J. Appl. Prob. **57**, 1005–1027 (2020)
doi:10.1017/jpr.2020.47

ON MODERATE DEVIATIONS IN POISSON APPROXIMATION

QINGWEI LIU,* ** AND
AIHUA XIA,* *** University of Melbourne

Abstract

In this paper we first use the distribution of the number of records to demonstrate that the right tail probabilities of counts of rare events are generally better approximated by the right tail probabilities of a Poisson distribution than those of the normal distribution. We then show that the moderate deviations in Poisson approximation generally require an adjustment and, with suitable adjustment, we establish better error estimates of the moderate deviations in Poisson approximation than those in [18]. Our estimates contain no unspecified constants and are easy to apply. We illustrate the use of the theorems via six applications: Poisson-binomial distribution, the matching problem, the occupancy problem, the birthday problem, random graphs, and 2-runs. The paper complements the works [16], [8], and [18].

Keywords: Stein–Chen method; Poisson approximation; moderate deviation

2010 Mathematics Subject Classification: Primary 60F05
Secondary 60E15

1. Introduction

An exemplary moderate deviation theorem is as follows (see [29, page 228]). Let X_i, $1 \le i \le n$, be independent and identically distributed (i.i.d.) random variables with $\mathbb{E}(X_1) = 0$ and $\mathrm{var}(X_1) = 1$. If, for some $t_0 > 0$,

$$\mathbb{E}e^{t_0|X_1|} \le c_0 < \infty, \tag{1.1}$$

then there exist positive constants c_1 and c_2 depending on c_0 and t_0 such that

$$\frac{\mathbb{P}(n^{-1/2}\sum_{i=1}^n X_i \ge z)}{1-\Phi(z)} = 1 + \mathrm{O}(1)\frac{1+z^3}{\sqrt{n}}, \qquad 0 \le z \le c_1 n^{1/6}, \tag{1.2}$$

where $\Phi(z)$ is the distribution function of the standard normal, $|\mathrm{O}(1)| \le c_2$. However, since the pioneering work [15], it has been shown [9] that, for the counts of rare events, Poisson distribution provides a better approximation. For example, the distribution of the number of records [23, 30] in Example 1.1 below can be better approximated by the Poisson distribution having the same mean than by a normal distribution [22]. Moreover, a suitable refinement of the Poisson distribution can further improve the performance of the approximation [10, 11].

The right tail probabilities of counts of rare events are often needed in statistical inference, but these probabilities are so small that the error estimates in approximations of distributions of the counts are usually of no use because the bounds are often larger than the probabilities

Received 18 November 2019; revision received 20 April 2020.
* Postal address: School of Mathematics and Statistics, University of Melbourne, VIC 3010, Australia.
** Email address: qingweil@student.unimelb.edu.au
*** Email address: aihuaxia@unimelb.edu.au

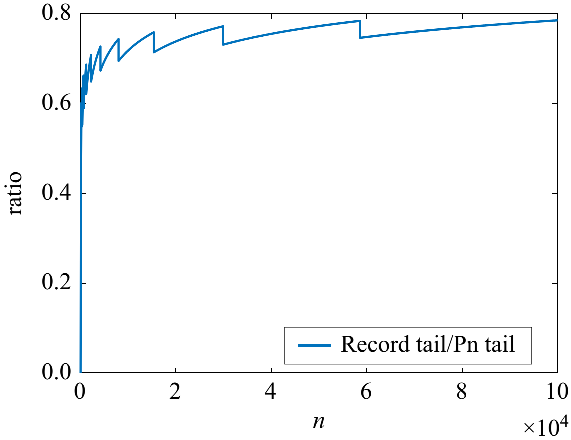

FIGURE 1: $\mathrm{Pn}(\lambda_n)$.

of interest. Hence it is of practical interest to consider their approximations via moderate deviations in Poisson approximation in a similar fashion to (1.2). However, there is not much progress in the general framework except the special cases in [16], [8], [18], [34], and [13]. This is partly due to the fact that the tail behaviour of a Poisson distribution is significantly different from that of a normal distribution, and this fact was observed by Gnedenko [25] in the context of extreme value theory. In particular, Gnedenko [25] concluded that the Poisson distribution does not belong to any domain of attraction, while the normal distribution belongs to the domain of attraction of the Gumbel distribution.

Example 1.1. We use the distribution of the number of records to explain the difference of moderate deviations between Poisson and normal approximations. More precisely, let $\{\eta_i\colon 1 \le i \le n\}$ be i.i.d. random variables with a continuous cumulative distribution function. As the value of η_1 is always a record, for $2 \le i \le n$, we say η_i is a record if $\eta_i > \max_{1\le j\le i-1} \eta_j$. We define the indicator random variable

$$I_i := \mathbf{1}\left[\eta_i > \max_{1\le j\le i-1} \eta_j\right],$$

that is, $I_i = 1$ if a new record occurs at time i and $I_i = 0$ otherwise. Our interest is in the distribution of $S_n := \sum_{i=2}^n I_i$, denoted by $\mathscr{L}(S_n)$. Dwass [23] and Rényi [30] stated that $\mathbb{E}I_i = 1/i$, $\{I_i\colon 2 \le i \le n\}$ are independent, so

$$\lambda_n := \mathbb{E}S_n = \sum_{i=2}^n \frac{1}{i}, \qquad \sigma_n^2 := \mathrm{var}(S_n) = \sum_{i=2}^n \frac{1}{i}\left(1 - \frac{1}{i}\right).$$

We use $\mathrm{Pn}(\lambda)$ to stand for the Poisson distribution with mean λ, $\mathrm{Pn}(\lambda)(A) := \mathbb{P}(Y \in A)$ for $Y \sim \mathrm{Pn}(\lambda)$, and $N(\mu, \sigma^2)$ to stand for the normal distribution with mean μ and variance σ^2.

Let $\nu_n := \lambda_n + x \cdot \sigma_n$, and we consider approximations of $\mathbb{P}(S_n \ge \nu_n)$ by moderate deviations based on $\mathrm{Pn}(\lambda_n)$ [8, 18] and $N_n \sim N(\lambda_n, \sigma_n^2)$. For $x = 3$, Figures 1, 2, and 4, respectively, are the plots of the ratios $\mathbb{P}(S_n \ge \nu_n)/\mathrm{Pn}(\lambda_n)([\nu_n, \infty))$, $\mathbb{P}(S_n \ge \nu_n)/\mathbb{P}(N_n \ge \nu_n)$, and $\mathbb{P}(S_n \ge \nu_n)/\mathrm{Pn}(\sigma_n^2)([\nu_n, \infty))$ for the range of $n \in [3, 10^5]$. As observed in [11], Poisson and normal approximations to $\mathscr{L}(S_n)$, respectively, are of order $\mathrm{O}((\ln n)^{-1})$ and $\mathrm{O}((\ln n)^{-1/2})$, and

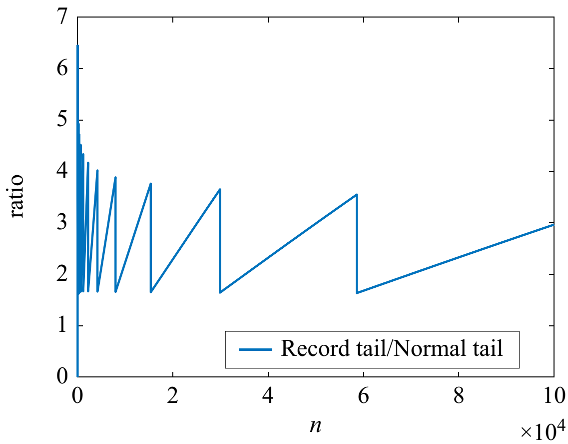

FIGURE 2. $N(\lambda_n, \sigma_n^2)$ without correction.

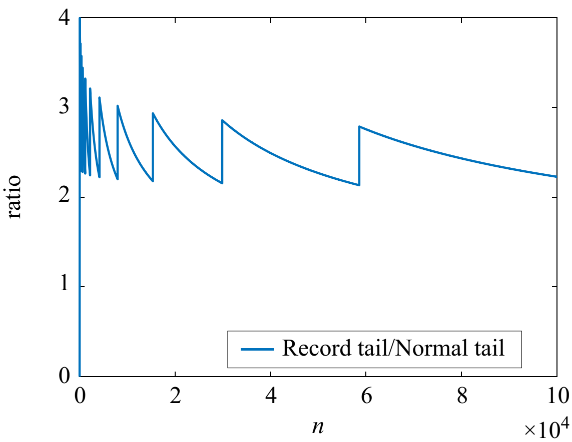

FIGURE 3. $N(\lambda_n, \sigma_n^2)$ with correction.

hence the numerical studies confirm that approximation by the Poisson distribution is better than that by the normal distribution. In fact, it appears that the speed of convergence of $\mathbb{P}(S_n \geq \nu_n)/\mathbb{P}(N_n \geq \nu_n)$ to 1 as $n \to \infty$ is too slow to be of practical use. In the context of normal approximation to the distribution of integer-valued random variables, a common practice is to introduce a 0.5 correction, giving the ratios $\mathbb{P}(S_n \geq \nu_n)/\mathbb{P}(N_n \geq \lceil \nu_n \rceil - 0.5)$, where $\lceil x \rceil$ is the smallest integer that is not less than x. Figure 3 is the plot of the ratios and we can see that the ratios are still far away from the limit of 1. Finally, the difference between Figure 1 and Figure 4 shows that a minor change of the mean of the approximating Poisson can change the quality of moderate deviation approximation significantly, further highlighting the difficulty of obtaining sharp bounds in theoretical studies in the area.

