Bernoulli **20**(3), 2014, 1404–1431
DOI: 10.3150/13-BEJ527

Discretized normal approximation by Stein's method

XIAO FANG

[1] *Department of Statistics and Applied Probability, National University of Singapore, 6 Science Drive 2, Singapore 117546, Republic of Singapore. E-mail: stafx@nus.edu.sg*

We prove a general theorem to bound the total variation distance between the distribution of an integer valued random variable of interest and an appropriate discretized normal distribution. We apply the theorem to 2-runs in a sequence of i.i.d. Bernoulli random variables, the number of vertices with a given degree in the Erdös–Rényi random graph, and the uniform multinomial occupancy model.

Keywords: discretized normal approximation; exchangeable pairs; local dependence; size biasing; Stein coupling; Stein's method

1. Introduction and the main result

Let S be a sum of independent random variables. The Berry-Esseen theorem gives a bound on the Kolmogorov distance between the distribution of S and the normal distribution with the same mean and variance as S.

Theorem 1.1 (Berry [5], Esseen [11]). *Assume $S = \sum_{i=1}^n X_i$ where $\{X_1, \ldots, X_n\}$ are independent random variables with $\mathbb{E}X_i = \mu_i$, $\mathrm{Var}X_i = \sigma_i^2$, $\mathbb{E}|X_i - \mu_i|^3 = \gamma_i$. Let $\mu = \sum_{i=1}^n \mu_i$, $\sigma^2 = \sum_{i=1}^n \sigma_i^2$, $\gamma = \sum_{i=1}^n \gamma_i$. Then,*

$$d_K\big(\mathcal{L}(S), N\big(\mu, \sigma^2\big)\big) \le c\gamma/\sigma^3, \tag{1.1}$$

where c is an absolute constant and

$$d_K\big(\mathcal{L}(X), \mathcal{L}(Y)\big) = \sup_{z\in\mathbb{R}}\big|\mathbb{P}(X \le z) - \mathbb{P}(Y \le z)\big|.$$

From (1.1), if $\sigma^{-2} = \mathrm{O}(1/n)$ and $\gamma = \mathrm{o}(n^{3/2})$, then

$$d_K\big(\mathcal{L}(S), N\big(\mu, \sigma^2\big)\big) \to 0 \qquad \text{as } n \to \infty. \tag{1.2}$$

A stronger distance, the total variation between two distributions, is defined as

$$d_{\mathrm{TV}}\big(\mathcal{L}(X), \mathcal{L}(Y)\big) = \sup_{A\subset\mathbb{R}}\big|\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)\big|. \tag{1.3}$$

If S is integer valued, the convergence in (1.2) is no longer valid under total variation distance because

$$d_{\mathrm{TV}}\big(\mathcal{L}(S), N\big(\mu, \sigma^2\big)\big) = 1 \qquad \forall n \ge 1. \tag{1.4}$$

1350-7265

Equation (1.4) follows by taking A to be the set of integers in the definition of total variation distance. Therefore, we need to find limiting distributions other than $N(\mu,\sigma^2)$ if small total variation distance is desired. Several alternatives have been studied, e.g., translated Poisson distribution [16,17], shifted binomial distribution [18] and a new family of discrete distributions [14]. A more natural limiting distribution, discretized normal distribution $N^d(\mu,\sigma^2)$, is defined to be supported on the integer set $\mathbb{Z}$ and have probability mass function at any integer $z\in\mathbb{Z}$ as

$$\mathbb{P}\bigl(z-\tfrac{1}{2}\le Z_{\mu,\sigma^2}<z+\tfrac{1}{2}\bigr), \tag{1.5}$$

where Z_{μ,σ^2} is a Gaussian variable with mean μ and variance σ^2.

Using Stein's method, Chen and Leong [7] (see also Theorem 7.4 of [6]) proved a bound on $d_{\mathrm{TV}}(\mathcal{L}(S),N^d(\mu,\sigma^2))$ for sums of independent integer valued random variables. Stein's method was introduced by Stein [20], and has become an important approach in proving distributional approximations because of its power in handling dependence within random variables. We refer to [1] for an introduction to Stein's method.

Chen and Leong [7] used the zero-bias coupling approach in Stein's method to obtain their result. In this paper, we develop a different approach in Stein's method for discretized normal approximation. Our approach not only recovers the result of Chen and Leong [7], but also works for general integer valued random variables. We work under the framework of Stein coupling, a concept introduced by Chen and Röllin [8] under which normal approximation results can be proved.

Definition 1.2. *Let S be a random variable with mean μ. We say a triple of square-integrable random variables (S,S',G) is a Stein coupling if*

$$\mathbb{E}\bigl\{Gf\bigl(S'\bigr)-Gf(S)\bigr\}=\mathbb{E}(S-\mu)f(S) \tag{1.6}$$

for all f such that the above expectations exist.

The above definition is adapted from [8] and includes many of the coupling structures employed in Stein's method such as local dependence, exchangeable pairs, and size biasing. These coupling structures are discussed in Section 2. Under the framework of Stein coupling, we obtain the following theorem.

Theorem 1.3. *Let S be an integer valued random variable with mean μ and finite variance σ^2. Suppose we can construct a Stein coupling (S,S',G). Then, with $D=S'-S$,*

$$\begin{aligned} &d_{\mathrm{TV}}\bigl(\mathcal{L}(S),N^d\bigl(\mu,\sigma^2\bigr)\bigr)\\ &\quad\le\frac{2}{\sigma^2}\sqrt{\operatorname{Var}\bigl(\mathbb{E}(GD|S)\bigr)}+\sqrt{\frac{\pi}{8}}\frac{\mathbb{E}|GD^2|}{\sigma^3}+\frac{\sqrt{\mathbb{E}G^2D^4}}{\sigma^3}\\ &\qquad+\frac{1}{2\sigma^2}\mathbb{E}\bigl[\bigl(\bigl|GD^2\bigr|+|GD|\bigr)d_{\mathrm{TV}}\bigl(\mathcal{L}(S|\mathcal{F}),\mathcal{L}(S+1|\mathcal{F})\bigr)\bigr], \end{aligned} \tag{1.7}$$

where $\mathcal{F}$ is a σ-field such that $\sigma(G,D)\subset\mathcal{F}$ where $\sigma(\cdot)$ denotes the σ-field generated by a random variable.

Remark 1.4. The discretization defined in (1.5) has no loss of generality. For example, one may define another discretized normal distribution $\tilde{N}^d(\mu, \sigma^2)$ with probability mass function at z as

$$\mathbb{P}(z \le Z_{\mu,\sigma^2} < z+1).$$

Then,

$$\begin{aligned} d_{\mathrm{TV}}\big(N^d\big(\mu,\sigma^2\big), \tilde{N}^d\big(\mu,\sigma^2\big)\big) &= d_{\mathrm{TV}}\big(N^d\big(\mu,\sigma^2\big), N^d\big(\mu-\tfrac{1}{2},\sigma^2\big)\big) \\ &\le d_{\mathrm{TV}}\big(N\big(\mu,\sigma^2\big), N\big(\mu-\tfrac{1}{2},\sigma^2\big)\big) \\ &\le c/\sigma, \end{aligned}$$

where c is an absolute constant. It can be seen from (3.9) in the proof of Theorem 1.3 that the bound (1.7) will only differ by a constant factor if one changes the limiting distribution from $N^d(\mu, \sigma^2)$ to $\tilde{N}^d(\mu, \sigma^2)$.

Remark 1.5. The first three terms in the bound (1.7) are comparable to those appearing in the upper bounds of the Kolmogorov or Wasserstein distance for normal approximations (see, e.g., Corollary 2.2 of [8]). The last term in the bound (1.7) arises because we are working in the total variation distance. It is easy to see that such a term must appear by considering the case when S has support restricted to the even integers. Also in bounding this term, we choose appropriate $\mathcal{F}$ so that $d_{\mathrm{TV}}(\mathcal{L}(S|\mathcal{F}), \mathcal{L}(S+1|\mathcal{F}))$ is relatively easy to bound, yet of the same order as $d_{\mathrm{TV}}(\mathcal{L}(S|G, D), \mathcal{L}(S+1|G, D))$.

Röllin and Ross [19] provided a general method of bounding $d_{\mathrm{TV}}(\mathcal{L}(V), \mathcal{L}(V+1))$ for a given integer valued random variable V. It is our main tool for bounding the last term in the bound (1.7).

Lemma 1.6 (Röllin and Ross [19]). *For a given integer valued random variable V, if we can construct an exchangeable pair (V, V') (i.e., $\mathcal{L}(V, V') = \mathcal{L}(V', V)$) so that $P(V - V' = 1) \neq 0$, then*

$$\begin{aligned} & d_{\mathrm{TV}}\big(\mathcal{L}(V), \mathcal{L}(V+1)\big) \\ &\quad \le \frac{\sqrt{\mathrm{Var}(\mathbb{E}(I(V-V'=1)|V))} + \sqrt{\mathrm{Var}(\mathbb{E}(I(V-V'=-1)|V))}}{P(V-V'=1)}. \end{aligned} \tag{1.8}$$

Remark 1.7. To apply Lemma 1.6, we need to construct exchangeable pairs such that the bound in (1.8) is small. A useful method to construct such exchangeable pairs when V is a function of independent random variables is as follows. Suppose $V = f(X_1, \dots, X_n)$ where $\{X_1, \dots, X_n\}$ are independent. Let I be an independent uniform random index from $\{1, \dots, n\}$. Given I, let X_I' be an independent copy of X_I. Define $V' = f(X_1, \dots, X_I' \dots, X_n)$. Then (V, V') is an exchangeable pair. We will use this construction in all the applications considered in this paper.

The remaining of the paper is organized as follows. In Section 2, we show the utility of Theorem 1.3 by adapting it to local dependence, exchangeable pairs, and size biasing, and bounding the total variation distance for discretized normal approximations for 2-runs in a sequence of i.i.d. Bernoulli random variables, the number of vertices with a given degree in the Erdös–Rényi random graph, and the uniform multinomial occupancy model. In Section 3, we give the proof of Theorem 1.3.

2. Applications

In this section, we apply Theorem 1.3 to prove discretized normal approximation results for integer valued random variables with different dependence structures including local dependence, exchangeable pairs, and size biasing.

2.1. Local dependence

A typical setting of local dependence is as follows. Let $S = \sum_{i=1}^n X_i$ be a sum of integer valued random variables with $\mathbb{E}X_i = \mu_i$, $\mu = \sum_{i=1}^n \mu_i$ and $\mathrm{Var}(S) = \sigma^2$. Suppose for each $i \in \{1, \dots, n\}$, there exist neighborhoods $A_i, B_i \subset \{1, \dots, n\}$ such that X_i is independent of $\{X_j : j \notin A_i\}$, and $\{X_j : j \in A_i\}$ is independent of $\{X_j : j \notin B_i\}$. It can be verified as in Section 3.2 of [8] that

$$\big(S, S', G\big) = \Big(S, S - \sum_{j\in A_I} (X_j - \mu_j), -n(X_I - \mu_I)\Big)$$

is a Stein coupling where I is a uniform random index from $\{1, \dots, n\}$ and independent of $\{X_1, \dots, X_n\}$. Theorem 1.3 has the following corollary for local dependence.

