Statistica Sinica **23** (2013), 1523-1540
doi:http://dx.doi.org/10.5705/ss.2012.203s

MODERATE DEVIATIONS IN POISSON APPROXIMATION: A FIRST ATTEMPT

Louis H. Y. Chen[1], Xiao Fang[1,2] and Qi-Man Shao[3]

[1] *National University of Singapore,* [2] *Stanford University and*
[3] *Chinese University of Hong Kong*

Abstract: Poisson approximation using Stein's method has been extensively studied in the literature. The main focus has been on bounding the total variation distance. This paper is a first attempt on moderate deviations in Poisson approximation for right-tail probabilities of sums of dependent indicators. We obtain results under certain general conditions for local dependence as well as for size-bias coupling. These results are then applied to independent indicators, 2-runs, and the matching problem.

Key words and phrases: Local dependence, moderate deviations, Poisson approximation, size-bias coupling, Stein's method.

1. Introduction

Poisson approximation using Stein's method has been applied to many areas, ranging from computer science to computational biology. The main focus has been on bounding the total variation distance between the distribution of a sum of dependent indicators and the Poisson distribution with the same mean.

Broadly speaking, there are two main approaches to Poisson approximation, the local approach and the size-bias coupling approach. The local approach was first studied by Chen (1975) and developed further by Arratia, Goldstein, and Gordon (1989, 1990), who presented Chen's results in a form which is easy to use, and applied them to a wide range of problems including problems in extreme values, random graphs and molecular biology. The size-bias coupling approach dates back to Barbour (1982) in his work on Poisson approximation for random graphs. Barbour, Holst, and Janson (1992) presented a systematic development of monotone couplings and applied their results to random graphs and many combinatorial problems. A recent review of Poisson approximation by Chatterjee, Diaconis, and Meckes (2005) used Stein's method of exchangeable pairs to study classical problems in combinatorial probability. They also reviewed a size-bias coupling of Stein (1986, p.93).

Although there is a vast literature on Poisson approximation, relatively little has been done on such refinements as moderate deviations. For sums of independent indicators, moderate deviations have been studied by Barbour, Holst, and Janson (1992), Chen and Choi (1992), and Barbour, Chen, and Choi (1995). The latter two actually considered the more general problem of unbounded function approximation and deduced moderate deviations as a special case. However no such results seem to have been obtained for dependent indicators, probably due to the fact that unbounded function approximation becomes much harder for dependent indicators. Although moderate deviations is a special case of unbounded function approximation, it is of a similar nature as the latter and, as such, it is also a difficult problem for dependent indicators.

This paper is a first attempt on moderate deviations in Poisson approximation for dependent indicators. We take both the local and the size-bias coupling approach. Under the local approach we consider locally dependent indicators. Under the size-bias coupling approach we consider size-bias coupling, which generalizes the monotone couplings of Barbour, Holst, and Janson (1992) and the size-bias coupling of Stein (1986). In both approaches, we consider moderate deviations for right-tail probabilities under certain general conditions.

This paper is organized as follows. Section 2 contains the main theorems. In Section 3, we apply our main theorems to Poisson-binomial trials, 2-runs in a sequence of i.i.d. Bernoulli random variables, and the matching problem. As far as we know, the results for the last two applications are new. In Section 4 we prove the main theorems.

2. Main Theorems

In this section, we state two general theorems on moderate deviations in Poisson approximation, one under local dependence and the other under size-bias coupling. Let $|\cdot|$ denote the Euclidean norm or cardinality.

2.1. Local dependence

Local dependence is a widely used dependence structure for Poisson approximation. We refer to Arratia, Goldstein, and Gordon (1989, 1990) for results on the total variation distance and applications. Here we prove a moderate deviation result. Let $X_i, i \in \mathcal{J}$, be random indicators indexed by $\mathcal{J}$. Let $W = \sum_{i\in\mathcal{J}} X_i$,

$$p_i = P(X_i = 1), \quad \text{and} \quad \lambda = \sum_{i\in\mathcal{J}} p_i > 0 \; . \tag{2.1}$$

Suppose for each $i \in \mathcal{J}$, there exists a subset B_i of $\mathcal{J}$ such that X_i is independent of $\{X_j : j \notin B_i\}$. The subset B_i is called a dependence neighborhood of X_i.

Assume that

$$\max_{i\in\mathcal{J}}|B_i|\le m,\quad \max_{j\in\mathcal{J}}|\{i:j\in B_i\}|\le m, \tag{2.2}$$

and, for some $\delta,\theta>0$,

$$E(\sum_{i\in\mathcal{J}}\sum_{j\in B_i\backslash\{i\}}X_iX_j|W=w)\le\delta w^2 \text{ for } w\le\theta. \tag{2.3}$$

Let $\tilde{p}=\max_{i\in\mathcal{J}}p_i$.

Theorem 1. *Let $W=\sum_{i\in\mathcal{J}}X_i$ be a sum of locally dependent random indicators with dependence neighborhoods B_i satisfying* (2.2) *and* (2.3)*. Then there exist absolute positive constants c,C such that for $k\ge\lambda$ satisfying*

$$k\le\frac{\theta}{Cm},\quad \tilde{p}(1+\xi^2)+\delta\lambda(1+\xi^2+\frac{\xi^3}{\sqrt{\lambda}})\le\frac{c}{m^2}, \tag{2.4}$$

where $\xi=(k-\lambda)/\sqrt{\lambda}$, we have

$$\Big|\frac{P(W\ge k)}{P(Y\ge k)}-1\Big|$$
$$\le Cm^2\Big\{\tilde{p}(1+\xi^2)+\delta\lambda(1+\xi^2+\frac{\xi^3}{\sqrt{\lambda}})\Big\}+C(1\wedge\frac{1}{\lambda})m^2\exp(-\frac{c\theta}{m}), \tag{2.5}$$

where $Y\sim Poi(\lambda)$.

