Alea 6, 415–433 (2009)

Univariate approximations in the infinite occupancy scheme

A. D. Barbour

Universität Zürich
Institut für Mathematik, Winterthurerstrasse 190, CH–8057 Zürich, Switzerland
E-mail address: a.d.barbour@math.uzh.ch; http://user.math.uzh.ch/barbour/

Abstract. In the classical occupancy scheme with infinitely many boxes, n balls are thrown independently into boxes $1, 2, \ldots$, with fixed probabilities p_j, $j \geq 1$. We establish approximations to the distributions of the summary statistics K_n, the number of occupied boxes, and $K_{n,r}$, the number of boxes containing exactly r balls, within the family of translated Poisson distributions. These are shown to be of ideal order as $n \to \infty$, with respect both to total variation distance and to the approximation of point probabilities. The proof is probabilistic, making use of a translated Poisson approximation theorem of Röllin (2005).

1. Introduction

In the classical occupancy scheme with infinitely many boxes, n balls are thrown independently into boxes $1, 2, \ldots$, with probability p_j of hitting box j, $j \geq 1$, where $p_1 \geq p_2 \geq \ldots > 0$ and $\sum_{j=1}^{\infty} p_j = 1$. The summary statistics K_n, the number of occupied boxes, and $K_{n,r}$, the number of boxes containing exactly r balls, have been widely studied. Central limit theorems were established by Karlin (1967), under a regular variation condition, and Dutko (1989) showed that K_n is asymptotically normal, assuming only the necessary condition that its variance tends to infinity with n. A full discussion of this and many more aspects of the problem can be found in Gnedin et al. (2007); see also Barbour and Gnedin (2009), in which multivariate approximation of the $K_{n,r}$ is treated.

As regards the accuracy of the central limit approximation, Hwang and Janson (2008) show that the point probabilities $\mathbf{P}[K_n = t]$ are uniformly approximated by the point probabilities of the integer discretization of the normal distribution

Received by the editors February 19, 2009; accepted November 8, 2009.

2000 *Mathematics Subject Classification.* 60F05, 60C05.

Key words and phrases. occupancy, translated Poisson approximation, total variation distance, local limit approximation.

A.D. Barbour gratefully acknowledges financial support from Schweizerischer Nationalfonds Projekt Nr. 20-117625/1.

$\mathcal{N}(\mu_n, \sigma_n^2)$, where $\mu_n := \mathbb{E}K_n$ and $\sigma_n^2 := \mathrm{Var}\, K_n$. The accuracy of their approximation is of order $O(1/\sigma_n^2)$, provided only that $\sigma_n^2 \to \infty$ as $n \to \infty$. This is the same accuracy as would be expected for sums of independent indicator random variables, and is thus a remarkably precise result. Earlier results, for instance those of Mirakhmedov (1989, 1992), in a more complicated setting, require additional assumptions. However, the proof in Hwang and Janson (2008) involves long and delicate analysis of the corresponding generating functions. In this paper, working within the context of a fixed sequence of p_j's, we derive an analogous approximation by purely probabilistic arguments, complement this result with a distributional approximation in total variation, and investigate the quantities $K_{n,r}$ as well.

The approach that we take begins with the well–known observation that, if the fixed value n were replaced by a Poisson distributed random number with mean n, then the numbers N_j of balls in the boxes $j = 1, 2, \ldots$ would be independent Poisson random variables. Approximations of the kind to be discussed would then be immediate, from the theory of sums of independent Bernoulli random variables. The essence of the problem lies in the dependence introduced by fixing n. One way of relaxing this dependence is to disregard the first few boxes, for which the result is essentially known, and to use the fact that the number of balls falling in the remaining boxes is now random. Indeed, defining $j_n \geq 1$ in such a way that

$$p_{j_n - 1} \;\geq\; 4n^{-1} \log n \;>\; p_{j_n}, \tag{1.1}$$

it is immediate that

$$\mathbf{P}[N_j \geq 1 \text{ for all } j \leq j_n - 1] \;\geq\; 1 - \frac{n}{4\log n}\Big(1 - \frac{4\log n}{n}\Big)^n \;\geq\; 1 - n^{-3},$$

so that, except on a set of probability at most n^{-3}, we have

$$\sum_{j=1}^{j_n-1} I_j \;=\; j_n - 1, \tag{1.2}$$

where $I_j := I[N_j \geq 1]$. Furthermore, a simple Poisson approximation argument, due to Le Cam (1960) and Michel (1987), can now be used to get a sharp description of the distribution of the remaining elements in the sum $K_n := \sum_{j \geq 1} I_j$, since

$$d_{\mathrm{TV}}(\mathcal{L}(N_j,\, j \geq j_n), \mathcal{L}(L_j,\, j \geq j_n)) \;\leq\; P_n \;:=\; \sum_{j \geq j_n} p_j,$$

where $(L_j,\, j \geq j_n)$ are *independent* Poisson random variables with means $\mathbb{E}L_j = np_j$: see Barbour and Gnedin (2009, Sec. 2). This means that the random sequences $(I_j,\, j \geq j_n)$ and $(I[L_j \geq 1],\, j \geq j_n)$ can be constructed to be identical, except on a set of probability at most P_n, so that, except on a set of probability at most $n^{-3} + P_n$, the distribution of K_n agrees with that of a sum of independent indicators, the first $j_n - 1$ of which are equal to 1. Hence a discretized central limit theorem and uniform approximation of point probabilities follow, using $\mathcal{N}(\mu_n, \sigma_n^2)$ as basis, with accuracies $O(\sigma_n^{-1} + n^{-3} + P_n)$ and $O(\sigma_n^{-2} + n^{-3} + P_n)$ respectively, and analogous results are also true for the statistics $K_{n,r}$.

The drawback to this very simple approach is that it need not be the case that, for instance, $P_n = O(\sigma_n^{-2})$. For example, Karlin's case of regular variation allows the possibility of having $\sigma_n^2 \asymp n^\beta$, for any given β, $0 < \beta < 1$. In such cases, $P_n \asymp (n^{-1} \log n)^{1-\beta}$, so that $P_n = O(\sigma_n^{-2})$ is *not* true if $\beta > 1/2$, and $P_n = O(\sigma_n^{-1})$

is not true if $\beta > 2/3$. To get the result of Hwang and Janson (2008), we in general need something sharper.

Our approach involves a technique analogous to that above, discarding a set of indices for which the outcome is essentially known, and using the randomness in the remainder. Foregoing the total independence of the above scheme, which costs too much to achieve, we instead construct a *conditionally independent* sequence of Binomial random variables within the problem, and use these to provide the necessary refinement. The way in which this can be done is described in Röllin (2005). There, and in this paper too, we use translations of Poisson distributions as approximations, instead of discretized normal distributions, though, to the accuracies being considered, they are equivalent: the translated Poisson distribution $\mathrm{TP}\,(\mu, \sigma^2)$ is defined to be that of the sum of an *integer* a and a Poisson $\mathrm{Po}\,(\lambda)$–distributed random variable, with λ and a so chosen that $a + \lambda = \mu$ and $\sigma^2 \leq \lambda < \sigma^2 + 1$.

Using this approach, we are able to prove the following two theorems. We use d_{TV} to denote the total variation distance between distributions:

$$d_{\mathrm{TV}}(P, Q) \;:=\; \sup_A |P(A) - Q(A)|,$$

and d_{loc} to denote the local distance (point metric) between distributions on the integers:

$$d_{\mathrm{loc}}(P, Q) \;:=\; \sup_{j \in \mathbb{Z}} |P\{j\} - Q\{j\}|.$$

We define j_0 so that

$$\sum_{j \geq j_0 - 1} p_j \;\geq\; 1/2 \;>\; \sum_{j \geq j_0} p_j =: P_0, \tag{1.3}$$

and let $n_0 \geq 3$ be such that j_n, defined in (1.1), satisfies $j_n \geq j_0$ for all $n \geq n_0$, and also that $n_0/\log^2 n_0 \geq 16/P_0$.

Theorem 1.1. *If* $\mu_n := \mathbb{E}K_n$ *and* $\sigma_n^2 := \mathrm{Var}\,K_n$*, then*

$$\begin{aligned} d_{\mathrm{TV}}(\mathcal{L}(K_n), \mathrm{TP}\,(\mu_n, \sigma_n^2)) &= O(\sigma_n^{-1}); \\ d_{\mathrm{loc}}(\mathcal{L}(K_n), \mathrm{TP}\,(\mu_n, \sigma_n^2)) &= O(\sigma_n^{-2}), \end{aligned}$$

uniformly in $n \geq n_0$.

Theorem 1.2. *For* $r \geq 1$*, setting* $\mu_{n,r} := \mathbb{E}K_{n,r}$ *and* $\sigma_{n,r}^2 := \mathrm{Var}\,K_{n,r}$*, we have*

$$\begin{aligned} d_{\mathrm{TV}}(\mathcal{L}(K_{n,r}), \mathrm{TP}\,(\mu_{n,r}, \sigma_{n,r}^2)) &= O(\sigma_{n,r}^{-1}); \\ d_{\mathrm{loc}}(\mathcal{L}(K_{n,r}), \mathrm{TP}\,(\mu_{n,r}, \sigma_{n,r}^2)) &= O(\sigma_{n,r}^{-2}), \end{aligned}$$

uniformly in $n \geq \max\{n_0, e^{r/4}, 2r\}$.