Corollary 2.1. *Under the above setting, assume that for every $i \in \{1, \dots, n\}$, $|N(B_i)| \le \theta$ where $N(B_i) = \{j \in \{1, \dots, n\} : A_j \cap B_i \ne \emptyset\}$ and $|\cdot|$ denotes cardinality. Let*

$$\xi_i = \frac{X_i - \mu_i}{\sigma}, \qquad \eta_i = \sum_{j\in A_i} \xi_j.$$

Then,

$$\begin{aligned} &d_{\mathrm{TV}}\big(\mathcal{L}(S), N^d\big(\mu, \sigma^2\big)\big) \\ &\quad \le 2\sqrt{\theta \sum_{i=1}^n \mathbb{E}\xi_i^2\eta_i^2} + \sqrt{\frac{\pi}{8}} \sum_{i=1}^n \mathbb{E}\big|\xi_i\eta_i^2\big| + \sqrt{n \sum_{i=1}^n \mathbb{E}\xi_i^2\eta_i^4} \\ &\qquad + \frac{1}{2}\sum_{i=1}^n \mathbb{E}\big[\big(\sigma\big|\xi_i\eta_i^2\big| + |\xi_i\eta_i|\big) d_{\mathrm{TV}}\big(\mathcal{L}(S|\mathcal{F}_i), \mathcal{L}(S+1|\mathcal{F}_i)\big)\big], \end{aligned} \tag{2.1}$$

where $\mathcal{F}_i$ is a σ-field such that $\sigma(X_j : j \in A_i) \subset \mathcal{F}_i$.

Proof. Let I be a uniform random index from $\{1,\dots,n\}$ and independent of $\{X_1,\dots,X_n\}$. Let $G=-n(X_I-\mu_I)$, $D=-\sum_{j\in A_I}(X_j-\mu_j)$, and let $X_{A_i}=\{X_j: j\in A_i\}$. We bound the right-hand side of (1.7) as follows. From the definition of neighborhoods A_i, B_i, the inequality $\operatorname{Cov}(X,Y)\le(\mathbb{E}X^2+\mathbb{E}Y^2)/2$ and the bound $|N(B_i)|\le\theta$, we have

$$
\begin{aligned}
&\operatorname{Var}\big(\mathbb{E}(GD|S)\big)\le\operatorname{Var}\big(\mathbb{E}\big(GD|\{X_1,\dots,X_n\}\big)\big)\\
&\quad=\operatorname{Var}\Bigg(\sum_{i=1}^n(X_i-\mu_i)\sum_{j\in A_i}(X_j-\mu_j)\Bigg)\\
&\quad\le\sum_{i,i':X_{A_i},X_{A_{i'}}\text{ not independent}}\operatorname{Cov}\Bigg((X_i-\mu_i)\sum_{j\in A_i}(X_j-\mu_j),(X_{i'}-\mu_{i'})\sum_{j'\in A_{i'}}(X_{j'}-\mu_{j'})\Bigg)\\
&\quad\le\sum_{i,i':X_{A_i},X_{A_{i'}}\text{ not independent}}\Bigg\{\frac{\mathbb{E}[(X_i-\mu_i)\sum_{j\in A_i}(X_j-\mu_j)]^2}{2}\\
&\qquad\qquad+\frac{\mathbb{E}[(X_{i'}-\mu_{i'})\sum_{j'\in A_{i'}}(X_{j'}-\mu_{j'})]^2}{2}\Bigg\}\\
&\quad\le\theta\sum_{i=1}^n\mathbb{E}\Bigg[(X_i-\mu_i)\sum_{j\in A_i}(X_j-\mu_j)\Bigg]^2\\
&\quad=\sigma^4\theta\sum_{i=1}^n\mathbb{E}\xi_i^2\eta_i^2.
\end{aligned}
$$

Moreover,

$$
\mathbb{E}|GD|=\sigma^2\sum_{i=1}^n\mathbb{E}|\xi_i\eta_i|,\qquad \mathbb{E}\big|GD^2\big|=\sigma^3\sum_{i=1}^n\mathbb{E}\big|\xi_i\eta_i^2\big|,\qquad \mathbb{E}G^2D^4=n\sigma^6\sum_{i=1}^n\mathbb{E}\xi_i^2\eta_i^4.
$$

The corollary is proved by applying the above bounds in (1.7) with $\mathcal{F}=\sigma(I,\mathcal{F}_I)$. □

We remark that in the case that S is a sum of independent integer valued random variables, a modification of the arguments from intermediate terms in the proof of Theorem 1.3 yields a result similar to Theorem 7.4 of [6].

2.1.1. *2-runs*

We provide a concrete example of local dependence here. Let $\zeta_1,\dots,\zeta_n$ be independent and identically distributed Bernoulli variables with $\mathbb{P}(\zeta_1=1)=1-\mathbb{P}(\zeta_1=0)=p$ where $p\in(0,1)$. Suppose $n\ge 7$. Let $X_i=\zeta_i\zeta_{i+1}$ and $S=\sum_{i=1}^n X_i$. Here and in the rest of this example, indices outside $\{1,\dots,n\}$ are understood as one plus their residues mod n. We can apply Corollary 2.1 with $A_i=\{i-1,i,i+1\}$, $B_i=\{i-2,\dots,i+2\}$, so that $\theta=7$. The mean and variance of S

can be calculated as

$$\mu = \mathbb{E}S = np^2, \qquad \sigma^2 = \operatorname{Var}(S) = n\bigl(p^2 + 2p^3 - 3p^4\bigr). \tag{2.2}$$

Applying (2.1) with $\mathcal{F}_i = \sigma(\zeta_{i-1}, \zeta_i, \zeta_{i+1}, \zeta_{i+2})$, along with the upper bounds $|\xi_i| \le 1/\sigma$, $|\eta_i| \le 3/\sigma$, we have

$$d_{\mathrm{TV}}\bigl(\mathcal{L}(S), N^d\bigl(\mu, \sigma^2\bigr)\bigr) \le c'_p \frac{1}{\sqrt{n}} + c''_p \sup_{a,b\in\{0,1\}} d_{\mathrm{TV}}\bigl(\mathcal{L}(V_{a,b}), \mathcal{L}(V_{a,b}+1)\bigr),$$

where c'_p, c''_p are constants depending on p and with $m = n-4$ and $a, b \in \{0, 1\}$ given,

$$V_{a,b} = a\zeta_1 + \sum_{j=2}^{m} \zeta_{j-1}\zeta_j + b\zeta_m.$$

Regarding $V_{a,b} = f(\zeta_1, \ldots, \zeta_m)$, we define $V'_{a,b} = f(\zeta_1, \ldots, \zeta'_I, \ldots, \zeta_m)$ where I is uniformly chosen from $\{1, \ldots, m\}$, independent of $\{\zeta_1, \ldots, \zeta_m\}$ and given I, ζ'_I is an independent copy of ζ_I. From Remark 1.7, $(V_{a,b}, V'_{a,b})$ is an exchangeable pair. Since given $\{\zeta_1, \ldots, \zeta_m\}$ and $I = i$,

$$\bigl\{V_{a,b} - V'_{a,b} = 1\bigr\} = \begin{cases} \bigl\{a + \zeta_2 = 1, \zeta_1 = 1, \zeta'_1 = 0\bigr\}, & i = 1, \\ \bigl\{b + \zeta_{m-1} = 1, \zeta_m = 1, \zeta'_m = 0\bigr\}, & i = m, \\ \bigl\{\zeta_{i-1} + \zeta_{i+1} = 1, \zeta_i = 1, \zeta'_i = 0\bigr\}, & 2 \le i \le m-1, \end{cases}$$

we have

$$\begin{aligned} &\mathbb{E}\bigl(I\bigl(V_{a,b} - V'_{a,b} = 1\bigr)|\{\zeta_1, \ldots, \zeta_m\}\bigr) \\ &\quad = \frac{1-p}{m}\Biggl[I(a + \zeta_2 = 1, \zeta_1 = 1) + I(b + \zeta_{m-1} = 1, \zeta_m = 1) \\ &\qquad + \sum_{i=2}^{m-1} I(\zeta_{i-1} + \zeta_{i+1} = 1, \zeta_i = 1)\Biggr]. \end{aligned} \tag{2.3}$$

Taking expectation on both sides of (2.3) and lower bounding the right-hand side by the last term lead to

$$\mathbb{P}\bigl(V_{a,b} - V'_{a,b} = 1\bigr) \ge \frac{2(n-6)}{n-4} p^2(1-p)^2.$$

In calculating the variance of the right-hand side of (2.3), we use the fact that each indicator is only correlated with at most two other indicators. Therefore,

$$\begin{aligned} \sqrt{\operatorname{Var}\bigl(\mathbb{E}\bigl(I\bigl(V_{a,b} - V'_{a,b} = 1\bigr)|V_{a,b}\bigr)\bigr)} &\le \sqrt{\operatorname{Var}\bigl(\mathbb{E}\bigl(I\bigl(V_{a,b} - V'_{a,b} = 1\bigr)|\{\zeta_1, \ldots, \zeta_m\}\bigr)\bigr)} \\ &\le \frac{1-p}{n-4}\sqrt{3(n-4)}. \end{aligned}$$

Similarly,

$$\sqrt{\mathrm{Var}\big(\mathbb{E}\big(I\big(V_{a,b}-V'_{a,b}=-1\big)|V_{a,b}\big)\big)}\le \frac{p}{n-4}\sqrt{3(n-4)}.$$

Applying Lemma 1.6, we have

$$d_{\mathrm{TV}}\big(\mathcal{L}(V_{a,b}),\mathcal{L}(V_{a,b}+1)\big)\le \frac{\sqrt{3(n-4)}}{2(n-6)p^2(1-p)^2}.$$

Therefore, we have proved the following proposition.

Proposition 2.2. *For $n\ge 2$, let $\zeta_1,\dots,\zeta_n$ be independent and identically distributed Bernoulli variables with $\mathbb{P}(\zeta_1=1)=1-\mathbb{P}(\zeta_1=0)=p$ where $p\in(0,1)$. Let $X_i=\zeta_i\zeta_{i+1}$ and $S=\sum_{i=1}^n X_i$. We have*

$$d_{\mathrm{TV}}\big(\mathcal{L}(S),N^d\big(\mu,\sigma^2\big)\big)\le c_p/\sqrt{n}, \tag{2.4}$$

where μ and σ^2 are defined as in (2.2) and c_p is a constant depending on p.

We remark that the above argument also applies to k-runs for $k>2$ with straightforward modifications, for example, enlarging the neighborhoods A_i and B_i, changing the definition of $\mathcal{F}_i$, etc.

Total variation approximation for 2-runs was studied by Barbour and Xia [3] and Röllin [16] using the translated Poission approximation. Barbour and Xia [3] assumed some extra conditions on p to obtain a bound on the total variation distance between $\mathcal{L}(S)$ and a translated Poisson distribution. Although the result in [16] is of the same order as the bound in (2.4) in terms of n and applies for all p, the approach used was different from ours.

2.2. Exchangeable pairs

A systematic introduction on the exchangeable pair approach can be found in Stein [21]. The basic setting is as follows. Let (S,S') be an exchangeable pair (i.e., $\mathcal{L}(S,S')=\mathcal{L}(S',S)$) of integer valued random variables with $\mathbb{E}S=\mu$, $\mathrm{Var}(S)=\sigma^2$. Suppose we have the approximate linearity condition,

$$\mathbb{E}\big(S-S'|S\big)=\lambda(S-\mu)+\sigma\mathbb{E}(R|S), \tag{2.5}$$

for a positive number λ and a random variable R. A simple modification of Theorem 1.3 yields the following corollary for exchangeable pairs.

Corollary 2.3. *Let (S,S') be an exchangeable pair of integer valued random variables satisfying (2.5). Let $\mu=\mathbb{E}S$, $\sigma^2=\mathrm{Var}(S)$. We have*

$$\begin{aligned}&d_{\mathrm{TV}}\big(\mathcal{L}(S),N^d\big(\mu,\sigma^2\big)\big)\\&\quad\le\left(\sqrt{\frac{\pi}{2}}+2\right)\frac{\sqrt{\mathbb{E}R^2}}{\lambda}+\frac{\sqrt{\mathrm{Var}(\mathbb{E}((S'-S)^2|S))}}{\lambda\sigma^2}\end{aligned} \tag{2.6}$$

$$+\sqrt{\frac{\pi}{8}}\frac{\mathbb{E}|S'-S|^3}{2\lambda\sigma^3}+\frac{\sqrt{\mathbb{E}|S'-S|^6}}{2\lambda\sigma^3}$$
$$+\frac{1}{4\lambda\sigma^2}\mathbb{E}\big[\big(|S'-S|^3+(S'-S)^2\big)d_{\mathrm{TV}}\big(\mathcal{L}(S|\mathcal{F}),\mathcal{L}(S+1|\mathcal{F})\big)\big],$$

where $\mathcal{F}$ is a σ-field such that $\sigma(S'-S)\subset\mathcal{F}$.