Remark 1. The main difficulty in applying Theorem 1 is to verify the condition (2.3). Intuitively, if for many $i\in\mathcal{J},j\in B_i\backslash\{i\}$, $p_{ji}:=P(X_j=1|X_i=1)$ is large, then given $W=w$, the w 1's tend to appear in clusters, which makes the left-hand side of (2.3) large (bounded by w^2 in the extreme case). If p_{ji} is small, then the w 1's tend to be distributed widely, making the left-hand side of (2.3) small (0 in the extreme case). It is a challenge to replace the δ in (2.3) by a quantity involving only $\{p_i,p_{ji}:i\in\mathcal{J},j\in B_i\backslash\{i\}\}$.

2.2. Size-bias coupling

Baldi, Rinott, and Stein (1989) and Goldstein and Rinott (1996) used size-bias coupling to prove normal approximation results by Stein's method. In the context of Stein's method for Poisson approximation, size-bias coupling was used implicitly by Stein (1986, p.93), Barbour (1982), Barbour, Holst, and Janson (1992, p.23) and Chatterjee, Diaconis, and Meckes (2005, p.93). The following definition of size-bias distribution can be found in Goldstein and Rinott (1996).

Definition 1. For W a non-negative random variable with $EW=\lambda>0$, W^s has a W-size biased distribution if

$$EWf(W)=\lambda Ef(W^s) \tag{2.6}$$

for all functions f such that the expectations exist.

We take W to be a non-negative integer-valued random variable, in particular, a sum of random indicators. If we can couple W with W^s on the same probability space, then we have a bound on the total variation distance between $\mathcal{L}(W)$ and a Poisson distribution.

Theorem 2. *Let W be a non-negative integer-valued random variable with $EW = \lambda > 0$. If W^s is defined on the same probability space as W with a W-size biased distribution, then*

$$\|\mathcal{L}(W) - Poi(\lambda)\|_{TV} \leq (1 - e^{-\lambda})E|W + 1 - W^s|. \tag{2.7}$$

Proof. Let $h(w) = I(w \in A)$ for $w \in \mathbb{Z}_+$, where A is any given subset of $\mathbb{Z}_+$. Let f_h be the bounded solution (unique except at $w = 0$) to the Stein equation

$$\lambda f(w+1) - wf(w) = h(w) - Eh(Y) \tag{2.8}$$

where $Y \sim Poi(\lambda)$. It is known that (see, for example, Barbour, Holst, and Janson (1992, p.7)

$$\Delta f_h := \sup_{j \in \mathbb{Z}_+, j \geq 1} |f_h(j+1) - f_h(j)| \leq \lambda^{-1}(1 - e^{-\lambda}). \tag{2.9}$$

From (2.8) and the fact that W^s is coupled with W and has the W-size biased distribution, we have

$$\begin{aligned}
|P(W \in A) - P(Y \in A)| &= |\lambda E f_h(W+1) - EWf_h(W)| \\
&= \lambda |E(f_h(W+1) - f_h(W^s))| \\
&\leq \lambda \Delta f_h E|W + 1 - W^s| \\
&\leq (1 - e^{-\lambda})E|W + 1 - W^s|,
\end{aligned}$$

where the first inequality is obtained by writing $f_h(W+1) - f_h(W^s)$ as a telescoping sum and using the definition of Δf_h, along with the fact that $W^s \geq 1$. The second inequality follows from (2.9). Taking supremum over A yields (2.7).

Similar results as Theorem 2 can be found in Barbour, Holst, and Janson (1992) and Chatterjee, Diaconis, and Meckes (2005). In order for the bound (2.7) to be useful, we need to couple W with W^s such that $E|W+1-W^s|$ is small. A general way of constructing such size-bias couplings for sums of random indicators is as follows; see, for example, Goldstein and Rinott (1996). Let $\mathbb{X} = \{X_i\}_{i \in \mathcal{J}}$ be $\{0, 1\}$-valued random variables with $P(X_i = 1) = p_i$, $\lambda = \sum_{i \in \mathcal{J}} p_i$, and let $W = \sum_{i \in \mathcal{J}} X_i$. Let I be independent of $\mathbb{X}$ with $P(I = i) = p_i/\lambda$. Given $i \in \mathcal{J}$, construct $\mathbb{X}^i = \{X_j^i\}_{j \in \mathcal{J}}$ on the same probability space as $\mathbb{X}$ such that

$$\mathcal{L}(X_j^i : j \in \mathcal{J}) = \mathcal{L}(X_j : j \in \mathcal{J} | X_i = 1).$$

Then $W^s = \sum_{j \in \mathcal{J}} X_j^I$ has the W-size biased distribution.

Theorem 3. *Let W be a non-negative integer-valued random variable with $EW = \lambda > 0$. Let W^s be defined on the same probability space as W with a W-size biased distribution. Assume that $\Delta := W + 1 - W^s \in \{-1, 0, 1\}$ and that there are non-negative constants δ_1, δ_2 such that*

$$P(\Delta = -1 \mid W) \leq \delta_1, \quad P(\Delta = 1 \mid W) \leq \delta_2 W. \tag{2.10}$$

For integers $k \geq \lambda$, let $\xi = (k - \lambda)/\sqrt{\lambda}$. Then there exist absolute positive constants c, C, such that for $(\delta_1 + \delta_2\lambda)(1 + \xi^2) \leq c$, we have

$$\Big|\frac{P(W \geq k)}{P(Y \geq k)} - 1\Big| \leq C(\delta_1 + \delta_2\lambda)(1 + \xi^2), \tag{2.11}$$

where $Y \sim Poi(\lambda)$.

The conditions of Theorem 3 do not hold for all size-bias couplings. Nevertheless, in Section 3, we are able to apply Theorem 3 to prove moderate deviation results for Poisson-binomial trials and the matching problem. It is possible to replace the upper bounds in (2.10) by any polynomial function of W, resulting in a change of the upper bound in (2.11). However we will not pursue this in this paper.

3. Applications

In this section, we apply our main results to Poisson-binomial trials, 2-runs in a sequence of i.i.d. indicators and the matching problem.