Example 1.1 shows that the distribution of the counts of rare events often has a heavier right tail than that of the corresponding normal distribution; approximations by the moderate

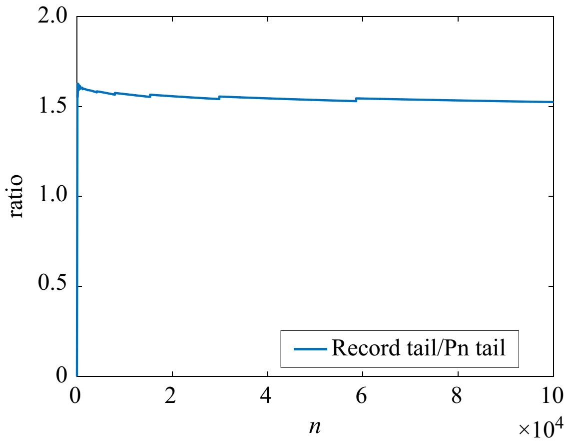

FIGURE 4. $\mathrm{Pn}(\sigma_n^2)$.

deviations in the normal distribution are generally inferior to those by the moderate deviations in the Poisson distribution. Example 1.2 says that the parameter of the approximating Poisson distribution suggested in [16], [8], and [18] is not optimal, and some adjustment can significantly improve the quality of approximations by the moderate deviations in the Poisson distribution.

Example 1.2. With $0 < p < 1$, let $W_n \sim \mathrm{Bi}(n, p)$, $Y_n \sim \mathrm{Pn}(np)$, and $Z \sim N(0, 1)$. Then, for a fixed $x > 0$,

$$\lim_{n\to\infty} \frac{\mathbb{P}(W_n \geq np + x\sqrt{np(1-p)})}{\mathbb{P}(Y_n \geq np + x\sqrt{np(1-p)})} = \frac{\mathbb{P}(Z \geq x)}{\mathbb{P}(Z \geq x\sqrt{1-p})},$$

which systematically deviates from 1 as x moves away from 0. The systematic bias can be removed by introducing an adjustment into the approximate models: for a fixed $x > 0$,

$$\lim_{n\to\infty} \frac{\mathbb{P}(W_n \geq np + x\sqrt{np(1-p)})}{\mathbb{P}(Y_n \geq np + x\sqrt{np})} = 1$$

or equivalently, with $Y_n{}' \sim \mathrm{Pn}(np(1-p))$,

$$\lim_{n\to\infty} \frac{\mathbb{P}(W_n \geq np + x\sqrt{np(1-p)})}{\mathbb{P}(Y_n' \geq np(1-p) + x\sqrt{np(1-p)})} = 1.$$

Example 1.2 suggests that it is more suitable to approximate the right tail probabilities by looking at the number of standard variations away from the mean, which is essentially the original idea of the translated (shifted) Poisson approximation [7, 31, 32]. In this paper we show that it is indeed better to approximate the right tail probabilities via the moderate deviations in the translated Poisson distribution.

Our approach does not rely on the boundedness of the Radon–Nikodým derivative as in [16] and [8] or the tacit assumption of well-behaved tail probabilities as in [18]; see Remark 2.3 for more details. For the case of Poisson-binomial, we show in Proposition 3.2 that our approach works for the case when the maximum of the success probabilities of the Bernoulli random variables is not small, such as the distribution of the number of records.

The paper is organised as follows. In Section 2 we state the main results in the context of local dependence, size-biased distribution, and discrete zero-biased distribution. In Section 3 we illustrate the accuracy of our bounds with six examples. The proofs of the main results are postponed to Section 4, where we also establish Stein's factors for Poisson moderate deviations in Lemma 4.1.

2. The main results

In this section we state three theorems on moderate deviations in Poisson approximation: the first is under a local dependent structure, the second is with respect to the size-biased distribution, and the last is in terms of the discrete zero-biased distribution.

We first consider a class of non-negative integer-valued random variables $\{X_i\colon i\in\mathcal{I}\}$ satisfying the local dependent structure (LD2) in [17] (see also [2] for its origin). For ease of reading, we quote the definition of (LD2) below.

(LD2) For each $i\in\mathcal{I}$, there exists an $A_i\subset B_i\subset\mathcal{I}$ such that X_i is independent of $\{X_j\colon j\in A_i^c\}$ and $\{X_i\colon i\in A_i\}$ is independent of $\{X_j\colon j\in B_i^c\}$.

We set

$$W=\sum_{i\in\mathcal{I}}X_i,\quad Z_i=\sum_{j\in A_i}X_j,\quad Z_i'=\sum_{j\in B_i}X_j,\quad W_i=W-Z_i,\quad W_i'=W-Z_i'.$$

We write

$$\mu_i=\mathbb{E}(X_i),\qquad \mu=\mathbb{E}(W),\qquad \sigma^2=\mathrm{Var}(W).$$

As suggested in Example 1.2, we consider $Y\sim\mathrm{Pn}(\lambda)$ approximation to $W-a$ with $|\lambda-\sigma^2|$ being not too large and $a=\mu-\lambda$ being an integer, so that k in $\mathbb{P}(W-a\ge k)$ and $\mathbb{P}(Y\ge k)$ is in terms of the number of standard deviations of W. In principle, the constant a is chosen to minimise the error of approximation. However, our theory is formulated in such a flexible way that other choices of λ and a are also acceptable. The three most useful choices of a are $a=0$, $a=\lfloor\mu-\sigma^2\rfloor$, and $a=\lceil\mu-\sigma^2\rceil$, where $\lfloor\cdot\rfloor$ stands for the largest integer in $(-\infty,\cdot]$.

Theorem 2.1. *With the set-up in the preceding paragraph, assume that $\{X_i\colon i\in\mathcal{I}\}$ satisfies (LD2) and, for each i, there exists a σ-algebra $\mathcal{F}_i$ such that $\{X_j\colon j\in B_i\}$ is $\mathcal{F}_i$-measurable. Define*

$$\theta_i:=\operatorname{ess\,sup}\max_j\mathbb{P}(W=j\mid\mathcal{F}_i),$$

where ess sup V *is the essential supremum of the random variable V. Then, for integer $a<\mu$, $\lambda=\mu-a$, and positive integer $k>\lambda$, we have*

$$\begin{aligned}\left|\frac{\mathbb{P}(W-a\ge k)}{\mathbb{P}(Y\ge k)}-1\right|\le C_2(\lambda,k)\sum_{i\in\mathcal{I}}&\theta_i\{|\mathbb{E}(X_i-\mu_i)Z_i|\mathbb{E}(Z_i')\\&+\mathbb{E}[|X_i-\mu_i|Z_i(Z_i'-Z_i/2-1/2)]\}\\&+C_1(\lambda,k)|\lambda-\sigma^2|+\mathbb{P}(W-a<-1),\end{aligned}\tag{2.1}$$

where, with $F(j)=\mathbb{P}(Y\le j)$, $\overline{F}(j)=\mathbb{P}(Y\ge j)$,

$$C_1(\lambda,k):=\frac{F(k-1)}{k\mathbb{P}(Y=k)}\left\{1-\min\left(\frac{F(k-2)}{F(k-1)}\cdot\frac{\lambda}{k-1},\frac{\overline{F}(k+1)}{\overline{F}(k)}\cdot\frac{k}{\lambda}\right)\right\},\tag{2.2}$$

$$C_2(\lambda,k):=\frac{F(k-1)}{k\mathbb{P}(Y=k)}\left(2-\frac{F(k-2)}{F(k-1)}\cdot\frac{\lambda}{k-1}-\frac{\overline{F}(k+1)}{\overline{F}(k)}\cdot\frac{k}{\lambda}\right).\tag{2.3}$$

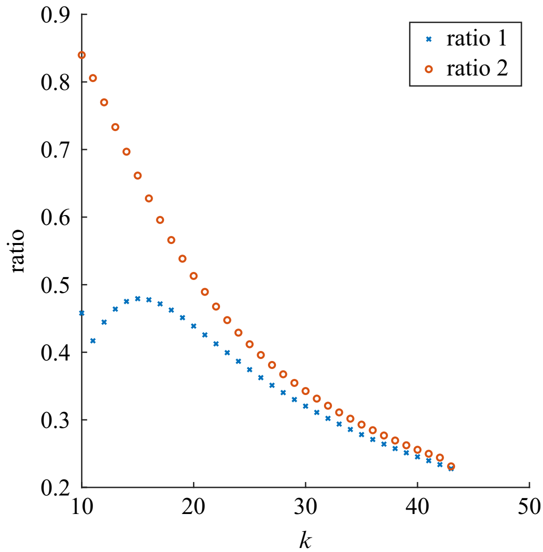

FIGURE 5. Performance of the bound.

Remark 2.1. Both C_1 and C_2 can be numerically computed in applications and they cannot generally be improved (see the proofs below). They are better than the 'naive' counterparts $(1 - e^{-\lambda})/(\lambda\mathbb{P}(Y \geq k))$ derived through the total variation bounds in [6] and [9]. Figure 5 provides details of

$$\text{ratio } i := C_i(\lambda, k)/[(1 - e^{-\lambda})/(\lambda\mathbb{P}(Y \geq k))], \quad i = 1, 2,$$

for $\lambda = 10$, k from 10 to 43. We would like to mention that for large k and/or large λ, the tail probabilities are so small that the calculation using MATLAB produces unstable results since accumulated computation errors often exceed the tail probabilities, and hence more powerful computational tools are needed to achieve the required accuracy or one has to resort to known approximations to the Poisson right tails and point probabilities.

Remark 2.2. Due to the discrete nature of Poisson distribution, it seems impossible to analytically simplify C_1 and C_2 at negligible costs for the diverse range of $k > \lambda$.

Remark 2.3. If λ is chosen reasonably close to σ^2 so that $\lambda - \sigma^2$ is bounded, then θ_i in the bound (2.1) converges to 0 when σ^2 converges to ∞. Our bound does not rely on the Radon–Nikodým derivative of $\mathscr{L}(W)$ with respect to $\text{Pn}(\lambda)$, which is the crucial ingredient in [16] and [8]. On the other hand, the tacit assumption of [18] is that

$$\sup_{\lambda \leq r \leq k} \frac{\mathbb{P}(W \geq r)}{\mathbb{P}(Y \geq r)} \quad \text{for } W \text{ and } Y$$

in Theorem 2.1 is well-behaved and this assumption is hard to verify. The bound (2.1), although relatively crude, does not rely on this assumption and covers more general cases.