Proof. We follow the proof of Theorem 1.3 with minor modification. Let $G=\frac{1}{2\lambda}(S'-S)$ and $D=S'-S$. From the exchangeability of (S,S'),

$$\mathbb{E}G\big(f(S')+f(S)\big)=0.$$

By (2.5) and the above equality,

$$\mathbb{E}(S-\mu)f(S)=\mathbb{E}\big\{Gf(S')-Gf(S)\big\}-\frac{\sigma}{\lambda}\mathbb{E}f(S)R.$$

Therefore, (3.6) has an extra term $\sigma\mathbb{E}f_h(S)R/\lambda$, which is bounded by $\sqrt{\pi/2}\mathbb{E}|R|/\lambda$ from (3.4). Moreover, from the exchangeability of (S,S') and (2.5),

$$\begin{aligned}\mathbb{E}GD&=\frac{1}{2\lambda}\mathbb{E}(S'-S)^2\\&=\frac{1}{2\lambda}\big[\mathbb{E}(S'-S)S'-\mathbb{E}(S'-S)S\big]\\&=\frac{1}{\lambda}\mathbb{E}(S-S')S=\frac{1}{\lambda}\mathbb{E}(S-S')(S-\mu)\\&=\sigma^2+\sigma\mathbb{E}\big((S-\mu)R\big)/\lambda.\end{aligned}$$

Hence instead of (3.7),

$$|R_1|\le\frac{2}{\sigma^2}\left(\sqrt{\mathrm{Var}\big(\mathbb{E}(GD|S)\big)}+\frac{\sigma}{\lambda}\mathbb{E}\big|(S-\mu)R\big|\right)\le\frac{\sqrt{\mathrm{Var}(\mathbb{E}((S'-S)^2|S))}}{\lambda\sigma^2}+\frac{2}{\lambda}\sqrt{\mathbb{E}R^2}.$$

Corollary 2.3 follows from Theorem 1.3 and the above arguments. □

A special case worth mentioning is when the exchangeable pair (S,S') satisfies $|S-S'|\le 1$. Examples of such exchangeable pairs include binary expansion of a random integer [Diaconis [10]] and anti-voter model [Rinott and Rotar [15]]. The following result shows that under this special assumption, bounding the total variation distance requires no more effort than bounding the Kolmogorov distance.

Corollary 2.4. *Let (S,S') be an exchangeable pair of integer valued random variables satisfying the approximate linearity condition (2.5). In addition, suppose $|S-S'|\le 1$. Then we* *have*

$$d_{\mathrm{TV}}\big(\mathcal{L}(S), N^d\big(\mu,\sigma^2\big)\big) \\ \leq \left(\sqrt{\frac{\pi}{2}}+2\right)\frac{\sqrt{\mathbb{E}R^2}}{\lambda}+\frac{\sqrt{\mathrm{Var}(\mathbb{E}((S'-S)^2|S))}}{\lambda\sigma^2}+\frac{\sqrt{\pi/8}+1}{2\lambda\sigma^3}, \tag{2.7}$$

where μ and σ^2 are the mean and variance of S.

Proof. Let $G=\frac{1}{2\lambda}(S'-S)$, $D=S'-S$. Then for $h\in\mathcal{H}$ defined in (3.2),

$$\begin{aligned}
&\mathbb{E}G\int_0^D\big(h(S+t)-h(S)\big)\,\mathrm{d}t \\
&=\frac{1}{2\lambda}\mathbb{E}\big(S'-S\big)\int_0^{S'-S}\big(h(S+t)-h(S)\big)\,\mathrm{d}t \\
&=\frac{1}{2\lambda}\mathbb{E}\bigg[\int_0^1\big(h(S+t)-h(S)\big)\,\mathrm{d}t\,I\big(S'-S=1\big) \\
&\qquad -\int_0^{-1}\big(h(S+t)-h(S)\big)\,\mathrm{d}t\,I\big(S'-S=-1\big)\bigg] \qquad (2.8)\\
&=\frac{1}{4\lambda}\mathbb{E}\big[\big(h(S+1)-h(S)\big)I\big(S'-S=1\big)+\big(h(S-1)-h(S)\big)I\big(S'-S=-1\big)\big] \\
&=\frac{1}{4\lambda}\mathbb{E}\big[\big(h\big(S'\big)-h(S)\big)I\big(S'-S=1\big)-\big(h(S)-h\big(S'\big)\big)I\big(S-S'=1\big)\big] \\
&=0.
\end{aligned}$$

We used the exchangeability of (S,S') in the last equality. From (2.8), the upper bound in (3.9) can be replaced by 0. Therefore, the bound on $d_{\mathrm{TV}}(\mathcal{L}(S),N^d(\mu,\sigma^2))$ can be deduced similarly as Corollary 2.3 except that we do not have the last term on the right-hand side of (2.6). □

Remark 2.5. Under the condition of Corollary 2.4, Röllin [17] obtained a bound on the total variation distance between $\mathcal{L}(S)$ and a translated Poisson distribution. His result, together with the triangle inequality and easy bounds on the total variation distance between the translated Poisson distribution and the discretized normal distribution, yields a similar bound as (2.7).

2.3. Size biasing

Size biasing was first introduced in the context of Stein's method by Goldstein and Rinott [13]. For S being a nonnegative integer valued random variable with mean μ, we say S^s has the S-size biased distribution if

$$\mathbb{E}Sf(S)=\mathbb{E}\mu f\big(S^s\big)$$

for all f such that the above expectations exist. If in addition S^s is defined on the same probability space as S, then

$$\big(S, S', G\big) = \big(S, S^s, \mu\big) \tag{2.9}$$

is a Stein coupling. Theorem 1.3 has the following corollary for size biasing which easily follows from (2.9).

Corollary 2.6. *Let S be a nonnegative integer valued random variable with mean μ and variance σ^2. Let S^s be defined on the same probability space and have the S-size biased distribution. Then*

$$\begin{aligned} &d_{\mathrm{TV}}\big(\mathcal{L}(S), N^d\big(\mu, \sigma^2\big)\big) \\ &\quad \le \frac{2\mu}{\sigma^2}\sqrt{\operatorname{Var}\big(\mathbb{E}\big(S^s - S|S\big)\big)} + \sqrt{\frac{\pi}{8}}\frac{\mu}{\sigma^3}\mathbb{E}\big|S^s - S\big|^2 + \frac{\mu}{\sigma^3}\sqrt{\mathbb{E}\big|S^s - S\big|^4} \\ &\qquad + \frac{\mu}{2\sigma^2}\mathbb{E}\big[\big(\big|S^s - S\big|^2 + \big|S^s - S\big|\big)d_{\mathrm{TV}}\big(\mathcal{L}(S|\mathcal{F}), \mathcal{L}(S+1|\mathcal{F})\big)\big], \end{aligned} \tag{2.10}$$

where $\mathcal{F}$ is a σ-field such that $\sigma(S^s - S) \subset \mathcal{F}$.

Next, we apply Corollary 2.6 to bound the total variation distance for discretized normal approximations for the number of vertices with a given degree in the Erdös–Rényi random graph, and the uniform multinomial occupancy model. These two models were recently studied by Goldstein [12] and Bartroff and Goldstein [4], respectively. They obtained the same bound for the Kolmogorov distance using the inductive size bias coupling technique introduced by Goldstein [12].

2.3.1. *Number of vertices with a given degree in the Erdös–Rényi random graph*

Let $G(n, p_n)$ be an Erdös–Rényi random graph with vertex set $\{1, \dots, n\}$ and edge probability $p_n \in (0, 1)$. Let S_n be the number of vertices with a given degree $d \ge 0$ in $G(n, p_n)$. The asymptotic normality of S_n was proved in [2] when $np_n \to \theta > 0$. Under the condition

$$\begin{aligned} &\text{there exist } 0 < \theta' \le \theta_n \le \theta'' < \infty, n_0 > 0 \text{ such that} \\ &p_n = \theta_n/(n-1) \qquad \text{for all } n \ge n_0, \end{aligned} \tag{2.11}$$

Goldstein [12] proved a bound on the Kolmogorov distance between the distribution of S_n and $N(\mu_n, \sigma_n^2)$,

$$d_K\big(\mathcal{L}(S_n), N\big(\mu_n, \sigma_n^2\big)\big) \le c_d/\sqrt{n},$$

where μ_n and σ_n^2 are the mean and variance of S_n, respectively. Here and in the rest of this example, let $c_d = c(d, \theta', \theta'', n_0)$ denote positive constants which may depend on $d, \theta', \theta'', n_0$. In the following proposition, we prove a bound on the total variation distance between the distribution of S_n and $N^d(\mu_n, \sigma_n^2)$.

Proposition 2.7. *Let $G(n, p_n)$, $n \geq 2$, be a sequence of Erdös–Rényi random graphs satisfying (2.11). Let S_n be the number of vertices with a given degree d in $G(n, p_n)$. We have*

$$d_{\mathrm{TV}}\big(\mathcal{L}(S_n), N^d\big(\mu_n, \sigma_n^2\big)\big) \leq c_d/\sqrt{n}. \tag{2.12}$$

Proof. Since the total variation distance is always bounded by 1, for $n < \max\{n_0, 8\}$, (2.12) holds true by choosing $c_d = \max\{\sqrt{n_0}, 2\sqrt{2}\}$. Therefore, we assume $n \geq \max\{n_0, 8\}$ in the rest of the proof.

In [12], it was proved that under condition (2.11),

$$\frac{n}{c_d} \leq \mu_n \leq c_d n, \qquad \frac{n}{c_d} \leq \sigma_n^2 \leq c_d n. \tag{2.13}$$

Let $\deg(i)$ denote the degree of vertex i. Then S_n can be expressed as

$$S_n = \sum_{i=1}^{n} I\big(\deg(i) = d\big).$$

Following the construction of size bias coupling in Goldstein and Rinott [13], let I be uniformly chosen from $\{1, \dots, n\}$ and independent of $G(n, p_n)$. If $\deg(I) = d$, then we define $G^s(n, p_n)$, the size biased graph, to be the same as $G(n, p_n)$. If $\deg(I) > d$, then we obtain $G^s(n, p_n)$ from $G(n, p_n)$ by removing $\deg(I) - d$ edges chosen uniformly at random from the edges that connect to I in $G(n, p_n)$. If $\deg(I) < d$, then we obtain $G^s(n, p_n)$ from $G(n, p_n)$ by connecting I to $d - \deg(I)$ vertices chosen uniformly at random from those not connected to I in $G(n, p_n)$. Let S_n^s be the number of vertices with degree d in the graph $G^s(n, p_n)$. It was proved in [13] that S_n^s has the S_n-size biased distribution and

$$\mathrm{Var}\big(\mathbb{E}\big(S_n^s - S_n | S_n\big)\big) \leq c_d/n. \tag{2.14}$$