3.1. Poisson-binomial trials

Let $X_i, i \in \mathcal{J}$, be independent with $P(X_i = 1) = p_i = 1 - P(X_i = 0)$. Set $\lambda = \sum_{i\in\mathcal{J}} p_i$ and $\tilde{p} = \sup_{i\in\mathcal{J}} p_i$. Let $W = \sum_{i\in\mathcal{J}} X_i$. Following the construction in Section 2.2, W^s in (2.6) can be constructed as $W^s = W - X_I + 1$, where I is independent of $\{X_i : i \in \mathcal{J}\}$ and $P(I = i) = p_i/\lambda$ for each $i \in \mathcal{J}$. Therefore, $\Delta = W + 1 - W^s = X_I$ and condition (2.10) is satisfied with $\delta_1 = 0, \delta_2 = \tilde{p}/\lambda$. Applying Theorem 3, there exist absolute positive constants c, C such that

$$\Big|\frac{P(W \geq k)}{P(Y \geq k)} - 1\Big| \leq C\tilde{p}(1 + \xi^2) \tag{3.1}$$

for integers $k \geq \lambda$ and $\tilde{p}(1+\xi^2) \leq c$ where $Y \sim Poi(\lambda)$ and $\xi = (k-\lambda)/\sqrt{\lambda}$. The range $\tilde{p}(1 + \xi^2) \leq c$ is optimal for the i.i.d. case where $p_i = \tilde{p}$ for all $i \in \mathcal{J}$ (see Theorem 9.D of Barbour, Holst, and Janson (1992, p.188) and Corollary 4.3 of Barbour, Chen, and Choi (1995)).

Remark 2. The moderate deviation result (3.1) also follows from Theorem 1 for sums of locally dependent random variables.

3.2. 2-runs

Let $\{\xi_1, \ldots, \xi_n\}$ be i.i.d. *Bernoulli*(p) variables with $n > 10$, $p < 1/2$. For each $i \in \{1, \ldots, n\}$, let $X_i = \xi_i \xi_{i+1}$ where $\xi_{j+n} = \xi_{j-n} = \xi_j$ for any integer $j \in \{1, \ldots, n\}$. Take $W = \sum_{i=1}^n X_i$ with mean $\lambda = np^2$. Then W is a sum of locally dependent random variables with $m = 3$ where m is defined in (2.2). For each $i \in \{1, \ldots, n\}$ and any positive integer $w \leq cnp$ for some sufficiently small constant $c < 1/50$ to be chosen later, we write

$$\begin{aligned}
&P(X_i = 1, X_{i+1} = 1, W = w) \\
&\quad = \sum_{\substack{m_1 \geq 0, m_2 \geq 1 \\ m_1 + m_2 < w}} P(X_{i-m_1} = \cdots = X_{i+m_2} = 1, X_{i-m_1-1} = X_{i+m_2+1} = 0, W = w) \\
&\quad =: \sum_{\substack{m_1 \geq 0, m_2 \geq 1 \\ m_1 + m_2 < w}} a_{m_1, m_2},
\end{aligned}$$

where the sum is over integers. By writing

$$a_{m_1, m_2} = p^{m_1 + m_2 + 2}(1-p)^2 P\Big(\sum_{i=1}^{n-(m_1+m_2+5)} X_i = w - (m_1 + m_2 + 1) \Big),$$

we have for $m_1 + m_2 + 1 < w$,

$$\frac{a_{m_1, m_2+1}}{a_{m_1, m_2}} = p \frac{P(\sum_{i=1}^{n-(m_1+m_2+6)} X_i = w - (m_1 + m_2 + 2))}{P(\sum_{i=1}^{n-(m_1+m_2+5)} X_i = w - (m_1 + m_2 + 1))} \leq Cp \frac{w}{\lambda} \tag{3.2}$$

for some positive constant C. The last inequality is proved by observing that for each event

$$\{X_i = x_i : 1 \leq i \leq n - (m_1 + m_2 + 6)\} \text{ with } \sum_{i=1}^{n-(m_1+m_2+6)} x_i = w - (m_1 + m_2 + 2),$$

we can change one of the $\ldots 000 \ldots$ to $\ldots 010 \ldots$ and let $x_{n-(m_1+m_2+5)} = 0$, thus resulting in an event

$$\{X_i = x_i : 1 \leq i \leq n - (m_1 + m_2 + 5)\} \text{ with } \sum_{i=1}^{n-(m_1+m_2+5)} x_i = w - (m_1 + m_2 + 1),$$

the probability of which is at least $c_1 p^2$ times the probability of the original event for an absolute positive constant c_1. Summing over the probabilities of all the events obtained in this way, and correcting for the multiple counts, yields the inequality in (3.2). By choosing c to be small,

$$\frac{a_{m_1, m_2+1}}{a_{m_1, m_2}} \leq \frac{1}{4}.$$

Similarly,

$$\frac{a_{m_1+1,m_2}}{a_{m_1,m_2}} \le \frac{1}{4}.$$

Therefore,

$$P(X_i = 1, X_{i+1} = 1, W = w) \le C a_{0,1} \le C p^3 P\Big(\sum_{i=1}^{n-6} X_i = w-2\Big).$$

Similar to (3.2),

$$P\Big(\sum_{i=1}^{n-6} X_i = w-2\Big) \le C\Big(\frac{w^2}{\lambda^2}\Big) P(W = w).$$

Therefore,

$$\begin{aligned}\sum_{i=1}^{n} \sum_{j=i-1,i+1} E(X_i X_j | W = w) &= 2n \frac{P(X_i = X_{i+1} = 1, W = w)}{P(W = w)} \\ &\le C n p^3 \frac{w^2}{\lambda^2} = \frac{C}{np} w^2\end{aligned}$$

for $w \le cnp$ with sufficiently small c. Applying Theorem 1, there exist absolute positive constants c, C, such that for $k \ge \lambda$ and $p + p\xi^2 + \xi^3/\sqrt{n} \le c$, where $\xi = (k-\lambda)/\sqrt{\lambda}$,

$$\Big|\frac{P(W \ge k)}{P(Y \ge k)} - 1\Big| \le C\Big(p + p\xi^2 + \frac{\xi^3}{\sqrt{n}}\Big), \tag{3.3}$$

where $Y \sim Poi(\lambda)$. We remark that if $\lambda \asymp O(1)$, then the range of ξ is of order $O(n^{1/6})$.