Corollary 2.1. *For the sum of independent non-negative integer-valued random variables* $W = \sum_{i\in\mathcal{I}} X_i$, *let* $\theta_i = \max_j \mathbb{P}(W - X_i = j)$, $\mu_i = \mathbb{E}X_i$, $\mu = \sum_{i\in\mathcal{I}} \mu_i$, $\sigma^2 = \mathrm{Var}(W)$. *For any integer* $a < \mu$, *let* $\lambda = \mu - a$, $Y \sim \mathrm{Pn}(\lambda)$. *Then, for* $k > \lambda$,

$$\left|\frac{\mathbb{P}(W-a\ge k)}{\mathbb{P}(Y\ge k)} - 1\right| \le C_2(\lambda, k)\sum_{i\in\mathcal{I}} \theta_i\left\{\mu_i|\mathbb{E}[X_i(X_i-\mu_i)]| + \frac{1}{2}\mathbb{E}[|X_i-\mu_i|X_i(X_i-1)]\right\}$$
$$+ C_1(\lambda, k)|\lambda - \sigma^2| + \mathbb{P}(W - a < -1).$$

Remark 2.4. We leave $\mathbb{P}(W - a < -1)$ in the upper bound (2.1) because the current approach cannot remove it from the bound. Nevertheless, it is no more than 1 and converges to zero exponentially fast with suitable choice of a. For the sum of independent non-negative integer-valued random variables in Corollary 2.1, if a is at least less than μ by a few σs, we can use [20, Theorem 2.7] to obtain

$$\mathbb{P}(W - a < -1) \le \exp\left(-\frac{(\mu - a + 2)^2}{2\sum_{i\in\mathcal{I}} \mathbb{E}(X_i^2)}\right). \tag{2.4}$$

For any non-negative random variable W with mean $\mu \in (0, \infty)$ and distribution $\mathrm{d}F(w)$, the *W-size-biased* distribution [1, 21] is given by

$$\mathrm{d}F^s(w) = \frac{w\mathrm{d}F(w)}{\mu}, \quad w \ge 0,$$

or equivalently by the characterising equation

$$\mathbb{E}[Wg(W)] = \mu\mathbb{E}g(W^s) \quad \text{for all } g \text{ with } \mathbb{E}|Wg(W)| < \infty.$$

Theorem 2.2. *Let* W *be a non-negative integer-valued random variable with mean* μ *and variance* σ^2, *and let* $a < \mu$ *be an integer,* $\lambda = \mu - a$. *Then, for integer* $k > \lambda$, *we have*

$$\left|\frac{\mathbb{P}(W-a\ge k)}{\mathbb{P}(Y\ge k)} - 1\right| \le C_1(\lambda, k)\{\mu\mathbb{E}|W + 1 - W^s| + |\mu - \lambda|\} + \mathbb{P}(W - a < -1), \tag{2.5}$$

where $Y \sim \mathrm{Pn}(\lambda)$.

Remark 2.5. Theorem 2.2 improves [18, Theorem 3] in a number of ways, with less restrictive conditions and no unspecified constants.

The next theorem is based on the discrete zero-biased distribution defined in [26] and the approach is very similar to that in [19]. For an integer-valued random variable V with mean μ and finite variance σ^2, we say that $V^\star$ has the discrete V-zero-biased distribution [26, Definition 2.1] if, for all bounded functions $g\colon \mathbb{Z} := \{0, \pm 1, \pm 2, \ldots\} \to \mathbb{R}$ with $\mathbb{E}|Vg(V)| < \infty$,

$$\mathbb{E}[(V-\mu)g(V)] = \sigma^2\mathbb{E}\Delta g(V^\star),$$

where $\Delta f(i) := f(i+1) - f(i)$.

Theorem 2.3. *Let* W *be a non-negative integer-valued random variable with mean* μ, *variance* σ^2, *let* $a < \mu$ *be an integer, and let* $W^\star$ *have the discrete* W*-zero-biased distribution and be defined on the same probability space as* W. *Set* $R = W^\star - W$ *and define*

$$\theta_R = \max_j \mathbb{P}(W = j \mid R).$$

Then, for integer $k > \lambda$, with $\lambda = \mu - a > 0$, we have

$$\left| \frac{\mathbb{P}(W - a \geq k)}{\mathbb{P}(Y \geq k)} - 1 \right| \leq C_2(\lambda, k)\sigma^2 \mathbb{E}[|R|\theta_R] + C_1(\lambda, k)|\lambda - \sigma^2|\lambda^{-1} + \mathbb{P}(W - a < -1), \quad (2.6)$$

where $Y \sim \mathrm{Pn}(\lambda)$.

3. Examples

As many applications of Poisson approximation rely on size-biased distributions, we begin with a review of some facts about size biasing.

Size biasing has been of considerable interest for many decades (see [9], [33], [3], and references therein). In the context of the sum of Bernoulli random variables, its size biasing is particularly simple. More precisely, if $\{X_i : i \in \mathcal{I}\}$ is a family of Bernoulli random variables with $\mathbb{P}(X_i = 1) = p_i$, then the size-biased distribution of $W = \sum_{i \in \mathcal{I}} X_i$ is

$$W^s = \sum_{j \neq I} X_j^{(I)} + 1, \quad (3.1)$$

where

$$\mathscr{L}(\{X_j^{(i)} : j \in \mathcal{I}\}) = \mathscr{L}(\{X_j : j \in \mathcal{I}\} \mid X_i = 1),$$

and I is a random element independent of $\{\{X_j^{(i)} : j \in \mathcal{I}\} : i \in \mathcal{I}\}$ having distribution $\mathbb{P}(I = i) = p_i/(\mathbb{E}W)$, $i \in \mathcal{I}$. Moreover, $\{X_i : i \in \mathcal{I}\}$ are said to be negatively related (resp. positively related) [9, page 24] if one can construct $\{\{X_j^{(i)} : j \in \mathcal{I}\} : i \in \mathcal{I}\}$ such that $X_j^{(i)} \leq$ (resp. $\geq$) X_j for all $j \neq i$. When $\{X_i : i \in \mathcal{I}\}$ are negatively related, we have

$$\mathbb{E}|W + 1 - W^s| = \mathbb{E}(W + 1 - W^s) = \mu^{-1}(\mu - \sigma^2), \quad (3.2)$$

where $\mu = \mathbb{E}W$ and $\sigma^2 = \mathrm{var}(W)$. On the other hand, if $\{X_i : i \in \mathcal{I}\}$ are positively related, then

$$\begin{aligned}
\mathbb{E}|W + 1 - W^s| &= \mathbb{E}\left| \sum_{j \neq I} \left(X_j^{(I)} - X_j\right) - X_I \right| \\
&\leq \mathbb{E}\left\{ \sum_{j \neq I} \left(X_j^{(I)} - X_j\right) + X_I \right\} \\
&= \mathbb{E}(W^s - W - 1) + 2\mu^{-1} \sum_{i \in \mathcal{I}} p_i^2 \\
&= \mu^{-1}(\sigma^2 - \mu) + 2\mu^{-1} \sum_{i \in \mathcal{I}} p_i^2. \quad (3.3)
\end{aligned}$$

3.1. Poisson-binomial trials

Let $\{X_i, 1 \leq i \leq n\}$ be independent Bernoulli random variables with $\mathbb{P}(X_i = 1) = p_i \in (0, 1)$, $W = \sum_{i=1}^n X_i$, $\mu = \mathbb{E}W$, and $\mu_2 = \sum_{i=1}^n p_i^2$. When $\tilde{p} := \max_{1 \leq i \leq n} p_i \to 0$, the large deviation of W is investigated in [16] and [8] with precise asymptotic order. We give two results for this particular case without the assumption $\tilde{p}$ being small: the first is a direct consequence of the general results in Section 2 and the second is based on our approach using a more fine-tuned analysis and well-studied properties of the tail behaviour of W.

Proposition 3.1. *Recalling C_1 and C_2 in (2.2) and (2.3), for any integer $k > \mu$ we have*

$$\left|\frac{\mathbb{P}(W \ge k)}{\mathrm{Pn}(\mu)([k, \infty))} - 1\right| \le C_1(\mu, k)\mu_2 \tag{3.4}$$

and, with $a = \lfloor \mu_2 \rfloor$ and $\lambda := \mu - a$,

$$\begin{aligned}&\left|\frac{\mathbb{P}(W - a \ge k)}{\mathrm{Pn}(\lambda)([k, \infty))} - 1\right| \\ &\quad \le \frac{C_2(\lambda, k)\sum_{i=1}^n p_i^2(1-p_i)}{1 \vee \sqrt{(\sum_{i=1}^n p_i \wedge (1-p_i) - 1/4)\pi/2}} + C_1(\lambda, k)|\lambda - \sigma^2| + \mathrm{e}^{-(\lambda+2)^2/(2\mu)}.\end{aligned} \tag{3.5}$$

Proof. The claim (3.4) is a consequence of Theorem 2.2 with $a = 0$ and $\mu\mathbb{E}|W + 1 - W^s| = \sum_{i=1}^n p_i^2$, as shown in (3.2).

The bound (3.5) is a special case of Corollary 2.1. Since $\mathscr{L}(W_i)$ is unimodal, Corollary 1.6 of [28] says that

$$\begin{aligned}\theta_i &= d_{\mathrm{TV}}(W_i, W_i + 1) \\ &\le 1 \wedge \left\{\sqrt{\frac{2}{\pi}}\left(\frac{1}{4} + \sum_{j \ne i} p_j \wedge (1-p_j)\right)^{-1/2}\right\} \\ &\le 1 \wedge \left\{\sqrt{\frac{2}{\pi}}\left(\sum_{i=1}^n p_i \wedge (1-p_i) - 1/4\right)^{-1/2}\right\}.\end{aligned} \tag{3.6}$$

On the other hand $\mathbb{E}(X_i^2) = p_i$, and hence the upper bound (3.5) is an immediate consequence of Corollary 2.1 and (2.4).

