Applications of size biased couplings for concentration of measures

Subhankar Ghosh and Larry Goldstein *

University of Southern California

January 4, 2010

Abstract

Let Y be a nonnegative random variable with mean μ and finite positive variance σ^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

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

were shown to hold for some explicit A and B in [14], and were applied to the lightbulb process of [34]. Those concentration of measure results are applied here to a number of new examples where one can construct a bounded coupling of Y^s to Y: 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 $\mathbb{R}^d$. Concentration of measure results are also provided for certain instances where the coupling of Y^s to Y is not bounded, in particular, for the number of isolated vertices in the Erdös-Rényi random graph model, and the infinitely divisible and compound Poisson distributions that satisfy a bounded moment generating function condition, the latter with applications to the generalized variance of a random matrix.

1 Introduction

Theorem 1.1, from [14], demonstrates that the existence of a bounded size bias coupling to a nonnegative variable Y implies bounds for the amount of concentration of the distribution of Y. In this work we explore a spectrum of consequences of Theorem 1.1 when bounded couplings can be constructed, as well as show that concentration of measure results can be also be produced using unbounded couplings.

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 [13], Section I.4, and the method of constructing unbiased ratio estimators in [26]. Size biased couplings are used in Stein's method for normal approximation (see,

*Department of Mathematics, University of Southern California, Los Angeles, CA 90089, USA,subhankg@usc.edu and larry@usc.edu

2000 *Mathematics Subject Classification*: Primary 60E15; Secondary 60C05,60D05.

Keywords: Large deviations, size biased couplings, Stein's method.

for instance, [37] and [3]), 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 [24] for a detailed overview. Specifically regarding the use of techniques related to Stein's method, using the Stein equation (see [37]) along with the Cramér transform, Raič [33] obtained large deviation bounds for certain graph related statistics. Chatterjee [8] 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 [8] using the exchangeable pair of Stein's method (see [38]). Considering the complementary method, Ghosh and Goldstein [14] 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, and also provide results that show how unbounded couplings can give rise to similar inequalities. First recall that for a given nonnegative random variable Y with finite nonzero mean μ, 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 μ and σ^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.*

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

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

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

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

In typical examples the variable Y is indexed by n, and the ones we consider have the property that the ratio μ/σ^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 [14], the number of lightbulbs switched on at the terminal time in the lightbulb process was shown to obey 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 4, on a somewhat case by case basis, we also consider cases where the coupling of Y^s to Y is unbounded. Our examples include the number of isolated vertices in the Erdös-Rényi random graph model, and some infinitely divisible and compound Poisson distributions; we apply the latter to obtain results for the generalized variance of a random matrix. As Theorem 1.1 shows for the bounded coupling case, additional information is available when the coupling is monotone; this additional condition holds for the m runs and isolated vertices examples, as well as the infinitely divisible and compound Poisson distributions considered.

In Section 2, we review the methods in [17] 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 [17] 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. [2]. The construction depends on being able to size bias a collection of nonnegative random variables in a given coordinate, as described in the following definition. 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 $\mathcal{A}$ be an arbitrary index set and let $\{X_\alpha : \alpha \in \mathcal{A}\}$ be a collection of nonnegative random variables with finite, nonzero expectations $EX_\alpha = \mu_\alpha$ and joint distribution $dF(\mathbf{x})$. For $\beta \in \mathcal{A}$, we say that $\mathbf{X}^\beta = \{X_\alpha^\beta : \alpha \in \mathcal{A}\}$ has the $\mathbf{X}$ size bias distribution in coordinate β 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 β if and only if

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

Now letting $f(\mathbf{X}) = g(X_\beta)$ for some function g one recovers (1), showing that the β^{th} coordinate of $\mathbf{X}^\beta$, that is, X_β^β, has the X_β 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_β, 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

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

Hence, to generate $\mathbf{X}^\beta$ with distribution dF^β, first generate a variable X_β^β with the X_β size bias distribution, then, when $X_\beta^\beta = x$, generate the remaining variables according to their original conditional distribution given that the β^{th} coordinate takes on the value x.

Definition 2.1 and the following proposition from Section 2 of [17] will be applied in the subsequent constructions; the reader is referred there for the simple proof.

Proposition 2.1. *Let $\mathcal{A}$ be an arbitrary index set, and let $\mathbf{X} = \{X_\alpha, \alpha \in \mathcal{A}\}$ be a collection of nonnegative random variables with finite means. For any subset $B \subset \mathcal{A}$, set*

$$X_B = \sum_{\beta \in B} X_\beta \quad \textit{and} \quad \mu_B = EX_B.$$

Suppose $B \subset \mathcal{A}$ with $0 < \mu_B < \infty$, and for $\beta \in B$ let $\mathbf{X}^\beta$ have the $\mathbf{X}$-size biased distribution in coordinate β as in Definition 2.1. If $\mathbf{X}^B$ has the mixture distribution

$$\mathcal{L}(\mathbf{X}^B) = \sum_{\beta \in B} \frac{\mu_\beta}{\mu_B} \mathcal{L}(\mathbf{X}^\beta),$$

then

$$EX_B f(\mathbf{X}) = \mu_B E f(\mathbf{X}^B)$$

for all real valued functions f for which these expectations exist. Hence, for any $A \subset \mathcal{A}$, if f is a function of $X_A = \sum_{\alpha \in A} X_\alpha$ only,

$$EX_B f(X_A) = \mu_B Ef(X_A^B) \quad \textit{where} \quad X_A^B = \sum_{\alpha \in A} X_\alpha^B. \tag{6}$$

Taking $A = B$ in (6) we have $EX_A f(X_A) = \mu_A Ef(X_A^A)$, and hence X_A^A has the X_A-size biased distribution, as in (1).

In our examples we use Proposition 2.1 and (5) to obtain a variable Y^s with the size bias distribution of Y, where $Y = \sum_{\alpha \in A} X_\alpha$, as follows. First choose a random index $I \in A$ with probability

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

Next generate X_I^I with the size bias distribution of X_I. If $I = \alpha$ and $X_\alpha^\alpha = x$, generating $\{X_\beta^\alpha : \beta \in A \setminus \{\alpha\}\}$ using the (original) conditional distribution

$$P(X_\beta, \beta \neq \alpha | X_\alpha = x),$$

the sum $Y^s = \sum_{\alpha \in A} X_\alpha^I$ has the Y size biased distribution.

3 Applications: bounded couplings

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, the covered volume in binomial coverage process. Without further mention we will use the fact that when (2) and (3) hold for some A and B then they also hold when these values are replaced by any larger ones, which may also be denoted by A and B.

3.1 Relatively ordered sub-sequences of a random permutation

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

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

where addition of elements of $\mathcal{V}$ is modulo n. We say the pattern τ 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 τ appears at α 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 π has a rising sequence of length m at position α. Rising sequences are studied in [7] in connection with card tricks and card shuffling.

Letting π be chosen uniformly from all permutations of $\{1, \ldots, n\}$, and X_α the indicator that τ appears at α,

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

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

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 α, following [15]. Let σ_α 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 π^α is the permutation π 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 τ. Now let

$$X_\beta^\alpha = X_\beta(\pi^\alpha(v), v \in \mathcal{V}_\beta),$$

the indicator that τ appears at position β in the reordered permutation π^α. As π^α and π 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

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

we may take $C = 2m - 1$ as the almost sure bound on the coupling of Y^s and Y.

