the bound in (2) decreases at rate $\exp(-ct^2)$ for some $c > 0$, and if $\sigma \to \infty$ as $n \to \infty$, the bound in (3) is of similar order, asymptotically.
In [8], the number of lightbulbs switched on at the terminal time in the lightbulb process was shown to obey the hypothesis of Theorem 1.1 and concentration of measure inequalities were obtained. In Section 3 we apply Theorem 1.1 to the number of relatively ordered subsequences of a random permutation, sliding window statistics including the number of $m$-runs in a sequence of coin tosses, the number of local maxima of a random function on the lattice, the number of urns containing exactly one ball in the uniform urn allocation model, and the volume covered by the union of $n$ balls placed uniformly over a volume $n$ subset of $\mathbb{R}^d$. In Section 2, we review the methods in [11] for the construction of size bias couplings in the presence of dependence, and then move to the examples.

## 2 Construction of size bias couplings

In this section we will review the discussion in [11] which gives a procedure for a construction of size bias couplings when $Y$ is a sum; the method has its roots in the work of Baldi et al. [1]. The construction depends on being able to size bias a collection of nonnegative random variables in a given coordinate, as described Definition 2.1. Letting $F$ be the distribution of $Y$, first note that the characterization (1) of the size bias distribution $F^s$ is equivalent to the specification of $F^s$ by its Radon Nikodym derivative

$$dF^s(x) = \frac{x}{\mu} dF(x). \tag{4}$$

**Definition 2.1.** *Let $\mathscr{A}$ be an arbitrary index set and let $\{X_\alpha : \alpha \in \mathscr{A}\}$ be a collection of nonnegative random variables with finite, nonzero expectations $EX_\alpha = \mu_\alpha$ and joint distribution $dF(\mathbf{x})$. For $\beta \in \mathscr{A}$, we say that $\mathbf{X}^\beta = \{X_\alpha^\beta : \alpha \in \mathscr{A}\}$ has the $\mathbf{X}$ size bias distribution in coordinate $\beta$ if $\mathbf{X}^\beta$ has joint distribution*

$$dF^\beta(\mathbf{x}) = x_\beta dF(\mathbf{x})/\mu_\beta.$$

Just as (4) is related to (1), the random vector $\mathbf{X}^\beta$ has the $\mathbf{X}$ size bias distribution in coordinate $\beta$ if and only if

$$E[X_\beta f(\mathbf{X})] = \mu_\beta E[f(\mathbf{X}^\beta)] \quad \text{for all functions } f \text{ for which these expectations exist.}$$

Letting $f(\mathbf{X}) = g(X_\beta)$ for some function $g$ one recovers (1), showing that the $\beta^{th}$ coordinate of $\mathbf{X}^\beta$, that is, $X_\beta^\beta$, has the $X_\beta$ size bias distribution.
The factorization

$$P(\mathbf{X} \in d\mathbf{x}) = P(\mathbf{X} \in d\mathbf{x} | X_\beta = x) P(X_\beta \in dx)$$

of the joint distribution of $\mathbf{X}$ suggests a way to construct $\mathbf{X}$. First generate $X_\beta$, a variable with distribution $P(X_\beta \in dx)$. If $X_\beta = x$, then generate the remaining variates $\{X_\alpha^\beta, \alpha \neq \beta\}$ with distribution $P(\mathbf{X} \in d\mathbf{x} | X_\beta = x)$. Now, by the factorization of $dF(\mathbf{x})$, we have

$$\begin{aligned} dF^\beta(\mathbf{x}) &= x_\beta dF(\mathbf{x})/\mu_\beta \qquad (5)\\ &= P(\mathbf{X} \in d\mathbf{x} | X_\beta = x) x_\beta P(X_\beta \in dx)/\mu_\beta = P(\mathbf{X} \in d\mathbf{x} | X_\beta = x) P(X_\beta^\beta \in dx). \end{aligned}$$

Hence, to generate $\mathbf{X}^\beta$ with distribution $dF^\beta$, first generate a variable $X_\beta^\beta$ with the $X_\beta$ size bias distribution, then, when $X_\beta^\beta = x$, generate the remaining variables according to their original conditional distribution given that the $\beta^{th}$ coordinate takes on the value $x$.
Definition 2.1 and the following special case of a proposition from Section 2 of [11] will be applied in the subsequent constructions; the reader is referred there for the simple proof.

**Proposition 2.1.** *Let $\mathscr{A}$ be an arbitrary index set, and let $\mathbf{X} = \{X_\alpha, \alpha \in \mathscr{A}\}$ be a collection of nonnegative random variables with finite means. Let $Y = \sum_{\beta \in A} X_\beta$ and assume $\mu_A = EY$ is finite and positive. Let $\mathbf{X}^\beta$ have the $\mathbf{X}$-size biased distribution in coordinate $\beta$ as in Definition 2.1. Let $I$ be a random index taking values in $A$ with distribution*

$$P(I = \beta) = \mu_\beta / \mu_A, \quad \beta \in A.$$

*Then if $\mathbf{X}^I$ has the mixture distribution $\sum_{\beta \in A} P(I = \beta) \mathscr{L}(\mathbf{X}^\beta)$, the variable $Y^s = \sum_{\alpha \in A} X_\alpha^I$ has the $Y$-sized biased distribution as in (1).*

In our examples we use Proposition 2.1 and the random index $I$, and (5), to obtain $Y^s$ by first generating $X_I^I$ with the size bias distribution of $X_I$, then, if $I = \alpha$ and $X_\alpha^\alpha = x$, generating $\{X_\beta^\alpha : \beta \in A \setminus \{\alpha\}\}$ according to the (original) conditional distribution $P(X_\beta, \beta \neq \alpha | X_\alpha = x)$.