One can also use Theorem 2.3 to obtain the same bound. More precisely, according to the construction of the discrete zero-biased distribution suggested in [26], let I be a random variable independent of $\{X_i, 1 \le i \le n\}$ with distribution $\mathbb{P}(I = i) = p_i(1-p_i)/\sigma^2$ for $1 \le i \le n$. Then we can write $W^\star = W - X_I$, giving $R = -X_I$. We then apply (3.6) to bound θ_R as

$$\theta_R = \max_j \mathbb{P}(W = j \mid R) \le \sqrt{\frac{2}{\pi}}\left(\sum_{i=1}^n p_i \wedge (1-p_i) - 1/4\right)^{-1/2},$$

and a routine calculation gives

$$\mathbb{E}|R| = \sum_{i=1}^n p_i^2(1-p_i)/\sigma^2.$$

Hence (3.5) follows from (2.6) and (2.4). □

Proposition 3.2. *Define*

$$M := M(p_1, \ldots, p_n) = \begin{cases}\mathrm{e}^{\mu} & \text{if } 0 < \mu < 1, \\ \mathrm{e}^{13/12}\sqrt{2\pi}(1 - \mu_2/\mu)^{-1/2} & \text{if } \mu \ge 1.\end{cases}$$

Then, for any integer k with $x := (k-\mu)/\sqrt{\mu} \geq 1$, we have

$$0 > \frac{\mathbb{P}(W \geq k)}{\mathrm{Pn}(\mu)([k, \infty))} - 1 > -2M(\mu_2/\mu)\left(x^2 + 1 + 4x\sqrt{\frac{1-\mathrm{e}^{-\mu}}{\mu}}\right). \tag{3.7}$$

The proof relies on more information on the solutions of Stein's equation and it is postponed to the end of Section 4. The bound (3.7) improves [18, (3.1)] in two respects: it contains no unspecified constants and it does not require $\tilde{p}$ to be small. For the distribution of the number S_n of records, the large deviation results in [8] do not apply. However, recalling that $\lambda_n = \sum_{i=2}^{n} (1/i)$, we apply Proposition 3.2 with the harmonic series $\lambda_n = \sum_{i=2}^{n} (1/i) \geq \ln n + \gamma - 1$ and the Riemann zeta function

$$\sum_{i=2}^{n} \frac{1}{i^2} \leq \sum_{i=2}^{\infty} \frac{1}{i^2} = \frac{\pi^2}{6} - 1$$

to get the following estimate.

Corollary 3.1. *For any integer k with $x := (k - \lambda_n)/\sqrt{\lambda_n} \geq 1$, we have*

$$0 > \frac{\mathbb{P}(S_n \geq k)}{\mathrm{Pn}(\lambda_n)([k, \infty))} - 1 > -\frac{2\mathrm{e}^{13/12}\sqrt{2\pi}(\pi^2/6 - 1)}{\sqrt{(\ln n + \gamma - 1)(\ln n + \gamma - \pi^2/6)}}\left(x^2 + 1 + \frac{4x}{\sqrt{\ln n + \gamma - 1}}\right),$$

where γ is Euler's constant.

Remark 3.1. We conjecture that, with $a = \lfloor \mu_2 \rfloor$ and $\lambda := \mu - a$, the bound in (3.5) can be significantly improved and the better estimate is likely dependent on the Radon–Nikodým derivative bound

$$\sup_{r \geq 0} \frac{\mathbb{P}(W - a = r)}{\mathrm{Pn}(\lambda)(\{r\})}.$$

3.2. Matching problem

For a fixed n, let π be a uniform random permutation of $\{1, \ldots, n\}$, and let

$$W = \sum_{j=1}^{n} \mathbf{1}_{\{j = \pi(j)\}}$$

be the number of fixed points in the permutation.

Proposition 3.3. *For the random variable W defined above and any integer $k \geq 2$, we have*

$$\left| \frac{\mathbb{P}(W \geq k)}{\mathrm{Pn}(1)([k, \infty))} - 1 \right| \leq \frac{2}{n} C_1(1, k). \tag{3.8}$$

Proof. In this case, the size-biased distribution $\mathscr{L}(W^s)$ can be coupled with W as follows [14]. Let I be uniformly distributed on $\{1, 2, \ldots, n\}$, independent of π, and define

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

Set $W^s = \sum_{j=1}^n \mathbf{1}_{\{j=\pi^s(j)\}}$. One can easily verify that W^s has the size-biased distribution of W. Also, we can check that $\mathbb{E}W = \mathrm{Var}(W) = 1$, giving $\mathbb{E}W^s = 2$. Let $\Delta = W + 1 - W^s$. Using the above construction of W^s, we can conclude that Δ takes values in $\{-1, 0, 1\}$ and $\mathbb{P}(\Delta = 1 \mid W) = W/n$. Since $\mathbb{E}\Delta = 0$, we have $\mathbb{P}(\Delta = 1) = \mathbb{P}(\Delta = -1)$, and $\mathbb{E}|\Delta| = 2/n$. On the other hand, $\lambda = \mu$ allows us to get rid of the second term in (2.5). By Theorem 2.2 with $a = 0$, $\lambda = \mu = 1$, the claim follows. □

Remark 3.2. The bound (3.8) contains no unknown constants and improves the bound of [18, Section 3.3].

3.3. Occupancy problem

The occupancy problem has a long history dating back to the early development of probability theory. General references on this subject can be found in classical treatments, e.g. [24, Vol. 1, Chapter 2] and [9, Chapter 6].

The occupancy problem can be formulated as follows. Let l balls be thrown independently of each other into n boxes uniformly. Let X_i be the indicator variable of the event that the ith box is empty, so the number of empty boxes can be written as $W = \sum_{i=1}^n X_i$. Noting that $p := \mathbb{E}X_i = (1 - 1/n)^l$, direct computation gives

$$\mu := \mathbb{E}W = np,$$

$$\sigma^2 := \mathrm{Var}(W) = \mu - \mu^2 + \mu(n-1)\left(1 - \frac{1}{n-1}\right)^l.$$

Proposition 3.4. *For the random variable W defined above and any integer $k > \mu$, we have*

$$\left|\frac{\mathbb{P}(W \geq k)}{\mathbb{P}(Y \geq k)} - 1\right| \leq C_1(\mu, k)\mu\left[\mu - (n-1)\left(1 - \frac{1}{n-1}\right)^l\right], \tag{3.9}$$

where $Y \sim \mathrm{Pn}(\mu)$.

Proof. For the sake of completeness, we provide the following proof, which is essentially a repeat of [9, page 23]. From the construction of W-size-biased distribution in (3.1), we can construct a coupling as follows. Let I be uniform on $\{1, \ldots, n\}$, that is, we randomly pick one box with equal probability. If the selected box is not empty, we redistribute all balls in the box randomly into the other $n-1$ boxes with equal probability $1/(n-1)$. Define $X_j^{(i)}$ as the indicator of the event that the box being selected is i, and after the redistribution, box j is empty. With this coupling in mind, one can verify that $\{X_i\}$ is negatively related, so it follows from (3.2) that

$$\mathbb{E}|W + 1 - W^s| = \mu - (n-1)\left(1 - \frac{1}{n-1}\right)^l.$$

Now, applying Theorem 2.2 with $a = 0$ yields (3.9). □

3.4. Birthday problem

The classical birthday problem is essentially a variant of the occupancy problem. For this reason, we throw l balls independently and equally likely into n boxes and let X_{ij} be the indicator random variable of the event that ball i and ball j fall into the same box. The number of pairs of balls going into the same boxes (i.e. the number of pairs of people having the same birthdays) can be written as $W = \sum_{i<j} X_{ij}$. Define $p = \mathbb{E}X_{ij} = 1/n$, so $\mu = \mathbb{E}W = \binom{l}{2}p$.

Chatterjee, Diaconis, and Meckes [14] gave the following construction of W^s: label the balls from 1 to l, randomly choose two balls J_1 and J_2, and move ball J_1 into the box that J_2 is in. Then W is the number of pairs of balls before the move while W^s is the number of pairs of balls after the move. Let E be the event that J_1 and J_2 are from the same box. When E occurs, $W^s = W$, so $|W + 1 - W^s| = 1$; otherwise J_1 and J_2 are from different boxes with B_1 and B_2 balls respectively, giving

$$W + 1 - W^s = B_1 - B_2.$$

Hence

$$\begin{aligned}\mathbb{E}|W + 1 - W^s| &= \mathbb{P}(E) + \mathbb{E}[|W + 1 - W^s| \mid E^c]\mathbb{P}(E^c) \\ &\le \frac{1}{n} + \mathbb{E}|B_1 - B_2| \\ &\le \frac{1}{n} + \mathbb{E}(B_1 + B_2) \\ &= \frac{1 + 2l}{n}.\end{aligned}$$

This, together with Theorem 2.2 and $a = 0$, gives the following proposition.

Proposition 3.5. *For the random variable W defined above and any integer $k > \mu$, we have*

$$\left|\frac{\mathbb{P}(W \ge k)}{\mathbb{P}(Y \ge k)} - 1\right| \le C_1(\mu, k)\mu\frac{1 + 2l}{n},$$

where $Y \sim \mathrm{Pn}(\mu)$.

3.5. Triangles in the Erdős–Rényi random graph

Let $G = G(n, p)$ be an Erdős–Rényi random graph on n vertices with edge probability p. Let K_n be the complete graph on n vertices, and let Γ be the set of all triangles in K_n. For $\alpha \in \Gamma$, let X_α be the indicator that there is a triangle in G at α, that is,

$$X_\alpha = \mathbf{1}_{\{\alpha \subset G\}}.$$

Therefore the number of triangles in G can be represented as $W = \sum_{\alpha\in\Gamma} X_\alpha$. Clearly, X_α is independent of X_β if α and β do not share a common edge. By analysing the numbers of shared edges, we obtain (see [33, page 255])

$$\mu = \mathbb{E}W = \binom{n}{3}p^3,$$

$$\sigma^2 = \mathrm{Var}(W) = \binom{n}{3}p^3[1 - p^3 + 3(n - 3)(p^2 - p^3)].$$

Proposition 3.6. *For the random variable W defined above and any integer $k > \mu$, we have*

$$\left|\frac{\mathbb{P}(W \ge k)}{\mathbb{P}(Y \ge k)} - 1\right| \le C_1(\mu, k)\mu(3(n - 3)p^2(1 - p) + p^3), \tag{3.10}$$

where $Y \sim \mathrm{Pn}(\mu)$.