Remark 3. Although the rate $O(n^{1/6})$ may not be optimal, we have not seen a result like (3.3) in the literature. Our argument for 2-runs can be extended to study k-runs for $k \ge 3$.

3.3. Matching problem

For a positive integer n, let π be a uniform random permutation of $\{1, \ldots, n\}$. Let $W = \sum_{i=1}^{n} \delta_{i\pi(i)}$ be the number of fixed points in π. In Chatterjee, Diaconis, and Meckes (2005), W^s satisfying (2.6) was constructed as follows. First pick I uniformly from $\{1, \ldots, n\}$, and then set

$$\pi^s(j) = \begin{cases} I & \text{if } j = I, \\ \pi(I) & \text{if } j = \pi^{-1}(I), \\ \pi(j) & \text{otherwise.} \end{cases}$$

Take $W^s = \sum_{i=1}^n \delta_{i\pi^s(i)}$. With $\Delta = W + 1 - W^s$, we have

$$P(\Delta = 1|W) = \frac{W}{n}, \quad P(\Delta = -1|W) = \frac{E(2a_2|W)}{n} \le \frac{2}{n},$$

where a_2 is the number of transpositions of π, and the last inequality follows since

$$E(2a_2|W) = \frac{n - W}{n - W - 1} \le 2$$

for $n - W \ge 2$, and $E(2a_2|W) = 0$ for $n - W \le 1$. By Theorem 3 with $\lambda = 1$, there exist absolute positive constants c, C such that for all positive integers k satisfying $k^2/n \le c$,

$$\Big|\frac{P(W \ge k)}{P(Y \ge k)} - 1\Big| \le \frac{Ck^2}{n}.$$

We remark that the order $O(1/n)$ is the same as that of the total variation bounds in Barbour, Holst, and Janson (1992) and Chatterjee, Diaconis, and Meckes (2005). As remarked in those papers, this order is not optimal; it is an open problem to prove the actual order $O(2^n/n!)$ using Stein's method.

4. Proofs

We use c, C, to denote absolute positive constants whose values may be different at each appearance.

Lemma 1. *For any integer $w \ge \lambda > 0$,*

$$\sum_{j=0}^{\infty} \lambda^j \frac{w!(j+1)}{(j+w+1)!} \le C. \tag{4.1}$$

Proof. We first bound λ^j by w^j. Next, by expanding the product $(w+j+1) \times \cdots \times (w+1)$ in terms of w and then bounding it below by w^{j+1} and $cj^4 w^{j-1}$, respectively, in the expansion, we have

$$\begin{aligned}\sum_{j=0}^{\infty} \lambda^j \frac{w!(j+1)}{(j+w+1)!} &\le \sum_{j=0}^{\infty} w^j \frac{j+1}{(w+j+1) \times \cdots \times (w+1)} \\ &\le \sum_{j \le \sqrt{w}} \frac{j+1}{w} + \sum_{j > \sqrt{w}} \frac{j+1}{cj^4/w} \le C,\end{aligned}$$

as desired.

Lemma 2. *Let $Y \sim Poi(\lambda)$ with $\lambda > 0$. Then we have*

$$P(Y \geq k) \geq c > 0 \quad \text{for all integer } k < \lambda, \tag{4.2}$$

$$\frac{P(Y \geq k)}{P(Y \geq k-1)} \geq \frac{\lambda}{\lambda + k} \quad \text{for all integer } k \geq 1, \tag{4.3}$$

$$P(Y \geq k) \leq P(Y = k)\frac{k+1}{k-\lambda+1} \quad \text{for all integer } k > \lambda - 1. \tag{4.4}$$

Proof. The inequality in (4.2) is trivial when $\lambda < 1$ or $1 \leq \lambda \leq C$ for some absolute constant C. When $\lambda > C$, we can use normal approximation to prove (4.2).

For (4.3), noting that

$$\begin{aligned} P(Y \geq k) &= P(Y = k)(1 + \frac{\lambda}{k+1} + \frac{\lambda^2}{(k+1)(k+2)} + \cdots) \\ &\geq \frac{\lambda + k + 1}{k+1} P(Y = k), \end{aligned}$$

we have

$$\frac{P(Y \geq k)}{P(Y \geq k-1)} = 1 - \frac{P(Y = k-1)}{P(Y \geq k-1)} \geq 1 - \frac{k}{\lambda + k} = \frac{\lambda}{\lambda + k}.$$

The inequality in (4.4) follows by observing that

$$\begin{aligned} P(Y \geq k) &= P(Y = k)(1 + \frac{\lambda}{k+1} + \frac{\lambda^2}{(k+1)(k+2)} + \cdots) \\ &\leq P(Y = k)(1 + \frac{\lambda}{k+1} + \frac{\lambda^2}{(k+1)^2} + \cdots) \\ &= P(Y = k)\frac{k+1}{k-\lambda+1}. \end{aligned}$$

The bounded solution f_h (unique except at $w = 0$) to the Stein equation

$$\lambda f(w+1) - w f(w) = h(w) - Eh(Y), \tag{4.5}$$

where $Y \sim Poi(\lambda)$ and $h(w) = I\{w \geq k\}$ for fixed integer $k \geq \lambda > 0$, is

$$\begin{aligned} f_h(w) &= -\frac{e^{\lambda}(w-1)!}{\lambda^w} E(h(Y) - Eh(Y)) I\{Y \geq w\} \\ &= \begin{cases} -\frac{e^{\lambda}(w-1)!}{\lambda^w}(1 - P(Y \geq k))P(Y \geq w), & w \geq k, \\ -\frac{e^{\lambda}(w-1)!}{\lambda^w} P(Y \geq k) P(Y \leq w-1), & 0 < w \leq k. \end{cases} \end{aligned}$$

Although $f_h(0)$ does not enter into consideration, we set $f_h(0) := f_h(1)$.