Regarding the mean μ of Y, clearly for any τ, 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 $0 \leq k \leq m-1$, 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 $0 \leq k \leq m-1$ 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 π of length $m-k$ of which both X_α and $X_{\alpha+k}$ are a function. If $I_k = 1$ then a given, common, relative order is demanded for this same length of π, and relative orders also for the two segments of length k on which exactly one of X_α and X_β depend, and so, in total a relative order on $m - k + 2k = m + k$ values of π, and therefore

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

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

$$\begin{aligned}
\sigma^2 &= \sum_{\alpha\in\mathcal{V}} \mathrm{Var}(X_\alpha) + \sum_{\alpha\neq\beta} \mathrm{Cov}(X_\alpha, X_\beta) \\
&= \sum_{\alpha\in\mathcal{V}} \mathrm{Var}(X_\alpha) + \sum_{\alpha\in\mathcal{V}} \sum_{\beta:1\leq|\alpha-\beta|\leq m-1} \mathrm{Cov}(X_\alpha, X_\beta) \\
&= \sum_{\alpha\in\mathcal{V}} \mathrm{Var}(X_\alpha) + 2\sum_{\alpha\in\mathcal{V}} \sum_{k=1}^{m-1} \mathrm{Cov}(X_\alpha, X_{\alpha+k}) \\
&= n\mathrm{Var}(X_1) + 2n\sum_{k=1}^{m-1} \mathrm{Cov}(X_1, X_{1+k}) \\
&= n\left(\frac{1}{m!} - \frac{1}{(m!)^2}\right) + 2n\sum_{k=1}^{m-1}\left(\frac{I_k}{(m+k)!} - (\frac{1}{m!})^2\right) \\
&= n\left(\frac{1}{m!}\left(1 - \frac{2m-1}{m!}\right) + 2\sum_{k=1}^{m-1}\frac{I_k}{(m+k)!}\right).
\end{aligned}$$

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

$$\sigma^2 \geq \frac{n}{m!}\left(1 - \frac{2m-1}{m!}\right) \tag{9}$$

is attained at the permutation

$$\tau(j) = \begin{cases} 1 & j = 1 \\ j+1 & 2 \le j \le m-1 \\ 2 & j = m \end{cases}$$

which has $I_k = 0$ for all $1 \le k \le m-1$. In particular, the bound (3) of Theorem 1.1 holds with

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

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 α when X_α 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

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

the collection $\mathbf{X}^\alpha = \{X_\beta^\alpha, \beta \in \mathcal{V}\}$ has the $\mathbf{X}$ distribution biased in direction α.

Furthermore, with I chosen proportional to EX_α, 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 α,

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

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 α, 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 $n \ge m \ge 1$, 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 $\alpha \in \mathcal{V}$ set

$$\mathcal{V}_\alpha = \{v \in \mathcal{V} : \alpha \le v \le \alpha + m - 1\}.$$

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 α we have $X_\alpha \le 1$ and

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

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 [35], the authors develop smooth function bounds for normal approximation for the case of 2-runs. Note that the construction given in Lemma 3.1 for this case is monotone, as for any i, letting

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

the number of m runs of $\{\xi'_j\}_{i=1}^n$, that is $Y^s = \sum_{i=1}^n \xi'_i \xi'_{i+1} \cdots \xi'_{i+m-1}$, is at least Y.

For the mean of Y clearly $\mu = np^m$. For the variance, now letting $n \ge 2m$ and using the fact that non-overlapping segments of the sequence are independent,

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

For the covariances,

$$\begin{aligned} \mathrm{Cov}(\xi_i \cdots \xi_{i+m-1}, \xi_{i+j} \cdots \xi_{i+j+m-1}) &= E(\xi_i \cdots \xi_{i+j-1} \xi_{i+j} \cdots \xi_{i+m-1} \xi_{i+m} \cdots \xi_{i+j+m-1}) - p^{2m} \\ &= p^{m+j} - p^{2m}, \end{aligned}$$

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

Size biasing the number of local extrema on graphs, for the purpose of normal approximation, was studied in [2] and [15]. For a given graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$, let $\mathcal{G}_v = \{\mathcal{V}_v, \mathcal{E}_v\}, v \in \mathcal{V}$, be a collection of isomorphic subgraphs of $\mathcal{G}$ such that $v \in \mathcal{V}_v$ and for all $v_1, v_2 \in \mathcal{V}$ the isomorphism from $\mathcal{G}_{v_1}$ to $\mathcal{G}_{v_2}$ maps v_1 to v_2. Let $\{C_g, g \in \mathcal{V}\}$ be a collection of independent and identically distributed random variables, and let X_v be defined by

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

Then the sum $Y = \sum_{v \in \mathcal{V}} X_v$ counts the number local maxima. In general one may define the neighbor distance d between two vertices $v, w \in \mathcal{V}$ by

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

Then for $v \in \mathcal{V}$ and $r = 0, 1, \ldots,$

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

is the set of vertices of $\mathcal{V}$ at distance at most r from v. We suppose that the given isomorphic graphs are of this form, that is, that there is some r such that $\mathcal{V}_v = \mathcal{V}_v(r)$ for all $v \in \mathcal{V}$. 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 $\mathcal{E} = \{\{v, w\} : d(v, w) = 1\}$; in this case d is the L^1 norm

$$d(v,w) = \sum_{i=1}^{p} |v_i - w_i|.$$

Considering the case where we call vertex v a local extreme value if the value C_v exceeds the values C_w over the immediate neighbors w of v, we take

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

the 1 accounting for v itself, and then $2p$ 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 $1/|\mathcal{V}_v|$ of having the largest value in its neighborhood, for the mean μ of Y we have

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

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 $d(v, w) = 1$,

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

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

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

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 $U_{m-k+1},\ldots,U_m$

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

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

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

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,

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

We conclude that (2) of Theorem 1.1 holds with $A = C\mu/\sigma^2$ and $B = C/2\sigma$ with μ, σ^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 of interest is the number of nonempty urns, for which Kolmogorov distance bounds to the normal were obtained in [12] and [32]. In [12], bounds were obtained for the uniform case where $p_i = 1/m$ for all $i = 1, \ldots, m$, while the bounds in [32] hold for the nonuniform case as well. In [30] 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 [30], 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

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

We first consider the uniform case. A construction in [30] 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 $\mathcal{B}$, 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 $\mathcal{B}$ be a Bernoulli variable with success probability $P(\mathcal{B} = 1) = \pi_{M_I}$, 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 $\mathcal{B} = 1$, 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 [30] 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 μ and σ^2 can be found in Theorem II.1.1 of [22],

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

Hence with μ and σ^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 μ and variance σ^2 obey

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

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 [30] 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 also using Theorem 2.4 in [30] for a bound on σ^2, 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 [16]. Given a collection $\mathcal{U} = \{U_1, U_2, \ldots, U_n\}$ 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 ρ centered at these n points, and S the number of balls isolated at distance ρ, that is, those points for which none of the other $n-1$ points lie within distance ρ. The random variables V and S are of fundamental interest in stochastic geometry, see [20] and [29]. If $n\to\infty$ and ρ remains fixed, both V and S satisfy a central limit theorem [20, 27, 31]. The L^1 distance of V, properly standardized, to the normal is studied in [10] 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 [16] 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

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

We will derive concentration of measure inequalities for V and S with the help of the bounded size biased couplings in [16].