# 3 Applications

We now consider the application of Theorem 1.1 to derive concentration of measure results for the number of relatively ordered subsequences of a random permutation, the number of $m$-runs in a sequence of coin tosses, the number of local extrema on a graph, the number of nonisolated balls in an urn allocation model, and the covered volume in a binomial coverage process. Without further mention we will use that when (2) and (3) hold for some $A$ and $B$ then they also hold when these values are replaced by larger ones, also denoted by $A$ and $B$, and that moment generating functions of bounded random variables are everywhere finite.

## 3.1 Relatively ordered sub-sequences of a random permutation

For $n \geq m \geq 3$, let $\pi$ and $\tau$ be permutations of $\mathcal{V} = \{1, \ldots, n\}$ and $\{1, \ldots, m\}$, respectively, and let

$$\mathcal{V}_\alpha = \{\alpha, \alpha+1, \ldots, \alpha+m-1\} \quad \text{for } \alpha \in \mathcal{V}, \tag{6}$$

where addition of elements of $\mathcal{V}$ is modulo $n$. We say the pattern $\tau$ appears at location $\alpha \in \mathcal{V}$ if the values $\{\pi(v)\}_{v \in \mathcal{V}_\alpha}$ and $\{\tau(v)\}_{v \in \mathcal{V}_1}$ are in the same relative order. Equivalently, the pattern $\tau$ appears at $\alpha$ if and only if $\pi(\tau^{-1}(v) + \alpha - 1), v \in \mathcal{V}_1$ is an increasing sequence. When $\tau = \iota_m$, the identity permutation of length $m$, we say that $\pi$ has a rising sequence of length $m$ at position $\alpha$. Rising sequences are studied in [2] in connection with card tricks and card shuffling.

Letting $\pi$ be chosen uniformly from all permutations of $\{1, \ldots, n\}$, and $X_\alpha$ the indicator that $\tau$ appears at $\alpha$,

$$X_\alpha(\pi(v), v \in \mathcal{V}_\alpha) = 1(\pi(\tau^{-1}(1) + \alpha - 1) < \cdots < \pi(\tau^{-1}(m) + \alpha - 1)),$$

the sum $Y = \sum_{\alpha \in \mathcal{V}} X_\alpha$ counts the number of $m$-element-long segments of $\pi$ that have the same relative order as $\tau$.

For $\alpha \in \mathcal{V}$ we may generate $\mathbf{X}^\alpha = \{X_\beta^\alpha, \beta \in \mathcal{V}\}$ with the $\mathbf{X} = \{X_\beta, \beta \in \mathcal{V}\}$ distribution size biased in direction $\alpha$, following [9]. Let $\sigma_\alpha$ be the permutation of $\{1, \ldots, m\}$ for which

$$\pi(\sigma_\alpha(1) + \alpha - 1) < \cdots < \pi(\sigma_\alpha(m) + \alpha - 1),$$

and set

$$\pi^\alpha(v) = \begin{cases} \pi(\sigma_\alpha(\tau(v - \alpha + 1)) + \alpha - 1), & v \in \mathcal{V}_\alpha \\ \pi(v) & v \notin \mathcal{V}_\alpha. \end{cases}$$

In other words $\pi^\alpha$ is the permutation $\pi$ with the values $\pi(v), v \in \mathcal{V}_\alpha$ reordered so that $\pi^\alpha(\gamma)$ for $\gamma \in \mathcal{V}_\alpha$ are in the same relative order as $\tau$. Now let $X_\beta^\alpha = X_\beta(\pi^\alpha(v), v \in \mathcal{V}_\beta)$, the indicator that $\tau$ appears at position $\beta$ in the reordered permutation $\pi^\alpha$. As $\pi^\alpha$ and $\pi$ agree except perhaps for the $m$ values in $\mathcal{V}_\alpha$, we have

$$X_\beta^\alpha = X_\beta(\pi(v), v \in \mathcal{V}_\beta) \quad \text{for all } |\beta - \alpha| \geq m.$$

Hence, as

$$|Y^\alpha - Y| \le \sum_{|\beta-\alpha|\le m-1} |X^\alpha_\beta - X_\beta| \le 2m-1. \tag{7}$$

we may take $C = 2m-1$ as the almost sure bound on the coupling of $Y^s$ and $Y$.
Regarding the mean $\mu$ of $Y$, clearly for any $\tau$, as all relative orders of $\pi(v), v \in \mathcal{V}_\alpha$ are equally likely,

$$EX_\alpha = 1/m! \quad \text{and therefore} \quad \mu = n/m!. \tag{8}$$

To compute the variance, for $0 \le k \le m-1$, let $I_k$ be the indicator that $\tau(1),\ldots,\tau(m-k)$ and $\tau(k+1),\ldots,\tau(m)$ are in the same relative order. Clearly $I_0 = 1$, and for rising sequences, as $\tau(j) = j$, $I_k = 1$ for all $k$. In general for $0 \le k \le m-1$ we have $X_\alpha X_{\alpha+k} = 0$ if $I_k = 0$, as the joint event in this case demands two different relative orders on the segment of $\pi$ of length $m-k$ of which both $X_\alpha$ and $X_{\alpha+k}$ are a function. If $I_k = 1$ then a given, common, relative order is demanded for this same length of $\pi$, and relative orders also for the two segments of length $k$ on which exactly one of $X_\alpha$ and $X_\beta$ depend, and so, in total a relative order on $m-k+2k = m+k$ values of $\pi$, and therefore

$$EX_\alpha X_{\alpha+k} = I_k/(m+k)! \quad \text{and} \quad \text{Cov}(X_\alpha, X_{\alpha+k}) = I_k/(m+k)! - 1/(m!)^2.$$

As the relative orders of non-overlapping segments of $\pi$ are independent, now taking $n \ge 2m$, the variance $\sigma^2$ of $Y$ is given by

$$\begin{aligned}
\sigma^2 &= n\text{Var}(X_1) + 2n\sum_{k=1}^{m-1}\text{Cov}(X_1, X_{1+k}) \\
&= n\left(\frac{1}{m!} - \frac{1}{(m!)^2}\right) + 2n\sum_{k=1}^{m-1}\left(\frac{I_k}{(m+k)!} - (\frac{1}{m!})^2\right) \\
&= n\left(\frac{1}{m!}\left(1 - \frac{2m-1}{m!}\right) + 2\sum_{k=1}^{m-1}\frac{I_k}{(m+k)!}\right).
\end{aligned} \tag{9}$$