For $w \geq k$,

$$\begin{aligned}\frac{f_h(w) - f_h(w+1)}{1 - P(Y \geq k)} &= \frac{e^\lambda w!}{\lambda^{w+1}} P(Y \geq w+1) - \frac{e^\lambda (w-1)!}{\lambda^w} P(Y \geq w) \\ &= \sum_{j=w+1}^{\infty} \frac{w!}{j!} \lambda^{j-w-1} - \sum_{j=w}^{\infty} \frac{(w-1)!}{j!} \lambda^{j-w} \\ &= \sum_{j=0}^{\infty} \lambda^j \Big(\frac{w!}{(j+w+1)!} - \frac{(w-1)!}{(j+w)!} \Big) \\ &= -\sum_{j=0}^{\infty} \lambda^j \frac{(w-1)!(j+1)}{(j+w+1)!},\end{aligned}$$

and hence by (4.1),

$$0 < f_h(w+1) - f_h(w) \leq \frac{C}{w} \quad \text{for } w \geq k. \tag{4.6}$$

For $0 \leq w \leq k-1$,

$$\frac{f_h(w) - f_h(w+1)}{P(Y \geq k)} = g_1(w),$$

where

$$g_1(w) = \frac{e^\lambda w!}{\lambda^{w+1}} P(Y \leq w) - \frac{e^\lambda (w-1)!}{\lambda^w} P(Y \leq w-1) \tag{4.7}$$

and $g_1(0) := 0$.

Let W be a non-negative integer-valued random variable with $EW = \lambda > 0$, and let $Y \sim Poi(\lambda)$. Define

$$\eta_k := \sup_{\lambda \leq r \leq k} \frac{P(W \geq r)}{P(Y \geq r)}. \tag{4.8}$$

By (4.2),

$$\sup_{0 \leq r \leq k} \frac{P(W \geq r)}{P(Y \geq r)} \leq \eta_k + C. \tag{4.9}$$

Lemma 3. *The function g_1 is non-negative, non-decreasing and*

$$g_1(w) \leq \frac{1}{\lambda} + \frac{(w-1)!(w-\lambda)_+}{\lambda^{w+1}} e^\lambda \tag{4.10}$$

for all $w \geq 1$ where x_+ denotes the positive part of x.

Proof. For $w \geq 1$, $g_1(w)$ can be expressed as

$$\frac{e^\lambda w!}{\lambda^{w+1}} P(Y \leq w) - \frac{e^\lambda (w-1)!}{\lambda^w} P(Y \leq w-1)$$

$$
\begin{aligned}
&= \frac{e^{\lambda}}{\lambda^{w+1}} \int_{\lambda}^{\infty} x^{w} e^{-x} dx - \frac{e^{\lambda}}{\lambda^{w}} \int_{\lambda}^{\infty} x^{w-1} e^{-x} dx \\
&= e^{\lambda} \int_{1}^{\infty} x^{w-1}(x-1) e^{-\lambda x} dx \\
&= \int_{0}^{\infty} x(1+x)^{w-1} e^{-\lambda x} dx,
\end{aligned}
$$

from which g_1 is non-negative and non-decreasing. Also for $w \geq 1$,

$$
\begin{aligned}
&\frac{e^{\lambda} w!}{\lambda^{w+1}} P(Y \leq w) - \frac{e^{\lambda}(w-1)!}{\lambda^{w}} P(Y \leq w-1) \\
&\quad = \frac{e^{\lambda} w!}{\lambda^{w+1}} P(Y = w) + \big(\frac{e^{\lambda} w!}{\lambda^{w+1}} - \frac{e^{\lambda}(w-1)!}{\lambda^{w}}\big) P(Y \leq w-1) \\
&\quad \leq \frac{1}{\lambda} + \frac{(w-1)!(w-\lambda)_{+}}{\lambda^{w+1}} e^{\lambda}.
\end{aligned}
$$

Lemma 4. *For any non-negative and non-decreasing function $g : \{0, 1, 2, \dots\} \to \mathbb{R}$ and any $k \geq 0$, we have*

$$
Eg(W \wedge k) \leq C(\eta_k + 1) Eg(Y \wedge k). \tag{4.11}
$$

Proof. Write

$$
g(W \wedge k) = g(0) + \sum_{j=1}^{k} (g(j) - g(j-1)) I(W \geq j).
$$

From (4.9) and the fact that g is non-decreasing, we have

$$
\begin{aligned}
Eg(W \wedge k) &\leq g(0) + C(\eta_k + 1) \sum_{j=1}^{k} (g(j) - g(j-1)) P(Y \geq j) \\
&= C(\eta_k + 1) Eg(Y \wedge k).
\end{aligned}
$$

Lemma 5. *For all $k \geq 0$, we have*

$$
Eg_1((W+1) \wedge k) \leq C(\eta_k + 1)\big(\frac{1}{\lambda} + \frac{(k+1-\lambda)_{+}^{2}}{\lambda^{2}}\big), \tag{4.12}
$$

$$
E[(W \wedge k) g_1(W \wedge k)] \leq C(\eta_k + 1)\big(1 + \frac{(k-\lambda)_{+}^{2}}{\lambda}\big), \tag{4.13}
$$

$$
E[(W \wedge k)^2 g_1(W \wedge k)] \leq C(\eta_k + 1)\big(\lambda + (k-\lambda)_{+}^{2} + \frac{(k-\lambda)_{+}^{3}}{\lambda}\big). \tag{4.14}
$$

Proof. The case $k = 0$ is trivial. Let $k \geq 1$. For any $p \in \{0, 1\}, q \geq 0$, by (4.11) and (4.10),

$$\begin{aligned}
&E\big[((W+p)\wedge k)^q g_1((W+p)\wedge k)\big] \\
&\quad \leq C(\eta_k+1)E\big[((Y+p)\wedge k)^q g_1((Y+p)\wedge k)\big] \\
&\quad \leq C(\eta_k+1)\big(\frac{k^q}{\lambda} + A(k,p,q) + B(k,q)\big),
\end{aligned}$$

where

$$\begin{aligned}
A(k,p,q) &= E\big[\frac{(Y+p)^q (Y+p-1)!(Y+p-\lambda)_+}{\lambda^{Y+p+1}} e^{\lambda} I(1-p \leq Y \leq k-1)\big], \\
B(k,q) &= \frac{k^q (k-1)!(k-\lambda)_+}{\lambda^{k+1}} e^{\lambda} P(Y \geq k).
\end{aligned}$$