Proof. The following proof is a special version of the general argument in [9, page 89]. Since X_α and X_β are independent if α and β have no common edges, a size-biased distribution of W can be constructed as follows. Let

$$X_\beta^{(\alpha)} := \mathbf{1}_{\{\beta \subset G \cup \alpha\}}, \quad \beta \in \Gamma.$$

Then

$$\mathscr{L}(\{X_\beta^{(\alpha)}, \beta \neq \alpha\}) = \mathscr{L}(\{X_\beta, \beta \neq \alpha\} \mid X_\alpha = 1).$$

Here the union of graphs is in the sense of set operations on their vertices and edges. Let I be a random element taking values in Γ with equal probability and independent of $\mathscr{L}(\{X_\beta^{(\alpha)}, \alpha, \beta\})$. Then we can write

$$W^s = \sum_{\beta \neq I} X_\beta^{(I)} + 1.$$

Because $X_\beta^{(\alpha)} \geq X_\beta$ for all $\beta \in \Gamma$, (3.3) implies

$$\mathbb{E}|W + 1 - W^s| \leq \mu^{-1}(\sigma^2 - \mu + 2\mu p^3) = 3(n-3)p^2(1-p) + p^3.$$

The claim follows from Theorem 2.2 with $a = 0$. □

Remark 3.3. Since $\mu = \binom{n}{3} p^3$, if $p = \mathrm{O}(1/n)$, then the error bound (3.10) is of the same order $\mathrm{O}(1/n)$.

3.6. 2-runs

Let $\{\xi_i, \ldots, \xi_n\}$ be i.i.d. *Bernoulli*(p) random variables with $n \geq 9$, $p < 2/3$. For each $1 \leq i \leq n$, define $X_i = \xi_i \xi_{i+1}$ and, to avoid edge effects, we define $\xi_{j+n} = \xi_j$ for $-3 \leq j \leq n$. The number of 2-runs in the Bernoulli sequence is defined as $W = \sum_{i=1}^n X_i$. Then $\mu = np^2$ and variance $\sigma^2 = np^2(1-p)(3p+1)$.

Proposition 3.7. *For any integer $k > \mu$,*

$$\left| \frac{\mathbb{P}(W_n \geq k)}{\mathrm{Pn}(\mu)([k, \infty))} - 1 \right| \leq C_1(\mu, k) np^3 (2 - p). \tag{3.11}$$

If $a := \lfloor np^3(3p-2) \rfloor$, $\lambda = \mu - a$, then for any integer $k > \lambda$,

$$\left| \frac{\mathbb{P}(W_n - a \geq k)}{\mathrm{Pn}(\lambda)([k, \infty))} - 1 \right| \leq C_2(\lambda, k) \frac{9.2 np^2(1+5p)}{\sqrt{(n-8)(1-p)^3}} + C_1(\lambda, k)(1 \wedge \lambda). \tag{3.12}$$

Proof. For (3.11), we apply Theorem 2.2 with $a = 0$,

$$X_j^{(i)} = \begin{cases} X_j & \text{if } |j - i| \geq 2, \\ \xi_j & \text{if } j = i - 1, \\ \xi_{j+1} & \text{if } j = i + 1, \\ 1 & \text{if } j = i, \end{cases}$$

I a uniform random variable on $\{1, \ldots, n\}$ independent of $\{X_j^{(i)}\}$, and

$$W^s = \sum_{j \neq I} X_j^{(I)} + 1,$$

giving

$$\begin{aligned}\mathbb{E}|W+1-W^s| &= \mathbb{E}|X_{I-1}+X_I+X_{I+1}-\xi_{I-1}-\xi_{I+2}|\\ &= \mathbb{E}|\xi_{i-1}\xi_i+\xi_i\xi_{i+1}+\xi_{i+1}\xi_{i+2}-\xi_{i-1}-\xi_{i+2}|\\ &= p(2-p).\end{aligned}$$

à propos of (3.12), we make use of Theorem 2.1. To this end, let $A_i=\{i-1, i, i+1\}$, $B_i=\{i-2, i-1, i, i+1, i+2\}$, and $\mathcal{F}_i=\sigma\{\xi_j\colon i-2\le j\le i+3\}$. Then Lemma 5.1 of [7] with $\alpha_j=0$ or 1 for $j=i-2, \ldots, i+5$ gives

$$\theta_i\le d_{\mathrm{TV}}(W, W+1\mid \mathcal{F}_i)\le \frac{2.3}{\sqrt{(n-8)p^2(1-p)^3}}.$$

On the other hand, $\mathbb{E}(Z_i')=5p^2$, $|\mathbb{E}((X_i-\mu_i)Z_i)|\le \mathbb{E}(Z_i)=3p^2$,

$$\mathbb{E}[|X_i-\mu_i|Z_i(Z_i'-Z_i/2-1/2)]\le \mathbb{E}[Z_i(Z_i'-Z_i/2-1/2)]=4p^3+5p^4,$$

and $|\lambda-\sigma^2|\lambda^{-1}\le 1\wedge(\lambda^{-1})$, $a=\lfloor np^3(3p-2)\le 0\rfloor$, $\lambda\ge\sigma^2$, and hence $\mathbb{P}(W-a<-1)=0$ and (3.12) follows from Theorem 2.1 by collecting these terms.

4. The proofs of the main results

The celebrated Stein–Chen method [15] is based on the observation that a non-negative random variable $Y\sim \mathrm{Pn}(\lambda)$ if and only if $\mathbb{E}[\lambda f(Y+1)-Yf(Y)]=0$ for all bounded functions $f\colon \mathbb{Z}_+ := \{0, 1, 2, \ldots\}\to\mathbb{R}$, leading to a Stein identity for Poisson approximation as

$$\lambda f(j+1)-jf(j)=h(j)-\mathrm{Pn}(\lambda)\{h\},\quad j\ge 0, \tag{4.1}$$

where $\mathrm{Pn}(\lambda)\{h\}:=\mathbb{E}h(Y)$. Since $f(0)$ plays no role in Stein's equation, we set $f(0)=f(1)$ and $f(j)=0$ for $j<0$. The following lemma plays a key role in the proofs of the main results, and it enables us to circumvent checking the moment condition (1.1), which seems to be inevitable in the existing procedure for proving moderate deviation theorems.

Lemma 4.1. *For fixed $k\in\mathbb{Z}_+$, let $h=\mathbf{1}_{[k,\infty)}$. With $\pi_\cdot=\mathrm{Pn}(\lambda)(\{\cdot\})$, $\Delta f(i)=f(i+1)-f(i)$, and $\Delta^2 f=\Delta(\Delta f)$, the solution $f:=f_h$ of the Stein equation (4.1) has the following properties:*

(i) $\|f\|:=\sup_{i\in\mathbb{Z}_+}|f(i)|=C_0(\lambda, k)\mathrm{Pn}(\lambda)\{h\}$, *where*

$$C_0(\lambda, k):=\frac{F(k-1)}{k\pi_k},$$

(ii) *$\Delta f(i)$ is negative and decreasing in $i\le k-1$ and positive and decreasing in $i\ge k$,*

(iii)

$$\|\Delta f\|_{k-}:=\sup_{i\le k-1}|\Delta f(i)|=C_{1-}(\lambda, k)\mathrm{Pn}(\lambda)\{h\}$$

and

$$\|\Delta f\|_{k+}:=\sup_{i\ge k}|\Delta f(i)|=C_{1+}(\lambda, k)\mathrm{Pn}(\lambda)\{h\},$$

where

$$C_{1-}(\lambda, k):=\frac{F(k-1)}{k\pi_k}\left(1-\frac{F(k-2)}{F(k-1)}\cdot\frac{\lambda}{k-1}\right)$$

and

$$C_{1+}(\lambda, k) := \frac{F(k-1)}{k\pi_k}\left(1 - \frac{\overline{F}(k+1)}{\overline{F}(k)} \cdot \frac{k}{\lambda}\right),$$

(iv)

$$\|\Delta f\| := \sup_{i\in\mathbb{Z}_+} |\Delta f(i)| = C_1(\lambda, k)\mathrm{Pn}(\lambda)\{h\}$$

and

$$\|\Delta^2 f\| := \sup_{i\in\mathbb{Z}_+} |\Delta^2 f(i)| = C_2(\lambda, k)\mathrm{Pn}(\lambda)\{h\},$$

where C_1 and C_2 are defined in (2.2) *and* (2.3).

For $k > \lambda$, death rates are bigger than the birth rate, so it seems intuitively obvious that τ_k^- is stochastically less than or equal to τ_{k-2}^+ for such k. In view of representation (4.11) and $f(k) < 0$ as shown in (4.6), this is equivalent to $C_{1-}(\lambda, k) > C_{1+}(\lambda, k)$, leading to the following conjecture.