Clearly Var$(Y)$ is maximized for the identity permutation $\tau(k) = k, k = 1,\ldots,m$, as $I_m = 1$ for all $1 \le m \le m-1$, and as mentioned, this case corresponds to counting the number of rising sequences. In contrast, the variance lower bound given when $I_k = 0$ for all $1 \le k \le m-1$ is

$$\sigma^2 \ge \frac{n}{m!}\left(1 - \frac{2m-1}{m!}\right), \quad \text{attained at the permutation} \quad \tau(j) = \begin{cases} 1 & j = 1 \\ j+1 & 2 \le j \le m-1 \\ 2 & j = m \end{cases}.$$

Hence, the bound (3) of Theorem 1.1 holds where $\mu$ and $\sigma^2$ are given in (8) and (9), respectively, and

$$A = \frac{2m-1}{1 - \frac{2m-1}{m!}} \quad \text{and} \quad B = \frac{2m-1}{2\sqrt{\frac{n}{m!}\left(1 - \frac{2m-1}{m!}\right)}}.$$

## 3.2 Local Dependence

The following lemma shows how to construct a collection of variables $\mathbf{X}^\alpha$ having the $\mathbf{X}$ distribution biased in direction $\alpha$ when $X_\alpha$ is some function of a subset of a collection of independent random variables.

**Lemma 3.1.** *Let $\{C_g, g \in \mathcal{V}\}$ be a collection of independent random variables, and for each $\alpha \in \mathcal{V}$ let $\mathcal{V}_\alpha \subset \mathcal{V}$ and $X_\alpha = X_\alpha(C_g, g \in \mathcal{V}_\alpha)$ be a nonnegative random variable with a nonzero, finite expectation. Then if $\{C_g^\alpha, g \in \mathcal{V}_\alpha\}$ has distribution*

$$dF^\alpha(c_g, g \in \mathcal{V}_\alpha) = \frac{X_\alpha(c_g, g \in \mathcal{V}_\alpha)}{EX_\alpha(C_g, g \in \mathcal{V}_\alpha)} dF(c_g, g \in \mathcal{V}_\alpha)$$

*and is independent of $\{C_g, g \in \mathcal{V}\}$, letting*

$$X_\beta^\alpha = X_\beta(C_g^\alpha, g \in \mathcal{V}_\beta \cap \mathcal{V}_\alpha, C_h, h \in \mathcal{V}_\beta \cap \mathcal{V}_\alpha^c),$$

*the collection $\mathbf{X}^\alpha = \{X_\beta^\alpha, \beta \in \mathcal{V}\}$ has the $\mathbf{X}$ distribution biased in direction $\alpha$.*
*Furthermore, with $I$ chosen proportional to $EX_\alpha$, independent of the remaining variables, the sum*

$$Y^s = \sum_{\beta \in \mathcal{V}} X_\beta^I$$

*has the $Y$ size biased distribution, and when there exists $M$ such that $X_\alpha \le M$ for all $\alpha$,*

$$|Y^s - Y| \le bM \quad \textit{where} \quad b = \max_\alpha |\{\beta : \mathcal{V}_\beta \cap \mathcal{V}_\alpha \ne \emptyset\}|. \tag{10}$$

*Proof.* By independence, the random variables

$$\{C_g^\alpha, g \in \mathcal{V}_\alpha\} \cup \{C_g, g \notin \mathcal{V}_\alpha\} \quad \text{have distribution} \quad dF^\alpha(c_g, g \in \mathcal{V}_\alpha) dF(c_g, g \notin \mathcal{V}_\alpha).$$

Thus, with $\mathbf{X}^\alpha$ as given, we find

$$\begin{aligned} EX_\alpha f(\mathbf{X}) &= \int x_\alpha f(\mathbf{x}) dF(c_g, g \in \mathcal{V}) \\ &= EX_\alpha \int f(\mathbf{x}) \frac{x_\alpha dF(c_g, g \in \mathcal{V}_\alpha)}{EX_\alpha(C_g, g \in \mathcal{V}_\alpha)} dF(c_g, g \notin \mathcal{V}_\alpha) \\ &= EX_\alpha \int f(\mathbf{x}) dF^\alpha(c_g, g \in \mathcal{V}_\alpha) dF(c_g, g \notin \mathcal{V}_\alpha) \\ &= EX_\alpha Ef(\mathbf{X}^\alpha). \end{aligned}$$

That is, $\mathbf{X}^\alpha$ has the $\mathbf{X}$ distribution biased in direction $\alpha$, as in Definition 2.1.
The claim on $Y^s$ follows from Proposition 2.1, and finally, since $X_\beta = X_\beta^\alpha$ whenever $\mathcal{V}_\beta \cap \mathcal{V}_\alpha = \emptyset$,

$$|Y^s - Y| \le \sum_{\beta : \mathcal{V}_\beta \cap \mathcal{V}_I \ne \emptyset} |X_\beta^I - X_\beta| \le bM.$$