Assume $d \ge 1$ and $n \ge 4$. Denote the mean and variance of V by μ_V and σ_V^2, respectively, and likewise for S, leaving their dependence on n and ρ implicit. Let $\pi_d = \pi^{d/2}/\Gamma(1+d/2)$, the volume of the unit sphere in $\mathbb{R}^d$, and for fixed ρ let $\phi = \pi_d\rho^d$. For $0 \le r \le 2$ 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

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

From [16], 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

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

It is shown in [16], 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 $|V^s - V| \le \phi$. 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 μ_V and σ_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 κ_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 [16],

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

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

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

4 Applications: unbounded couplings

One of the major drawbacks of Theorem 1.1 is the hypothesis that $|Y^s - Y|$ be almost surely bounded with probability one. In this section we derive concentration of measure inequalities for two examples where $Y^s - Y$ is not bounded: the number of isolated vertices in the Erdös-Rényi random graph model, and the nonnegative infinitely divisible distributions with certain associated moment generating functions which satisfy a boundedness condition. For the latter, compound Poisson distributions will be our main illustration.

The arguments that follow make use of the inequality

$$\frac{e^y - e^x}{y - x} = \int_0^1 e^{ty+(1-t)x} dt \leq \int_0^1 (te^y + (1-t)e^x) dt = \frac{e^y + e^x}{2} \quad \text{for all } x \neq y, \tag{21}$$

which holds by the convexity of the exponential function;

4.1 Number of isolated vertices in the Erdös Rényi random graph model

Let $K_{n,p}$ be the random graph on the vertices $\mathcal{V} = \{1, 2, \ldots, n\}$, with the indicators X_{vw} of the presence of edges between two unequal vertices v and w being independent Bernoulli $p \in (0, 1)$ variables, and $X_{vv} = 0$ for all $v \in \mathcal{V}$. Recall that the degree of a vertex $v \in \mathcal{V}$ is the number of edges incident on v,

$$d(v) = \sum_{w \in \mathcal{V}} X_{vw}. \tag{22}$$

The problem of approximating the distribution of the number of vertices v with degree $d(v) = d$ for some fixed d was considered in [6], and a smooth function bound to the multivariate normal for a vector whose components count the number of vertices of some fixed degrees was given in [17].

Here we study the number of isolated vertices $Y_{n,p}$ of $K_{n,p}$, that is, those vertices which have no incident edges, given by

$$Y_{n,p} = \sum_{v \in \mathcal{V}} 1(d(v) = 0).$$

In [23], the mean μ and variance σ^2 of $Y_{n,p}$ are given as

$$\mu_{n,p} = n(1-p)^{n-1} \quad \text{and} \quad \sigma^2_{n,p} = n(1-p)^{n-1}(1 + np(1-p)^{n-2} - (1-p)^{n-2}), \tag{23}$$

where also Kolmogorov distance bounds to the normal were obtained, and asymptotic normality shown when

$$n^2 p \to \infty \quad \text{and} \quad np - \log(n) \to -\infty.$$

O'Connell [28] shows an asymptotic large deviation principle holds for $Y_{n,p}$. Raič [33] obtained nonuniform large deviation bounds in some generality for random variables W with $E(W) = 0$ and $\mathrm{Var}(W) = 1$ of the form,

$$\frac{P(W \geq t)}{1 - \Phi(t)} \leq e^{t^3 \beta(t)/6}(1 + Q(t)\beta(t)) \quad \text{for all } t \geq 0, \tag{24}$$

where $\Phi(t)$ denotes the distribution function of a standard normal variate and $Q(t)$ is a quadratic in t. Although in general the expression for $\beta(t)$ is not simple, when W is $Y_{n,p}$ properly standardized and $np \to c$ as $n \to \infty$, then (24) holds for all n sufficiently large with

$$\beta(t) = \frac{C_1}{\sqrt{n}} \exp\left(\frac{C_2 t}{\sqrt{n}} + C_3(e^{C_4 t/\sqrt{n}} - 1)\right)$$

for some constants C_1, C_2, C_3 and C_4. For t of order $n^{1/2}$, for instance, the function $\beta(t)$ will be small as $n \to \infty$, allowing an approximation of the deviation probability $P(W \geq t)$ by the normal, to within some factors. Theorem 4.1 below, by contrast, provides a non-asymptotic bound, that is, not relying on any limiting relations between n and p, with explicit constants, which hold for every n. Moreover, the bound is of order e^{-at^2} over some range of t, and of worst case order e^{-bt}, for the right tail by (27), and e^{-ct^2} by (26) for the left tail, where a, b and c are explicit, with the bounds holding for all $t \in \mathbb{R}$.

For notational ease, we keep the dependence on n and p implicit in the sequel.

Theorem 4.1. *Let K denote the random graph on n vertices where each edge is present with probability $p \in (0,1)$, independently of all other edges, and let Y denote the number of isolated vertices in K. Then for all $t > 0$,*

$$P\left(\frac{Y-\mu}{\sigma} \geq t\right) \leq \inf_{\theta \geq 0} \exp(-\theta t + H(\theta)) \quad \text{where} \quad H(\theta) = \frac{\mu}{2\sigma^2}\int_0^\theta s\gamma_s ds, \tag{25}$$

with the mean μ and variance σ^2 of Y given in (23), and

$$\gamma_s = 2e^{2s}\left(1 + \frac{pe^s}{1-p}\right)^n + \beta + 1 \quad \text{where } \beta = (1-p)^{-n}.$$

For the left tail, for all $t > 0$,

$$P\left(\frac{Y-\mu}{\sigma} \leq -t\right) \quad \leq \quad \exp\left(-\frac{t^2}{2}\frac{\sigma^2}{\mu(\beta+1)}\right). \tag{26}$$

Remark 4.1. *Though the minimization in (25) is admittedly cumbersome, useful bounds may be obtained by restricting the minimization to $\theta \in [0, \theta_0]$ for some θ_0. In this case, as γ_s is an increasing function of s, we have*

$$H(\theta) \leq \frac{\mu}{4\sigma^2}\gamma_{\theta_0}\theta^2 \quad \text{for } \theta \in [0, \theta_0].$$

The quadratic $-\theta t + \mu\gamma_{\theta_0}\theta^2/(4\sigma^2)$ in θ is minimized at $\theta = 2t\sigma^2/(\mu\gamma_{\theta_0})$. When this value falls in $[0, \theta_0]$ we obtain the first bound in (27), while otherwise setting $\theta = \theta_0$ yields the second.