From the construction of $G^s(n, p_n)$, at most $|\deg(I) - d| + 1$ vertices have different degrees in $G(n, p_n)$ and $G^s(n, p_n)$. Therefore,

$$\big|S_n^s - S_n\big| \leq \big|\deg(I) - d\big| + 1. \tag{2.15}$$

Given I, $\deg(I) \sim \mathrm{Binomial}(n-1, p_n)$. This, together with (2.11), implies that for any positive integer $k \leq 4$,

$$\mathbb{E}\deg(I)^k \leq c_d. \tag{2.16}$$

From (2.15) and (2.16),

$$\mathbb{E}\big|S_n^s - S_n\big|^k \leq c_d, \qquad k \leq 4. \tag{2.17}$$

Applying (2.13), (2.14) and (2.17) in (2.10), the proof will be complete after we show that

$$\mathbb{E}\big[\big(\big|S_n^s - S_n\big|^2 + \big|S_n^s - S_n\big|\big) d_{\mathrm{TV}}\big(\mathcal{L}(S_n|\mathcal{F}), \mathcal{L}(S_n + 1|\mathcal{F})\big)\big] \leq c_d/\sqrt{n} \tag{2.18}$$

for a σ-field $\mathcal{F}$ such that $\sigma(S_n^s - S_n) \subset \mathcal{F}$. For a given I, define

$$A_I = \{I\} \cup \{j : e_{Ij} = 1 \text{ or } e_{Ij}^s = 1\}, \qquad B_I = \{k \notin A_I : e_{kj} = 1 \text{ for some } j \in A_I\},$$

where e_{uv} (e_{uv}^s) is the indicator that there is an edge connecting u and v in $G(n, p_n)$ ($G^s(n, p_n)$). Let

$$\mathcal{F} = \sigma\big(I, A_I, B_I, \{e_{uv} : u \in A_I, v \in A_I \cup B_I\}, \{e_{Iv}^s : v \in A_I\}\big). \tag{2.19}$$

From the construction of $G^s(n, p_n)$, we have $\sigma(S_n^s - S_n) \subset \mathcal{F}$. Let $|\cdot|$ denote cardinality when the argument is a set. From (2.15), (2.16) and $|A_I| = \max(\deg(I), d) + 1$,

$$\begin{aligned}
&\mathbb{E}\big(|S_n^s - S_n|^2 + |S_n^s - S_n|\big) I\big(|A_I| > \sqrt{n}\big) \\
&\quad \le 2\mathbb{E}|A_I|^2 I\big(|A_I| > \sqrt{n}\big) \le \frac{2}{\sqrt{n}}\mathbb{E}|A_I|^3 \\
&\quad = \frac{2}{\sqrt{n}}\mathbb{E}\big(\max\big(\deg(I), d\big) + 1\big)^3 \\
&\quad \le c_d/\sqrt{n}.
\end{aligned}$$

Similarly,

$$\begin{aligned}
&\mathbb{E}\big(|S_n^s - S_n|^2 + |S_n^s - S_n|\big) I\big(|B_I| > \sqrt{n}\big) \\
&\quad \le 2\mathbb{E}|A_I|^2 |B_I|/\sqrt{n} \le 2\mathbb{E}|A_I|^2\big[\mathbb{E}\big(|B_I| | I, A_I\big)\big]/\sqrt{n} \le c_d \mathbb{E}|A_I|^3/\sqrt{n} \le c_d/\sqrt{n},
\end{aligned}$$

where we used $\mathbb{E}(|B_I| | I, A_I) \le c_d |A_I|$, which is from the fact that the expected degree of a given vertex is bounded by c_d under condition (2.11). Therefore, to prove (2.18), we only need to prove

$$\begin{aligned}
&\mathbb{E}\big[\big(|S_n^s - S_n|^2 + |S_n^s - S_n|\big) I\big(|A_I|, |B_I| \le \sqrt{n}\big) d_{\mathrm{TV}}\big(\mathcal{L}(S_n|\mathcal{F}), \mathcal{L}(S_n + 1|\mathcal{F})\big)\big] \\
&\quad \le c_d/\sqrt{n},
\end{aligned} \tag{2.20}$$

where $\mathcal{F}$ was defined in (2.19). Given $\mathcal{F}$ with $|A_I|, |B_I| \le \sqrt{n}$, we define a random graph $G^{\mathcal{F}}$ with vertex set $\{1, \ldots, n\}$ by letting $e_{uv}^{\mathcal{F}} = e_{uv}$ for $u \in A_I, v \in \{1, \ldots, n\}$, and letting $e_{uv}^{\mathcal{F}}$ be independent Bernoulli(p_n) random variables for $u, v \in (A_I)^c$ where $e^{\mathcal{F}}$ is the edge indicator for $G^{\mathcal{F}}$. Let $V^{\mathcal{F}} = \sum_{i=1}^n I(\deg^{\mathcal{F}}(i) = d)$ be the number of vertices with degree d in $G^{\mathcal{F}}$. Then $\mathcal{L}(V^{\mathcal{F}}) = \mathcal{L}(S_n|\mathcal{F})$, which follows from $\mathcal{L}(G^{\mathcal{F}}) = \mathcal{L}(G(n, p_n)|\mathcal{F})$.

In the following we fix a given $\mathcal{F}$ with $|A_I|, |B_I| \le \sqrt{n}$, and prove

$$d_{\mathrm{TV}}\big(\mathcal{L}\big(V^{\mathcal{F}}\big), \mathcal{L}\big(V^{\mathcal{F}} + 1\big)\big) \le c_d/\sqrt{n}. \tag{2.21}$$

For ease of notation, we suppress the superscript $\mathcal{F}$, that is, let $G = G^{\mathcal{F}}$, $V = V^{\mathcal{F}}$, $e = e^{\mathcal{F}}$, $\deg = \deg^{\mathcal{F}}$. To bound $d_{\mathrm{TV}}(\mathcal{L}(V), \mathcal{L}(V + 1))$ using Lemma 1.6 and Remark 1.7, we construct an exchangeable pair (V, V') by uniformly choosing $J \ne K$ from $C_I := (A_I \cup B_I)^c$ and independently resampling e_{JK} to be e'_{JK}. Writing out all four possibilities for $\{V-V'=1\}$,

$$\begin{aligned} I\big(V-V'=1\big) = (1-e_{JK})e'_{JK}\big\{&I\big(\deg(J)=d\big)I\big(\deg(K)\neq d-1,d\big) \\ &+ I\big(\deg(J)\neq d-1,d\big)I\big(\deg(K)=d\big)\big\} \\ + e_{JK}\big(1-e'_{JK}\big)\big\{&I\big(\deg(J)=d\big)I\big(\deg(K)\neq d,d+1\big) \\ &+ I\big(\deg(J)\neq d,d+1\big)I\big(\deg(K)=d\big)\big\}. \end{aligned} \tag{2.22}$$

Let $m=|C_I|\geq n-2\sqrt{n}\geq 2$ (recall $n\geq 8$), and let ξ_1,ξ_2 be independent Binomial$(|B_I|+m-2,p_n)$ random variables. Taking expectation on both sides of (2.22), lower bounding the right-hand side by its first term and observing that $\deg(J)$ and $\deg(K)$ are independent $\sim$ Binomial$(|B_I|+m-2,p_n)$ given that J is not connected to K and $J,K\in C_I$ (thus not connected to A_I), we have

$$\begin{aligned} \mathbb{P}\big(V-V'=1\big) &\geq \mathbb{E}(1-e_{JK})e'_{JK}I\big(\deg(J)=d\big)I\big(\deg(K)\neq d-1,d\big) \\ &= \frac{1}{m(m-1)}\sum_{j,k\in C_I:j\neq k}(1-p_n)p_n\mathbb{P}(\xi_1=d)\mathbb{P}(\xi_2\neq d-1,d). \end{aligned}$$

From (2.11) and $n-2\sqrt{n}\leq |B_I|+m-2\leq n$, we have, for some positive constant c_d,

$$p_n\geq\frac{c_d}{n},\qquad 1-p_n\geq c_d,\qquad \mathbb{P}(\xi_1=d)\geq c_d,\qquad \mathbb{P}(\xi_2\neq d-1,d)\geq c_d.$$

Therefore,

$$\mathbb{P}\big(V-V'=1\big)\geq c_d/n. \tag{2.23}$$

Next, we obtain an upper bound of $\mathrm{Var}(\mathbb{E}(I(V-V'=1)|V))$. By taking expectation with respect to J,K first and then writing the variance of a sum as a sum of covariances, we have

$$\begin{aligned} &\mathrm{Var}\big(\mathbb{E}\big((1-e_{JK})e'_{JK}I\big(\deg(J)=d\big)I\big(\deg(K)\neq d-1,d\big)|V\big)\big) \\ &\quad\leq \mathrm{Var}\big(\mathbb{E}\big((1-e_{JK})e'_{JK}I\big(\deg(J)=d\big)I\big(\deg(K)\neq d-1,d\big)|G,\mathcal{F}\big)\big) \\ &\quad\leq \frac{c_d}{n^4}\mathrm{Var}\Big[\sum_{j,k\in C_I:j\neq k}(1-e_{jk})e'_{jk}I\big(\deg(j)=d\big)I\big(\deg(k)\neq d-1,d\big)\Big] \\ &\quad= \frac{c_d}{n^4}\sum_{\substack{j,k,j',k'\in C_I:\\ j\neq k,j'\neq k',|\{j,k,j',k'\}|=2}}\mathrm{Cov}\big[(1-e_{jk})e'_{jk}I\big(\deg(j)=d\big)I\big(\deg(k)\neq d-1,d\big), \\ &\qquad\qquad (1-e_{j'k'})e'_{j'k'}I\big(\deg\big(j'\big)=d\big)I\big(\deg\big(k'\big)\neq d-1,d\big)\big] \\ &\qquad+\frac{c_d}{n^4}\sum_{\substack{j,k,j',k'\in C_I:\\ j\neq k,j'\neq k',|\{j,k,j',k'\}|=3}}\mathrm{Cov}\big[(1-e_{jk})e'_{jk}I\big(\deg(j)=d\big)I\big(\deg(k)\neq d-1,d\big), \\ &\qquad\qquad (1-e_{j'k'})e'_{j'k'}I\big(\deg\big(j'\big)=d\big)I\big(\deg\big(k'\big)\neq d-1,d\big)\big] \end{aligned}$$

$$+\frac{c_d}{n^4}\sum_{\substack{j,k,j',k'\in C_I:\\ |\{j,k,j',k'\}|=4}} \operatorname{Cov}\big[(1-e_{jk})e'_{jk}I\big(\deg(j)=d\big)I\big(\deg(k)\neq d-1,d\big),$$
$$(1-e_{j'k'})e'_{j'k'}I\big(\deg\big(j'\big)=d\big)I\big(\deg\big(k'\big)\neq d-1,d\big)\big].$$

Since $\mathbb{E}e'_{jk}\le c_d/n$, the first two terms in the above bound are bounded by c_d/n^3. To bound the last term, for any $j,k,j',k'\in C_I$ with $|\{j,k,j',k'\}|=4$, let C be the event that there is no edge connecting $\{j,k\}$ and $\{j',k'\}$ and define

$$\begin{aligned}
a_{jk} &= (1-e_{jk})e'_{jk}I\big(\deg(j)=d\big)I\big(\deg(k)\neq d-1,d\big),\\
\alpha &= \mathbb{E}\big[(1-e_{jk})e'_{jk}I\big(\deg(j)=d\big)I\big(\deg(k)\neq d-1,d\big)|C\big],\\
\beta &= \mathbb{E}\big[(1-e_{jk})e'_{jk}I\big(\deg(j)=d\big)I\big(\deg(k)\neq d-1,d\big)\big].
\end{aligned}$$