Note that the variances σ_n^2 and $\sigma_{n,r}^2$ in the above theorems cannot necessarily be replaced by their simpler 'poissonized' versions $\tilde{\sigma}_n^2$ and $\tilde{\sigma}_{n,r}^2$ from the model in which a Poisson $\mathrm{Po}\,(n)$–distributed number of balls are thrown. For instance, in Theorem 1.1, the difference $(\tilde{\sigma}_n/\sigma_n - 1)$ may tend to zero more slowly than σ_n^{-1}, in which case it would dominate the error in the corresponding approximation. Note also that the implied constants in Theorems 1.1 and 1.2 are the same for *all* choices of the p_j's, as long as the quantities j_n, j_0 and P_0 determined from (1.1) and (1.3) are such that $j_n \geq j_0$, and that $P_0 \geq 16 \log^2 n/n$. To this extent, the p_j's can also be allowed to depend on n.

Röllin's theorem and our construction are set out in Section 2, together with the general scheme of the proofs. The details for the two theorems are then given in Sections 3 and 4. Some useful technical results are collected in the appendix.

2. The basic method

We begin with the following theorem from Röllin (2005). Let W be an integer valued random variable, with mean μ and variance σ^2, and let M be some random element. Define

$$\begin{aligned} \mu_M &:= \mathbb{E}(W \mid M); & \sigma_M^2 &:= \mathrm{Var}\,(W \mid M); & \tau^2 &:= \mathrm{Var}\,(\mu_M); \\ \rho^2 &:= \mathbb{E}(\sigma_M^2); & \nu^2 &:= \mathrm{Var}\,(\sigma_M^2); & U &:= \tau^{-1}(\mu_M - \mu). \end{aligned} \tag{2.1}$$

Of course, $\sigma^2 = \rho^2 + \tau^2$.

Theorem 2.1. *Suppose that, for some $\varepsilon > 0$,*

$$|\mathbb{E}\{f'(U) - Uf(U)\}| \;\le\; \varepsilon\|f''\| \tag{2.2}$$

for all bounded functions f with bounded second derivative; $\|\cdot\|$ denotes the supremum norm. Then there exist universal constants R_1 and R_2 such that

$$\begin{aligned} & d_{\mathrm{TV}}(\mathcal{L}(W), \mathrm{TP}\,(\mu,\sigma^2)) \\ &\quad \le\; \mathbb{E}\{d_{\mathrm{TV}}(\mathcal{L}(W \mid M), \mathrm{TP}\,(\mu_M,\sigma_M^2))\} + R_1\frac{1}{\rho}\Big\{1 + \frac{\nu}{\rho} + \frac{\varepsilon\tau^3}{\sigma^2}\Big\}; \\ & d_{\mathrm{loc}}(\mathcal{L}(W), \mathrm{TP}\,(\mu,\sigma^2)) \\ &\quad \le\; \mathbb{E}\{d_{\mathrm{loc}}(\mathcal{L}(W \mid M), \mathrm{TP}\,(\mu_M,\sigma_M^2))\} + R_2\frac{1}{\rho^2}\Big\{1 + \frac{\nu^2}{\rho^2} + \frac{\varepsilon\tau^3}{\sigma^2}\Big\}. \end{aligned}$$

Values of the constants are given in Röllin (2005). Note that (2.2) is exactly what has to be established for the simplest smooth metric standard normal approximation to $\mathcal{L}(U)$, using Stein's method. For U a sum of independent random variables, ε would typically be the Lyapounov ratio, and thus the quantity $\sigma^{-2}\tau^3\varepsilon$ would be bounded by an average of the ratios of third to second moments of the summands.

The theorem is useful provided that $\mathcal{L}(W \mid M)$ is such that it is well approximated for each value of M by the translated Poisson distribution with its mean and variance as parameters. This is the case, for instance, for sums of independent Bernoulli random variables, as well as for many sums of independent integer valued random variables, as noted in Röllin (2005). Here is the result that we shall use in what follows.

Theorem 2.2. *Suppose that $\mathcal{L}(W \mid M)$ is the distribution of a sum $\sum_{j\ge1} I_j(M)$ of independent Bernoulli random variables with probabilities $p_j(M)$ such that $\mu_M := \sum_{j\ge1} p_j(M) < \infty$ a.s.; write $\sigma_M^2 := \sum_{j\ge1} p_j(M)(1-p_j(M))$, $\rho^2 := \mathbb{E}(\sigma_M^2)$ and $\nu^2 := \mathrm{Var}\,(\sigma_M^2)$. Suppose that $\nu^2 \le C\rho^2$ for some $C < \infty$. Then there exist universal constants C_1 and C_2 such that*

$$\begin{aligned} \mathbb{E}\{d_{\mathrm{TV}}(\mathcal{L}(W \mid M), \mathrm{TP}\,(\mu_M,\sigma_M^2))\} &\;\le\; \frac{4C}{\rho^2} + \frac{C_1\sqrt{2}}{\rho}; \\ \mathbb{E}\{d_{\mathrm{loc}}(\mathcal{L}(W \mid M), \mathrm{TP}\,(\mu_M,\sigma_M^2))\} &\;\le\; \frac{4C + 2C_2}{\rho^2}. \end{aligned}$$

Proof. Bounds of the form

$$\begin{aligned} d_{\rm TV}(\mathcal{L}(W \mid M), {\rm TP}\,(\mu_M, \sigma_M^2)) &\le \min\{C_1\sigma_M^{-1}, 1\}; \\ d_{\rm loc}(\mathcal{L}(W \mid M), {\rm TP}\,(\mu_M, \sigma_M^2)) &\le \min\{C_2\sigma_M^{-2}, 1\}, \end{aligned} \tag{2.3}$$

are given in Barbour (2009, Theorems 6.2 and 6.3), with $C_1 = 4$ and $C_2 = 280$. The former follows as in Barbour and Čekanavičius (2002, Theorem 3.1), and similar techniques can be used to establish the latter; see also Röllin (2005). Then, by Chebyshev's inequality, $\mathbf{P}[\sigma_M^2 < \frac{1}{2}\rho^2] \le 4C/\rho^2$. The bounds follow by taking expectations in (2.3). □

We now need to find a suitable collection of conditionally independent Bernoulli random variables. To do so, we start by observing, as before, that it is enough to consider indices $j \ge j_n$ in the sums, so we need only consider the distribution of $(N_j,\, j \ge j_n)$. We realize these random variables in two stages: first, we realize $M := (M_j,\, j \ge j_0)$ by throwing n balls independently into the boxes with indices $j \ge j_0$, with probability p_j/P_0 for box j, and then 'thinning' them independently with retention probability P_0, so that, conditionally on M, the $(N_j,\, j \ge j_0)$ are independent, with $N_j \sim {\rm Bi}\,(M_j, P_0)$. With this construction, it remains to evaluate the quantities appearing in Röllin's theorem, and to check that we have the right result. More specifically, we need to check that, for some constants C, C', C'',

$$\text{(i) } \nu^2 \le C\rho^2; \quad \text{(ii) } \rho^2 \ge C'\sigma^2, \quad \text{and} \quad \text{(iii) } \varepsilon \le C''\tau^{-3}\sigma^2, \tag{2.4}$$

uniformly in the stated ranges of n, for the random sums $W_n := \sum_{j \ge j_n} I[N_j \ge 1]$ and $W_{n,r} := \sum_{j \ge j_n} I[N_j = r]$, $r \ge 1$. Theorems 1.1 and 1.2 will then follow directly from Theorems 2.1 and 2.2.

The first two inequalities in (2.4) cause no great problems, since they involve only variance calculations, though care has to be taken with the correlations in Theorem 1.2, because the summands in

$$\mu_M := \sum_{j \ge j_n} \binom{M_j}{r} P_0^r (1 - P_0)^{M_j - r}$$

are not monotone functions of the (negatively associated) M_j. The main effort is required in evaluating ε for the third inequality. We now sketch the structure of this argument, leaving the details to the next two sections.