$$P(\frac{Y-\mu}{\sigma} \geq t) \leq \begin{cases} \exp(-\frac{t^2\sigma^2}{\mu\gamma_{\theta_0}}) & \text{for } t \in [0, \theta_0\mu\gamma_{\theta_0}/(2\sigma^2)] \\ \exp(-\theta_0 t + \frac{\mu\gamma_{\theta_0}\theta_0^2}{4\sigma^2}) & \text{for } t \in (\theta_0\mu\gamma_{\theta_0}/(2\sigma^2), \infty). \end{cases} \tag{27}$$

Though Theorem 4.1 is not an asymptotic, as it gives bounds for any specific n and p, when $np \to c$ as $n \to \infty$ we have

$$\frac{\sigma^2}{\mu} \to 1 + ce^{-c} - e^{-c}, \quad \beta + 1 \to e^c + 1 \quad \text{and} \quad \gamma_s \to 2e^{2s+ce^s} + e^c + 1 \quad \text{as } n \to \infty.$$

Hence, the left tail bound (26), for example, in this asymptotic behaves as

$$\lim_{n\to\infty} \exp\left(-\frac{t^2}{2}\frac{\sigma^2}{\mu(\beta+1)}\right) = \exp\left(-\frac{t^2}{2}\frac{1+ce^{-c}-e^{-c}}{e^c+1}\right).$$

Proof. We first review the construction of Y^s, having the Y size bias distribution, as given in [17]. Let K, a particular realization of $K(n,p)$, be given, and let Y be the number of isolated vertices for this realization. To size bias Y, choose one of the n vertices of K uniformly. If the chosen vertex, say V, is already isolated, we do nothing and set $K^s = K$. Otherwise obtain K^s by deleting all the edges connected to K. Then Y^s, the number of isolated vertices of K^s, has the Y size biased distribution.

To derive the needed properties of this coupling, let $N(v)$ be the set of neighbors of $v \in \mathcal{V}$, and $\mathcal{T}$ the collection of isolated vertices of K, that is, with $d(v)$, the degree of v, given in (22),

$$N(v) = \{w : X_{vw} = 1\} \quad \text{and} \quad \mathcal{T} = \{v : d(v) = 0\}.$$

Note that $Y = |\mathcal{T}|$. Since all edges incident to the chosen V are removed in order to form K^s, any neighbor of V which had degree one thus becomes isolated, and V also becomes isolated if it was not so earlier. As all others vertices are otherwise unaffected, as far as their being isolated or not, we have

$$Y^s - Y = d_1(V) + 1(d(V) \neq 0) \quad \text{where} \quad d_1(V) = \sum_{w \in N(V)} 1(d(w) = 1), \tag{28}$$

so in particular the coupling is monotone. Since $d_1(V) \leq d(V)$, (28) yields

$$Y^s - Y \leq d(V) + 1. \tag{29}$$

By (21), using that the coupling is monotone, for $\theta \geq 0$ we have

$$\begin{aligned} E(e^{\theta Y^s} - e^{\theta Y}) &\leq \frac{\theta}{2} E\left((Y^s - Y)(e^{\theta Y^s} + e^{\theta Y})\right) \\ &= \frac{\theta}{2} E\left(\exp(\theta Y)(Y^s - Y)\left(\exp(\theta(Y^s - Y)) + 1\right)\right) \\ &= \frac{\theta}{2} E\left\{\exp(\theta Y) E((Y^s - Y)(\exp(\theta(Y^s - Y)) + 1)|\mathcal{T})\right\}. \end{aligned} \tag{30}$$

Now using that $Y^s = Y$ when $V \in \mathcal{T}$, and (29), we have

$$\begin{aligned} & E((Y^s - Y)(\exp(\theta(Y^s - Y)) + 1)|\mathcal{T}) \\ &\leq E((d(V) + 1)(\exp(\theta(d(V) + 1)) + 1)1(V \notin \mathcal{T})|\mathcal{T}) \\ &\leq e^\theta E\left(\left(d(V)e^{\theta d(V)} + e^{\theta d(V)} + d(V)\right) 1(V \notin \mathcal{T})|\mathcal{T}\right) + 1. \end{aligned} \tag{31}$$

Note that since V is chosen independently of K,

$$\mathcal{L}(d(V)1(V \notin \mathcal{T})|\mathcal{T}) = P(V \notin \mathcal{T})\mathcal{L}(\mathit{Bin}(n - 1 - Y, p)|\, \mathit{Bin}(n - 1 - Y, p) > 0) + P(V \in \mathcal{T})\delta_0, \tag{32}$$

where δ_0 is point mass at zero. By (32), and that the mass function of the conditioned binomial there is

$$P(d(V) = k|\mathcal{T}, V \notin \mathcal{T}) = \begin{cases} \binom{n-1-Y}{k} \frac{p^k(1-p)^{n-1-Y-k}}{1-(1-p)^{n-1-Y}} & \text{for } 1 \leq k \leq n - 1 - Y \\ 0 & \text{otherwise,} \end{cases}$$

it can be easily verified that the conditional moment generating function of $d(V)$ and its first derivative are bounded by

$$\begin{aligned} E(e^{\theta d(V)}1(V \notin \mathcal{T})|\mathcal{T}) &\leq \frac{(pe^\theta + 1 - p)^{n-1-Y} - (1 - p)^{n-1-Y}}{1 - (1 - p)^{n-1-Y}} \quad \text{and} \\ E(d(V)e^{\theta d(V)}1(V \notin \mathcal{T})|\mathcal{T}) &\leq \frac{(n - 1 - Y)(pe^\theta + 1 - p)^{n-2-Y}pe^\theta}{1 - (1 - p)^{n-1-Y}}. \end{aligned}$$

By the mean value theorem applied to the function $f(x) = x^{n-1-Y}$, for some $\xi \in (1 - p, 1)$ we have

$$1 - (1 - p)^{n-1-Y} = f(1) - f(1 - p) = (n - 1 - Y)p\xi^{n-2-Y} \geq (n - 1 - Y)p(1 - p)^n.$$

Hence, recalling $\theta \geq 0$,

$$\begin{aligned} E(d(V)e^{\theta d(V)}1(V \notin \mathcal{T})|\mathcal{T}) &\leq \frac{(n - 1 - Y)(pe^\theta + 1 - p)^n pe^\theta}{1 - (1 - p)^{n-1-Y}} \\ &\leq \frac{(n - 1 - Y)(pe^\theta + 1 - p)^n pe^\theta}{(n - 1 - Y)p(1 - p)^n} \\ &= \alpha_\theta \quad \text{where} \quad \alpha_\theta = e^\theta \left(1 + \frac{pe^\theta}{1 - p}\right)^n. \end{aligned} \tag{33}$$

Similarly applying the mean value theorem to $f(x) = (x+1-p)^{n-1-Y}$, for some $\xi \in (0, pe^\theta)$ we have

$$\begin{aligned} E(e^{\theta d(V)} 1(V \notin \mathcal{T})|\mathcal{T}) &\leq \frac{(n-1-Y)(\xi+(1-p))^{n-2-Y} pe^\theta}{1-(1-p)^{n-1-Y}} \\ &\leq \frac{(n-1-Y)(pe^\theta+(1-p))^{n-2-Y} pe^\theta}{1-(1-p)^{n-1-Y}} \\ &\leq \alpha_\theta, \end{aligned} \tag{34}$$

as in (33).