This completes the proof. □

### 3.2.1 Sliding $m$ window statistics

For $n \geq m \geq 1$, let $\mathcal{V} = \{1, \ldots, n\}$ considered modulo $n$, $\{C_g : g \in \mathcal{V}\}$ i.i.d. real valued random variables, and for each $\alpha \in \mathcal{V}$ let $\mathcal{V}_\alpha$ be as in (6). Then for $X : \mathbb{R}^m \to [0,1]$, say, Lemma 3.1 may be applied to the sum $Y = \sum_{\alpha \in \mathcal{V}} X_\alpha$ of the $m$-dependent sequence $X_\alpha = X(C_\alpha, \ldots, C_{\alpha+m-1})$, formed by applying the function $X$ to the variables in the '$m$-window' $\mathcal{V}_\alpha$. As for all $\alpha$ we have $X_\alpha \leq 1$ and

$$\max_\alpha |\{\beta : \mathcal{V}_\beta \cap \mathcal{V}_\alpha \neq \emptyset\}| = 2m - 1,$$

we may take $C = 2m - 1$ in Theorem 1.1, by Lemma 3.1.
For a concrete example let $Y$ be the number of $m$ runs of the sequence $\xi_1, \xi_2, \ldots, \xi_n$ of $n$ i.i.d Bernoulli($p$) random variables with $p \in (0,1)$, given by $Y = \sum_{i=1}^n X_i$ where $X_i = \xi_i \xi_{i+1} \cdots \xi_{i+m-1}$, with the periodic convention $\xi_{n+k} = \xi_k$. In [23], the authors develop smooth function bounds for normal approximation for $Y$. Note that the construction given in Lemma 3.1 for this case is monotone, as for any $i$, size biasing the Bernoulli variables $\xi_j$ for $j \in \{i, \ldots, i+m-1\}$ by setting

$$\xi_j' = \begin{cases} \xi_j & j \notin \{i, \ldots, i+m-1\} \\ 1 & j \in \{i, \ldots, i+m-1\}, \end{cases}$$

the number $Y^s = \sum_{i=1}^n \xi_i' \xi_{i+1}' \cdots \xi_{i+m-1}'$ of $m$ runs of $\{\xi_j'\}_{i=1}^n$ is at least $Y$.
The mean $\mu$ of $Y$ is clearly $np^m$. For the variance, now letting $n \geq 2m$ and using the fact that non-overlapping segments of the sequence are independent,

$$\begin{aligned} \sigma^2 &= \sum_{i=1}^n \text{Var}(\xi_i \xi_{i+1} \cdots \xi_{i+m-1}) + 2 \sum_{i<j} \text{Cov}(\xi_i \cdots \xi_{i+m-1}, \xi_j \cdots \xi_{j+m-1}) \\ &= np^m(1-p^m) + 2 \sum_{i=1}^n \sum_{j=1}^{m-1} \text{Cov}(\xi_i \cdots \xi_{i+m-1}, \xi_{i+j} \cdots \xi_{i+j+m-1}). \end{aligned}$$

For the covariances $\text{Cov}(\xi_i \cdots \xi_{i+m-1}, \xi_{i+j} \cdots \xi_{i+j+m-1})$ one obtains

$$E(\xi_i \cdots \xi_{i+j-1} \xi_{i+j} \cdots \xi_{i+m-1} \xi_{i+m} \cdots \xi_{i+j+m-1}) - p^{2m} = p^{m+j} - p^{2m},$$

and therefore

$$\sigma^2 = np^m \left( (1-p^m) + 2 \left( \frac{p - p^m}{1-p} - (m-1)p^m \right) \right) = np^m \left( 1 + 2\frac{p-p^m}{1-p} - (2m-1)p^m \right).$$

Hence (2) and (3) of Theorem 1.1 hold with

$$A = \frac{2m-1}{1 + 2\frac{p-p^m}{1-p} - (2m-1)p^m} \quad \text{and} \quad B = \frac{2m-1}{2\sqrt{np^m \left(1 + 2\frac{p-p^m}{1-p} - (2m-1)p^m\right)}}.$$

### 3.2.2 Local extrema on a lattice

Let $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ be a given graph, and for every $v \in \mathcal{V}$ let $\mathcal{V}_v \subset \mathcal{V}$ be a collection of vertices depending on $v$; we think of $\mathcal{V}_v$ as some 'neighborhood' of the vertex $v$. Let $\{C_g, g \in \mathcal{V}\}$ be a collection of independent and identically distributed random variables, and let $X_v$ be the indicator that vertex $v$ corresponds to a local maximum value with respect to the neighborhood $\mathcal{V}_v$, that is

$$X_v(C_w, w \in \mathcal{V}_v) = \mathbf{1}(C_v > C_w, w \in \mathcal{V}_v), \quad v \in \mathcal{V}.$$

The sum $Y = \sum_{v \in \mathcal{V}} X_v$ counts the number of local maxima. Size biased couplings to $Y$, for the purpose of normal approximation, were studied in [1] and [9].
In general one may define the neighbor distance $d$ between two vertices $v, w \in \mathcal{V}$ by

$$d(v,w) = \min\{n : \text{there } \exists\ v_0, \ldots, v_n \text{ in } \mathcal{V} \text{ so that } v_0 = v, v_n = w \text{ and } (v_k, v_{k+1}) \in \mathcal{E} \text{ for } k = 0, \ldots, n\},$$

and for $r \in \mathbb{N}$, the $r$ neighborhood of $v \in \mathcal{V}$ consisting of vertices at distance at most $r$ from $v$,

$$\mathcal{V}_v(r) = \{w \in \mathcal{V} : d(w,v) \le r\}.$$

We consider the case where there is some $r$ such that the graphs $\mathcal{G}_v = \{\mathcal{V}_v, \mathcal{E}_v\}, v \in \mathcal{V}$, where $\mathcal{V}_v = \mathcal{V}_v(r)$ and $\mathcal{E}_v = \{\{w,u\} \subset \mathcal{V}_v(r), \{w,v\} \in \mathcal{E}\}$, are isomorphic, and the isomorphism from $\mathcal{G}_{v_1}$ to $\mathcal{G}_{v_2}$ maps $v_1$ to $v_2$. Then if $d(v_1,v_2) > 2r$, and $(w_1,w_2) \in \mathcal{V}_{v_1} \times \mathcal{V}_{v_2}$, rearranging

$$2r < d(v_1,v_2) \le d(v_1,w_1) + d(w_1,w_2) + d(w_2,v_2)$$

and using $d(v_i,w_i) \le r, i = 1,2$, yields $d(w_1,w_2) > 0$. Hence,

$$d(v_1,v_2) > 2r \quad \text{implies} \quad \mathcal{V}_{v_1} \bigcap \mathcal{V}_{v_2} = \emptyset, \quad \text{so by (10) we may take} \quad b = \max_v |\mathcal{V}_v(2r)|. \tag{11}$$