Take $z(l)$, $l \ge 0$, to be either ${\rm Bi}\,(l, P_0)\{[1, \infty)\}$ or ${\rm Bi}\,(l, P_0)\{r\}$, as appropriate, (zero if $l = 0$). Then define the quantity U that we wish to address by $U := \sum_{j \ge j_n} Y_j$, where

$$\zeta_j := \mathbb{E}(z(M_j)), \quad y_j(l) := z(l) - \zeta_j \quad \text{and} \quad Y_j := \tau^{-1} y_j(M_j). \tag{2.5}$$

Thus U is a sum of mean zero, weakly dependent random variables. In order to approach (2.2), we begin by writing

$$\mathbb{E}\{Uf(U)\} = \sum_{j \ge j_n} \mathbb{E}\{Y_j f(U)\} = \tau^{-1} \sum_{j \ge j_n} \sum_{l \ge 0} q_j(l) y_j(l) \mathbb{E}\{f(U_j^{(n-l)} + \tau^{-1} y_j(l))\}, \tag{2.6}$$

where $q_j(l) := \mathbf{P}[M_j = l]$ and

$$U_j^{(m)} := \tau^{-1} \sum_{\substack{s \ge j_n \\ s \ne j}} y_s(M_{js}^{(m)}), \tag{2.7}$$

and where

$$M_{j\cdot}^{(m)} \ := \ (M_{js},\, s \geq j_n, s \neq j) \ \sim \ \mathrm{MN}\,(m;\, (p_s/P_{0j},\, s \geq j_n, s \neq j)) \qquad (2.8)$$

has the multinomial distribution of m balls thrown independently into the boxes with indices $(s \geq j_n, s \neq j)$ with probabilities $(p_s/P_{0j},\, s \geq j_n, s \neq j)$, with $P_{0j} := P_0 - p_j \geq 3P_0/4$. We need to show that the expression in (2.6) is close to $\mathbb{E}\{f'(U)\}$.

As a first step, we use Taylor development to discard all but the constant and linear terms in $\mathbb{E}\{f(U_j^{(n-l)} + \tau^{-1}y_j(l))\}$, establishing that

$$(1) \qquad \Big|\tau^{-1} \sum_{j \geq j_n} \sum_{l \geq 0} q_j(l) y_j(l) \Big\{ \mathbb{E}f(U_j^{(n-l)} + \tau^{-1}y_j(l)) - \mathbb{E}f(U_j^{(n-l)}) - \tau^{-1}y_j(l)\mathbb{E}f'(U_j^{(n-l)}) \Big\}\Big| \ \leq \ k_1\sigma^2\tau^{-3}\|f''\|.$$

The next step is to remove the l-dependence in the constant term, replacing $U_j^{(n-l)}$ by $U_j^{(n)}$. To make the computations, we realize $U_j^{(n-l)}$ and $U_j^{(n)}$ on the same probability space by writing $M_{j\cdot}^{(n)} = M_{j\cdot}^{(n-l)} + Z_{j\cdot}^{(l)}$, where $M_{j\cdot}^{(n-l)}$ and $Z_{j\cdot}^{(l)}$ are independent, and distributed as $M_{j\cdot}^{(m)}$ in (2.8), with $m = n - l$ and $m = l$, respectively; and then defining $U_j^{(n-l)}$ and $U_j^{(n)}$ as before, using (2.7). Using this representation, we then show that

$$(2) \qquad \Big|\tau^{-1} \sum_{j \geq j_n} \sum_{l \geq 0} q_j(l) y_j(l) \Big\{ \mathbb{E}f(U_j^{(n-l)}) - \mathbb{E}f(U_j^{(n)}) - \mathbb{E}[f'(U_j^{(n-l)})(U_j^{(n-l)} - U_j^{(n)})] \Big\}\Big| \ \leq \ k_2\sigma^2\tau^{-3}\|f''\|.$$

Although this has introduced a further term $\mathbb{E}[f'(U_j^{(n-l)})(U_j^{(n-l)} - U_j^{(n)})]$ involving l, there is simplification because $\mathbb{E}f(U_j^{(n)})$ is multiplied by $\sum_{l\geq 0} q_j(l)y_j(l) = \mathbb{E}Y_j = 0$, and hence drops out.

We now simplify what is left by showing that

$$(3) \qquad \Big|\tau^{-1} \sum_{j \geq j_n} \sum_{l \geq 0} q_j(l) y_j(l) \Big\{ \mathbb{E}[f'(U_j^{(n-l)})(U_j^{(n-l)} - U_j^{(n)})] - \mathbb{E}[f'(U_j^{(n)})]\mathbb{E}(U_j^{(n-l)} - U_j^{(n)}) \Big\}\Big| \ \leq \ k_3\sigma^2\tau^{-3}\|f''\|.$$

As a result of this, the quantity $\mathbb{E}f(U_j^{(n-l)})$ in (1) has been replaced by a multiple of $\mathbb{E}f'(U_j^{(n)})$, with errors of the desired order, which is a useful step in approaching the intended goal of $\mathbb{E}f'(U)$. There is also the quantity $\mathbb{E}f'(U_j^{(n-l)})$ appearing in (1), but this is easily reduced to one involving only $\mathbb{E}f'(U_j^{(n)})$, too:

$$(4) \quad \Big|\tau^{-1} \sum_{j \geq j_n} \sum_{l \geq 0} q_j(l) y_j^2(l) \{\mathbb{E}f'(U_j^{(n-l)}) - \mathbb{E}f'(U_j^{(n)})\}\Big| \ \leq \ k_4\sigma^2\tau^{-3}\|f''\|.$$

At this point, we have thus established that

$$\Big|\mathbb{E}Uf(U) - \tau^{-2} \sum_{j \geq j_n} \kappa_j \mathbb{E}f'(U_j^{(n)})\Big| \ \leq \ (k_1 + k_2 + k_3 + k_4)\sigma^2\tau^{-3}\|f''\|, \qquad (2.9)$$

with

$$\kappa_j \ := \ \sum_{l \geq 0} q_j(l) y_j(l) \{y_j(l) - \tau\mathbb{E}(U_j^{(n)} - U_j^{(n-l)})\}, \qquad (2.10)$$

and, for example by taking $f(x) = x$,

$$1 \;=\; \mathbb{E}U^2 \;=\; \tau^{-2} \sum_{j \ge j_n} \kappa_j.$$

In parallel with the above reduction starting from (2.6), we now start with

$$\begin{aligned} \mathbb{E}f'(U) &= \tau^{-2} \sum_{j \ge j_n} \kappa_j \mathbb{E}f'(U) \\ &= \tau^{-2} \sum_{j \ge j_n} \kappa_j \sum_{l \ge 0} q_j(l) \mathbb{E}f'(U_j^{(n-l)} + \tau^{-1} y_j(l)), \end{aligned} \tag{2.11}$$

and make two rather simpler steps, first proving that

$$(5)\quad \Bigl|\tau^{-2} \sum_{j \ge j_n} \kappa_j \sum_{l \ge 0} q_j(l) \{\mathbb{E}f'(U_j^{(n-l)} + \tau^{-1} y_j(l)) - \mathbb{E}f'(U_j^{(n-l)})\}\Bigr| \;\le\; k_5 \sigma^2 \tau^{-3} \|f''\|,$$

and then that

$$(6)\quad \Bigl|\tau^{-2} \sum_{j \ge j_n} \kappa_j \sum_{l \ge 0} q_j(l) \{\mathbb{E}f'(U_j^{(n-l)}) - \mathbb{E}f'(U_j^{(n)})\}\Bigr| \;\le\; k_6 \sigma^2 \tau^{-3} \|f''\|.$$

Putting these two into (2.11), it follows that

$$\Bigl|\mathbb{E}f'(U) - \tau^{-2} \sum_{j \ge j_n} \kappa_j \mathbb{E}f'(U_j^{(n)})\Bigr| \;\le\; (k_5 + k_6) \sigma^2 \tau^{-3} \|f''\|, \tag{2.12}$$

and combining this with (2.9) yields

$$|\mathbb{E}\{f'(U) - U f(U)\}| \;\le\; \varepsilon \|f''\|, \tag{2.13}$$

with $\sigma^{-2}\tau^3\varepsilon \le \sum_{t=1}^{6} k_t$ bounded, as required.

3. The argument for K_n

We begin by noting, for future reference, that we have

$$\begin{aligned} \bar{p}_n \;&:=\; \max_{j \ge j_n} p_j \;\le\; 4n^{-1} \log n \;\le\; P_0/4 \;\le\; 1/8; \\ n\bar{p}_n^2 \;&\le\; 16 n^{-1} \log^2 n \;\le\; P_0, \end{aligned} \tag{3.1}$$

whenever $n \ge n_0$, and that $\beta := (1 - P_0/2) \ge 3/4$. We use c and c' to denote generic universal constants, not depending on n or the p_j's.

For K_n, we have $\mathcal{L}(W_n \,|\, M)$ that of a sum of indicator random variables $I_j(M)$, $j \ge j_n$, with probabilities

$$\{1 - (1 - P_0)^{M_j}\} \;=:\; z(M_j);$$

recall (2.5). Hence $\sigma_M^2 = \sum_{j \ge j_n} z(M_j)(1 - z(M_j))$, and

$$\rho^2 \;=\; \mathbb{E}\sigma_M^2 \;=\; \sum_{j \ge j_n} \mathbb{E}\{(1 - P_0)^{M_j} - (1 - P_0)^{2M_j}\}.$$

Applying Lemma 5.1 (iv) with $x = \sqrt{1 - P_0}$, and using the fact that $n\bar{p}_n^2 \le P_0$, now immediately gives the lower bound

$$\rho^2 \;\ge\; c_\rho \sum_{j \ge j_n} e^{-np_j} \min\{1, np_j\}, \tag{3.2}$$

where $c_\rho = c(\sqrt{1-P_0})e^{-2P_0}$, and $c(\cdot)$ is as in Lemma 5.1. On the other hand, because the N_j are negatively associated,

$$\sigma^2 \le \sum_{j\ge j_n} \mathrm{Var}\, I[N_j \ge 1] = \sum_{j\ge j_n} \{1-(1-p_j)^n\}(1-p_j)^n \le \sum_{j\ge j_n} e^{-np_j} \min\{1, np_j\}.$$

It thus follows that $\rho^2 \ge c_\rho \sigma^2$, establishing (2.4) (ii).