From the conditional independence between a_{jk} and $a_{j'k'}$ given C, $\mathbb{P}(C^c)\le c_d/n$ and $\mathbb{E}e'_{jk}\le c_d/n$, we have

$$\begin{aligned}
&\big|\operatorname{Cov}\big[(1-e_{jk})e'_{jk}I\big(\deg(j)=d\big)I\big(\deg(k)\neq d-1,d\big),\\
&\quad (1-e_{j'k'})e'_{j'k'}I\big(\deg\big(j'\big)=d\big)I\big(\deg\big(k'\big)\neq d-1,d\big)\big]\big|\\
&= |\mathbb{E}a_{jk}a_{j'k'}-\mathbb{E}a_{jk}\mathbb{E}a_{j'k'}|\\
&= \big|\mathbb{E}a_{jk}a_{j'k'}I(C)+\mathbb{E}a_{jk}a_{j'k'}I\big(C^c\big)-\beta^2\big|\\
&= \big|\mathbb{E}(a_{jk}|C)\mathbb{E}(a_{j'k'}|C)\mathbb{P}(C)-\beta^2+\mathbb{E}a_{jk}a_{j'k'}I\big(C^c\big)\big|\\
&\le \big|\alpha^2-\beta^2\big|+\alpha^2\mathbb{P}\big(C^c\big)+\mathbb{E}a_{jk}a_{j'k'}I\big(C^c\big)\\
&\le 2\mathbb{E}e'_{jk}|\alpha-\beta|+\big(\mathbb{E}e'_{jk}\big)^2\frac{c_d}{n}\\
&\le c_d|\alpha-\beta|/n+c_d/n^3.
\end{aligned}$$

Let

$$R=(1-e_{jk})I\big(\deg(j)=d\big)I\big(\deg(k)\neq d-1,d\big).$$

We have

$$\begin{aligned}
\alpha-\beta &= \big(\mathbb{E}e'_{jk}\big)\big(\mathbb{E}(R|C)-\mathbb{E}RI(C)-\mathbb{E}RI\big(C^c\big)\big)\\
&= \big(\mathbb{E}e'_{jk}\big)\mathbb{P}\big(C^c\big)\big(\mathbb{E}(R|C)-\mathbb{E}\big(R|C^c\big)\big).
\end{aligned}$$

Since $\mathbb{E}e'_{jk}\le c_d/n$ and $\mathbb{P}(C^c)\le c_d/n$, we have $|\alpha-\beta|\le c_d/n^2$. Therefore,

$$\operatorname{Var}\big(\mathbb{E}\big((1-e_{JK})e'_{JK}I\big(\deg(J)=d\big)I\big(\deg(K)\neq d-1,d\big)|V\big)\big)\le c_d/n^3.$$

After bounding the variances of the other terms appearing in $\mathbb{E}(I(V - V' = 1)|V)$ by the same argument, we conclude that

$$\operatorname{Var}\big(\mathbb{E}\big(I\big(V - V' = 1\big)|V\big)\big) \le c_d/n^3. \tag{2.24}$$

Similarly,

$$\operatorname{Var}\big(\mathbb{E}\big(I\big(V - V' = -1\big)|V\big)\big) \le c_d/n^3. \tag{2.25}$$

Applying (2.23), (2.24) and (2.25) in (1.8), we obtain (2.21), which yields (2.20). □

2.3.2. *Uniform multinomial occupancy model*

We consider the uniform multinomial occupancy model studied by Bartroff and Goldstein [4], to which we refer for the literature on this and related problems. Let $n \ge d \ge 2, m \ge 2$ be integers. Let S be the number of urns having occupancy d when n balls are uniformly distributed among m urns. Bartroff and Goldstein [4] proved

$$d_K\big(\mathcal{L}(S), N\big(\mu, \sigma^2\big)\big) \le \frac{c_d(1 + (n/m)^3)}{\sigma},$$

where μ, σ^2 are the mean and variance of S given by

$$\mu = m\binom{n}{d}\frac{1}{m^d}\left(1 - \frac{1}{m}\right)^{n-d}, \tag{2.26}$$

$$\sigma^2 = \mu - \mu^2 + m(m-1)\binom{n}{d, d, n-2d}\frac{1}{m^{2d}}\left(1 - \frac{2}{m}\right)^{n-2d} \tag{2.27}$$

and c_d is a constant only depending on d. Using Corollary 2.6, we will prove the following bound on the total variation distance between the distribution of S and $N^d(\mu, \sigma^2)$.

Proposition 2.8. *Let $n \ge d \ge 2, m \ge 2$ be positive integers. Let S be the number of urns containing d balls when n balls are uniformly distributed among m urns. Then, with μ, σ^2 given by* (2.26), (2.27), *we have*

$$d_{\mathrm{TV}}\big(\mathcal{L}(S), N^d\big(\mu, \sigma^2\big)\big) \le \frac{c_d(1 + (n/m)^3)}{\sigma}, \tag{2.28}$$

where c_d is a constant only depending on d.

Remark 2.9. Our approach should also work for the cases $d = 0, 1$ if one could prove similar results as Lemma 3.2 and (3.21) of [4] and (2.57) below. However, we do not pursue it here.

Proof. We follow the construction of size bias coupling in [4]. For a given $i \in \{1, \ldots, m\}$, we define m-dimensional random vectors $\mathbf{M}_n, \mathbf{M}_n^i$ as follows. Let $\langle \mathbf{M} \rangle_i$ be the vector obtained by deleting the ith component of $\mathbf{M}$. First, we define the ith components of $\mathbf{M}_n, \mathbf{M}_n^i$ to be

$M_n(i) \sim \text{Binomial}(n, 1/m)$, $M_n^i(i) = d$. Next, let $\mathbf{M}'_{n,i}, \mathbf{R}_n^i$ be m-dimensional random vectors conditionally independent given $M_n(i)$ such that $M'_{n,i}(i) = R_n^i(i) = 0$ and

$$\mathcal{L}\big(\langle \mathbf{M}'_{n,i}\rangle_i | M_n(i)\big) = \text{Multinomial}\big(n - \max\{M_n(i), d\}, m-1\big)$$

and

$$\mathcal{L}\big(\langle \mathbf{R}_n^i\rangle_i | M_n(i)\big) = \text{Multinomial}\big(\big|d - M_n(i)\big|, m-1\big), \tag{2.29}$$

where for positive integers x and y, Multinomial(x, y) denotes the distribution of the numbers of balls in y urns when x balls are uniformly distributed among them. Finally, let

$$\langle \mathbf{M}_n\rangle_i = \langle \mathbf{M}'_{n,i}\rangle_i + I\big(M_n(i) < d\big)\langle \mathbf{R}_n^i\rangle_i$$

and

$$\langle \mathbf{M}_n^i\rangle_i = \langle \mathbf{M}'_{n,i}\rangle_i + I\big(M_n(i) > d\big)\langle \mathbf{R}_n^i\rangle_i.$$

From the above construction,

$$\mathcal{L}(\mathbf{M}_n) = \text{Multinomial}(n, m), \qquad \mathcal{L}\big(\mathbf{M}_n^i\big) = \mathcal{L}\big(\mathbf{M}_n | M_n(i) = d\big).$$

Therefore, the number of urns having occupancy d in the uniform multinomial occupancy model can be written as

$$S = \sum_{j=1}^m I\big(M_n(j) = d\big).$$

Define

$$S^s = \sum_{j=1}^m I\big(M_n^I(j) = d\big),$$

where I is uniformly distributed over $\{1, \dots, m\}$ and independent of all other variables. It was proved in [4] that S^s has the S-size biased distribution. We are now ready to apply Corollary 2.6. In the rest of this proof, let c_d denote absolute constants which may depend on d, and let $|\cdot|$ denote cardinality when the argument is a set.

By 4(a) of Lemma 3.2 and (3.21) of [4], for fixed d, there exists a constant r'_d such that if $\frac{\sigma}{1+(n/m)^3} \ge r'_d$, then

$$\sqrt{\text{Var}\big(\mathbb{E}\big(S^s - S|S\big)\big)} \le c_d \frac{1 + (n/m)^3}{\sqrt{n}}. \tag{2.30}$$

By (3.18), (3.17), (3.16) and 4(a) of Lemma 3.2 of [4], there exists another constant r''_d such that if $\sigma \ge r''_d$, then

$$n \le 2m \log m, \qquad \frac{\mu}{\sigma^2} \le c_d, \qquad \sigma^2 \le c_d n, \qquad n > \max\big\{(d+1)^2, 100\big\}. \tag{2.31}$$

Let $r_d := r'_d \vee r''_d$. The range of n and m can be divided into two parts:

(i) $\frac{\sigma}{1+(n/m)^3} < r_d$,
(ii) $\frac{\sigma}{1+(n/m)^3} \geq r_d$.

Since the total variation distance is always bounded by 1, (2.28) holds true in case (i). Therefore, in the rest of the proof we only need to consider case (ii), where (2.30) and (2.31) hold.

Since $\mathbf{M}_n, \mathbf{M}_n^I$ differ by at most $|M_n(I) - d| + 1$ components,

$$\left|S^s - S\right| \leq \left|M_n(I) - d\right| + 1. \tag{2.32}$$

Recall that given I, $M_n(I) \sim \text{Binomial}(n, 1/m)$. From the bounds on the moments of binomial distributions,

$$\mathbb{E}\left|S^s - S\right|^k \leq c_d\left(1 + \left(\frac{n}{m}\right)^k\right), \qquad k \leq 4. \tag{2.33}$$

The first three terms on the right-hand side of (2.10) are bounded by $c_d \frac{1+(n/m)^3}{\sigma}$ from (2.30), (2.31) and (2.33). Therefore, to prove Proposition 2.8, we only need to show that

$$\mathbb{E}\big[\big(\left|S^s - S\right|^2 + \left|S^s - S\right|\big) d_{\mathrm{TV}}\big(\mathcal{L}(S|\mathcal{F}), \mathcal{L}(S+1|\mathcal{F})\big)\big] \leq c_d \frac{1 + (n/m)^3}{\sigma} \tag{2.34}$$

for a σ-field $\mathcal{F}$ such that $\sigma(S^s - S) \subset \mathcal{F}$. Such a σ-field can be chosen as

$$\mathcal{F} = \sigma\big\{I, M_n(I), \mathbf{R}_n^I, \big\{M_n(j) : R_n^I(j) > 0\big\}\big\}$$

from the construction of $\mathbf{M}_n$ and $\mathbf{M}_n^I$. Write

$$\begin{aligned}
&\mathbb{E}\big[\big(\left|S^s - S\right|^2 + \left|S^s - S\right|\big) d_{\mathrm{TV}}\big(\mathcal{L}(S|\mathcal{F}), \mathcal{L}(S+1|\mathcal{F})\big)\big] \\
&\quad = \mathbb{E}\bigg[\big(\left|S^s - S\right|^2 + \left|S^s - S\right|\big) \\
&\qquad \times I\bigg(M_n(I) + \sum_{j : R_n^I(j) > 0} M_n(j) > \sqrt{n}\bigg) d_{\mathrm{TV}}\big(\mathcal{L}(S|\mathcal{F}), \mathcal{L}(S+1|\mathcal{F})\big)\bigg] \\
&\qquad + \mathbb{E}\bigg[\big(\left|S^s - S\right|^2 + \left|S^s - S\right|\big) \\
&\qquad\quad \times I\bigg(M_n(I) + \sum_{j : R_n^I(j) > 0} M_n(j) \leq \sqrt{n}\bigg) d_{\mathrm{TV}}\big(\mathcal{L}(S|\mathcal{F}), \mathcal{L}(S+1|\mathcal{F})\big)\bigg].
\end{aligned} \tag{2.35}$$

By the construction of $\mathbf{R}_n^I$, (2.29),

$$\left|\big\{j : R_n^I(j) > 0\big\}\right| \leq \left|d - M_n(I)\right|.$$

Also for each j such that $R_n^I(j) > 0$,

$$\mathbb{E}\big(M_n(j)|I, M_n(I), \mathbf{R}_n^I\big) \le M_n(I) + \mathbb{E}B_{n,1/(m-1)} \qquad \big(B_{n,p} \sim \text{Binomial}(n, p)\big). \tag{2.36}$$