For example, for $p \in \{1,2,\ldots\}$ and $n \ge 5$ consider the lattice $\mathcal{V} = \{1,\ldots,n\}^p$ modulo $n$ in $\mathbb{Z}^p$ and $\mathcal{E} = \{\{v,w\} : \sum_{i=1}^p |v_i - w_i| = 1\}$; in this case $d$ is the $L^1$ norm. To consider the case where we call vertex $v$ a local maximum if $C_v$ exceeds the values $C_w$ over the immediate neighbors $w$ of $v$, we take $r = 1$ and obtain

$$\mathcal{V}_v = \mathcal{V}_v(1) \quad \text{and that} \quad |\mathcal{V}_v(1)| = 1 + 2p,$$

the 1 accounting for $v$ itself, and $2p$ for the number of neighbors at distance 1 from $v$, which differ from $v$ by either $+1$ or $-1$ in exactly one coordinate.
Lemma 3.1, (11), and $|X_v| \le 1$ yield

$$|Y^s - Y| \le \max_v |\mathcal{V}_v(2)| = 1 + 2p + \left(2p + 4\binom{p}{2}\right) = 2p^2 + 2p + 1, \tag{12}$$

where the 1 counts $v$ itself, the $2p$ again are the neighbors at distance 1, and the term in the parenthesis accounting for the neighbors at distance 2, $2p$ of them differing in exactly one coordinate by $+2$ or $-2$, and $4\binom{p}{2}$ of them differing by either $+1$ or $-1$ in exactly two coordinates. Note that we have used the assumption $n \ge 5$ here, and continue to do so below.
Now letting $C_v$ have a continuous distribution, without loss of generality we can assume $C_v \sim \mathcal{U}[0,1]$. As any vertex has chance $1/|\mathcal{V}_v| = 1/(2p+1)$ of having the largest value in its neighborhood, $\mu = EY$ satisfies

$$\mu = \frac{n}{2p+1}. \tag{13}$$

To begin the calculation of the variance, note that when $v$ and $w$ are neighbors they cannot both be maxima, so $X_v X_w = 0$ and therefore, for $d(v,w) = 1$,

$$\text{Cov}(X_v, X_w) = -(EX_v)^2 = -\frac{1}{(2p+1)^2}.$$

If the distance between $v$ and $w$ is 3 or more, $X_v$ and $X_w$ are functions of disjoint sets of independent variables, and hence are independent.

When $d(w,v)=2$ there are two cases, as $v$ and $w$ may have either 1 or 2 neighbors in common, and

$$EX_vX_w = P(U > U_j, V > V_j, j=1,\ldots,m-k \quad \text{and} \quad U > U_j, V > U_j, j = m-k+1,\ldots,m),$$

where $m$ is the number of vertices over which $v$ and $w$ are extreme, so $m=2p$, and $k=1$ and $k=2$ for the number of neighbors in common. For $k=1,2,\ldots$, letting $M_k=\max\{U_{m-k+1},\ldots,U_m\}$, as the variables $X_v$ and $X_w$ are conditionally independent given $U_{m-k+1},\ldots,U_m$

$$\begin{aligned} E(X_vX_w|U_{m-k+1},\ldots,U_m) &= P(U>U_j, j=1,\ldots,m|U_{m-k+1},\ldots,U_m)^2 \\ &= \frac{1}{(m-k+1)^2}(1-M_k^{m-k+1})^2, \end{aligned} \tag{14}$$

as

$$\begin{aligned} P(U>U_j, j=1,\ldots,m|U_{m-k+1},\ldots,U_m) &= \int_{M_k}^1\int_0^u\cdots\int_0^u du_1\cdots du_{m-k}du \\ &= \int_{M_k}^1 u^{m-k}du \\ &= \frac{1}{m-k+1}(1-M_k^{m-k+1}). \end{aligned}$$

Since $P(M_k\le x)=x^k$ on $[0,1]$, we have

$$\begin{aligned} EM_k^{m-k+1} &= k\int_0^1 x^{m-k+1}x^{k-1}dx = \frac{k}{m+1} \quad \text{and} \\ E(M_k^{m-k+1})^2 &= k\int_0^1 x^{2(m-k+1)}x^{k-1}dx = \frac{k}{2m-k+2}. \end{aligned}$$

Hence, averaging (14) over $U_{m-k+1},\ldots,U_m$ yields

$$EX_vX_w = \frac{2}{(m+1)(2(m+1)-k)}.$$

For $n\ge 3$, when $m=2p$, for $k=1$ and 2 we obtain

$$\text{Cov}(X_v,X_w) = \frac{1}{(2p+1)^2(2(2p+1)-1)} \quad \text{and} \quad \text{Cov}(X_v,X_w) = \frac{2}{(2p+1)^2(2(2p+1)-2)}, \quad \text{respectively.}$$

For $n\ge 5$, of the $2p+4\binom{p}{2}$ vertices $w$ that are at distance 2 from $v$, $2p$ of them share 1 neighbor in common with $v$, while the remaining $4\binom{p}{2}$ of them share 2 neighbors. Hence,

$$\begin{aligned}
\sigma^2 &= \sum_{v\in V}\mathrm{Var}(X_v)+\sum_{v\neq w}\mathrm{Cov}(X_v,X_w)\\
&= \sum_{v\in V}\mathrm{Var}(X_v)+\sum_{d(v,w)=1}\mathrm{Cov}(X_v,X_w)+\sum_{d(v,w)=2}\mathrm{Cov}(X_v,X_w)\\
&= n\left(\frac{2p}{(2p+1)^2}-2p\frac{1}{(2p+1)^2}+2p\frac{1}{(2p+1)^2(2(2p+1)-1)}+4\binom{p}{2}\frac{2}{(2p+1)^2(2(2p+1)-2)}\right)\\
&= n\frac{2p}{(2p+1)^2}\left(\frac{1}{(2(2p+1)-1)}+\frac{2(p-1)}{(2(2p+1)-2)}\right)\\
&= n\left(\frac{4p^2-p-1}{(2p+1)^2(4p+1)}\right). \qquad (15)
\end{aligned}$$