For $\nu^2 = \mathrm{Var}\, \sigma_M^2$, we note that σ_M^2 is the difference of the random variables $s_1(M) := \sum_{j\ge j_n} (1-P_0)^{M_j}$ and $s_2(M) := \sum_{j\ge j_n}(1-P_0)^{2M_j}$, thus implying that $\nu^2 \le 2(\mathrm{Var}\, s_1(M) + \mathrm{Var}\, s_2(M))$. Since $(1-P_0)^l$ is decreasing in l, we can use the negative association of the M_j's to upper bound the variances:

$$\mathrm{Var}\, s_1(M) \le \sum_{j\ge j_n} \mathrm{Var}\,\{(1-P_0)^{M_j}\}; \qquad \mathrm{Var}\, s_2(M) \le \sum_{j\ge j_n} \mathrm{Var}\,\{(1-P_0)^{2M_j}\}.$$

Now both of these quantities can be bounded by using Lemma 5.1 (iv):

$$\mathrm{Var}\,\{(1-P_0)^{M_j}\} \le e^{-2\beta n p_j} \min\{1, 2\beta n p_j\},$$

and

$$\mathrm{Var}\,\{(1-P_0)^{2M_j}\} \le e^{-2\beta' n p_j} \min\{1, 2\beta' n p_j\},$$

with $\beta' := 4 - 6P_0 + 4P_0^2 - P_0^3$. Thus $\rho^{-2}\nu^2$ is uniformly bounded, establishing (2.4) (i). It thus remains to prove that $\varepsilon \le C''\tau^{-3}\sigma^2$ for some constant C'', and we are finished. To do this, we successively verify the inequalities (1) – (6) of Section 2.

To establish inequality (1), we note that its left hand side is bounded by

$$\tfrac{1}{2}\tau^{-3} \sum_{j\ge j_n} \sum_{l\ge 0} q_j(n) |y_j(l)|^3 \|f''\|. \tag{3.3}$$

Now $|y_j(l)| \le 1$, and

$$\sum_{l\ge 0} q_j(l) y_j^2(l) = \mathbb{E}\{(1-P_0)^{2M_j}\} - \{\mathbb{E}(1-P_0)^{M_j}\}^2,$$

with $M_j \sim \mathrm{Bi}\,(n, p_j/P_0)$. From Lemma 5.1 (iv) with $x = 1-P_0$, it follows that

$$\sum_{l\ge 0} q_j(l) y_j^2(l) \le e^{-2\beta n p_j} \min\{1, 2\beta n p_j\}. \tag{3.4}$$

Hence, from Lemma 5.4 (i),

$$\tau^{-3} \sum_{l\ge 0} q_j(l) |y_j(l)|^3 \le \tau^{-3} \sum_{j\ge j_n} n p_j e^{-2\beta n p_j} \le K_0^{(2\beta-1)} \sigma^2 \tau^{-3}.$$

By (3.3), this proves (1) with $k_1 = K_0^{(2\beta-1)}$.

For inequality (2), we have

$$|\mathbb{E}\{f(U_j^{(n)}) - f(U_j^{(n-l)}) - f'(U_j^{(n-l)})(U_j^{(n)} - U_j^{(n-l)})\}| \le \tfrac{1}{2}\|f''\| \mathbb{E}\{(U_j^{(n)} - U_j^{(n-l)})^2\}. \tag{3.5}$$

Now

$$\tau^2 \mathbb{E}\{(U_j^{(n)} - U_j^{(n-l)})^2\} \le \mathbb{E}\Big\{\Big(\sum_{\substack{s\ge j_n\\ s\ne j}} Z_{js}^{(l)} P_0 (1-P_0)^{M_{js}^{(n-l)}}\Big)^2\Big\},$$

and the collections of random variables $(Z_{js}^{(l)},\, s \ge j_n)$ and $((1-P_0)^{M_{js}^{(n-l)}},\, s \ge j_n)$ are independent, and each is composed of negatively correlated elements. Hence

$$\begin{aligned}\tau^2\mathbb{E}\{(U_j^{(n)} - U_j^{(n-l)})^2\} &\le P_0^2\Big(\sum_{\substack{s\ge j_n\\ s\ne j}} \mathbb{E}Z_{js}^{(l)}\,\mathbb{E}\Big\{(1-P_0)^{M_{js}^{(n-l)}}\Big\}\Big)^2 \\ &\qquad + P_0^2\sum_{\substack{s\ge j_n\\ s\ne j}} \mathbb{E}\{(Z_{js}^{(l)})^2\}\mathbb{E}\Big\{(1-P_0)^{2M_{js}^{(n-l)}}\Big\}.\end{aligned}$$

Now routine calculation gives

$$\begin{aligned} P_0\,\mathbb{E}Z_{js}^{(l)} \ &\le\ lP_0p_s/P_{0j}\ \le\ 2lp_s; \qquad & P_0^2\,\mathbb{E}\{(Z_{js}^{(l)})^2\}\ &\le\ 2lp_s(1+2lp_s);\\ \mathbb{E}\Big\{(1-P_0)^{M_{js}^{(n-l)}}\Big\}\ &\le\ e^{-(n-l)p_s}; \qquad & \mathbb{E}\Big\{(1-P_0)^{2M_{js}^{(n-l)}}\Big\}\ &\le\ e^{-2\beta(n-l)p_s},\end{aligned}$$

and hence, with crude simplifications,

$$\tau^2\mathbb{E}\{(U_j^{(n)} - U_j^{(n-l)})^2\}\ \le\ 10l^2e^{l\delta_n}\sum_{s\ge j_n} p_se^{-2\beta np_s}\ \le\ cl^2e^{l\delta_n}n^{-1}\sigma^2, \tag{3.6}$$

this last using (3.2) and Lemma 5.4 (i), where $\delta_n := 2\bar p_n$ and $c = 10(K_0^{(2\beta-1)}/c_\rho)$. Hence, putting (3.5) and (3.6) into (2), we obtain the bound

$$\begin{aligned}&\frac{c}{2}\|f''\|\tau^{-3}\sum_{j\ge j_n}\sum_{l\ge 0} q_j(l)|y_j(l)|l^2e^{l\delta_n}n^{-1}\sigma^2\\ &\qquad \le\ c'\tau^{-3}\sigma^2\|f''\|\exp\{\delta_n(3+n\bar p_n e/P_0)\}\sum_{j\ge j_n} e^{-np_j}p_j(1+np_j),\end{aligned}$$

by Lemma 5.1 (ii) and (iii), and this is uniformly of order $\tau^{-3}\sigma^2\|f''\|$ in the stated range of n, because

$$\sum_{j\ge j_n} p_j(1+np_j)e^{-np_j}\ \le\ P_n(1+e^{-1}) \quad\text{and}\quad \delta_n + n\delta_n\bar p_n\ \le\ 5P_0/4.$$

This establishes inequality (2).

For inequality (3), we begin by writing

$$\begin{aligned}&\mathbb{E}\{(U_j^{(n-l)} - U_j^{(n)})f'(U_j^{(n-l)})\}\\ &\quad = \ \mathbb{E}\{[\mathbb{E}(U_j^{(n-l)} - U_j^{(n)}\,|\,M_{j\cdot}^{(n-l)}) - \mathbb{E}(U_j^{(n-l)} - U_j^{(n)})](f'(U_j^{(n-l)}) - f'(\mathbb{E}U_j^{(n-l)}))\}\\ &\qquad\qquad - \mathbb{E}(U_j^{(n)} - U_j^{(n-l)})\mathbb{E}f'(U_j^{(n-l)});\end{aligned} \tag{3.7}$$

note that introducing $f'(\mathbb{E}U_j^{(n-l)})$ changes nothing, since it is multiplied by a quantity with mean zero. The first term we bound by

$$\|f''\|\sqrt{\mathrm{Var}\,[\mathbb{E}(U_j^{(n-l)} - U_j^{(n)}\,|\,M_{j\cdot}^{(n-l)})]}\sqrt{\mathrm{Var}\,U_j^{(n-l)}}. \tag{3.8}$$