For the first term on the right-hand side of (2.35), we bound the total variation distance by 1, and then apply (2.32), (2.36), and the bounds on the moments of binomial distributions,

$$\begin{aligned}
&\mathbb{E}\Big[\big(|S^s - S|^2 + |S^s - S|\big) I\Big(M_n(I) + \sum_{j:R_n^I(j)>0} M_n(j) > \sqrt{n}\Big) d_{\text{TV}}\big(\mathcal{L}(S|\mathcal{F}), \mathcal{L}(S+1|\mathcal{F})\big)\Big] \\
&\quad \le \frac{2}{\sqrt{n}} \mathbb{E}\big(|M_n(I) - d| + 1\big)^2 \Big(M_n(I) + \sum_{j:R_n^I(j)>0} M_n(j)\Big) \\
&\quad \le \frac{c_d}{\sqrt{n}} \mathbb{E}\Big\{1 + \big(M_n(I)\big)^3 + \big(1 + M_n(I)\big)^2 \mathbb{E}\Big(\sum_{j:R_n^I(j)>0} M_n(j)|I, M_n(I), \mathbf{R}_n^I\Big)\Big\} \qquad (2.37) \\
&\quad \le \frac{c_d}{\sqrt{n}} \mathbb{E}\big\{1 + \big(M_n(I)\big)^3 + \big(1 + M_n(I)\big)^3 \big(M_n(I) + \mathbb{E}B_{n,1/(m-1)}\big)\big\} \\
&\quad \le c_d \frac{1 + (n/m)^4}{\sqrt{n}}.
\end{aligned}$$

By observing that for $n \le m$, we have $1/\sqrt{n} \le c_d/\sigma$ from (2.31), and for $m < n \le 2m \log m$, we have (see equation (3.13) of [4] with $\varphi_d(n/m) \le 1$)

$$\sigma^2 \le c_d m \left(\frac{n}{m}\right)^d \mathrm{e}^{-n/m} \qquad \left(\text{therefore } \frac{1}{\sqrt{n}} \le \frac{c_d}{\sigma} \sqrt{\left(\frac{n}{m}\right)^{d-1} \mathrm{e}^{-n/m}}\right), \tag{2.38}$$

the bound in (2.37) can be further bounded by c_d/σ.

To bound the second term on the right-hand side of (2.35), for a given $\mathcal{F}$ with $M_n(I) + \sum_{j:R_n^I(j)>0} M_n(j) \le \sqrt{n}$, let V be the number of urns containing d balls when n_1 balls are uniformly distributed among m_1 urns where

$$\begin{aligned}
n_1 &= n - \Big(M_n(I) + \sum_{j:R_n^I(j)>0} M_n(j)\Big) \ge n - \sqrt{n} \\
&> d + 1 \qquad \big(\text{from } n > (d+1)^2 \text{ and } d \ge 2\big)
\end{aligned} \tag{2.39}$$

and

$$\begin{aligned}
m_1 &= m - 1 - \big|\{j : R_n^I(j) > 0\}\big| \ge m - 1 - \big|M_n(I) - d\big| \ge m - 1 - \sqrt{n} \\
&> 2 \qquad (\text{from } n > 100 \text{ and } n \le 2m \log m).
\end{aligned} \tag{2.40}$$

Then $d_{\mathrm{TV}}(\mathcal{L}(V), \mathcal{L}(V+1)) = d_{\mathrm{TV}}(\mathcal{L}(S|\mathcal{F}), \mathcal{L}(S+1|\mathcal{F}))$. To apply Lemma 1.6, we construct an exchangeable pair (V, V') by picking a ball uniformly from the n_1 balls and distributing it to an independently and uniformly chosen urn from the m_1 urns. Formally, let $\mathbf{M}_{n_1}$ be an m_1-dimensional random vector with distribution

$$\mathcal{L}(\mathbf{M}_{n_1}) = \mathrm{Multinomial}(n_1, m_1).$$

Given $\mathbf{M}_{n_1}$, define two independent random variables $J, K \in \{1, 2, \dots, m_1\}$ with probability mass functions

$$\mathbb{P}(J = j) = \frac{M_{n_1}(j)}{n_1}, \qquad \mathbb{P}(K = k) = \frac{1}{m_1}.$$

Given $\mathbf{M}_{n_1}, J, K$, if $J = K$, let $\mathbf{M}'_{n_1} = \mathbf{M}_{n_1}$, and if $J \neq K$, let $\mathbf{M}'_{n_1}$ be the m_1-dimensional vector with

$$M'_{n_1}(J) = M_{n_1}(J) - 1, \qquad M'_{n_1}(K) = M_{n_1}(K) + 1$$

and $M'_{n_1}(i) = M_{n_1}(i)$ for $i \neq J, K$. Define

$$V = \sum_{j=1}^{m_1} I\big(M_{n_1}(j) = d\big)$$

and

$$V' = \sum_{j=1}^{m_1} I\big(M'_{n_1}(j) = d\big).$$

From the above construction,

$$\begin{aligned}
&\mathbb{E}\big(I\big(V - V' = 1\big)|\mathbf{M}_{n_1}\big) \\
&\quad = \sum_{1 \le j \neq k \le m_1} \frac{M_{n_1}(j)}{m_1 n_1} \big[I\big(M_{n_1}(j) = d\big) I\big(M_{n_1}(k) \neq d-1, d\big) \\
&\qquad\qquad + I\big(M_{n_1}(j) \neq d, d+1\big) I\big(M_{n_1}(k) = d\big)\big], \\
&\mathbb{E}\big(I\big(V - V' = -1\big)|\mathbf{M}_{n_1}\big) \\
&\quad = \sum_{1 \le j \neq k \le m_1} \frac{M_{n_1}(j)}{m_1 n_1} \big[I\big(M_{n_1}(j) \neq d, d+1\big) I\big(M_{n_1}(k) = d-1\big) \\
&\qquad\qquad + I\big(M_{n_1}(j) = d+1\big) I\big(M_{n_1}(k) \neq d-1, d\big)\big].
\end{aligned} \tag{2.41}$$

Taking expectation on both sides of (2.41),

$$\begin{aligned}
&\mathbb{P}\big(V - V' = 1\big) \\
&\quad = \sum_{1\le j\neq k\le m_1}\bigg[\frac{d}{m_1n_1}\mathbb{P}\big(M_{n_1}(j) = d, M_{n_1}(k)\neq d-1, d\big) \\
&\qquad + \frac{1}{m_1n_1}\mathbb{E}M_{n_1}(j)I\big(M_{n_1}(j)\neq d, d+1\big)I\big(M_{n_1}(k) = d\big)\bigg].
\end{aligned}\tag{2.42}$$

Let $B_{n,p}\sim\text{Binomial}(n,p)$. Because binomial distributions do not concentrate on two positive integers, we claim that for a positive constant c_d,

$$\mathbb{P}(B_{n_1-d,1/(m_1-1)}\neq d-1, d)\ge c_d$$

and

$$\mathbb{E}B_{n_1-d,1/(m_1-1)}I(B_{n_1-d,1/(m_1-1)}\neq d, d+1)\ge c_d\frac{n_1}{m_1}.$$

In fact, recall that $d\ge 2$ and write out the binomial probabilities explicitly, we have

$$\begin{aligned}
&\mathbb{P}(B_{n_1-d,1/(m_1-1)} = d-1) + \mathbb{P}(B_{n_1-d,1/(m_1-1)} = d) \\
&\quad \le c_d\big[\mathbb{P}(B_{n_1-d,1/(m_1-1)} = d-2) + \mathbb{P}(B_{n_1-d,1/(m_1-1)} = d+1)\big],
\end{aligned}$$

and

$$\begin{aligned}
&d\mathbb{P}(B_{n_1-d,1/(m_1-1)} = d) + (d+1)\mathbb{P}(B_{n_1-d,1/(m_1-1)} = d+1) \\
&\quad \le c_d\big[(d-1)\mathbb{P}(B_{n_1-d,1/(m_1-1)} = d-1) + (d+2)\mathbb{P}(B_{n_1-d,1/(m_1-1)} = d+2)\big],
\end{aligned}$$

which lead to the claim. Therefore,

$$\begin{aligned}
&\mathbb{P}\big(M_{n_1}(j) = d, M_{n_1}(k)\neq d-1, d\big) \\
&\quad = \mathbb{P}(B_{n_1,1/m_1} = d)\mathbb{P}(B_{n_1-d,1/(m_1-1)}\neq d-1, d) \\
&\quad \ge c_d\mathbb{P}(B_{n_1,1/m_1} = d)
\end{aligned}\tag{2.43}$$

and

$$\begin{aligned}
&\mathbb{E}M_{n_1}(j)I\big(M_{n_1}(j)\neq d, d+1\big)I\big(M_{n_1}(k) = d\big) \\
&\quad = \mathbb{P}(B_{n_1,1/m_1} = d)\mathbb{E}B_{n_1-d,1/(m_1-1)}I(B_{n_1-d,1/(m_1-1)}\neq d, d+1) \\
&\quad \ge c_d\frac{n_1}{m_1}\mathbb{P}(B_{n_1,1/m_1} = d).
\end{aligned}\tag{2.44}$$

By (2.42), (2.43) and (2.44),

$$\mathbb{P}\big(V - V' = 1\big)\ge c_d\bigg(1+\frac{m_1}{n_1}\bigg)\mathbb{P}(B_{n_1,1/m_1} = d).\tag{2.45}$$

We proceed to bound $\mathrm{Var}(\mathbb{E}(I(V-V'=1)|V))$ and $\mathrm{Var}(\mathbb{E}(I(V-V'=-1)|V))$. From (2.41) and the inequality $\mathrm{Var}(X+Y)\le 2(\mathrm{Var}(X)+\mathrm{Var}(Y))$, we have

$$\begin{aligned}
&\mathrm{Var}\big(\mathbb{E}\big(I\big(V-V'=1\big)|V\big)\big)\\
&\quad\le \mathrm{Var}\big(\mathbb{E}\big(I\big(V-V'=1\big)|\mathbf{M}_{n_1}\big)\big)\\
&\quad\le \frac{2}{m_1^2n_1^2}\biggl[\mathrm{Var}\biggl(d\sum_{1\le j\neq k\le m_1} I\big(M_{n_1}(j)=d\big)I\big(M_{n_1}(k)\neq d-1,d\big)\biggr)\\
&\qquad\quad+\mathrm{Var}\biggl(\sum_{1\le j\neq k\le m_1} M_{n_1}(j)I\big(M_{n_1}(j)\neq d,d+1\big)I\big(M_{n_1}(k)=d\big)\biggr)\biggr].
\end{aligned}\tag{2.46}$$

Let

$$a_{n_1,m_1}(j,k):=I\big(M_{n_1}(j)=d\big)I\big(M_{n_1}(k)\neq d-1,d\big),\tag{2.47}$$

and let $U_l\in\{1,\dots,m_1\}$ denote the location of the lth ball. Applying the arguments in Bartroff and Goldstein [4] (page 17, equation (3.41) and (3.42)),

$$\begin{aligned}
&\mathrm{Var}\biggl(\sum_{1\le j\neq k\le m_1} a_{n_1-1,m_1}(j,k)\biggr)\\
&\quad\le n_1\mathbb{E}\biggl[\sum_{1\le k\le m_1,k\neq U_{n_1}}\big(a_{n_1,m_1,(n_1)}(U_{n_1},k)-a_{n_1,m_1}(U_{n_1},k)\big)\\
&\qquad\quad+\sum_{1\le j\le m_1,j\neq U_{n_1}}\big(a_{n_1,m_1,(n_1)}(j,U_{n_1})-a_{n_1,m_1}(j,U_{n_1})\big)\biggr]^2,
\end{aligned}\tag{2.48}$$

where $a_{n_1,m_1,(n_1)}(j,k)$ is the value of $a_{n_1,m_1}(j,k)$ when withholding ball n_1, that is,

$$a_{n_1,m_1,(n_1)}(j,k)=I\big(M_{n_1}^{(n_1)}(j)=d\big)I\big(M_{n_1}^{(n_1)}(k)\neq d-1,d\big)\tag{2.49}$$

with

$$M_{n_1}^{(n_1)}(j)=\begin{cases}M_{n_1}(j), & \text{if } j\neq U_{n_1},\\ M_{n_1}(j)-1, & \text{if } j=U_{n_1}.\end{cases}$$