Using (4.4), $B(k,q)$ is bounded by

$$B(k,q) \leq \frac{k^q}{\lambda}\frac{(k-\lambda)_+}{k}\frac{k+1}{k-\lambda+1} \leq \frac{k^q}{\lambda}.$$

The relevant special cases of the quantities $A(k,p,q)$ are

$$\begin{aligned}
A(k,1,0) &= \sum_{w=0}^{k-1} \frac{(w+1-\lambda)_+}{\lambda^2} \leq \frac{(k+1-\lambda)_+^2}{2\lambda^2}, \\
A(k,0,1) &= \sum_{w=1}^{k-1} \frac{(w-\lambda)_+}{\lambda} \leq \frac{(k-\lambda)_+^2}{2\lambda}, \\
A(k,0,2) &= \sum_{w=1}^{k-1} \frac{w(w-\lambda)_+}{\lambda} = \sum_{w=1}^{k-1} \big[(w-\lambda)_+ + \frac{(w-\lambda)_+^2}{\lambda}\big] \\
&\leq \frac{(k-\lambda)_+^2}{2} + \frac{(k-\lambda)_+^3}{3\lambda}.
\end{aligned}$$

Combining these bounds and observing that $(k-\lambda)_+ \leq C(\lambda + (k-\lambda)_+^2)$ yields the desired result.

We first prove Theorem 3, which is easier than Theorem 1.

Proof of Theorem 3. For fixed integer $k \geq \lambda$, let $h(w) = I\{w \geq k\}$. Observe that by (2.6), for general f,

$$E(\lambda f(W+1) - W f(W)) = \lambda E(f(W+1) - f(W^s)). \tag{4.15}$$

In particular, for $f := f_h$,

$$Eh(W) - Eh(Y) = \lambda E(f(W+1) - f(W^s)) := H_1 + H_2, \tag{4.16}$$

where

$$H_1 = \lambda E\big[(f(W+1) - f(W+2))I\{\Delta = -1\}\big],$$
$$H_2 = \lambda E\big[(f(W+1) - f(W))I\{\Delta = 1\}\big].$$

Using (2.10), the definition of η_k in (4.8), and the properties of f_h, H_1 is bounded by

$$\begin{aligned}
|H_1| &\le \lambda\delta_1 E\big[|f(W+1) - f(W+2)|(I(W+1 \ge k) + I(W+1 \le k-1))\big] \\
&\le \lambda\delta_1 \frac{CP(W \ge k-1)}{k} + \lambda\delta_1 P(Y \ge k) E\big[I(W+1 \le k-1) g_1(W+1)\big] \\
&\le \lambda\delta_1 \frac{CP(W \ge k-1)}{k} + \lambda\delta_1 P(Y \ge k) E g_1((W+1) \wedge (k-1)) \\
&\le CP(Y \ge k)\delta_1(\eta_k + 1) + CP(Y \ge k)\delta_1(1 + \frac{(k-\lambda)^2}{\lambda})(\eta_k + 1),
\end{aligned}$$

where we used (4.9), (4.3) and (4.12).

Similarly,

$$\begin{aligned}
|H_2| &\le \lambda\delta_2 E\big[W|f(W) - f(W+1)|(I(W \ge k) + I(W \le k-1))\big] \\
&\le C\lambda\delta_2 P(W \ge k) + \lambda\delta_2 P(Y \ge k) E\big[I(W \le k-1) W g_1(W)\big] \\
&\le C\lambda\delta_2 P(W \ge k) + \lambda\delta_2 P(Y \ge k) E\big[(W \wedge (k-1)) g_1(W \wedge (k-1))\big] \\
&\le CP(Y \ge k)\lambda\delta_2\eta_k + CP(Y \ge k)\delta_2(\lambda + (k-\lambda)^2)(\eta_k + 1).
\end{aligned}$$

by (4.8) and (4.13). Therefore,

$$\Big|\frac{P(W \ge k)}{P(Y \ge k)} - 1\Big| \le C(\eta_k + 1)(\delta_1 + \delta_2\lambda)(1 + \xi^2). \tag{4.17}$$

Since the right-hand side here is increasing in k, we have

$$\eta_k - 1 \le C(\eta_k + 1)(\delta_1 + \delta_2\lambda)(1 + \xi^2). \tag{4.18}$$

The bound in (2.11) is proved by solving this recursive inequality.

Proof of Theorem 1. From (4.5) and the definition of the neighborhood B_i, we have

$$\begin{aligned}
&P(W \ge k) - P(Y \ge k) \\
&\quad = \sum_{i \in \mathcal{J}} EX_i[f(V_i + 1) - f(W)] + \sum_{i \in \mathcal{J}} p_i E[f(W+1) - f(V_i + 1)] \\
&\quad =: H_3 + H_4,
\end{aligned}$$

where $V_i := \sum_{j \notin B_i} X_j$.