Since

$$\tau\mathbb{E}(U_j^{(n-l)} - U_j^{(n)}\,|\,M_{j\cdot}^{(n-l)})\ =\ \sum_{\substack{s\ge j_n\\ s\ne j}} (1-P_0)^{M_{js}^{(n-l)}}\{1-(1-p_sP_0/P_{0j})^l\}, \tag{3.9}$$

and since the $(M_{js}^{(n-l)},\, s \geq j_n)$ are negatively associated, it follows that

$$\begin{aligned}\tau^2 \mathrm{Var}\,[\mathbb{E}(U_j^{(n-l)} - U_j^{(n)} \,|\, M_{j\cdot}^{(n-l)})] &\leq 4l^2 \sum_{\substack{s \geq j_n \\ s \neq j}} p_s^2 e^{-2\beta(n-l)p_s} \\ &\leq 4l^2 e^{l\delta_n} n^{-1}/(2\beta e) \;=\; cl^2 e^{l\delta_n} n^{-1},\end{aligned}$$

for a suitable c. In much the same way, and using Lemma 5.1 (iv), we have

$$\tau^2 \mathrm{Var}\, U_j^{(n-l)} \;\leq\; \sum_{\substack{s \geq j_n \\ s \neq j}} \mathrm{Var}\,\{(1-P_0)^{M_{js}^{(n-l)}}\} \;\leq\; 2\frac{P_0}{P_{0j}} \sum_{\substack{s \geq j_n \\ s \neq j}} np_s e^{-2\beta(n-l)p_s} \;\leq\; ce^{l\delta_n}\sigma^2.$$

Hence the first term in (3.7) is bounded by

$$c\tau^{-2}\|f''\|\, le^{l\delta_n} n^{-1/2}\sigma, \tag{3.10}$$

for a suitable c. For the second, we replace $\mathbb{E}f'(U_j^{(n-l)})$ by $\mathbb{E}f'(U_j^{(n)})$:

$$|\mathbb{E}(U_j^{(n)} - U_j^{(n-l)})\{\mathbb{E}f'(U_j^{(n-l)}) - \mathbb{E}f'(U_j^{(n)})\}| \;\leq\; \|f''\|\mathbb{E}\{(U_j^{(n)} - U_j^{(n-l)})^2\}, \tag{3.11}$$

which is at most $c\tau^{-2}\|f''\| l^2 e^{l\delta_n} n^{-1}\sigma^2$, by (3.6). Putting these bounds into (3.7), it follows that the left hand side in (3) is at most

$$\begin{aligned} c\tau^{-3}\|f''\| \sum_{j \geq j_n} \sum_{l \geq 0} q_j(l)|y_j(l)| e^{l\delta_n}\{ln^{-1/2}\sigma + l^2 n^{-1}\sigma^2\} \\ \leq\; c'\tau^{-3}\|f''\| \Big\{n^{-1/2}\sigma \sum_{j \geq j_n} np_j e^{-np_j} + \sigma^2\Big\}, \end{aligned} \tag{3.12}$$

by using Lemma 5.1 (ii) and (iii), for suitable constants c and c'. But now

$$\sum_{j \geq j_n} np_j e^{-np_j} \;\leq\; \sqrt{K' n\sigma^2},$$

by Lemma 5.4 (iv), and this, together with (3.12), shows that (3) is satisfied.

For (4), we use the simple bound

$$|\mathbb{E}f'(U_j^{(n-l)}) - \mathbb{E}f'(U_j^{(n)})| \;\leq\; \|f''\|\,\mathbb{E}|U_j^{(n)} - U_j^{(n-l)}| \;\leq\; \tau^{-1} l\|f''\|. \tag{3.13}$$

This gives a bound for the left hand side of (4) of

$$\begin{aligned}\tau^{-3}\|f''\| \sum_{j \geq j_n} \sum_{l \geq 0} q_j(l) y_j^2(l) l &\leq \tau^{-3}\|f''\| \sum_{j \geq j_n} np_j\{e^{-2np_j} + e^{-2\beta np_j}\} \\ &\leq k_4 \tau^{-3}\|f''\|\sigma^2,\end{aligned}$$

by Lemma 5.4 (i); and hence we have proved (2.9).

For the remaining two inequalities, we observe that, from (2.10) and (3.4),

$$\kappa_j^+ \;:=\; \max\{\kappa_j, 0\} \;\leq\; 2\beta np_j e^{-2\beta np_j}, \tag{3.14}$$

whereas, from (3.9),

$$\kappa_j^- \;=\; |\min\{0, \kappa_j\}| \;\leq\; \sum_{l \geq 0} q_j(l)|y_j(l)| \sum_{s \geq j_n} 2lp_s e^{-(n-l)p_s} \;\leq\; cnp_j e^{-np_j} \sum_{s \geq j_n} p_s e^{-np_s}, \tag{3.15}$$

from Lemma 5.1 (ii) and (iii). Hence, for inequality (5), we obtain the bound

$$\begin{aligned} \tau^{-3}\|f''\| \sum_{j\ge j_n} |\kappa_j| \sum_{l\ge 0} q_j(l)|y_j(l)| \;&\le\; 2\tau^{-3}\|f''\| \sum_{j\ge j_n} |\kappa_j| e^{-np_j} \\ &\le\; c\tau^{-3}\|f''\| \sum_{j\ge j_n} np_j e^{-2np_j} \;\le\; k_5\tau^{-3}\sigma^2\|f''\|, \end{aligned} \tag{3.16}$$

by Lemma 5.4 (i), for a suitable k_5. For inequality (6), we start from the bound

$$\begin{aligned} &\tau^{-2}\|f''\| \sum_{j\ge j_n} |\kappa_j| \sum_{l\ge 0} q_j(l)\mathbb{E}|U_j^{(n)} - U_j^{(n-l)}| \\ &\le\; \tau^{-3}\|f''\| \sum_{j\ge j_n} |\kappa_j| \sum_{l\ge 0} q_j(l) 2l e^{l\delta_n} \sum_{\substack{s\ge j_n \\ s\ne j}} p_s e^{-np_s} \\ &\le\; c\tau^{-3}\|f''\| \sum_{j\ge j_n} |\kappa_j| np_j \sum_{s\ge j_n} p_s e^{-np_s}, \end{aligned}$$

again from (3.9) and Lemma 5.1 (ii), and substituting from (3.14) and (3.15) for $|\kappa_j|$ gives at most

$$\begin{aligned} &c\tau^{-3}\|f''\| \sum_{j\ge j_n} (np_j)^2 \Big\{ P_n e^{-2\beta np_j} + e^{-np_j}\Big(\sum_{s\ge j_n} p_s e^{-np_s}\Big)^2 \Big\} \\ &\le\; k_6\tau^{-3}\|f''\|\,\sigma^2, \end{aligned} \tag{3.17}$$

by Lemma 5.4 (i) and (iv). Since (3.16) and (3.17) together establish (2.12), we have completed the proof of (2.13), and hence of (2.4) (iii), thus proving Theorem 1.1.

4. The argument for $K_{n,r}$

Fix $r \ge 1$. We now require n to satisfy $4\log n \ge r-1$ and $n \ge 2r$. Then, with $p := p_{j_n-1} \ge 4n^{-1}\log n$, we have

$$\begin{aligned} \sum_{j<j_n} \mathbf{P}[N_j = r] \;&\le\; (j_n - 1)\binom{n}{r} p^r (1-p)^{n-r} \;\le\; n^r p^{r-1} e^{-(n-r)p}/r! \\ &\le\; n^{-3}(4\log n)^{r-1} e^r/r! \;\le\; c(\log n)^{r-1} n^{-3}, \end{aligned}$$

since $x^s e^{-x}$ is decreasing in $x \ge s$ and $4\log n \ge r-1$. Thus $\sum_{j<j_n} I[N_j = r] = 0$ except on a set of probability of order $O(n^{-3}(\log n)^{r-1})$, and we can restrict attention to $W_{n,r} := \sum_{j\ge j_n} I[N_j = r]$. We recall that $\beta := (1 - P_0/2) \ge 3/4$, and that

$$\bar{p}_n \;\le\; P_0/4 \;\le\; 1/8 \quad\text{and}\quad n\bar{p}_n^2 \;\le\; P_0,$$

whenever $n \ge n_0$. The generic constants c and c' are now allowed to depend on r.

For $K_{n,r}$, the distribution $\mathcal{L}(W_{n,r} \,|\, M)$ is that of a sum of indicator random variables $I_j(M)$, $j \ge j_n$, with probabilities

$$\binom{M_j}{r} P_0^r (1-P_0)^{M_j} \;=:\; z(M_j);$$

recall (2.5). The argument now runs much as before, but is complicated by the fact that $z(\cdot)$ is not monotonic in l. First, we have $\mu = \sum_{j\ge j_n} \mathbb{E}z(M_j) = \sum_{j\ge j_n} \zeta_j$, with $\zeta_j := \mathrm{Bi}\,(n, p_j)\{r\}$, whence, defining

$$\hat\mu_r \;:=\; \sum_{j\ge j_n} \frac{(np_j)^r e^{-np_j}}{r!},$$

it easily follows that

$$\exp\{-n\bar p_n^2 - n^{-1}r^2\} \;\le\; \mu/\hat\mu_r \;\le\; e^{r\bar p_n}, \tag{4.1}$$

for $n \ge 2r$, with both lower and upper estimates uniformly bounded away from zero and infinity in the chosen range of n: hence μ and $\hat\mu_r$ are uniformly of the same order.