Conjecture 4.1. *We conjecture that $C_{1-}(\lambda, k) > C_{1+}(\lambda, k)$ for all $k > \lambda$, and the gap increases exponentially as a function of $k - \lambda$.*

Proof of Lemma 4.1. We build our argument on the birth–death process representation of the solution

$$f(i) = -\int_0^\infty \mathbb{E}[h(Z_i(t)) - h(Z_{i-1}(t))]\,\mathrm{d}t, \quad \text{for } i \geq 1, \tag{4.2}$$

where $Z_n(t)$ is a birth–death process with birth rate λ, unit per capita death rate, and initial state $Z_n(0) = n$ [4, 5, 12]. For convenience we adopt the notation in [12]: for $i, j \in \mathbb{Z}_+$, define

$$\tau_{ij} = \inf\{t\colon Z_i(t) = j\}, \quad \tau_j^+ = \tau_{j,j+1}, \quad \tau_j^- = \tau_{j,j-1},$$

and

$$\overline{\tau_j^+} = \mathbb{E}(\tau_j^+), \quad \overline{\tau_j^-} = \mathbb{E}(\tau_j^-), \quad \pi_i = \mathrm{Pn}(\lambda)(\{i\}).$$

Applying Lemmas 2.1 and 2.2 of [12] with birth rate λ, death rate $\beta_i = i$, $A := [k, \infty)$, and $\pi(\cdot) = \sum_{l\in\cdot} \pi_l$, we have

$$f(i) = \overline{\tau_i^-}\,\pi(A \cap [0, i-1]) - \overline{\tau_{i-1}^+}\,\pi(A \cap [i, \infty)), \quad i \geq 1, \tag{4.3}$$

and for $j \in \mathbb{Z}_+$,

$$\overline{\tau_j^+} = \frac{F(j)}{\lambda\pi_j}, \quad \overline{\tau_j^-} = \frac{\overline{F}(j)}{j\pi_j}, \tag{4.4}$$

where, as in Theorem 2.1,

$$F(j) = \sum_{i=0}^{j} \pi_i, \quad \overline{F}(j) = \sum_{i=j}^{\infty} \pi_i. \tag{4.5}$$

One can easily simplify (4.3) to get

$$f(i) = \begin{cases} -\overline{\tau_{i-1}^+}\,\pi(A) & \text{for } i \leq k, \\ -\overline{\tau_i^-}\,F(k-1) & \text{for } i > k, \end{cases} \tag{4.6}$$

which, together with (4.4) and the balance equations

$$\lambda\pi_i = (i+1)\pi_{i+1} \quad \text{for all } i \in \mathbb{Z}_+, \tag{4.7}$$

implies

$$\Delta f(i) = \begin{cases} -\pi(A)\left(\dfrac{F(i)}{\lambda\pi_i} - \dfrac{F(i-1)}{\lambda\pi_{i-1}}\right) & \text{for } i \le k-1, \\ -(1-\pi(A))\left(\dfrac{\overline{F}(i+1)}{\lambda\pi_i} - \dfrac{\overline{F}(i)}{\lambda\pi_{i-1}}\right) & \text{for } i \ge k. \end{cases} \tag{4.8}$$

It follows from [12, Lemma 2.4] that for $i \ge 1$,

$$\frac{F(i)}{F(i-1)} \ge \frac{\lambda}{i} \ge \frac{\overline{F}(i+1)}{\overline{F}(i)},$$

which, together with (4.7), ensures

$$\Delta f(i) \le 0 \quad \text{for } i \le k-1, \tag{4.9}$$

$$\Delta f(i) \ge 0 \quad \text{for } i \ge k. \tag{4.10}$$

Hence $f(k) \le f(i) \le 0$, and combining (4.4), (4.5), and (4.6) gives

$$\|f\| = |f(k)| = \frac{F(k-1)}{k\pi_k}\pi(A),$$

as claimed in (i).

à propos of (ii), because of (4.9) and (4.10), it remains to show that Δf is decreasing in the two ranges. To this end, we will mainly rely on the properties of the solution (4.2). Let T be an exponential random variable with mean 1 and independent of birth–death process Z_{i-1}. Then Z_i can be represented as

$$Z_i(t) = Z_{i-1}(t) + \mathbf{1}_{\{T>t\}},$$

and hence we obtain from (4.2) and the strong Markov property in the second-to-last equality that

$$\begin{aligned} f(i) &= -\int_0^\infty \mathbb{E}[\,\mathbf{1}_{\{Z_{i-1}(t)+\mathbf{1}_{\{T>t\}}\ge k\}} - \mathbf{1}_{\{Z_{i-1}(t)\ge k\}}]\,\mathrm{d}t \\ &= -\mathbb{E}\int_0^\infty \mathrm{e}^{-t}\,\mathbf{1}_{\{Z_{i-1}(t)=k-1\}}\,\mathrm{d}t \\ &= -\mathbb{E}\left\{\int_{\tau_{i-1,k-1}}^\infty \mathrm{e}^{-t}\,\mathbf{1}_{\{Z_{i-1}(t)=k-1\}}\,\mathrm{d}t\right\} \\ &= -\mathbb{E}\{\mathrm{e}^{-\tau_{i-1,k-1}}\}\mathbb{E}\int_0^\infty \mathrm{e}^{-t}\,\mathbf{1}_{\{Z_{k-1}(t)=k-1\}}\,\mathrm{d}t \\ &= \mathbb{E}\mathrm{e}^{-\tau_{i-1,k-1}}f(k). \end{aligned}$$

This enables us to give another representation of (4.8) as

$$\Delta f(i) = f(k)(\mathbb{E}\mathrm{e}^{-\tau_{i,k-1}} - \mathbb{E}\mathrm{e}^{-\tau_{i-1,k-1}}), \tag{4.11}$$

and so

$$\Delta^2 f(i) = f(k)(\mathbb{E}\mathrm{e}^{-\tau_{i+1,k-1}} - 2\mathbb{E}\mathrm{e}^{-\tau_{i,k-1}} + \mathbb{E}\mathrm{e}^{-\tau_{i-1,k-1}}).$$

For $i \geq k$, using the strong Markov property again in the equalities below, we have

$$\begin{aligned}
&\mathbb{E}(\mathrm{e}^{-\tau_{i+1,k-1}} - 2\mathrm{e}^{-\tau_{i,k-1}} + \mathrm{e}^{-\tau_{i-1,k-1}}) \\
&\quad = \mathbb{E}\mathrm{e}^{-\tau_{i-1,k-1}}(\mathbb{E}\mathrm{e}^{-\tau_{i+1,i-1}} - 2\mathbb{E}\mathrm{e}^{-\tau_{i,i-1}} + 1) \\
&\quad = \mathbb{E}\mathrm{e}^{-\tau_{i-1,k-1}}(\mathbb{E}\mathrm{e}^{-\tau_{i+1,i}}\mathbb{E}\mathrm{e}^{-\tau_{i,i-1}} - 2\mathbb{E}\mathrm{e}^{-\tau_{i,i-1}} + 1) \\
&\quad \geq \mathbb{E}\mathrm{e}^{-\tau_{i-1,k-1}}(\mathbb{E}\mathrm{e}^{-\tau_{i,i-1}} - 1)^2 \geq 0,
\end{aligned}$$

where the inequality follows from

$$\begin{aligned}
\tau_{i,i-1} &= \inf\{t : Z_i(t) = i-1\} \\
&= \inf\{t : Z_i(t) + \mathbf{1}_{\{T>t\}} = i - 1 + \mathbf{1}_{\{T>t\}}\} \\
&\geq \inf\{t : Z_{i+1}(t) = i\} = \tau_{i+1,i}.
\end{aligned}$$

Similarly, for $i \leq k-2$, $\tau_{i-1,i}$ is stochastically less than or equal to $\tau_{i,i+1}$, so

$$\begin{aligned}
&\mathbb{E}(\mathrm{e}^{-\tau_{i+1,k-1}} - 2\mathrm{e}^{-\tau_{i,k-1}} + \mathrm{e}^{-\tau_{i-1,k-1}}) \\
&\quad = \mathbb{E}\mathrm{e}^{-\tau_{i+1,k-1}}(\mathbb{E}\mathrm{e}^{-\tau_{i-1,i+1}} - 2\mathbb{E}\mathrm{e}^{-\tau_{i,i+1}} + 1) \\
&\quad \geq \mathbb{E}\mathrm{e}^{-\tau_{i+1,k-1}}(\mathbb{E}\mathrm{e}^{-\tau_{i,i+1}} - 1)^2 \geq 0.
\end{aligned}$$

Hence $\Delta^2 f(i) \leq 0$ for $i \geq k$ and $i \leq k-2$, which concludes the proof of (ii).

In terms of (iii), we use (ii) to obtain

$$\begin{aligned}
\|\Delta f\|_{k-} &= |\Delta f(k-1)| \\
&= f(k-1) - f(k) \\
&= \pi(A)\frac{1}{\lambda}\left(\frac{F(k-1)}{\pi_{k-1}} - \frac{F(k-2)}{\pi_{k-2}}\right) \\
&= \pi(A)\frac{F(k-1)}{k\pi_k}\left(1 - \frac{F(k-2)}{F(k-1)} \cdot \frac{\lambda}{k-1}\right).
\end{aligned}$$

Likewise,

$$\begin{aligned}
\|\Delta f\|_{k+} &= |\Delta f(k)| \\
&= f(k+1) - f(k) \\
&= \frac{F(k-1)}{\lambda\pi_{k-1}}\pi(A) - \frac{\overline{F}(k+1)}{\lambda\pi_k}F(k-1) \\
&= \pi(A)\frac{F(k-1)}{k\pi_k}\left(1 - \frac{\overline{F}(k+1)}{\overline{F}(k)} \cdot \frac{k}{\lambda}\right).
\end{aligned}$$

Since (iv) is clearly an immediate consequence of (iii), (2.2), and (2.3), the proof of Lemma 4.1 is complete. □

Proof of Theorem 2.1. As in the proof of Lemma 4.1, we set $A=[k,\infty)$ and $h=\mathbf{1}_A$. Then

$$\mathbb{P}(W-a\geq k)-\mathbb{P}(Y\geq k)=\mathbb{E}h(W-a)-\mathrm{Pn}(\lambda)\{h\}.$$

Define

$$\begin{aligned}e_1 &:=\mathbb{E}(h(W-a)-\mathrm{Pn}(\lambda)\{h\})\,\mathbf{1}_{\{W-a<0\}}-\lambda f(0)\mathbb{P}(W-a=-1),\\ e_2 &:=\mathbb{E}(\lambda f(W-a+1)-(W-a)f(W-a)).\end{aligned}$$

Then it follows from (4.1) that

$$\mathbb{P}(W-a\geq k)-\mathbb{P}(Y\geq k)=e_1+e_2. \tag{4.12}$$

For the estimate of e_1, from $f(0)=f(1)$, we know that $\lambda f(0)=-\mathrm{Pn}(\lambda)\{h\}$, and thus

$$e_1=-\mathbb{P}(W-a<-1)\mathrm{Pn}(\lambda)\{h\},$$

which gives

$$|e_1|=\pi(A)\mathbb{P}(W-a<-1). \tag{4.13}$$

For the estimate of e_2, denoting $\tilde{f}(j):=f(j-a)$, we have

$$e_2=\mathbb{E}\{\lambda\Delta\tilde{f}(W)-(W-\mu)\tilde{f}(W)\}. \tag{4.14}$$

Using Lemma 4.1(ii), we have that $\Delta^2\tilde{f}(m)$ is negative for all m except $m=a+k-1$, which implies

$$-\sum_{m\neq k-1}\Delta^2 f(m)\leq\Delta^2 f(k-1)=\|\Delta^2 f\|$$

and

$$\begin{aligned}\mathbb{E}[\Delta^2\tilde{f}(W_i'+l)\mid\mathcal{F}_i] &\leq\Delta^2 f(k-1)\mathbb{P}[W_i'=k-1+a-l\mid\mathcal{F}_i]\leq\|\Delta^2 f\|\theta_i,\\ \mathbb{E}[\Delta^2\tilde{f}(W_i'+l)\mid\mathcal{F}_i] &\geq\sum_{m\neq k-1}\Delta^2 f(m)\mathbb{P}[W_i'=m+a-l\mid\mathcal{F}_i]\geq-\theta_i\|\Delta^2 f\|,\end{aligned}$$

and hence

$$|\mathbb{E}[\Delta^2\tilde{f}(W_i'+l)\mid\mathcal{F}_i]|\leq\|\Delta^2 f\|\theta_i. \tag{4.15}$$