Next, to handle the second to last term in (31) consider

$$E(d(V)1(V \notin \mathcal{T})|\mathcal{T}) \leq \frac{(n-1-Y)p}{1-(1-p)^{n-1-Y}} \leq \frac{(n-1-Y)p}{(n-1-Y)p(1-p)^n} = \beta \quad \text{where} \quad \beta = (1-p)^{-n}. \tag{35}$$

Applying inequalities (33),(34) and (35) to (31) yields

$$E((Y^s - Y)(\exp(\theta(Y^s - Y)) + 1)|\mathcal{T}) \leq \gamma_\theta \quad \text{where} \quad \gamma_\theta = 2e^\theta \alpha_\theta + \beta + 1. \tag{36}$$

Hence we obtain, using (30),

$$E(e^{\theta Y^s} - e^{\theta Y}) \leq \frac{\theta \gamma_\theta}{2} E(e^{\theta Y}) \quad \text{for all } \theta \geq 0.$$

Letting $m(\theta) = E(e^{\theta Y})$ thus yields

$$m'(\theta) = E(Ye^{\theta Y}) = \mu E(e^{\theta Y^s}) \leq \mu \left(1 + \frac{\theta \gamma_\theta}{2}\right) m(\theta). \tag{37}$$

Setting

$$M(\theta) = E(\exp(\theta(Y-\mu)/\sigma)) = e^{-\theta\mu/\sigma} m(\theta/\sigma),$$

differentiating and using (37), we obtain

$$\begin{aligned} M'(\theta) &= \frac{1}{\sigma} e^{-\theta\mu/\sigma} m'(\theta/\sigma) - \frac{\mu}{\sigma} e^{-\theta\mu/\sigma} m(\theta/\sigma) \\ &\leq \frac{\mu}{\sigma} e^{-\theta\mu/\sigma} (1 + \frac{\theta\gamma_\theta}{2\sigma}) m(\theta/\sigma) - \frac{\mu}{\sigma} e^{-\theta\mu/\sigma} m(\theta/\sigma) \\ &= e^{-\theta\mu/\sigma} \frac{\mu\theta\gamma_\theta}{2\sigma^2} m(\theta/\sigma) = \frac{\mu\theta\gamma_\theta}{2\sigma^2} M(\theta). \end{aligned} \tag{38}$$

Since $M(0) = 1$, (38) yields upon integration of $M'(s)/M(s)$ over $[0, \theta]$,

$$\log(M(\theta)) \leq H(\theta) \quad \text{so that} \quad M(\theta) \leq \exp(H(\theta)) \quad \text{where} \quad H(\theta) = \frac{\mu}{2\sigma^2} \int_0^\theta s\gamma_s ds.$$

Hence for $t \geq 0$,

$$P(\frac{Y-\mu}{\sigma} \geq t) \leq P(\exp(\frac{\theta(Y-\mu)}{\sigma}) \geq e^{\theta t}) \leq e^{-\theta t} M(\theta) \leq \exp(-\theta t + H(\theta)).$$

As the inequality holds for all $\theta \geq 0$, it holds for the θ achieving the minimal value, proving (25).

For the left tail bound let $\theta < 0$. Since $Y^s \geq Y$ and $\theta < 0$, using (21) and (29) we obtain

$$\begin{aligned} E(e^{\theta Y} - e^{\theta Y^s}) &\leq \frac{|\theta|}{2} E\left((e^{\theta Y} + e^{\theta Y^s})(Y^s - Y)\right) \\ &\leq |\theta| E(e^{\theta Y}(Y^s - Y)) \\ &= |\theta| E(e^{\theta Y} E(Y^s - Y|\mathcal{T})) \\ &\leq |\theta| E(e^{\theta Y} E((d(V)+1)1(V \notin \mathcal{T}))|\mathcal{T})). \end{aligned}$$

Applying (35) we obtain

$$E(e^{\theta Y} - e^{\theta Y^s}) \leq (\beta+1)|\theta| E(e^{\theta Y}),$$

and therefore

$$m'(\theta) = \mu E(e^{\theta Y^s}) \geq \mu \left(1 + (\beta+1)\theta\right) m(\theta).$$

Hence for $\theta < 0$,

$$\begin{aligned} M'(\theta) &= \frac{1}{\sigma} e^{-\theta\mu/\sigma} m'(\theta/\sigma) - \frac{\mu}{\sigma} e^{-\theta\mu/\sigma} m(\theta/\sigma) \\ &\geq \frac{\mu}{\sigma} e^{-\theta\mu/\sigma} ((1 + (\beta+1)\theta/\sigma) m(\theta/\sigma)) - \frac{\mu}{\sigma} e^{-\theta\mu/\sigma} m(\theta/\sigma) \\ &= \frac{\mu(\beta+1)\theta}{\sigma^2} M(\theta). \end{aligned}$$

Dividing by $M(\theta)$ and integrating over $[\theta, 0]$ yields

$$\log(M(\theta)) \leq \frac{\mu(\beta+1)\theta^2}{2\sigma^2}. \tag{39}$$

The inequality in (39) implies that for all $t > 0$ and $\theta < 0$,

$$P(\frac{Y-\mu}{\sigma} \leq -t) \leq \exp(\theta t + \frac{\mu(\beta+1)\theta^2}{2\sigma^2}).$$

Taking $\theta = -t\sigma^2/(\mu(\beta+1))$ we obtain (26). □ □

4.2 Infinitely divisible and compound Poisson distributions

The examples in this section generalize the application of Theorem 1.1 from the case where Y is Poisson with parameter $\lambda > 0$. In this case, Y admits a bounded coupling to a variable with its size bias distribution due to the characterization

$$E[Yf(Y)] = \lambda E[f(Y+1)] \quad \text{if and only if } Y \sim \text{Poisson}(\lambda), \tag{40}$$

which forms the basis of the Chen-Stein Poisson approximation method, see [9, 5]. In particular we may take $Y^s = Y + 1$, and, therefore $C = 1$. As the mean and variance for the Poisson are equal, and the coupling is monotone, applying Theorem 1.1 we obtain the following result.

Proposition 4.1. *If $Y \sim$ Poisson(λ), then for all $t > 0$,*

$$P\left(\frac{Y-\lambda}{\sqrt{\lambda}} \leq -t\right) \leq \exp\left(-\frac{t^2}{2}\right) \quad \textit{and} \quad P\left(\frac{Y-\lambda}{\sqrt{\lambda}} \geq t\right) \leq \exp\left(-\frac{t^2}{2+t\lambda^{-1/2}}\right).$$

The Poisson distribution is infinitely divisible, and also a special case of the compound Poisson distributions. We generalize Proposition 4.1 in these directions.

4.2.1 Infinitely divisible distributions

When Y is Poisson then by (40) $Y^s = Y + 1$ and we may write

$$Y^s = Y + X \tag{41}$$

with X and Y independent. Theorem 5.3 of [39] shows that if Y is nonnegative with finite mean then (41) holds if and only if Y is infinitely divisible. Hence, in this case, a coupling of Y to Y^s may be achieved by generating the independent variable X and adding it to Y. Since Y^s is always stochastically larger than Y we must have $X \geq 0$, and therefore this coupling is monotone. In addition $Y^s - Y = X$ so the coupling is bounded if and only if X is bounded. When X is unbounded, Theorem 4.2 provides concentration of measure inequalities for Y under appropriate growth conditions on two generating functions in Y and X. We assume without further mention that Y is nontrivial, and note that therefore the means of both Y and X are positive.