Now

$$\sigma_M^2 \;=\; \sum_{j\ge j_n} z(M_j)(1 - z(M_j)) \;\ge\; \sum_{j\ge j_n} z(M_j)(1 - z_r), \tag{4.2}$$

where $z_r := \max_{l\ge r} \binom{l}{r} P_0^r (1-P_0)^{l-r} < 1$, and hence

$$\rho^2 \;=\; \mathbb{E}\sigma_M^2 \;\ge\; \mu(1 - z_r). \tag{4.3}$$

For

$$\sigma^2 \;=\; \mathrm{Var}\,W_n \;=\; \sum_{j\ge j_n}\sum_{s\ge j_n} \{\mathbf{P}[N_j = N_s = r] - \mathbf{P}[N_j = r]\mathbf{P}[N_s = r]\},$$

we use Lemma 5.3 for $j \ne s$ to give

$$\mathbf{P}[N_j = N_s = r] - \mathbf{P}[N_j = r]\mathbf{P}[N_s = r] \;\le\; 2er(p_j + p_s)e^{4r\bar p_n}\mathbf{P}[N_j = r]\mathbf{P}[N_s = r],$$

and adding over j and s gives an upper bound of at most

$$c \sum_{j\ge j_n} p_j (np_j)^r e^{-np_j} \sum_{s\ge j_n} (np_s)^r e^{-np_s} \;\le\; c' P_n \hat\mu_r.$$

For $j = s$, the total contribution to the variance is at most $\sum_{j\ge j_n} \mathbf{P}[N_j = r] = \mu$. Hence, and from (4.3), we have

$$\sigma^2 \;\asymp\; \rho^2 \;\asymp\; \mu \;\asymp\; \hat\mu_r, \tag{4.4}$$

where the implied constants are universal for each r. This shows also that (2.4) (ii) holds.

For (2.4) (i), we take

$$\nu^2 \;:=\; \mathrm{Var}\,(\sigma_M^2) \;=\; \mathrm{Var}\Bigl(\sum_{j\ge j_n} z(M_j)(1 - z(M_j))\Bigr),$$

to which we can apply Lemma 5.3, noting that $0 \le z(l)(1-z(l)) \le \binom{l}{r} P_0^r (1-P_0)^{l-r}$. For $j \ne s$, this gives

$$\begin{aligned} &\mathrm{Cov}\,\{z(M_j)(1 - z(M_j)), z(M_s)(1 - z(M_s))\} \\ &\qquad \le\; c(p_j + p_s)(n(p_j + p_s) + 2r)(np_j)^r (np_s)^r e^{-n(p_j+p_s)}, \end{aligned}$$

by Lemma 5.2. Adding over j and s, this gives at most

$$\begin{aligned} c'\Bigl\{ &\sum_{j\ge j_n} p_j (np_j + 2r)(np_j)^r e^{-np_j} \sum_{s\ge j_n} (np_s)^r e^{-np_s} \\ &\qquad + \sum_{j\ge j_n} p_j (np_j)^r e^{-np_j} \sum_{s\ge j_n} (np_s)^{r+1} e^{-np_s}\Bigr\}, \end{aligned} \tag{4.5}$$

and this is at most $cP_n\hat{\mu}_r + K_{11}P_n\hat{\mu}_r$, by Lemma 5.4 (iii) and (v). The terms with $j = s$ give at most

$$\begin{aligned}\sum_{j\ge j_n}\mathbb{E}\{z^2(M_j)\} &\le \frac{P_0^{2r}}{(r!)^2}\mathbb{E}\Big\{[(M_j)_{(2r)} + (2r)_{(r)}(M_j)_{(r)}](1-P_0)^{2(M_j-r)}\Big\} \\ &\le c\{(np_j)^{2r} + (np_j)^r\}e^{-2\beta(n-r)p_j}, \qquad (4.6)\end{aligned}$$

by Lemma 5.1, and because $l_{(r)}^2 \le \binom{2r}{r}l_{(2r)} + (2r)_{(r)}l_{(r)}$. Adding over j, this gives at most a contribution of $c\hat{\mu}_r$, by Lemma 5.4 (ii). Thus we have shown that $\nu^2 \le c\sigma^2$, and (2.4) (i) is satisfied. It thus remains to show that $\varepsilon \le c\tau^{-3}\sigma^2$, and the proof is accomplished.

To establish inequality (1), we again observe that $|y_j(l)| := |z(l) - \mathbb{E}z(M_j)| \le 1$, and hence, recalling (3.3), that

$$\tfrac{1}{2}\tau^{-3}\|f''\|\sum_{j\ge j_n}\mathbb{E}|y_j(M_j)|^3 \;\le\; \tau^{-3}\|f''\|\sum_{j\ge j_n}\mathbb{E}z^2(M_j) \;\le\; c\tau^{-3}\|f''\|\,\hat{\mu}_r,$$

as for (4.6); so (1) holds, as required.

For (2), we recall (3.5). We then note that, for $u \ge r$,

$$\begin{aligned}|z(u+t) - z(u)| &= P_0^r\left|\binom{u}{r}(1-P_0)^{u-r} - \binom{u+t}{r}(1-P_0)^{u+t-r}\right| \\ &\le c\binom{u}{r}(1-P_0)^u, \qquad (4.7)\end{aligned}$$

for c a universal constant. From this, it follows that

$$\begin{aligned}\tau|U_j^{(n)} - U_j^{(n-l)}| \;\le\; &\sum_{\substack{s\ge j_n\\ s\ne j}}\Big\{cI[Z_{js}^{(l)} \ge 1]\binom{M_{js}^{(n-l)}}{r}(1-P_0)^{M_{js}^{(n-l)}} \\ &\qquad + \sum_{u=0}^{r-1} I[Z_{js}^{(l)} \ge r-u]\,I[M_{js}^{(n-l)} = u]\Big\}. \qquad (4.8)\end{aligned}$$

Since $(x_1 + \cdots + x_r)^2 \le r(x_1^2 + \cdots + x_r^2)$, we can bound $\tau^2\mathbb{E}(U_j^{(n)} - U_j^{(n-l)})^2$ by considering the r different sums separately.

First, for

$$\mathbb{E}\Big\{\Big(\sum_{\substack{s\ge j_n\\ s\ne j}} I[Z_{js}^{(l)} \ge 1]\binom{M_{js}^{(n-l)}}{r}(1-P_0)^{M_{js}^{(n-l)}}\Big)^2\Big\},$$

using the independence of $Z_{j\cdot}^{(l)}$ and $M_{j\cdot}^{(n-l)}$ and Lemma 5.2, and with $\delta_n = 2\bar{p}_n$ as before, the off-diagonal terms give at most

$$c\sum_{s\ge j_n}\sum_{t\ge j_n}(l^2p_sp_t)(np_s)^r(np_t)^r e^{-n(p_s+p_t)}e^{2\delta_n(2r+l)} \;\le\; c'l^2e^{2l\delta_n}\,n^{-1}P_n\hat{\mu}_r,$$

the last line using Lemma 5.4 (v). The terms with $j = s$ then contribute at most

$$c\sum_{s\ge j_n} lp_s(np_s)^r\{1 + (np_s)^r\}e^{-2\beta np_s}e^{2l\delta_n} \;\le\; c'le^{2l\delta_n}n^{-1}\hat{\mu}_r,$$

using Lemma 5.4 (ii). The contribution to $\tau^2\mathbb{E}(U_j^{(n)} - U_j^{(n-l)})^2$ from this first sum is thus no more than $cl^2e^{2l\delta_n}n^{-1}\hat{\mu}_r$.

For $0 \le u \le r-1$, we need to find similar bounds for

$$\mathbb{E}\Big\{\Big(\sum_{\substack{s\ge j_n\\ s\ne j}} I[Z_{js}^{(l)} \ge r-u]I[M_{js}^{(n-l)} = u]\Big)^2\Big\}.$$

Here, the off-diagonal terms contribute at most

$$\begin{aligned} c\sum_{s\ge j_n}\sum_{t\ge j_n}&(l^{2(r-u)}(p_sp_t)^{r-u}(np_s)^u(np_t)^u e^{-n(p_s+p_t)}e^{2\delta_n(2u+l)} \\ &\le\ c'(l/n)^{2(r-u)}e^{2l\delta_n}n\hat\mu_r, \end{aligned}$$

by Lemma 5.4 (v), and the diagonal terms give at most

$$c\sum_{s\ge j_n}(lp_s)^{r-u}(np_s)^u e^{-np_s}e^{2\delta_n(2u+l)}\ \le\ c'(l/n)^{r-u}e^{2l\delta_n}\hat\mu_r.$$

Since, in the above, $u\le r-1$ and $l\le n$, it follows that

$$\tau^2\mathbb{E}(U_j^{(n)} - U_j^{(n-l)})^2\ \le\ cl^2e^{2l\delta_n}n^{-1}\hat\mu_r. \tag{4.9}$$

Returning to (2), and once again recalling (3.5), we thus have a bound of

$$\begin{aligned} \tfrac12\|f''\|\,\tau^{-1}&\sum_{j\ge j_n}\sum_{l\ge0}q_j(l)|y_j(l)|\mathbb{E}(U_j^{(n)}-U_j^{(n-l)})^2 \\ &\le\ c\tau^{-3}\|f''\|\,\frac{\hat\mu_r}{n}\sum_{j\ge j_n}\mathbb{E}\{|y_j(M_j)|M_j^2e^{2M_j\delta_n}\} \\ &\le\ c'\tau^{-3}\|f''\|\,\frac{\hat\mu_r}{n}\sum_{j\ge j_n}(np_j)^r(1+(np_j)^2)e^{-np_j} \\ &\le\ c'\hat\mu_r\tau^{-3}\|f''\|\,(K_{r-1}+K_{r+1})P_n, \end{aligned}$$

from Lemma 5.4 (iii), and this, with (4.4), completes the proof of (2).