We bound H_4 first. Write $\{X_k : k \in B_i\} = \{X_{ij} : 1 \le j \le |B_i|\}$, where $|B_i|$ is the cardinality of B_i and $X_{i,|B_i|} := X_i$. Let

$$V_{ij} := V_i + \sum_{l=1}^{j-1} X_{il} + 1.$$

From the definition, if $X_{ij} = 1$, then $W \ge V_{ij}$. By the definitions of $\tilde{p}, m$ and the properties of f,

$$\begin{aligned}
|H_4| &\le \sum_{i\in\mathcal{J}} p_i E\big\{\sum_{j=1}^{|B_i|} X_{ij}\big|f(V_{ij}) - f(V_{ij}+1)\big|\big[I(V_{ij} \ge k) + I(V_{ij} \le k-1)\big]\big\} \\
&\le \tilde{p}E\Big\{\sum_{i\in\mathcal{J}}\sum_{j=1}^{|B_i|} X_{ij}\big[\frac{CI(V_{ij}\ge k)}{V_{ij}} + P(Y \ge k)g_1(V_{ij})I(V_{ij} \le k-1)\big]\Big\} \\
&\le \tilde{p}E\Big\{\sum_{i\in\mathcal{J}}\sum_{j=1}^{|B_i|} X_{ij}\big[\frac{CmI(W\ge k)}{W} \\
&\qquad\qquad + P(Y \ge k)g_1(W \wedge (k-1))I(W \le k+m)\big]\Big\} \\
&\le m\tilde{p}\Big\{CmP(W \ge k) \\
&\quad + P(Y \ge k)E\big[WI(k \le W \le k+m)g_1(k-1)\big] \\
&\quad + P(Y \ge k)E\big[(W \wedge (k-1))g_1(W \wedge (k-1))\big]\Big\}.
\end{aligned}$$

By (4.8), (4.10), (4.4) and (4.13),

$$|H_4| \le CP(Y \ge k)m^2\tilde{p}(\eta_k + 1)\big[1 + \frac{(k-\lambda)^2}{\lambda}\big].$$

Let $c_1 \ge 1$ be an absolute constant to be chosen later such that $c_1 km < \theta$. We have

$$\begin{aligned}
|H_3| &\le \sum_{i\in\mathcal{J}} E\Big\{X_i \sum_{j=1}^{|B_i|-1} X_{ij}\big|f(V_{ij}) - f(V_{ij}+1)\big| \\
&\qquad \times \big[I(W \le c_1 km) + I(c_1 km < W \le \theta) + I(W > \theta)\big]\Big\} \\
&=: H_{3,1} + H_{3,2} + H_{3,3}.
\end{aligned}$$

By (2.3), $H_{3,1}$ can be bounded similarly as for $|H_4|$ as

$$H_{3,1} \le \sum_{i\in\mathcal{J}} E\Big\{X_i \sum_{j=1}^{|B_i|-1} X_{ij}\big[\frac{CI(V_{ij}\ge k)}{V_{ij}} I(W \le c_1 km)$$

$$
\begin{aligned}
&\qquad\qquad\qquad +P(Y \geq k) g_1(V_{ij}) I(V_{ij} \leq k-1)\big]\Big\} \\
&\leq \sum_{i \in \mathcal{J}} E\Big\{X_i \sum_{j=1}^{|B_i|-1} X_{ij}\big[\frac{CmI(W \geq k)}{W} I(W \leq c_1 km) \\
&\qquad\qquad\qquad +P(Y \geq k) g_1(W \wedge (k-1)) I(W \leq k+m)\big]\Big\} \\
&\leq Cm\delta E\big[WI(k \leq W \leq c_1 km)\big] \\
&\quad +\delta P(Y \geq k) E\big[W^2 I(k \leq W \leq k+m) g_1(k-1)\big] \\
&\quad +\delta P(Y \geq k) E\big[W^2 I(1 \leq W \leq k-1) g_1(W)\big] \\
&\leq CP(Y \geq k)(\eta_k + 1)\delta m^2 (\lambda + (k-\lambda)^2 + \frac{(k-\lambda)^3}{\lambda}),
\end{aligned}
$$

where we used (4.14) in the last inequality. Similarly,

$$
\begin{aligned}
H_{3,2} &\leq Cm\delta EWI(c_1 km < W \leq \theta) \\
&\quad +CP(Y \geq k)(\eta_k + 1)\delta m^2 (\lambda + (k-\lambda)^2 + \frac{(k-\lambda)^3}{\lambda}).
\end{aligned}
$$

From (4.20) of Lemma 6, proved later, there exists an absolute positive constant C such that for $c_1 > C$ and $k < \theta/Cm$,

$$
\begin{aligned}
Cm\delta EWI(c_1 km < W \leq \theta) &\leq Cm\delta E[WI(W > c_1 km)] \\
&\leq Cm^2 \delta P(Y \geq k).
\end{aligned}
$$

By (4.20) and the upper bound $|f(w) - f(w+1)| \leq 1 \wedge \frac{1}{\lambda}$ for all integers $w \geq 1$ (see, for example, Barbour, Holst, and Janson (1992)),

$$
H_{3,3} \leq P(Y \geq k)(1 \wedge \frac{1}{\lambda}) m^2 \exp(-\frac{c\theta}{m}).
$$

Therefore,

$$
\begin{aligned}
|H_3| &\leq CP(Y \geq k)(\eta_k + 1)\delta m^2 (\lambda + (k-\lambda)^2 + \frac{(k-\lambda)^3}{\lambda}) \\
&\quad +P(Y \geq k)(1 \wedge \frac{1}{\lambda}) m^2 \exp(-\frac{c\theta}{m}).
\end{aligned}
$$

From the bounds on $|H_3|$ and $|H_4|$, we have

$$
\begin{aligned}
|\frac{P(W \geq k)}{P(Y \geq k)} - 1| &\leq C(\eta_k + 1) m^2 \Big\{\frac{\tilde{p}}{\lambda}(\lambda + (k-\lambda)^2) + \delta(\lambda + (k-\lambda)^2 + \frac{(k-\lambda)^3}{\lambda})\Big\} \\
&\quad +(1 \wedge \frac{1}{\lambda}) m^2 \exp(-C\theta).
\end{aligned}
$$

Since the right-hand side of this bound is increasing in k, we have

$$\eta_k - 1 \le C(\eta_k+1)m^2\Big\{\frac{\tilde{p}}{\lambda}(\lambda+(k-\lambda)^2)+\delta(\lambda+(k-\lambda)^2+\frac{(k-\lambda)^3}{\lambda})\Big\}$$
$$+(1\wedge\frac{1}{\lambda})m^2\exp(-\frac{c\theta}{m}).$$

Solving the above inequality yields Theorem 1.