Theorem 4.2. *Let Y have a nonnegative infinitely divisible distribution and suppose that there exists $\gamma > 0$ so that $E(e^{\gamma Y}) < \infty$. Let X have the distribution such that (41) holds when Y and X are independent, and assume $E(Xe^{\gamma X}) = C < \infty$. Letting $\mu = E(Y), \sigma^2 = Var(Y), \nu = E(X)$ and $K = (C+\nu)/2$, the following concentration of measure inequalities hold for all $t > 0$,*

$$P\left(\frac{Y-\mu}{\sigma} \geq t\right) \leq \begin{cases} \exp\left(-\frac{t^2\sigma^2}{2K\mu}\right) & \text{for } t \in [0, \gamma K\mu/\sigma^2) \\ \exp\left(-\gamma t + \frac{K\mu\gamma^2}{2\sigma^2}\right) & \text{for } t \in [\gamma K\mu/\sigma^2, \infty), \end{cases} \quad \text{and } P\left(\frac{Y-\mu}{\sigma} \leq -t\right) \leq \exp\left(-\frac{t^2\sigma^2}{2\nu\mu}\right).$$

Proof. The proof is similar to that of Theorem 4.1. Since $Y^s = Y + X$ with Y and X independent and $X \geq 0$, using (21) with $\theta \in (0, \gamma)$ we have,

$$\begin{aligned} E(e^{\theta Y^s} - e^{\theta Y}) &= E(e^{\theta(X+Y)} - e^{\theta Y}) \leq \frac{1}{2}E\left(\theta X(e^{\theta(X+Y)} + e^{\theta Y})\right) \\ &= \frac{\theta}{2}E\left(X(e^{\theta X}+1)e^{\theta Y}\right) = \frac{\theta}{2}E\left(X(e^{\theta X}+1)\right)E(e^{\theta Y}) \\ &\leq \frac{\theta}{2}(E(Xe^{\gamma X}) + E(X))E(e^{\theta Y}) \\ &= K\theta m(\theta) \quad \text{where } K = (C+\nu)/2 \text{ and } m(\theta) = E(e^{\theta Y}). \end{aligned}$$

Now adding $m(\theta)$ to both sides yields

$$E(e^{\theta Y^s}) \leq (1 + K\theta)m(\theta),$$

and therefore

$$m'(\theta) = E(Ye^{\theta Y}) = \mu E(e^{\theta Y^s}) \leq \mu(1 + K\theta)m(\theta). \tag{42}$$

Again, with $M(\theta)$ the moment generating function of $(Y-\mu)/\sigma$,

$$M(\theta) = Ee^{\theta(Y-\mu)/\sigma} = e^{-\theta\mu/\sigma}m(\theta/\sigma),$$

by (42) we have,

$$\begin{aligned} M'(\theta) &= -(\mu/\sigma)e^{-\theta\mu/\sigma}m(\theta/\sigma) + e^{-\theta\mu/\sigma}m'(\theta/\sigma)/\sigma \\ &\leq -(\mu/\sigma)e^{-\theta\mu/\sigma}m(\theta/\sigma) + (\mu/\sigma)e^{-\theta\mu/\sigma}\left(1 + K\frac{\theta}{\sigma}\right)m(\theta/\sigma) \\ &= (\mu/\sigma^2)K\theta M(\theta). \end{aligned} \tag{43}$$

Integrating, and using the fact that $M(0) = 1$ yields

$$M(\theta) \leq \exp\left(\frac{K\mu\theta^2}{2\sigma^2}\right) \quad \text{for } \theta \in (0, \gamma).$$

Hence for a fixed $t > 0$, for all $\theta \in (0, \gamma)$,

$$P\left(\frac{Y-\mu}{\sigma} \geq t\right) \leq e^{-\theta t}M(\theta) \leq \exp\left(-\theta t + \frac{K\mu\theta^2}{2\sigma^2}\right).$$

The infimum of the quadratic in the exponent is attained at $\theta = t\sigma^2/K\mu$. When this value lies in $(0,\gamma)$ we obtain the first, right tail bound, for t in the bounded interval, while setting $\theta = \gamma$ yields the second.

Moving on to the left tail bound, using (21) for $\theta < 0$ yields

$$E(e^{\theta Y} - e^{\theta Y^s}) \quad \leq \quad -\frac{\theta}{2}E((Y^s - Y)(e^{\theta Y} + e^{\theta Y^s})) \leq -\theta E(Xe^{\theta Y}) = -\theta E(X)E(e^{\theta Y}).$$

Rearranging we obtain

$$m'(\theta) = \mu E(e^{\theta Y^s}) \geq \mu(1 + \theta\nu)m(\theta).$$

Following calculations similar to (43) one obtains

$$M'(\theta) \geq (\mu/\sigma^2)\nu\theta M(\theta) \quad \text{for all } \theta < 0,$$

which upon integration over $[\theta, 0]$ yields

$$M(\theta) \leq \exp\left(\frac{\nu\mu\theta^2}{2\sigma^2}\right) \quad \text{for all } \theta < 0.$$

Hence for any fixed $t > 0$, for all $\theta < 0$,

$$P\left(\frac{Y-\mu}{\sigma} \leq -t\right) \leq e^{\theta t}M(\theta) \leq \exp\left(\theta t + \frac{\nu\mu\theta^2}{2\sigma^2}\right). \tag{44}$$

Substituting $\theta = -t\sigma^2/(\nu\mu)$ in (44) yields the lower tail bound, thus completing the proof. □ □

Though Theorem 4.2 applies in principle to all nonnegative infinitely divisible distributions with generating functions for Y and X that satisfy the given growth conditions, we now specialize to the subclass of compound Poisson distributions, over which it is always possible to determine the independent increment X. Not too much is sacrificed in narrowing the focus to this case, since a nonnegative infinitely divisible random variable Y has a compound Poisson distribution if and only if $P(Y = 0) > 0$.

4.2.2 Generalized variance

Let $\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_m$ be a sample of size $m = n + p + 1$ with distribution $\mathcal{N}_p(\mathbf{0}, \mathbf{\Sigma})$, the p-variate normal with mean $\mathbf{0}$ and covariance matrix $\mathbf{\Sigma}$. The usual estimate of $\mathbf{\Sigma}$ is given by the sample covariance matrix $V/(m-1)$ where

$$V = \sum_{i=1}^{m}(\mathbf{Y}_i - \overline{\mathbf{Y}})(\mathbf{Y}_i - \overline{\mathbf{Y}})^t \quad \text{with} \quad \overline{\mathbf{Y}} = \frac{1}{m}\sum_{i=1}^{m}\mathbf{Y}_i.$$

The sample covariance matrix follows the Wishart distribution $\mathcal{W}_p(m, \mathbf{\Sigma})$ with m degrees of freedom, and its determinant is known as the generalized variance, which gives, in some sense, an estimate of a one dimensional summary of the variability inherent in the underlying distribution generating the sample.