For inequality (3), recalling (3.7) and (3.8), we first need to bound the variance $\mathrm{Var}\,\{\mathbb{E}(U_j^{(n)} - U_j^{(n-l)}\,|\,M_{j\cdot}^{(n-l)})\}$. Now

$$\begin{aligned} \tau\mathbb{E}(U_j^{(n)} - U_j^{(n-l)}\,|\,M_{j\cdot}^{(n-l)}) &= \sum_{\substack{s\ge j_n\\ s\ne j}}\mathbb{E}(z(M_{js}^{(n)}) - z(M_{js}^{(n-l)})\,|\,M_{j\cdot}^{(n-l)}) \\ &=: \sum_{\substack{s\ge j_n\\ s\ne j}} g_s(M_{js}^{(n-l)}), \end{aligned}$$

where, from (4.7) and the independence of $Z_{j\cdot}^{(l)}$ and $M_{j\cdot}^{(n-l)}$,

$$|g_s(t)|\ \le\ \frac{lp_s}{P_{0j}}\binom{t}{r}(1-P_0)^tP_0^r, \tag{4.10}$$

but g_s is not non-negative. From Lemmas 5.3 and 5.2, the off-diagonal terms in the variance $\mathrm{Var}\,\{\sum_{s\ge j_n,\,s\ne j}g_s(M_{js}^{(n-l)})\}$ contribute at most

$$\begin{aligned} cl^2e^{2l\delta_n}&\sum_{s\ge j_n}\sum_{t\ge j_n}p_sp_t(np_s)^r(np_t)^re^{-n(p_s+p_t)} \\ &\times\{(p_s+p_t)(1+np_s+np_t)+n^{-1}(1+np_s)(1+np_t)+np_sp_t\}, \end{aligned}$$

and, using Lemma 5.4 (v), this can be bounded by $cl^2 e^{2l\delta_n} n^{-2} P_n \hat{\mu}_r$. The diagonal terms in turn yield at most

$$\sum_{\substack{s \ge j_n \\ s \ne j}} \mathrm{Var}\, g_s(M_{js}^{(n-l)}) \;\le\; cl^2 e^{2l\delta_n} \sum_{s \ge j_n} p_s^2 (np_s)^r (1+(np_s)^r) e^{-2\beta np_s} \;\le\; c' l^2 e^{2l\delta_n} n^{-1} P_n,$$

by Lemma 5.4 (iii). Since also $\hat{\mu}_r \le cn$, it follows that

$$\mathrm{Var}\,\{\mathbb{E}(U_j^{(n)} - U_j^{(n-l)} \,|\, M_{j\cdot}^{(n-l)})\} \;\le\; c\tau^{-2} l^2 e^{2l\delta_n} n^{-1} P_n.$$

For $\tau^2 \mathrm{Var}\, U_j^{(n-l)}$, the considerations are similar but easier, since we now have

$$0 \;\le\; z(t) \le \binom{t}{r} (1-P_0)^t P_0^r$$

in place of (4.10), and the contributions from both diagonal and off-diagonal terms are bounded by $e^{2l\delta_n} \hat{\mu}_r$. Hence, and recalling (3.7) and (3.8), we have arrived at a bound

$$\begin{aligned} &|\mathbb{E}\{[\mathbb{E}(U_j^{(n-l)} - U_j^{(n)} \,|\, M_{j\cdot}^{(n-l)}) - \mathbb{E}(U_j^{(n-l)} - U_j^{(n)})](f'(U_j^{(n-l)}) - f'(\mathbb{E}U_j^{(n-l)}))\}| \\ &\qquad \le\; c\tau^{-2} \|f''\|\, l e^{2l\delta_n} \sqrt{\hat{\mu}_r P_n / n}; \end{aligned} \tag{4.11}$$

the analogue of (3.11),

$$|\mathbb{E}(U_j^{(n)} - U_j^{(n-l)})\{\mathbb{E}f'(U_j^{(n-l)}) - \mathbb{E}f'(U_j^{(n)})\}| \;\le\; c\tau^{-2} \|f''\|\, l^2 e^{2l\delta_n} n^{-1} \hat{\mu}_r, \tag{4.12}$$

follows directly from (4.9). Hence, for (3), we have

$$\begin{aligned} &\Bigg| \tau^{-1} \sum_{j \ge j_n} \sum_{l \ge 0} q_j(l) y_j(l) \{\mathbb{E}[f'(U_j^{(n-l)})(U_j^{(n-l)} - U_j^{(n)})]\mathbb{E}[f'(U_j^{(n)})] \\ &\qquad\qquad - \mathbb{E}[f'(U_j^{(n)})]\mathbb{E}(U_j^{(n-l)} - U_j^{(n)})\} \Bigg| \\ &\qquad \le\; c\tau^{-3} \|f''\| \sum_{j \ge j_n} \mathbb{E}\{M_j^2 |y_j(M_j)| e^{2M_j \delta_n}\} (\sqrt{\hat{\mu}_r P_n / n} + n^{-1} \hat{\mu}_r) \\ &\qquad \le\; c'\tau^{-3} \|f''\| \Big\{ \sum_{j \ge j_n} (np_j)^{r+1} (1+np_j) e^{-np_j} \Big\} (\sqrt{\hat{\mu}_r P_n / n} + n^{-1} \hat{\mu}_r), \end{aligned}$$

and since

$$\Big\{ \sum_{j \ge j_n} (np_j)^{r+1} (1+np_j) e^{-np_j} \Big\}^2 \;\le\; cnP_n \hat{\mu}_r, \tag{4.13}$$

by Lemma 5.4 (v), we conclude that inequality (3) is indeed satisfied.

For inequality (4), we use the simple bound in (3.13), obtaining

$$\begin{aligned} &\Bigg| \tau^{-1} \sum_{j \ge j_n} \sum_{l \ge 0} q_j(l) y_j^2(l) \{\mathbb{E}f'(U_j^{(n-l)}) - \mathbb{E}f'(U_j^{(n)})\} \Bigg| \\ &\qquad \le\; \tau^{-3} \|f''\| \sum_{j \ge j_n} \mathbb{E}\{M_j y_j^2(M_j)\} \\ &\qquad \le\; c\tau^{-3} \|f''\| \sum_{j \ge j_n} (np_j)^r (1+(np_j)^{r+1}) e^{-2\beta np_j} \;\le\; c' \hat{\mu}_r \tau^{-3} \|f''\|, \end{aligned}$$

from Lemma 5.1 (iii), in much the same way as for (4.6). Hence we have now established (2.9).

For (5) and (6), we need the constants κ_j, for which we now have the bounds

$$\kappa_j^+ \ \le\ c(np_j)^r(1+(np_j)^r)e^{-2\beta np_j},$$

from (4.6), and

$$\begin{aligned}\kappa_j^- &\le c\mathbb{E}\{M_j|y_j(M_j)|e^{2M_j\delta_n}\}\sqrt{\hat\mu_r/n}\\ &\le c'(np_j)^r(1+np_j)e^{-np_j}\sqrt{\hat\mu_r/n},\end{aligned}$$

from (4.9). For inequality (5), this immediately gives a bound of

$$c\tau^{-3}\|f''\|\sum_{j\ge j_n}|\kappa_j|(np_j)^r e^{-np_j} \ \le\ c'\hat\mu_r\tau^{-3}\|f''\|,$$

using Lemma 5.4 (ii); for (6), we obtain the bound

$$c\tau^{-3}\|f''\|\sum_{j\ge j_n}|\kappa_j|np_j\sqrt{\hat\mu_r/n} \ \le\ c'\hat\mu_r\tau^{-3}\|f''\|,$$

where, for the contribution from κ_j^-, we again use Lemma 5.4 (v), much as for (4.13). This completes the proof of (2.12), and thus of Theorem 1.2.

5. Appendix

We collect several useful calculations, the first two of which need little proof. We write $m_{(s)} := m(m-1)\dots(m-s+1)$.