By the definition of U_l, given U_{n_1},

$$M_{n_1}(U_{n_1})-1\sim \mathrm{Binomial}\biggl(n_1-1,\frac{1}{m_1}\biggr).\tag{2.50}$$

Substituting (2.47) and (2.49) in (2.48), and then applying the inequality $\mathbb{E}(\sum_{i=1}^n X_i)^2 \le n\mathbb{E}(X_i)^2$ and (2.50), we have

$$
\begin{aligned}
&\operatorname{Var}\Biggl(\sum_{1\le j\ne k\le m_1} a_{n_1-1,m_1}(j,k)\Biggr)\\
&\quad\le n_1\mathbb{E}\Biggl\{\sum_{1\le k\le m_1, k\ne U_{n_1}}\bigl[I\bigl(M_{n_1}(U_{n_1})=d+1\bigr)I\bigl(M_{n_1}(k)\ne d-1,d\bigr)\\
&\qquad\qquad - I\bigl(M_{n_1}(U_{n_1})=d\bigr)I\bigl(M_{n_1}(k)\ne d-1,d\bigr)\bigr]\\
&\qquad + \sum_{1\le j\le m_1, j\ne U_{n_1}}\bigl[I\bigl(M_{n_1}(j)=d\bigr)I\bigl(M_{n_1}(U_{n_1})\ne d,d+1\bigr)\\
&\qquad\qquad - I\bigl(M_{n_1}(j)=d\bigr)I\bigl(M_{n_1}(U_{n_1})\ne d-1,d\bigr)\bigr]\Biggr\}^2\\
&\quad\le 2n_1m_1\Biggl\{\sum_{1\le k\le m_1}\mathbb{E}\bigl[I\bigl(M_{n_1}(U_{n_1})=d+1\bigr)I\bigl(M_{n_1}(k)\ne d-1,d\bigr)\\
&\qquad\qquad - I\bigl(M_{n_1}(U_{n_1})=d\bigr)I\bigl(M_{n_1}(k)\ne d-1,d\bigr)\bigr]^2\\
&\qquad + \sum_{1\le j\le m_1}\mathbb{E}\bigl[I\bigl(M_{n_1}(j)=d\bigr)I\bigl(M_{n_1}(U_{n_1})\ne d,d+1\bigr)\\
&\qquad\qquad - I\bigl(M_{n_1}(j)=d\bigr)I\bigl(M_{n_1}(U_{n_1})\ne d-1,d\bigr)\bigr]^2\Biggr\}\\
&\quad\le 4n_1m_1\bigl\{m_1\bigl[\mathbb{P}\bigl(M_{n_1}(U_{n_1})=d+1\bigr)+\mathbb{P}\bigl(M_{n_1}(U_{n_1})=d\bigr)\bigr]+2m_1\mathbb{P}\bigl(M_{n_1}(1)=d\bigr)\bigr\}\\
&\quad\le c_d n_1 m_1^2\bigl[\mathbb{P}(B_{n_1-1,1/m_1}=d-1 \text{ or } d)+\mathbb{P}(B_{n_1,1/m_1}=d)\bigr].
\end{aligned}
\tag{2.51}
$$

Next, let

$$
b_{n_1,m_1}(j,k) := M_{n_1}(j)I\bigl(M_{n_1}(j)\ne d,d+1\bigr)I\bigl(M_{n_1}(k)=d\bigr).
$$

By the same argument as for $\operatorname{Var}(\sum_{1\le j\ne k\le m_1} a_{n_1-1,m_1}(j,k))$,

$$
\begin{aligned}
&\operatorname{Var}\Biggl(\sum_{1\le j\ne k\le m_1} b_{n_1-1,m_1}(j,k)\Biggr)\\
&\quad\le n_1\mathbb{E}\Biggl\{\sum_{1\le k\le m_1, k\ne U_{n_1}}\bigl[\bigl(M_{n_1}(U_{n_1})-1\bigr)I\bigl(M_{n_1}(U_{n_1})\ne d+1,d+2\bigr)I\bigl(M_{n_1}(k)=d\bigr)\\
&\qquad\qquad - M_{n_1}(U_{n_1})I\bigl(M_{n_1}(U_{n_1})\ne d,d+1\bigr)I\bigl(M_{n_1}(k)=d\bigr)\bigr]
\end{aligned}
$$

$$+\sum_{1\le j\le m_1, j\neq U_{n_1}}\bigl[M_{n_1}(j)I\bigl(M_{n_1}(j)\neq d,d+1\bigr)I\bigl(M_{n_1}(U_{n_1})=d+1\bigr)$$

$$-M_{n_1}(j)I\bigl(M_{n_1}(j)\neq d,d+1\bigr)I\bigl(M_{n_1}(U_{n_1})=d\bigr)\bigr]\Bigr\}^2 \tag{2.52}$$

$$\le c_d n_1 m_1 \mathbb{E}\Bigl\{\sum_{1\le k\le m_1, k\neq U_{n_1}}\bigl[\mathbb{P}(B_{n_1-1,1/m_1}=d)\mathbb{E}(1+B_{n_1-d-1,1/(m_1-1)})^2\bigr]$$

$$+\sum_{1\le j\le m_1, j\neq U_{n_1}}\bigl[\mathbb{P}(B_{n_1-1,1/m_1}=d)\mathbb{E}B^2_{n_1-d-1,1/(m_1-1)}$$

$$+\mathbb{P}(B_{n_1-1,1/m_1}=d-1)\mathbb{E}B^2_{n_1-d,1/(m_1-1)}\bigr]\Bigr\}$$

$$\le c_d n_1 m_1^2\biggl[\biggl(\frac{n_1}{m_1}+\Bigl(\frac{n_1}{m_1}\Bigr)^2\biggr)\mathbb{P}(B_{n_1-1,1/m_1}=d-1)+\biggl(1+\Bigl(\frac{n_1}{m_1}\Bigr)^2\biggr)\mathbb{P}(B_{n_1-1,1/m_1}=d)\biggr].$$

From (2.46) and the bounds (2.51) and (2.52),

$$\begin{aligned}&\mathrm{Var}\bigl(\mathbb{E}\bigl(I\bigl(V-V'=1\bigr)|V\bigr)\bigr)\\ &\quad\le\frac{c_d}{n_1}\biggl(1+\Bigl(\frac{n_1}{m_1}\Bigr)^2\biggr)\bigl[\mathbb{P}(B_{n_1,1/m_1}=d-1\text{ or }d)+\mathbb{P}(B_{n_1+1,1/m_1}=d)\bigr] \qquad (2.53)\\ &\quad\le\frac{c_d}{n_1}\biggl(1+\Bigl(\frac{n_1}{m_1}\Bigr)^2\biggr)\mathbb{P}(B_{n_1,1/m_1}=d-1\text{ or }d).\end{aligned}$$

The last inequality follows from

$$\mathbb{P}(B_{n_1+1,1/m_1}=d)\le c_d\mathbb{P}(B_{n_1,1/m_1}=d)$$

by writing our these probabilities explicitly.

By the same argument as in proving (2.46), (2.51) and (2.52),

$$\begin{aligned}&\mathrm{Var}\bigl(\mathbb{E}\bigl(I\bigl(V-V'=-1\bigr)|V\bigr)\bigr)\\ &\quad\le\frac{2}{m_1^2n_1^2}\biggl[\mathrm{Var}\biggl((d+1)\sum_{1\le j\neq k\le m_1}I\bigl(M_{n_1}(j)=d+1\bigr)I\bigl(M_{n_1}(k)\neq d-1,d\bigr)\biggr)\\ &\qquad+\mathrm{Var}\biggl(\sum_{1\le j\neq k\le m_1}M_{n_1}(j)I\bigl(M_{n_1}(j)\neq d,d+1\bigr)I\bigl(M_{n_1}(k)=d-1\bigr)\biggr)\biggr] \qquad (2.54)\\ &\quad\le\frac{c_d}{n_1}\mathbb{P}(B_{n_1,1/m_1}=d\text{ or }d+1)\\ &\qquad+\frac{c_d}{n_1}\biggl[\biggl(\frac{n_1}{m_1}+\Bigl(\frac{n_1}{m_1}\Bigr)^2\biggr)\mathbb{P}(B_{n_1,1/m_1}=d-2)+\biggl(1+\Bigl(\frac{n_1}{m_1}\Bigr)^2\biggr)\mathbb{P}(B_{n_1,1/m_1}=d-1)\biggr].\end{aligned}$$

Applying Lemma 1.6 with (2.45), (2.53) and (2.54), we obtain

$$
\begin{aligned}
&d_{\mathrm{TV}}\bigl(\mathcal{L}(V), \mathcal{L}(V+1)\bigr)\\
&\quad\le \frac{c_d}{\sqrt{n_1}} \frac{1}{(1+m_1/n_1)\mathbb{P}(B_{n_1,1/m_1}=d)}\\
&\qquad\times\Biggl\{\biggl(\sqrt{\frac{n_1}{m_1}}+\frac{n_1}{m_1}\biggr)\sqrt{\mathbb{P}(B_{n_1,1/m_1}=d-2)}\\
&\qquad\quad+\biggl(1+\frac{n_1}{m_1}\biggr)\sqrt{\mathbb{P}(B_{n_1,1/m_1}=d-1 \text{ or } d)}\\
&\qquad\quad+\sqrt{\mathbb{P}(B_{n_1,1/m_1}=d+1)}\Biggr\}\\
&\quad\le c_d\biggl(1+\sqrt{\frac{n_1}{m_1}}\biggr)\frac{1}{\sqrt{m_1(n_1/m_1)^d(1-1/m_1)^{n_1-d}}}.
\end{aligned}
$$

The last inequality was obtained by writing out the binomial probabilities explicitly. For example,

$$
\begin{aligned}
&\frac{c_d}{\sqrt{n_1}}\frac{\sqrt{n_1/m_1}+(n_1/m_1)}{1+m_1/n_1}\frac{\sqrt{\mathbb{P}(B_{n_1,1/m_1}=d-2)}}{\mathbb{P}(B_{n_1,1/m_1}=d)}\\
&\quad=\frac{c_d}{\sqrt{n_1}}\frac{\sqrt{n_1/m_1}+(n_1/m_1)}{1+m_1/n_1}\frac{\sqrt{\binom{n_1}{d-2}m_1^{-(d-2)}(1-1/m_1)^{n_1-(d-2)}}}{\binom{n_1}{d}m_1^{-d}(1-1/m_1)^{n_1-d}}\\
&\quad\le\frac{c_d}{\sqrt{n_1}}\frac{\sqrt{n_1/m_1}+(n_1/m_1)}{1+m_1/n_1}\frac{\sqrt{(n_1/m_1)^{d-2}(1-1/m_1)^{n_1-d}}}{(n_1/m_1)^d(1-1/m_1)^{n_1-d}}\\
&\quad\le\frac{c_d}{\sqrt{n_1}}\frac{\sqrt{n_1/m_1}+n_1/m_1}{1+m_1/n_1}\frac{m_1/n_1}{\sqrt{(n_1/m_1)^d(1-1/m_1)^{n_1-d}}}\\
&\quad\le c_d\biggl(1+\sqrt{\frac{n_1}{m_1}}\biggr)\frac{1}{\sqrt{m_1(n_1/m_1)^d(1-1/m_1)^{n_1-d}}}.
\end{aligned}
$$