For the next lemma, we need a Bennett-Hoeffding inequality. Let $\{\xi_i, 1\le i\le n\}$ be independent random variables. Assume that $E\xi_i \le 0$, $\xi_i \le a(a>0)$ for each $1\le i\le n$, and $\sum_{i=1}^n E\xi_i^2 \le B_n^2$. Then for $x>0$

$$P(\sum_{i=1}^n \xi_i \ge x) \le \exp(-\frac{B_n^2}{a^2}\{(1+\frac{ax}{B_n^2})\log(1+\frac{ax}{B_n^2})-\frac{ax}{B_n^2}\}).$$

In particular, for $x > 4B_n^2/a$

$$P(\sum_{i=1}^n \xi_i \ge x) \le \exp(-\frac{x}{2a}\log(1+\frac{ax}{B_n^2})). \tag{4.19}$$

Lemma 6. *Let W be defined as in Theorem 1. Then there exists an absolute constant C such that for $\theta > Ckm$, we have*

$$EWI(W>x) \le Cm\exp(-\frac{x}{8m}\log(1+\frac{x}{2m\lambda})). \tag{4.20}$$

Proof. We follow the proof of Lemma 8.2 in Shao and Zhou (2012). Separate $\mathcal{J}$ into $\mathcal{J}_l, 1\le l\le m$, such that for each l, $X_i, i\in\mathcal{J}_l$ are independent. This can be done by coloring $\{X_i : i\in\mathcal{J}\}$ one by one, and in step j we color X_j such that it is independent of those $\{X_i : i<j\}$ with the same color. The total number of colors used can be controlled by m because of (2.2). Write $W_l = \sum_{i\in\mathcal{J}_l} X_i$. Then for $y>0$,

$$\begin{aligned}EWI(W>2ym) &= 2ymP(W>2ym)+2m\int_y^\infty P(W>2tm)dt\\ &\le 2E(W-ym)^+ + 2\int_y^\infty \frac{1}{t}E(W-tm)_+dt\\ &\le 2\sum_{1\le l\le m}E(W_l-y)_+ + 2\sum_{1\le l\le m}\int_y^\infty \frac{1}{t}E(W_l-t)_+dt\end{aligned}$$

For $s > 5\lambda_l := 5\sum_{i\in\mathcal{J}_l} p_i$, by (4.19),

$$P(W_l > s) \le \exp(-\frac{s}{4}\log(1+\frac{s}{\lambda_l})).$$

For $t \geq y > 5\lambda_l$,

$$
\begin{aligned}
E(W_l - t)_+ &= \int_t^\infty P(W_l > s) ds \\
&\leq \int_t^\infty \exp(-\frac{s}{4}\log(1+\frac{s}{\lambda_l}))ds \\
&\leq 4\exp(-\frac{t}{4}\log(1+\frac{t}{\lambda_l})),
\end{aligned}
$$

$$
\begin{aligned}
\int_y^\infty \frac{1}{t} E(W_l - t)_+ dt &\leq 4\int_y^\infty \frac{1}{t}\exp(-\frac{t}{4}\log(1+\frac{t}{\lambda_l}))dt \\
&\leq \frac{16}{y}\exp(-\frac{y}{4}\log(1+\frac{y}{\lambda_l})).
\end{aligned}
$$

Combining these inequalities yields

$$
EWI(W > 2ym) \leq 8m\exp(-\frac{y}{4}\log(1+\frac{y}{\lambda}))(1+\frac{4}{y}). \tag{4.21}
$$

Acknowledgements

Louis Chen and Xiao Fang were partially supported by Grant C-389-000-010-101 at the National University of Singapore. Part of the revision was done when Xiao Fang was visiting Stanford University supported by NUS-Overseas Postdoctoral Fellowship from the National University of Singapore. Qi-Man Shao was partially supported by Hong Kong RGC GRF-602608, 603710 and CUHK2130344. The authors thank two referees for their valuable comments and suggestions that significantly improved the exposition of the paper.

References

Arratia, R., Goldstein, L. and Gordon, L. (1989). Two moments suffice for Poisson approximations: the Chen-Stein method. *Ann. Probab.* **17**, 9-25.

Arratia, R., Goldstein, L. and Gordon, L. (1990). Poisson approximation and the Chen-Stein method. With comments and a rejoinder by the authors. *Statist. Sci.* **5**, 403-434.

Baldi, P., Rinott, Y. and Stein, C. (1989). A normal approximation for the number of local maxima of a random function on a graph. *Probability, Statistics, and Mathematics*, 59-81. Academic Press, Boston, MA.

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

Barbour, A. D., Chen, L. H. Y. and Choi, K. P. (1995). Poisson approximation for unbounded functions, I: Independent summands. *Statist. Sinica* **2**, 749-766.

Barbour, A. D., Holst, L. and Janson, S. (1992). *Poisson Approximation.* Oxford Science Publications, Oxford.

Chatterjee, S., Diaconis, P. and Meckes, E. (2005). Exchangeable pairs and Poisson approximation. *Probab. Surv.* **2**, 64-106.

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

Chen, L. H. Y. and Choi, K. P. (1992). Some asymptotic and large deviation results in Poisson approximation. *Ann. Probab.* **20**, 1867-1876.

Goldstein, L. and Rinott, Y. (1996). Multivariate normal approximations by Stein's method and size bias couplings. *J. Appl. Probab.* **33**, 1-17.

Shao, Q. M. and Zhou, W. X. (2012). Cramér type moderate deviation theorems for self-normalized processes. Preprint.

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

Department of Mathematics, National University of Singapore, 10 Lower Kent Ridge Road, Singapore 119076, Republic of Singapore.

E-mail: matchyl@nus.edu.sg

Department of Statistics and Applied Probability, National University of Singapore, 6 Science Drive 2, Singapore 117546, Republic of Singapore.
Department of Statistics, Sequoia Hall, 390 Serra Mall, Stanford University, Stanford, CA 94305-4065, USA.

E-mail: stafx@nus.edu.sg

Department of Statistics, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong, P.R. China.

E-mail: qmshao@cuhk.edu.hk

(Received July 2012; accepted May 2013)