By taking

$$\theta:=\lambda-\sigma^2,$$

we have from (4.14) that

$$
\begin{aligned}
e_2 &= \theta\mathbb{E}\Delta\tilde{f}(W) + \mathbb{E}\{\sigma^2\Delta\tilde{f}(W) - (W-\mu)\tilde{f}(W)\} \\
&= \theta\mathbb{E}\Delta\tilde{f}(W) + \mathbb{E}\left\{\sigma^2\Delta\tilde{f}(W) - \sum_{i\in\mathcal{I}}(X_i-\mu_i)\tilde{f}(W)\right\} \\
&= \theta\mathbb{E}\Delta\tilde{f}(W) + \sigma^2\mathbb{E}\Delta\tilde{f}(W) - \sum_{i\in\mathcal{I}}\mathbb{E}\{(X_i-\mu_i)(\tilde{f}(W)-\tilde{f}(W_i))\} \\
&= \theta\mathbb{E}\Delta\tilde{f}(W) + \sigma^2\mathbb{E}\Delta\tilde{f}(W) - \sum_{i\in\mathcal{I}}\mathbb{E}\left\{(X_i-\mu_i)\left(\sum_{j=0}^{Z_i-1}\Delta\tilde{f}(W_i+j)\right)\right\} \\
&= \theta\mathbb{E}\Delta\tilde{f}(W) + \sigma^2\mathbb{E}\Delta\tilde{f}(W) - \sum_{i\in\mathcal{I}}\mathbb{E}[(X_i-\mu_i)Z_i]\mathbb{E}\Delta\tilde{f}(W_i') \\
&\quad - \sum_{i\in\mathcal{I}}\mathbb{E}\left\{(X_i-\mu_i)\sum_{j=0}^{Z_i-1}[\Delta\tilde{f}(W_i+j)-\Delta\tilde{f}(W_i')]\right\} \\
&= \theta\mathbb{E}\Delta\tilde{f}(W) + \sum_{i\in\mathcal{I}}\mathbb{E}[(X_i-\mu_i)Z_i]\mathbb{E}[\Delta\tilde{f}(W)-\Delta\tilde{f}(W_i')] \\
&\quad - \sum_{i\in\mathcal{I}}\mathbb{E}\left\{(X_i-\mu_i)\sum_{j=0}^{Z_i-1}[\Delta\tilde{f}(W_i+j)-\Delta\tilde{f}(W_i')]\right\} \\
&= \theta\mathbb{E}\Delta\tilde{f}(W) + \sum_{i\in\mathcal{I}}\mathbb{E}[(X_i-\mu_i)Z_i]\mathbb{E}\left[\sum_{j=0}^{Z_i'-1}\Delta^2\tilde{f}(W_i'+j)\right] \\
&\quad - \sum_{i\in\mathcal{I}}\mathbb{E}\left\{(X_i-\mu_i)\sum_{j=0}^{Z_i-1}\sum_{l=0}^{Z_i'-Z_i+j-1}\Delta^2\tilde{f}(W_i'+l)\right\} \\
&= \theta\mathbb{E}\Delta\tilde{f}(W) + \sum_{i\in\mathcal{I}}\mathbb{E}[(X_i-\mu_i)Z_i]\mathbb{E}\left[\sum_{j=0}^{Z_i'-1}\mathbb{E}(\Delta^2\tilde{f}(W_i'+j)\mid\mathcal{F}_i)\right] \\
&\quad - \sum_{i\in\mathcal{I}}\mathbb{E}\left\{(X_i-\mu_i)\sum_{j=0}^{Z_i-1}\sum_{l=0}^{Z_i'-Z_i+j-1}\mathbb{E}(\Delta^2\tilde{f}(W_i'+l)\mid\mathcal{F}_i)\right\},
\end{aligned}
\tag{4.16}
$$

where the third-to-last equality is because $\sum_{i\in\mathcal{I}}\mathbb{E}[(X_i-\mu_i)Z_i]=\sigma^2$ and (X_i, Z_i) is independent of W_i', and the last equality is due to the assumption that $\{X_j : j\in B_i\}$ is $\mathcal{F}_i$-measurable. Using (4.15) in (4.16), we obtain

$$
|e_2| \le \|\Delta f\||\theta| + \|\Delta^2 f\|\sum_{i\in\mathcal{I}}\theta_i\{|\mathbb{E}(X_i-\mu_i)Z_i|\mathbb{E}(Z_i') + \mathbb{E}[|X_i-\mu_i|Z_i(Z_i'-Z_i/2-1/2)]\}.
\tag{4.17}
$$

Now, combining Lemma 4.1(iii, iv), (4.12), (4.13), and (4.17) gives (2.1). □

Proof of Corollary 2.1. Under the setting of the local dependence, the claim follows from Theorem 2.1 by taking $Z_i = Z_i' = X_i$. □

Proof of Theorem 2.2. Recall the Stein representation (4.12) and the estimate (4.13). It remains to tackle (4.14). However,

$$\begin{aligned} e_2 &= \mathbb{E}(\lambda\tilde{f}(W+1) - \mu\tilde{f}(W^s) + a\tilde{f}(W)) \\ &= \mu\mathbb{E}(\tilde{f}(W+1) - \tilde{f}(W^s)) + (\lambda-\mu)\mathbb{E}\Delta\tilde{f}(W), \end{aligned}$$

and thus

$$\begin{aligned} |e_2| &\le \|\Delta f\|(\mu\mathbb{E}|W+1-W^s| + |\lambda-\mu|) \\ &\le \mathrm{Pn}(\lambda)\{h\}[C_1(\lambda,k)(\mu\mathbb{E}|W+1-W^s| + |\lambda-\mu|)]. \end{aligned} \tag{4.18}$$

Hence, combining (4.12), (4.13), and (4.18) completes the proof. □

Proof of Theorem 2.3. Again we make use of the Stein representation (4.12) and the estimate (4.13) so that it suffices to deal with (4.14). To this end, we have

$$\begin{aligned} e_2 &= \mathbb{E}(\lambda\Delta\tilde{f}(W) - (W-\mu)\tilde{f}(W)) \\ &= \mathbb{E}(\lambda\Delta\tilde{f}(W) - \sigma^2\Delta\tilde{f}(W^\star)) \\ &= \mathbb{E}((\lambda-\sigma^2)\Delta\tilde{f}(W) + \sigma^2(\Delta\tilde{f}(W) - \Delta\tilde{f}(W^\star))). \end{aligned}$$

However, with $R = W^\star - W$,

$$\begin{aligned} &\mathbb{E}[\Delta\tilde{f}(W) - \Delta\tilde{f}(W^\star)] \\ &\quad = -\mathbb{E}\left\{\sum_{j=0}^{R-1}\mathbb{E}(\Delta^2\tilde{f}(W+j))\,\mathbf{1}_{R>0} - \sum_{j=1}^{-R}\mathbb{E}(\Delta^2\tilde{f}(W-j))\,\mathbf{1}_{R<0}\right\} \\ &\quad = -\mathbb{E}\left\{\sum_{j=0}^{R-1}\mathbb{E}(\Delta^2\tilde{f}(W+j)\mid R)\,\mathbf{1}_{R>0} - \sum_{j=1}^{-R}\mathbb{E}(\Delta^2\tilde{f}(W-j)\mid R)\,\mathbf{1}_{R<0}\right\}, \end{aligned}$$

and a similar argument for (4.15) ensures

$$|\mathbb{E}(\Delta^2\tilde{f}(W+j)\mid R)| \le \|\Delta^2 f\|\theta_R,$$

and hence

$$|e_2| \le |\lambda-\sigma^2|\|\Delta f\| + \sigma^2\|\Delta^2 f\|\mathbb{E}[|R|\theta_R]. \tag{4.19}$$

The claim follows from combining (4.12), (4.13), and (4.19) and using Lemma 4.1(iii, iv). □

Proof of Proposition 3.2. The first inequality of (3.7) is a direct consequence of [27]. For the second inequality, let $h = \mathbf{1}_{[k,\infty)}$ and f be the solution of the Stein identity (4.1) with $\lambda = \mu$, setting $W_i = W - X_i$, $Y \sim \mathrm{Pn}(\mu)$, the following argument is standard (see [9, page 6]) and we repeat it for ease of reading:

$$\begin{aligned}
&\mathbb{P}(W \geq k) - \mathbb{P}(Y \geq k)\\
&\quad = \mathbb{E}\{\mu f(W+1) - Wf(W)\}\\
&\quad = \mu\mathbb{E}f(W+1) - \sum_{i=1}^{n} \mathbb{E}\{X_i f(W)\}\\
&\quad = \mu\mathbb{E}f(W+1) - \sum_{i=1}^{n} p_i \mathbb{E}\{f(W_i+1)\}\\
&\quad = \sum_{i=1}^{n} p_i^2 \mathbb{E}\Delta f(W_i+1). \qquad (4.20)
\end{aligned}$$

For any non-negative integer-valued random variable U such that the following expectations exist, summation by parts gives

$$\mathbb{E}g(U+1) = \sum_{j=1}^{\infty} \Delta g(j)\mathbb{P}(U \geq j) + g(1).$$