We conclude that (2) of Theorem 1.1 holds with $A=C\mu/\sigma^2$ and $B=C/2\sigma$ with $\mu$, $\sigma^2$ and $C$ given by (13), (15) and (12), respectively, that is,

$$A=\frac{(2p+1)(4p+1)(2p^2+2p+1)}{4p^2-p-1} \quad \text{and} \quad B=\frac{2p^2+2p+1}{2\sqrt{n\left(\frac{4p^2-p-1}{(2p+1)^2(4p+1)}\right)}}.$$

## 3.3 Urn allocation

In the classical urn allocation model $n$ balls are thrown independently into one of $m$ urns, where, for $i=1,\ldots,m$, the probability a ball lands in the $i^{th}$ urn is $p_i$, with $\sum_{i=1}^m p_i=1$. A much studied quantity is the number of nonempty urns, for which Kolmogorov distance bounds to the normal were obtained in [6] and [21]. In [6], bounds were obtained for the uniform case where $p_i=1/m$ for all $i=1,\ldots,m$, while the bounds in [21] hold for the nonuniform case as well. In [19] the author considers the normal approximation for the number of isolated balls, that is, the number of urns containing exactly one ball, and obtains Kolmogorov distance bounds to the normal. Using the coupling provided in [19], we derive right tail inequalities for the number of non-isolated balls, or, equivalently, left tail inequalities for the number of isolated balls.

For $i=1,\ldots,n$ let $X_i$ denote the location of ball $i$, that is, the number of the urn into which ball $i$ lands. The number $Y$ of non-isolated balls is given by

$$Y=\sum_{i=1}^n \mathbf{1}(M_i>0) \quad \text{where} \quad M_i=-1+\sum_{j=1}^n \mathbf{1}(X_j=X_i).$$

We first consider the uniform case. A construction in [19] produces a coupling of $Y$ to $Y^s$, having the $Y$ size biased distribution, which satisfies $|Y^s-Y|\le 2$. Given a realization of $\mathbf{X}=\{X_1,X_2,\ldots,X_n\}$, the coupling proceeds by first selecting a ball $I$, uniformly from $\{1,2,\ldots,n\}$, and independently of $\mathbf{X}$. Depending on the outcome of a Bernoulli variable $\mathscr{B}$, whose distribution depends on the number of balls found in the urn containing $I$, a different ball $J$ will be imported into the urn that contains ball $I$. In some additional detail, let $\mathscr{B}$ be a Bernoulli variable with success probability $P(\mathscr{B}=1)=\pi_{M_I}$, where

$$\pi_k=\begin{cases}\frac{P(N>k|N>0)-P(N>k)}{P(N=k)(1-k/(n-1))} & \text{if } 0\le k\le n-2\\ 0 & \text{if } k=n-1,\end{cases}$$

with $N \sim \text{Bin}(1/m, n-1)$. Now let $J$ be uniformly chosen from $\{1,2,\ldots,n\} \setminus \{I\}$, independent of all other variables. Lastly, if $\mathscr{B}=1$, move ball $J$ into the same urn as $I$. It is clear that $|Y'-Y| \le 2$, as at most the occupancy of two urns can affected by the movement of a single ball. We also note that if $M_I = 0$, which happens when ball $I$ is isolated, $\pi_0 = 1$, so that $I$ becomes no longer isolated after relocating ball $J$. We refer the reader to [19] for a full proof that this procedure produces a coupling of $Y$ to a variable with the $Y$ size biased distribution.

For the uniform case, the following explicit formulas for $\mu$ and $\sigma^2$ can be found in Theorem II.1.1 of [13],

$$\begin{aligned}
\mu &= n\left(1-\left(1-\frac{1}{m}\right)^{n-1}\right) \quad \text{and} \\
\sigma^2 &= (n-\mu) + \frac{(m-1)n(n-1)}{m}\left(1-\frac{2}{m}\right)^{n-2} - (n-\mu)^2 \\
&= n\left(1-\frac{1}{m}\right)^{n-1} + \frac{(m-1)n(n-1)}{m}\left(1-\frac{2}{m}\right)^{n-2} - n^2\left(1-\frac{1}{m}\right)^{2n-2}. \qquad (16)
\end{aligned}$$

Hence with $\mu$ and $\sigma^2$ as in (16), we can apply (3) of Theorem 1.1 for $Y$, the number of non isolated balls with $C=2$, $A=2\mu/\sigma^2$ and $B=1/\sigma$.

Taking limits in (16), if $m$ and $n$ both go to infinity in such a way that $n/m \to \alpha \in (0,\infty)$, the mean $\mu$ and variance $\sigma^2$ obey

$$\mu \asymp n(1-e^{-\alpha}) \quad \text{and } \sigma^2 \asymp n g(\alpha)^2 \text{ where } g(\alpha)^2 = e^{-\alpha} - e^{-2\alpha}(\alpha^2-\alpha+1) > 0 \text{ for all } \alpha \in (0,\infty),$$

where for positive functions $f$ and $h$ depending on $n$ we write $f \asymp h$ when $\lim_{n\to\infty} f/h = 1$.
Hence, in this limiting case $A$ and $B$ satisfy

$$A \asymp \frac{2(1-e^{-\alpha})}{e^{-\alpha} - e^{-2\alpha}(\alpha^2-\alpha+1)} \quad \text{and} \quad B \asymp \frac{1}{\sqrt{n}g(\alpha)}.$$