It can be shown that

$$\text{Det}(V) = \text{Det}(\mathbf{\Sigma})D_{p,n} \quad \text{where} \quad D_{p,n} = \prod_{k=0}^{p-1} X_k \tag{45}$$

where $X_k \sim \chi^2(n + k + 1)$ and $X_0, X_1, \ldots, X_{p-1}$ are mutually independent; the reader is referred to [36] for a proof. The case of $p = 2$ was treated by Wilks [40], who showed $2(D_{2,n})^{1/2} = 2(X_0X_1)^{1/2} \sim \chi^2(2(n+1))$. The exact distribution of $W = p(D_{p,n})^{1/p}$ is not simple in general and approximations can be helpful. Hoel [21] approximated W with $\chi^2(p(n+1))$, while Luk [25] approximated the distribution of W by the Gamma distribution $\Gamma(EW/2, 1/2)$ having the same mean as W. Here we parameterize the Gamma distribution $\Gamma(\alpha, \lambda)$ as the one with density

$$p(x; \alpha, \lambda) = \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)} \quad \text{for } x > 0.$$

Although a size bias coupling is available for W (see [25]), obtaining results by its direct application is elusive. However, the following stochastic order result, in conjunction with Theorem 4.2, can be used to indirectly obtain a concentration of measure result for the generalized variance. Recall that we say the random variable ξ_2 stochastically dominates ξ_1, and write $\xi_2 \succ \xi_1$, when

$$P(\xi_2 > t) \geq P(\xi_1 > t) \quad \text{for any } t \in \mathbb{R}. \tag{46}$$

From Gordon [19] we have

$$W_2 \succ W \succ W_1 \tag{47}$$

where

$$W_2 \sim \chi^2(p(n+1) + (p-1)(p-2)/2) \quad \text{and} \quad W_1 \sim \chi^2(p(n+1)).$$

Letting μ_1 and μ_2 be the means of W_1 and W_2, respectively, we have

$$\mu_1 = p(n+1) + (p-1)(p-2)/2 \quad \text{and} \quad \mu_2 = p(n+1), \tag{48}$$

and likewise for their variances

$$\sigma_1^2 = 2p(n+1) + (p-1)(p-2) \quad \text{and} \quad \sigma_2^2 = 2p(n+1), \tag{49}$$

from (47) and (46), we have for all $t \in \mathbb{R}$

$$P\left(\frac{W - \mu_2}{\sigma_2} \geq t\right) \leq P\left(\frac{W_2 - \mu_2}{\sigma_2} \geq t\right) \quad \text{and} \quad P\left(\frac{W - \mu_1}{\sigma_1} \geq t\right) \geq P\left(\frac{W_1 - \mu_1}{\sigma_1} \geq t\right). \tag{50}$$

In general, when $Y \sim \Gamma(\alpha, \lambda)$, then as

$$Ee^{\gamma Y} = \left(1 - \frac{\gamma}{\lambda}\right)^{-\alpha},$$

we have $Ee^{\gamma Y} < \infty$ for all $\gamma \in (0, \lambda)$. Additionally, the size bias distribution Y^s is given by $Y^s \sim \Gamma(\alpha+1, \lambda)$, so that (41) is satisfied with $X \sim \Gamma(1, \lambda)$, an exponential random variable with mean $1/\lambda$, for which

$$EXe^{\gamma X} = \frac{1}{\lambda}\left(1 - \frac{\gamma}{\lambda}\right)^{-2} \quad \text{for } \gamma \in (0, \lambda).$$

Hence the conclusions of Theorem 4.2 apply to Y taking any $\gamma \in (0, \lambda)$ and

$$C = \frac{1}{\lambda}\left(1 - \frac{\gamma}{\lambda}\right)^{-2}, \quad \nu = \frac{1}{\lambda}, \quad \text{and} \quad K = \frac{1}{2\lambda}\left((1 - \frac{\gamma}{\lambda})^{-2} + 1\right).$$

Specializing to the case $\lambda = 1/2$ yields inequalities for the distributions of W_1 and W_2. These inequalities may then be applied to W through (50). For example, from (50) and the first inequality in Theorem 4.2, taking $\gamma = 1/4$ for illustration, we conclude that, with μ_2 and σ_2^2 as in (48) and (49) respectively,

$$P\left(\frac{W - \mu_2}{\sigma_2} \geq t\right) \leq e^{-t^2/5} \quad \text{for } t \in [0, 5/8) \text{ and} \quad P\left(\frac{W - \mu_2}{\sigma_2} \geq t\right) \leq e^{-t/4 + 5/64} \quad \text{for } t \in [5/8, \infty).$$

4.2.3 Compound Poisson distribution

One important subfamily of the infinitely divisible distributions are the compound Poisson distributions, that is, those distributions that are given by

$$Y = \sum_{i=1}^{N} Z_i, \quad \text{where } N \sim \text{Poisson}(\lambda), \text{ and } \{Z_i\}_{i=1}^{\infty} \text{ are independent and distributed as } Z. \tag{51}$$

Compound Poisson distributions are popular in several applications, such as insurance mathematics, seismological data modelling, and reliability theory; the reader is referred to [4] for a detailed review.

Although Z is not in general required to be nonnegative, in order to be able to size bias Y we restrict ourselves to this situation. It is straightforward to verify that when the moment generating function $m_Z(\theta) = Ee^{\theta Z}$ of Z is finite, then the moment generating function $m(\theta)$ of Y is given by

$$m(\theta) = \exp(-\lambda(1 - m_Z(\theta))).$$

In particular $m(\theta)$ is finite whenever $m_Z(\theta)$ is finite. As Y in (51) is infinitely divisible the equality (41) holds for some X; the following lemma determines the distribution of X in this particular case, see [1] for a proof.

Lemma 4.1. *Let Y have the compound Poisson distribution as in (51) where Z is nonnegative and has finite, positive mean. Then*

$$Y^s = Y + Z^s,$$

has the Y size biased distribution, where Z^s has the Z size bias distribution and is independent of N and $\{Z_i\}_{i=1}^{\infty}$.

Proof. Let $\phi_V(u) = Ee^{iuV}$ for any random variable V. If V is nonnegative and has finite positive mean, using $f(y) = e^{iuy}$ in (1) results in

$$\phi_{V^s}(u) = \frac{1}{EV}\left(EVEe^{iuV^s}\right) = \frac{1}{EV}EVe^{iuV} = \frac{1}{iEV}\phi_V'(u). \tag{52}$$

It is easy to check that the characteristic function of the compound Poisson Y in (51) is given by

$$\phi_Y(u) = \exp(-\lambda(1 - \phi_Z(u))), \tag{53}$$

and letting $EZ = \vartheta$, that $EY = \lambda\vartheta$. Now applying (52) and (53) results in

$$\phi_{Y^s}(u) = \frac{1}{i\lambda\vartheta}\phi_Y'(u) = \frac{1}{i\vartheta}\phi_Y(u)\phi_Z'(u) = \phi_Y(u)\phi_{Z^s}(u).$$

$\square$

To illustrate Lemma 4.1, consider the Cramér-Lundberg model [11] from insurance mathematics. Suppose an insurance company starts with an initial capital u_0, and premium is collected at the constant rate α. Claims arrive according to a homogenous Poisson process $\{N_\tau\}_{\tau \geq 0}$ with rate λ, and the claim sizes are independent with common distribution Z. The aggregate claims Y_τ made by time $\tau \geq 0$ is therefore given by (51) with N and λ replaced by N_τ and λ_τ, respectively.