On the other hand, Proposition 2.1 of [8] ensures that

$$\frac{\mathbb{P}(W_i \geq j)}{\mathbb{P}(Y \geq j)} \leq \frac{\mathbb{P}(W \geq j)}{\mathbb{P}(Y \geq j)} \leq \sup_{r \geq 0} \frac{\mathbb{P}(W = r)}{\mathbb{P}(Y = r)} \leq M,$$

so using Lemma 4.1(ii), we have

$$\begin{aligned}
&\mathbb{E}\Delta f(W_i+1)\\
&\quad = \sum_{j=1}^{\infty} \Delta^2 f(j)\mathbb{P}(W_i \geq j) + \Delta f(1)\\
&\quad \geq M \sum_{j \geq 1, j \neq k-1} \Delta^2 f(j)\mathbb{P}(Y \geq j) + \Delta f(1)\\
&\quad = M\left\{\sum_{j=1}^{\infty} \Delta^2 f(j)\mathbb{P}(Y \geq j) + \Delta f(1)\right\} - M\Delta^2 f(k-1)\mathbb{P}(Y \geq k-1) + (1-M)\Delta f(1)\\
&\quad = M\mathbb{E}\Delta f(Y+1) - M\Delta^2 f(k-1)\mathbb{P}(Y \geq k-1) + (1-M)\Delta f(1)\\
&\quad > M\mathbb{E}\Delta f(Y+1) - M\Delta^2 f(k-1)\mathbb{P}(Y \geq k-1). \qquad (4.21)
\end{aligned}$$

However, by (4.2), since $\mathrm{Pn}(\mu)$ is the stationary distribution of Z_i, $Z_Y(t) \sim \mathrm{Pn}(\mu)$, leading to

$$\begin{aligned}
&\mathbb{E}\Delta f(Y+1)\\
&\quad = -\int_0^{\infty} \mathbb{E}[h(Z_{Y+2}(t)) - 2h(Z_{Y+1}(t)) + h(Z_Y(t))]\,\mathrm{d}t\\
&\quad = -\int_0^{\infty} \mathbb{E}[h(Y + \mathbf{1}_{\{T_1>t\}} + \mathbf{1}_{\{T_2>t\}}) - h(Y + \mathbf{1}_{\{T_1>t\}}) - h(Y + \mathbf{1}_{\{T_2>t\}}) + h(Y)]\,\mathrm{d}t\\
&\quad = -\int_0^{\infty} \mathrm{e}^{-2t}\mathbb{E}[\Delta^2 h(Y)]\,\mathrm{d}t = -\frac{1}{2}(\pi_{k-2} - \pi_{k-1}), \qquad (4.22)
\end{aligned}$$

where T_1, T_2 are i.i.d. $\exp(1)$ random variables independent of Y. Combining (4.20), (4.21), and (4.22), we have

$$\frac{\mathbb{P}(W \geq k)}{\mathbb{P}(Y \geq k)} - 1 > -\frac{1}{2} M\mu_2 \frac{\pi_{k-2} - \pi_{k-1}}{\mathbb{P}(Y \geq k)} - M\mu_2 \Delta^2 f(k-1) \frac{\mathbb{P}(Y \geq k-1)}{\mathbb{P}(Y \geq k)}. \tag{4.23}$$

For the first term of (4.23), using [9, Proposition A.2.1 (ii)], we obtain

$$\begin{aligned} \frac{\pi_{k-2} - \pi_{k-1}}{\mathbb{P}(Y \geq k)} &= \frac{k}{\mu} \cdot \frac{\pi_k}{\mathbb{P}(Y \geq k)} \cdot \frac{k-1-\mu}{\mu} \\ &\leq \frac{4(k-\mu)}{\mu} \cdot \frac{k-1-\mu}{\mu} \\ &\leq 4x^2/\mu. \end{aligned} \tag{4.24}$$

For the second term of (4.23), we use the crude estimate of

$$\Delta^2 f(k-1) \leq 2\|\Delta f\| \leq 2(1-\mathrm{e}^{-\mu})/\mu$$

(see [9, Lemma 1.1.1] or Remark 2.1), so applying [9, Proposition A.2.1 (ii)] again,

$$\begin{aligned} &\Delta^2 f(k-1) \frac{\mathbb{P}(Y \geq k-1)}{\mathbb{P}(Y \geq k)} \\ &\quad \leq \frac{2(1-\mathrm{e}^{-\mu})}{\mu} \left(1 + \frac{\pi_k}{\mathbb{P}(Y \geq k)} \cdot \frac{k}{\mu}\right) \\ &\quad \leq \frac{2(1-\mathrm{e}^{-\mu})}{\mu} \left(1 + \frac{4(k-\mu)}{\mu}\right) \\ &\quad \leq \frac{2}{\mu} \left(1 + 4x\sqrt{\frac{1-\mathrm{e}^{-\mu}}{\mu}}\right). \end{aligned} \tag{4.25}$$

The bound (3.7) follows by collecting (4.23), (4.24), and (4.25). □

Acknowledgements

We thank the anonymous referees for suggesting the ‘naive bound’ in Remark 2.1 and for comments leading to the improved version of the paper. We also thank Serguei Novak for email discussions about the quality of the bounds presented in the paper versus the ‘naive bound’. This work was supported in part by the Chinese Scholarship Council and in part by Australian Research Council grant DP190100613.

References

[1] Arratia, R. and Goldstein, L. (2010). Size bias, sampling, the waiting time paradox, and infinite divisibility: when is the increment independent? Available at arXiv:1007.3910.

[2] Arratia, R., Goldstein, L. and Gordon, L. (1989). Two moments suffice for Poisson approximations: the Chen–Stein method. *Ann. Prob.* **17**, 9–25.

[3] Arratia, R., Goldstein, L. and Kochman, F. (2013). Size bias for one and all. Available at arXiv:1308.2729.

[4] Barbour, A. D. (1988). Stein’s method and Poisson process convergence. *J. Appl. Prob* **25** (A), 175–184.

[5] Barbour, A. D. and Brown, T. C. (1992). Stein's method and point process approximation. *Stoch. Proc. Appl.* **43**, 9–31.
[6] Barbour, A. D. and Eagleson, G. K. (1984). Poisson convergence for dissociated statistics. *J. R. Statist. Soc. B [Statist. Methodology]* **46**, 397–402.
[7] Barbour, A. D. and Xia, A. (1999). Poisson perturbations. *ESAIM Prob. Statist.* **3**, 131–150.
[8] 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.
[9] Barbour, A. D., Holst, L. and Janson, S. (1992). *Poisson Approximation.* Oxford University Press.
[10] Borovkov, K. A. (1988). Refinement of Poisson approximation. *Theory Prob. Appl* **33**, 343–347.
[11] Borovkov, K. and Pfeifer, D. (1996). On improvements of the order of approximation in the Poisson limit theorem. *J. Appl. Prob.* **33**, 146–155.
[12] Brown, T. C. and Xia, A. (2001). Stein's method and birth–death processes. 1373–1403.
[13] Čekanavičius, V. and Vellaisamy, P. (2019). On large deviations for sums of discrete m-dependent random variables. *Stochastics* **91** (8), 1092–1108.
[14] Chatterjee, S., Diaconis, P. and Meckes, E. (2005). Exchangeable pairs and Poisson approximation. *Prob. Surv.* **2**, 64–106.
[15] Chen, L. H. Y. (1975). Poisson approximation for dependent trials. *Ann. Prob.* **3**, 534–545.
[16] Chen, L. H. Y. and Choi, K. P. (1992). Some asymptotic and large deviation results in Poisson approximation. *Ann. Prob.* **20**, 1867–1876.
[17] Chen, L. H. Y. and Shao, Q.-M. (2004). Normal approximation under local dependence. *Ann. Prob.* **32**, 1985–2028.
[18] Chen, L. H. Y., Fang, X. and Shao, Q.-M. (2013). Moderate deviations in Poisson approximation: a first attempt. *Statist. Sinica.* **23**, 1523–1540.
[19] Chen, L. H. Y., Fang, X. and Shao, Q.-M. (2013). From Stein identities to moderate deviations. *Ann. Prob.* **41**, 262–293.
[20] Chung, F. and Lu, L. (2006). *Complex Graphs and Networks* (CBMS Regional Conference Series in Mathematics **107**). American Mathematical Society.
[21] Cochran, W. (1977). *Sampling Techniques*. John Wiley & Sons.
[22] Deheuvels, P. and Pfeifer, D. (1988). On a relationship between Uspensky's theorem and Poisson approximations. *Ann. Inst. Statist. Math.* **40** (4), 671–681.
[23] Dwass, M. (1960). Some k-sample rank order tests. In *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, ed. I. Olkin, pp. 198–202. Stanford University Press, CA.
[24] Feller, W. (1968). *An Introduction to Probability Theory and its Applications*, vols 1 and 2, 3rd edn. John Wiley & Sons.
[25] Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. of Math.* **44**, 423–453.
[26] Goldstein, L. and Xia, A. (2006). Zero biasing and a discrete central limit theorem. *Ann. Prob* **34**, 1782–1806.
[27] Hoeffding, W. (1956). On the distribution of the number of successes in independent trials. *Ann. Math. Statist.* **27**, 713–721.
[28] Mattner, L. and Roos, B. (2007). A shorter proof of Kanter's Bessel function concentration bound. *Prob. Theory Rel. Fields* **139**, 191–205.
[29] Petrov, V. V. (1975). *Sums of Independent Random Variables*. Springer.
[30] Rényi, A. (1962). Théorie des éléments saillants d'une suite d'observations. *Ann. Fac. Sci. Univ. Clermont-Ferrand No.* **8**, 7–13.
[31] Röllin, A. (2005). Approximation of sums of conditionally independent variables by the translated Poisson distribution. *Bernoulli* **11**, 1115–1128.
[32] Röllin, A. (2007). Translated Poisson approximation using exchangeable pair couplings. *Ann. Appl. Prob.* **17**, 1596–1614.
[33] Ross, N. (2011). Fundamentals of Stein's method. *Prob. Surv.* **8**, 210–293.
[34] Tan, Y., Lu, Y. and Xia, C. (2018). Relative error of scaled Poisson approximation via Stein's method. Available at arXiv:1810.04300.