In the nonuniform case similar results hold with some additional conditions. Letting

$$||p|| = \sup_{1\le i\le m} p_i \quad \text{and} \quad \gamma = \gamma(n) = \max(n||p||, 1),$$

in [19] it is shown that when $||p|| \le 1/11$ and $n \ge 83\gamma^2(1+3\gamma+3\gamma^2)e^{1.05\gamma}$, there exists a coupling such that

$$|Y^s - Y| \le 3 \quad \text{and} \quad \frac{\mu}{\sigma^2} \le 8165\gamma^2 e^{2.1\gamma}.$$

Now, using the bound $\sigma^2 \ge (7776)^{-1}\gamma^{-2}e^{-2.1\gamma}n^2\sum_i p_i^2$ from (2.14) of Theorem 2.4 in [19] to yield $B$, we find that (3) of Theorem 1.1 holds with

$$A = 24,495\,\gamma^2 e^{2.1\gamma} \quad \text{and} \quad B = \frac{1.5\sqrt{7776}\,\gamma e^{1.05\gamma}}{n\sqrt{\sum_{i=1}^m p_i^2}}.$$

### 3.4 An application to coverage processes

We consider the following coverage process, and associated coupling, from [10]. Given a collection $\mathscr{U} = \{U_1, U_2, \ldots, U_n\}$ of independent, uniformly distributed points in the $d$ dimensional torus of volume $n$, that is, the cube $C_n = [0, n^{1/d})^d \subset \mathbb{R}^d$ with periodic boundary conditions, let $V$ denote the total volume of the union of the $n$ balls of fixed radius $\rho$ centered at these $n$ points, and $S$ the number of balls isolated at distance $\rho$, that is, those points for which none of the other $n-1$ points lie within distance $\rho$. The random variables $V$ and $S$ are of fundamental interest in stochastic geometry, see [12] and [18]. If $n \to \infty$ and $\rho$ remains fixed, both $V$ and $S$ satisfy a central limit theorem [12, 17, 20]. The $L^1$ distance of $V$, properly standardized, to the normal is studied in [4] using Stein's method. The quality of the normal approximation to the distributions of both $V$ and $S$, in the Kolmogorov metric, is studied in [10] using Stein's method via size bias couplings.

In more detail, for $x \in C_n$ and $r > 0$ let $B_r(x)$ denote the ball of radius $r$ centered at $x$, and $B_{i,r} = B(U_i, r)$. The covered volume $V$ and number of isolated balls $S$ are given, respectively, by

$$V = \text{Volume}(\bigcup_{i=1}^{n} B_{i,\rho}) \quad \text{and} \quad S = \sum_{i=1}^{n} \mathbf{1}\{(\mathcal{U}_n \cap B_{i,\rho} = \{U_i\}\}. \tag{17}$$

We will derive concentration of measure inequalities for $V$ and $S$ with the help of the bounded size biased couplings in [10].

Assume $d \geq 1$ and $n \geq 4$. Denote the mean and variance of $V$ by $\mu_V$ and $\sigma_V^2$, respectively, and likewise for $S$, leaving their dependence on $n$ and $\rho$ implicit. Let $\pi_d = \pi^{d/2}/\Gamma(1+d/2)$, the volume of the unit sphere in $\mathbb{R}^d$, and for fixed $\rho$ let $\phi = \pi_d \rho^d$. For $0 \leq r \leq 2$ let $\omega_d(r)$ denote the volume of the union of two unit balls with centers $r$ units apart. We have $\omega_1(r) = 2 + r$, and

$$\omega_d(r) = \pi_d + \pi_{d-1} \int_0^r (1 - (t/2)^2)^{(d-1)/2} dt, \quad \text{for } d \geq 2.$$

From [10], the means of $V$ and $S$ are given by

$$\mu_V = n\left(1 - (1 - \phi/n)^n\right) \quad \text{and} \quad \mu_S = n(1 - \phi/n)^{n-1}, \tag{18}$$

and their variances by

$$\sigma_V^2 = n \int_{B_{2\rho}(\mathbf{0})} \left(1 - \frac{\rho^d \omega_d(|y|/\rho)}{n}\right)^n dy + n(n - 2^d \phi)\left(1 - \frac{2\phi}{n}\right)^n - n^2(1 - \phi/n)^{2n}, \tag{19}$$

and

$$\begin{aligned} \sigma_S^2 \quad = \quad & n(1 - \phi/n)^{n-1}(1 - (1 - \phi/n)^{n-1}) \\ & +(n-1) \int_{B_{2\rho}(\mathbf{0}) \setminus B_\rho(\mathbf{0})} \left(1 - \frac{\rho^d \omega_d(|y|/\rho)}{n}\right)^{n-2} dy \\ & +n(n-1)\left(\left(1 - \frac{2^d \phi}{n}\right)\left(1 - \frac{2\phi}{n}\right)^{n-2} - \left(1 - \frac{\phi}{n}\right)^{2n-2}\right). \end{aligned} \tag{20}$$

It is shown in [10], by using a coupling similar to the one briefly described for the urn allocation problem in Section 3.3, that one can construct $V^s$ with the $V$ size bias distribution which satisfies $|V^s - V| \leq \phi$. Hence (2) of Theorem 1.1 holds for $V$ with

$$A_V = \frac{\phi \mu_V}{\sigma_V^2} \quad \text{and} \quad B_V = \frac{\phi}{2\sigma_V},$$

where $\mu_V$ and $\sigma_V^2$ are given in (18) and (19), respectively. Similarly, with $Y = n - S$ the number of non-isolated balls, it is shown that $Y^s$ with $Y$ size bias distribution can be constructed so that $|Y^s - Y| \le \kappa_d + 1$, where $\kappa_d$ denotes the maximum number of open unit balls in $d$ dimensions that can be packed so they all intersect an open unit ball in the origin, but are disjoint from each other. Hence (2) of Theorem 1.1 holds for $Y$ with

$$A_Y = \frac{(\kappa_d + 1)(n - \mu_S)}{\sigma_S^2} \quad \text{and} \quad B_Y = \frac{\kappa_d + 1}{2\sigma_S}.$$