Distributions for Z which are of interest for applications include the Gamma, Weibull, and Pareto, among others. For concreteness, if $Z \sim \text{Gamma}(\alpha, \beta)$ then $Z^s \sim \text{Gamma}(\alpha+1, \beta)$, and the mean ν of the increment Z^s, and the mean μ_τ and variance σ_τ^2 of Y_τ, are given by

$$\nu = \quad (\alpha+1)\beta, \quad \mu_\tau = \lambda\tau\alpha\beta \quad \text{and} \quad \sigma_\tau^2 = \lambda\tau\beta^2\alpha.$$

The conditions of Theorem 4.2 are satisfied with any $\gamma \in (0, 1/\beta)$ since $E(e^{\theta Y}) < \infty$ and $E(Z^s e^{\theta Z^s}) < \infty$ for all $\theta < 1/\beta$. Taking $\gamma = 1/(M\beta)$ for $M > 1$ for example, yields

$$C = E(Z^s e^{\gamma Z^s}) = (\alpha + 1)\beta(\frac{M}{M-1})^{\alpha+2}.$$

For instance, the lower tail bound of Theorem 4.2 now yields a bound on the probability that the aggregate claims by time τ will be 'small', of

$$P\left(\frac{Y_\tau - \mu_\tau}{\sigma_\tau} \leq -t\right) \leq \exp\left(-\frac{t^2}{2(\alpha+1)}\right).$$

It should be noted that in some applications one may be interested in Z which are heavy tailed, and hence do not satisfy the conditions in Theorem 4.2.

References

[1] Arratia, R., and Goldstein, L. Size bias, sampling, the waiting time paradox, and infinite divisibility: when is the increment independent? *Preprint*

[2] Baldi, P., Rinott, Y. and Stein, C. (1989). 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., Academic Press, 59-81.

[3] Barbour, A.D. and Chen, L.H.Y(2005). An Introduction to Stein's Method, Chen,L.H.Y and Barbour,A.D. eds,Lecture Notes Series No. 4, Institute for Mathematical Sciences, National University of Singapore, Singapore University Press and World Scientific 2005, 1-59.

[4] Barbour, A.D. and Chryssaphinou, O.(2001). Compound Poisson approximation: A user's guide, *Ann. Appl. Probab.*, **11**, 964-1002.

[5] Barbour, A.D., Holst, L., and Janson, S. (1992). Poisson Approximation, Oxford University Press.

[6] Barbour, A.D., Karoński, M. and Ruciński,A.(1989). A central limit theorem for decomposable random variables with applications to random graphs, *J. Combinatorial Theory B*, **47**, 125-145.

[7] Bayer, D. and Diaconis, P.(1992). Trailing the Dovetail Shuffle to its Lair. *Ann. of Appl. Probab.* **2**, 294-313.

[8] Chatterjee, S.(2007). Stein's method for concentration inequalities, *Probab. Theory Related Fields*, **138**, 305-321.

[9] Chen, L.H.Y (1975). Poisson approximation for dependent trials, *Ann. Probab.*, **3**, 534-545.

[10] Chatterjee, S.(2008) A new method of normal approximation. *Ann. Probab.*, **4**, 1584-1610.

[11] Embrechts, P. and Klüppelberg, C.(1993). Some aspects of insurance mathematics, *Th. Probab. Appl.*, **38**, 262-295.

[12] Englund, G.(1981). A remainder term estimate for the normal approximation in classical occupancy, *Ann. Probab.*, **9**, 684-692.

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

[14] Ghosh, S. and Goldstein, L.(2009). Concentration of measures via size biased couplings, to appear in *Probab. Th. Rel. Fields.*

[15] Goldstein, L.(2005). Berry Esseen bounds for combinatorial central limit theorems and pattern occurrences, using zero and size biasing, *Journal of Applied Probability*, **42**, 661-683.

[16] Goldstein, L. and Penrose, M.(2008). Normal approximation for coverage models over binomial point processes, preprint.

[17] Goldstein, L. and Rinott, Y.(1996). Multivariate normal approximations by Stein's method and size bias couplings, *Journal of Applied Probability*, **33**,1-17.

[18] Goldstein, L. and Zhang, H. (2009). A Berry Esseen theorem for the lightbulb problem. *Preprint*

[19] Gordon, L. (1989). Bounds for the distribution of the generalized variance, *Ann. Statist.*, **17**, 1684-1692.

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

[21] Hoel, P.G.(1937). A significance test for somponent analysis, *Ann. Math. Statist.*, **8**, 149-158.

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

[23] Kordecki, W.(1990). Normal approximation and isolated vertices in random graphs, *Random Graphs '87*, Karoński, M., Jaworski, J. and Ruciński, A. eds., John Wiley & Sons Ltd.,1990, 131-139.

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

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

[26] Midzuno, H. (1951). On the sampling system with probability proportionate to sum of sizes, *Annals of the Institute of Statistical Mathematics*, **2**, 99-108.

[27] Moran, P.A.P.(1973). The random volume of interpenetrating spheres in space, *J. Appl. Probab.*, **10**, 483-490.

[28] O'Connell, N.(1998). Some large deviation results for sparse random graphs, *Probab. Th. Rel. Fields*, **110**, 277-285.

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

[30] Penrose, M.(2009). Normal approximation for isolated balls in an urn allocation model, preprint.

[31] Penrose, M.D. and Yukich, J.E.(2001). Central limit theorems for some graphs in computational geometry, *Ann. Appl. Probab.*, **11**, 1005-1041.

[32] Quine, M.P. and Robinson, J.(1982). A Berry Esseen bound for an occupancy problem, *Ann. Probab*, **10**, 663-671.

[33] Raič, M.(2007). CLT related large deviation bounds based on Stein's method, *Adv. Appl. Prob.*, **39**, 731-752.

[34] Rao, C.R., Rao, B.M., and Zhang, H.(2007). One Bulb? Two Bulbs? How Many Bulbs Light Up? A Discrete Probability Problem Involving Dermal Patches, *Sankhyā*, **69**, pp. 137-161.

[35] Reinert, G. and Röllin, A.(2008). Multivariate normal approximation with Stein's method of exchangeable pairs under a general linearity condition, *Ann. Probab.*, to appear.

[36] Srivastava, M. and Khatri, C.(1979). An Introduction to Multivariate Statistics. North Holland Inc., New York.

[37] 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.

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

[39] Steutel, W.F.(1973). Some recent results in infinite divisibility. *Stoch. Proc. Appl.*, **1**, 125-143.

[40] Wilks, S.S.(1932). Certain generalizations in the analysis of variance. *Biometrika*, **24**, 471-494.

[41] Zhou, H. and Lange, K. (2009). Composition Markov chains of multinomial type. *Advances in Applied Probability*