From (2.39), (2.40) and $n\le 2m\log m$ in (2.31), we have

$$
n_1\asymp n,\qquad m_1\asymp m,\qquad \frac{n_1-d}{m_1^2}\le c_d, \tag{2.55}
$$

and hence,

$$
\biggl(1-\frac{1}{m_1}\biggr)^{n_1-d}\ge\frac{c_d}{(1+1/m_1)^{n_1-d}}\ge\frac{c_d}{\mathrm{e}^{(n_1-d)/m_1}}. \tag{2.56}
$$

By (2.38), (2.55) and (2.56),

$$
\begin{aligned}
& d_{\mathrm{TV}}\big(\mathcal{L}(V), \mathcal{L}(V+1)\big) \\
& \quad \le c_d\left(1+\sqrt{\frac{n_1}{m_1}}\right)\frac{1}{\sigma\sqrt{\mathrm{e}^{n/m}(1-1/m_1)^{n_1-d}}} \\
& \quad \le \frac{c_d(1+\sqrt{n/m})}{\sigma}\sqrt{\exp\left(\frac{n_1-d}{m_1}-\frac{n}{m}\right)} \le \frac{c_d(1+\sqrt{n/m})}{\sigma}.
\end{aligned}
\tag{2.57}
$$

This, together with (2.33), proves that the second term on the right-hand side of (2.35) is bounded by $c_d(1+(n/m)^{5/2})/\sigma$. Therefore, (2.34) is proved. □

3. Proof of Theorem 1.3

From the definition of $N^d(\mu,\sigma^2)$, (1.5), we have

$$
d_{\mathrm{TV}}\big(\mathcal{L}(S), N^d\big(\mu,\sigma^2\big)\big) = \sup_{h\in\mathcal{H}}\big|\mathbb{E}h(S) - \mathbb{E}h(Z_{\mu,\sigma^2})\big|, \tag{3.1}
$$

where Z_{μ,σ^2} is a Gaussian variable with mean μ and variance σ^2 and

$$
\mathcal{H} = \left\{h:\mathbb{R}\to\{0,1\}, h(x)=h(z) \text{ when } z-\frac{1}{2}\le x< z+\frac{1}{2} \text{ for } z\in\mathbb{Z}\right\}. \tag{3.2}
$$

For each $h\in\mathcal{H}$, consider the following Stein equation,

$$
\sigma^2 f'(s) - (s-\mu)f(s) = h(s) - \mathbb{E}h(Z_{\mu,\sigma^2}). \tag{3.3}
$$

It is known (see [9]) that there exists a bounded solution f_h to (3.3) and

$$
\|f_h\| \le \sqrt{\frac{\pi}{2}}\frac{1}{\sigma}, \qquad \|f_h'\| \le \frac{2}{\sigma^2}. \tag{3.4}
$$

By (3.1) and (3.3),

$$
d_{\mathrm{TV}}\big(\mathcal{L}(S), N^d\big(\mu,\sigma^2\big)\big) = \sup_{h\in\mathcal{H}}\big|\mathbb{E}\sigma^2 f_h'(S) - \mathbb{E}(S-\mu)f_h(S)\big|. \tag{3.5}
$$

Since (S, S', G) satisfies (1.6), we have

$$
\begin{aligned}
& \mathbb{E}\sigma^2 f_h'(S) - \mathbb{E}(S-\mu)f_h(S) \\
& \quad = \mathbb{E}\sigma^2 f_h'(S) - \mathbb{E}\big\{Gf_h\big(S'\big) - Gf_h(S)\big\} \\
& \quad = \mathbb{E}\sigma^2 f_h'(S) - \mathbb{E}GDf_h'(S) - \mathbb{E}G\int_0^D \big(f_h'(S+t) - f_h'(S)\big)\,\mathrm{d}t \\
& \quad = R_1 - R_2,
\end{aligned}
\tag{3.6}
$$

where

$$R_1 = \mathbb{E} f_h'(S)\big(\sigma^2 - GD\big),$$
$$R_2 = \mathbb{E} G \int_0^D \big(f_h'(S+t) - f_h'(S)\big)\,\mathrm{d}t.$$

From (1.6), $\mathbb{E}GD = \sigma^2$. This, along with (3.4), yields

$$|R_1| \le \frac{2\sqrt{\mathrm{Var}(\mathbb{E}(GD|S))}}{\sigma^2}. \tag{3.7}$$

For R_2, since f_h solves (3.3),

$$\begin{aligned} R_2 &= \mathbb{E} G \int_0^D \frac{1}{\sigma^2}\big((S+t-\mu) f_h(S+t) - (S-\mu) f_h(S) + h(S+t) - h(S)\big)\,\mathrm{d}t \\ &= \mathbb{E} G \int_0^D \frac{1}{\sigma^2}\big(t f_h(S+t) + (S-\mu)\big(f_h(S+t) - f_h(S)\big) + h(S+t) - h(S)\big)\,\mathrm{d}t. \end{aligned} \tag{3.8}$$

Using (3.4), the first two summands in (3.8) can be bounded by

$$\sqrt{\frac{\pi}{8}}\frac{1}{\sigma^3}\mathbb{E}\big|GD^2\big| + \frac{1}{\sigma^4}\mathbb{E}\big|GD^2(S-\mu)\big|.$$

From (3.2) and (1.3),

$$\begin{aligned} &\frac{1}{\sigma^2}\left|\mathbb{E} G \int_0^D \big(h(S+t) - h(S)\big)\,\mathrm{d}t\right| \\ &= \frac{1}{\sigma^2}\left|\mathbb{E} G \int_{-\infty}^{\infty} \big[I(0 \le t \le D) - I(D \le t < 0)\big]\big[\mathbb{E}^{\mathcal{F}}\big(h(S+t) - h(S)\big)\big]\,\mathrm{d}t\right| \\ &\le \frac{1}{\sigma^2}\mathbb{E}|G| \int_{-\infty}^{\infty} \big|I(0 \le t \le D) - I(D \le t < 0)\big|\big|\mathbb{E}^{\mathcal{F}}\big(h(S+t) - h(S)\big)\big|\,\mathrm{d}t \\ &\le \frac{1}{\sigma^2}\mathbb{E}|G| \int_{-\infty}^{\infty} \big|I(0 \le t \le D) - I(D \le t < 0)\big|\left(|t| + \frac{1}{2}\right) d_{\mathrm{TV}}\big(\mathcal{L}(S|\mathcal{F}), \mathcal{L}(S+1|\mathcal{F})\big)\,\mathrm{d}t \\ &\le \frac{1}{2\sigma^2}\mathbb{E}\big[\big(\big|GD^2\big| + |GD|\big) d_{\mathrm{TV}}\big(\mathcal{L}(S|\mathcal{F}), \mathcal{L}(S+1|\mathcal{F})\big)\big]. \end{aligned} \tag{3.9}$$

Therefore,

$$\begin{aligned} |R_2| \le{}& \sqrt{\frac{\pi}{8}}\frac{1}{\sigma^3}\mathbb{E}\big|GD^2\big| + \frac{\sqrt{\mathbb{E}G^2D^4}}{\sigma^3} \\ &+ \frac{1}{2\sigma^2}\mathbb{E}\big[\big(\big|GD^2\big| + |GD|\big) d_{\mathrm{TV}}\big(\mathcal{L}(S|\mathcal{F}), \mathcal{L}(S+1|\mathcal{F})\big)\big]. \end{aligned} \tag{3.10}$$

The theorem is proved by using (3.5), (3.6) and the bounds (3.7), (3.10).

Acknowledgements

This work is based on part of the Ph.D. thesis of the author. The author is thankful to his advisor, Louis H.Y. Chen, for his guidance and helpful discussions. The author would also like to thank a referee and the Associate Editor whose suggestions have significantly improved the presentation of this paper. This work is partially supported by Grant C-389-000-010-101 and Grant C-389-000-012-101 at the National University of Singapore.

References

[1] Barbour, A.D. and Chen, L.H.Y. (2005). An Introduction to Stein's Method. *Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap.* **4**. Singapore: Singapore Univ. Press.

[2] Barbour, A.D., Karoński, M. and Ruciński, A. (1989). A central limit theorem for decomposable random variables with applications to random graphs. *J. Combin. Theory Ser. B* **47** 125–145. MR1047781

[3] Barbour, A.D. and Xia, A. (1999). Poisson perturbations. *ESAIM Probab. Statist.* **3** 131–150. MR1716120

[4] Bartroff, J. and Goldstein, L. (2013). A Berry–Esseen bound for the uniform multinomial occupancy model. *Electron. J. Probab.* **18** no. 27, 29. MR3035755

[5] Berry, A.C. (1941). The accuracy of the Gaussian approximation to the sum of independent variates. *Trans. Amer. Math. Soc.* **49** 122–136. MR0003498

[6] Chen, L.H.Y., Goldstein, L. and Shao, Q.M. (2011). *Normal Approximation by Stein's Method. Probability and Its Applications (New York)*. Heidelberg: Springer. MR2732624

[7] Chen, L.H.Y. and Leong, Y.K. (2010). From zero-bias to discretized normal approximation. Personal communication.

[8] Chen, L.H.Y. and Röllin, A. (2010). Stein couplings for normal approximation. Preprint. Available at http://arxiv.org/abs/1003.6039v2.

[9] Chen, L.H.Y. and Shao, Q.M. (2005). Stein's method for normal approximation. In *An Introduction to Stein's Method* (A.D. Barbour and L.H.Y. Chen, eds.). *Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap.* **4** 1–59. Singapore: Singapore Univ. Press. MR2235448

[10] Diaconis, P. (1977). The distribution of leading digits and uniform distribution mod 1. *Ann. Probability* **5** 72–81. MR0422186

[11] Esseen, C.G. (1942). On the Liapounoff limit of error in the theory of probability. *Ark. Mat. Astr. Fys.* **28A** 1–19. MR0011909

[12] Goldstein, L. (2013). A Berry–Esseen bound with applications to vertex degree counts in the Erdős–Rényi random graph. *Ann. Appl. Probab.* **23** 617–636. MR3059270

[13] Goldstein, L. and Rinott, Y. (1996). Multivariate normal approximations by Stein's method and size bias couplings. *J. Appl. Probab.* **33** 1–17. MR1371949

[14] Goldstein, L. and Xia, A. (2006). Zero biasing and a discrete central limit theorem. *Ann. Probab.* **34** 1782–1806. MR2271482

[15] Rinott, Y. and Rotar, V. (1997). On coupling constructions and rates in the CLT for dependent summands with applications to the antivoter model and weighted U-statistics. *Ann. Appl. Probab.* **7** 1080–1105. MR1484798

[16] Röllin, A. (2005). Approximation of sums of conditionally independent variables by the translated Poisson distribution. *Bernoulli* **11** 1115–1128. MR2189083

[17] Röllin, A. (2007). Translated Poisson approximation using exchangeable pair couplings. *Ann. Appl. Probab.* **17** 1596–1614. MR2358635

[18] Röllin, A. (2008). Symmetric and centered binomial approximation of sums of locally dependent random variables. *Electron. J. Probab.* **13** 756–776. MR2399295

[19] Röllin, A. and Ross, N. (2012). Local limit theorems via Landau-Kolmogorov inequalities. Preprint. Available at http://arxiv.org/abs/1011.3100v2.

[20] Stein, C. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability* (*Univ. California*, *Berkeley*, *Calif.*, 1970/1971), *Vol. II*: *Probability Theory* 583–602. Berkeley, CA: Univ. California Press. MR0402873

[21] Stein, C. (1986). *Approximate Computation of Expectations*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **7**. Hayward, CA: IMS. MR0882007

Received January 2012 and revised February 2013