To see how the $A_V, A_Y$ and $B_V, B_Y$ behave as $n \to \infty$, let

$$J_{r,d}(\rho) = d\pi_d \int_0^r \exp(-\rho^d \omega_d(t)) t^{d-1} dt,$$

and define

$$\begin{aligned} g_V(\rho) &= \rho^d J_{2,d}(\rho) - (2^d \phi + \phi^2) e^{-2\phi} \quad \text{and} \\ g_S(\rho) &= e^{-\phi} - (1 + (2^d - 2)\phi + \phi^2) e^{-2\phi} + \rho^d (J_{2,d}(\rho) - J_{1,d}(\rho)). \end{aligned}$$

Then, again from [10],

$$\begin{aligned} \lim_{n\to\infty} n^{-1}\mu_V = \lim_{n\to\infty} (1 - n^{-1}\mu_S) &= 1 - e^{-\phi}, \\ \lim_{n\to\infty} n^{-1}\sigma_V^2 &= g_V(\rho) > 0, \quad \text{and} \\ \lim_{n\to\infty} n^{-1}\sigma_S^2 &= g_S(\rho) > 0. \end{aligned}$$

Hence, $B_V$ and $B_Y$ tend to zero at rate $n^{-1/2}$, and

$$\lim_{n\to\infty} A_V = \frac{\phi(1 - e^{-\phi})}{g_V(\rho)}, \quad \text{and} \quad \lim_{n\to\infty} A_Y = \frac{(\kappa_d + 1)(1 - e^{-\phi})}{g_S(\rho)}.$$

## References


[1] P. Baldi, Y. Rinott and C. Stein, A normal approximations for the number of local maxima of a random function on a graph, Probability, Statistics and Mathematics, Papers in Honor of Samuel Karlin, T. W. Anderson, K.B. Athreya and D. L. Iglehart eds. (1989), Academic Press, 59-81.

[2] D. Bayer and P. Diaconis, Trailing the dovetail shuffle to its lair, Ann. Appl. Probab. **2** (1992), no. 2, 294–313. MR1161056 (93d:60014)

[3] S. Chatterjee, Stein's method for concentration inequalities, Probab. Theory Related Fields **138** (2007), no. 1-2, 305–321. MR2288072 (2008e:60038)

[4] S. Chatterjee, A new method of normal approximation, Ann. Probab. **36** (2008), no. 4, 1584–1610. MR2435859 (2009j:60043)

[5] L.H.Y. Chen, L. Goldstein and Q.M. Shao. Normal Approximation by Stein's Method.(2010). Springer, New York.

[6] G. Englund, A remainder term estimate for the normal approximation in classical occupancy, Ann. Probab. **9** (1981), no. 4, 684–692. MR0624696 (82j:60015)

[7] FELLER, W.(1966). An Introduction to Probability and its Applications, volume II. Wiley.

[8] S. Ghosh and L. Goldstein. Concentration of measures via size biased cou- plings, (2009). To appear in Probab. Th. Rel. Fields.

[9] L. Goldstein, Berry-Esseen bounds for combinatorial central limit theorems and pattern occurrences, using zero and size biasing, J. Appl. Probab. **42** (2005), no. 3, 661–683. MR2157512 (2006g:60035)

[10] L. Goldstein and M. D. Penrose, Normal approximation for coverage models over binomial point processes, Ann. Appl. Probab. **20** (2010), no. 2, 696–721. MR2650046

[11] L. Goldstein and Y. Rinott, Multivariate normal approximations by Stein's method and size bias couplings, J. Appl. Probab. **33** (1996), no. 1, 1–17. MR1371949 (97d:60038)

[12] P. Hall. Introduction to the theory of coverage processes, (1988). John Wiley, New York.

[13] V.F. Kolchin, B.A. Sevast'yanov and V.P. Chistyakov. Random Allocations, (1978). Winston, Washington D.C.

[14] M. Ledoux. The concentration of measure phenomenon, (2001). Amer. Math. Soc., Providence, RI.

[15] H. Luk. Stein's Method for the Gamma Distribution and Related Statistical Applications, Ph.D Thesis, (1994). Department of Mathematics, University of Southern California, USA.

[16] H. Midzuno, On the sampling system with probability proportionate to sum of sizes, Ann. Inst. Statist. Math., Tokyo **3** (1952), 99–107. MR0050840 (14,390d)

[17] P. A. P. Moran, The random volume of interpenetrating spheres in space, J. Appl. Probability **10** (1973), 483–490. MR0353412 (50 #5895)

[18] M. Penrose. Random geometric graphs, (2003). Oxford University Press, Oxford.

[19] M. D. Penrose, Normal approximation for isolated balls in an urn allocation model, Electron. J. Probab. **14** (2009), no. 74, 2156–2181.MR2550296 (2010k:60094)

[20] M. D. Penrose and J. E. Yukich, Central limit theorems for some graphs in computational geometry, Ann. Appl. Probab. **11** (2001), no. 4, 1005–1041. MR1878288 (2002k:60068)

[21] M. P. Quine and J. Robinson, A Berry-Esseen bound for an occupancy problem, Ann. Probab. **10** (1982), no. 3, 663–671. MR0659536 (83i:60027)

[22] M. Raič, CLT-related large deviation bounds based on Stein's method, Adv. in Appl. Probab. **39** (2007), no. 3, 731–752. MR2357379 (2009d:60069)

[23] G. Reinert and A. Röllin, Multivariate normal approximation with Stein's method of exchangeable pairs under a general linearity condition, Ann. Probab. **37** (2009), no. 6, 2150–2173. MR2573554

[24] C. Stein. Approximate Computation of Expectations, (1986). Institute of Mathematical Statistics, Hayward, CA.