Lemma 5.1. *If $M \sim \mathrm{Bi}\,(m,p)$, then for any $x>0$ and $0\le s\le m$,*

$$(i)\quad \mathbb{E}\{M_{(s)}x^M\} \ =\ m_{(s)}(xp)^s(1+p(x-1))^{m-s}.$$

In particular, if $x=e^\delta$, where $0\le\delta\le\delta_0\le 1$, and if $(1-P)e^{\delta_0}\le 1$, then

$$\begin{aligned}&(ii)\quad \mathbb{E}\{M_{(s)}x^M\} \ \le\ (mp)^s\exp\{\delta_0(s+mpe)\};\\ &(iii)\quad \mathbb{E}\{M_{(s)}[(1-P)e^\delta]^M\}\\ &\qquad\qquad \le\ (mp(1-P))^s e^{-(m-s)pP}\exp\{\delta_0[s+mpe(1-P)]\}.\end{aligned}$$

Furthermore, for $0\le x\le 1$ and $p\le 1/2$, we have

$$\begin{aligned}(iv)\quad c(x)e^{-2mp^2}\min\{1,mp\} &\le e^{mp(1-x^2)}\{\mathbb{E}x^{2M}-(\mathbb{E}x^M)^2\}\\ &\le \min\{1,mp(1-x^2)\},\end{aligned}$$

where $c(x) := \min\{(1-e^{-(1-x)^2}),(1-x)^2e^{-(1-x)^2})\}$.

Proof. We prove only (iv). From (i), we have

$$\mathbb{E}x^{2M}-(\mathbb{E}x^M)^2 \ =\ \{1-p(1-x^2)\}^m\Big\{1-\Big(1-\frac{p(1-p)(1-x)^2}{1-p(1-x^2)}\Big)^m\Big\}.$$

The upper bound follows immediately, using the fact that $1-p\le 1-p(1-x^2)$. The lower bound

$$e^{-mp(1-x^2)-2mp^2}\{1-e^{-mp(1-x)^2}\}$$

also uses the fact that $p\le 1/2$, and the argument is completed in standard fashion. □

Lemma 5.2. *Let* $(L, M, m-L-M) \sim \mathrm{MN}\,(m;\, p, q, 1-p-q)$ *be trinomially distributed. Then*

$$\mathbb{E}\{L_{(u)}M_{(v)}w^L x^M\} \;=\; m_{(u+v)}(wp)^u(xq)^v(1+p(w-1)+q(x-1))^{m-u-v}.$$

In particular, if $0 \le w, x \le e^{\delta}$*, where* $0 \le \delta \le \delta_0 \le 1$*, and if* $(1-P)e^{\delta_0} \le 1$*, then*

$$\begin{aligned}
&\mathbb{E}\{L_{(u)}M_{(v)}w^L x^M\} \;\le\; (mp)^u(mq)^v \exp\{\delta_0[(u+v)+m(p+q)e]\};\\
&\mathbb{E}\{L_{(u)}M_{(v)}[(1-P)e^{\delta}]^{L+M}\}\\
&\quad \le\; (mp(1-P))^u(mq(1-P))^v e^{-(m-u-v)(p+q)P}\\
&\qquad\quad \times \exp\{\delta_0[(u+v)+m(p+q)e(1-P)]\}.
\end{aligned}$$

Lemma 5.3. *Let* $(L, M, m-L-M) \sim \mathrm{MN}\,(m;\, p, q, 1-p-q)$ *be trinomial, where* $p+q \le \delta \le 1/4$*, and let the functions* f, g, h, k *satisfy* $0 \le f(l) \le h(l)$ *and* $0 \le g(l) \le k(l)$ *for* $l \in \mathbb{Z}_+$*. Then*

$$\begin{aligned}
&\mathrm{Cov}\,(f(L), g(M)) \;\le\; C_1\\
&\quad := e(p+q)\{\mathbb{E}(Lh(L)e^{2L\delta})\mathbb{E}(k(M)e^{2M\delta}) + \mathbb{E}(h(L)e^{2L\delta})\mathbb{E}(Mk(M)e^{2M\delta})\}.
\end{aligned}$$

If f *and* g *are not nonnegative, but* $|f|$ *and* $|g|$ *are bounded as above, then*

$$\mathrm{Cov}\,(f(L), g(M)) \;\;\le\;\; C_1 + 2m^{-1}\mathbb{E}(Lh(L))\mathbb{E}(Mk(M)) + \frac{4m}{3}pq\mathbb{E}h(L)\mathbb{E}k(M).$$

Proof. From the multinomial formulae, we have

$$\begin{aligned}
&f(u)g(v)\{\mathbf{P}[L=u, M=v] - \mathbf{P}[L=u]\mathbf{P}[M=v]\}\\
&\quad = \frac{f(u)g(v)}{u!v!}p^u q^v\{m_{(u+v)}(1-p-q)^{m-u-v} - m_{(u)}m_{(v)}(1-p)^{m-u}(1-q)^{m-v}\}\\
&\quad \le\; f(u)g(v)\mathbf{P}[L=u]\mathbf{P}[M=v]\{(1-p-q)^{-(u+v)} - 1\} \qquad (5.1)\\
&\quad \le\; h(u)k(v)\mathbf{P}[L=u]\mathbf{P}[M=v](p+q)(u+v)\exp\{2(p+q)(u+v+1)\},
\end{aligned}$$

where the last inequality uses $p+q \le 1/4$. The first part of the lemma now follows.

For the second part, (5.1) should be replaced by

$$\begin{aligned}
&|f(u)g(v)|\mathbf{P}[L=u]\mathbf{P}[M=v]\\
&\quad \left\{\left|(1-p-q)^{-(u+v)} - 1\right| + \left|\frac{(m-u)_{(v)}}{m_{(v)}} - 1\right| + \left|\left(1 - \frac{pq}{(1-p)(1-q)}\right)^m - 1\right|\right\},
\end{aligned}$$

after which we use the bounds

$$\left|\frac{(m-u)_{(v)}}{m_{(v)}} - 1\right| \;\le\; \frac{2uv}{m}; \qquad \left|\left(1 - \frac{pq}{(1-p)(1-q)}\right)^m - 1\right| \;\le\; 4mpq/3.$$

□

Lemma 5.4. *Let* p_s*,* $s \ge j$*, be nonnegative numbers summing to* $P \le 1$*, and define*

$$\sigma_n^2(r) \;:=\; \sum_{s\ge j}(np_s)^r e^{-np_s}, \quad r \ge 1; \qquad \sigma_n^2(0) \;:=\; \sum_{s\ge j}\min(np_s, 1)e^{-np_s}.$$

Then there exist universal constants $K_r^{(\alpha)}$, K_u, K_{uv} and K' such that, for any integers $u \geq v \geq 0$ and for any $\alpha > 0$,

$$(i) \quad \sum_{s \geq j} (np_s)^{u+1} e^{-(1+\alpha)np_s} \;\leq\; K_0^{(\alpha)} \sigma_n^2(0);$$

$$(ii) \quad \sum_{s \geq j} (np_s)^{u+r} e^{-(1+\alpha)np_s} \;\leq\; K_r^{(\alpha)} \sigma_n^2(r);$$

$$(iii) \quad \sum_{s \geq j} (np_s)^{u+1} e^{-np_s} \;\leq\; K_u nP;$$

$$(iv) \quad \Bigl(\sum_{s \geq j} np_s e^{-np_s}\Bigr)^2 \;\leq\; K' n \sigma_n^2(0);$$

$$(v) \quad \sum_{s \geq j} \sum_{t \geq j} (np_s)^{r+u} (np_t)^{r+v} e^{-n(p_s+p_t)} \;\leq\; K_{uv} nP \sigma_n^2(r).$$

Proof. The first inequality reflects the fact that $x^{u+1}e^{-(1+\alpha)x} \leq xe^{-x}$ for $0 \leq x \leq 1$, whereas $x^{u+1}e^{-(1+\alpha)x} \leq e^{-x} \sup_{z \geq 1}\{ze^{-\alpha z}\}$: thus we can take $K^{(\alpha)} = 1/e\alpha$. The second is similar in vein, but easier. The third inequality, and case $u = v = 0$ in the fifth, follow from

$$\sum_{s \geq j} (np_s)^{u+1} e^{-np_s} \;=\; n \sum_{s \geq j} p_s (np_s)^u e^{-np_s} \;\leq\; nP(u/e)^u.$$

For the fifth with $u \geq 1$, we write the sum as

$$n^2 \sum_{s \geq j} p_s (np_s)^{r+u-1} e^{-np_s} \sum_{t \geq j} p_t [(np_t)^{r+u-1} e^{-np_t}]^{\frac{r+v-1}{r+u-1}} \exp\Bigl\{-np_t \frac{u-v}{r+u-1}\Bigr\},$$

and use Cauchy–Schwarz to yield the upper bound

$$\begin{aligned} &n^2 P \sum_{s \geq j} p_s (np_s)^{2r+u+v-2} \exp\Bigl\{-np_s \frac{2r+u+v-2}{r+u-1}\Bigr\} \\ &\quad \leq\; nP \sum_{s \geq j} (np_s)^r e^{-np_s} \max_{x \geq 0}\{x^{r+u+v-1} \exp\{-x(r+v-1)/(r+u-1)\}\}, \end{aligned}$$

noting that $r + u - 1 \geq 1$. For the fourth part, Cauchy–Schwarz gives

$$\Bigl(\sum_{s \geq j} np_s e^{-np_s}\Bigr)^2 \;\leq\; n \sum_{s \geq j} np_s e^{-2np_s} \;\leq\; \sum_{s \geq j} \min\{np_s, e^{-1}\} e^{-np_s}.$$

Acknowledgement This work was carried out during a visit to the Institute for Mathematical Sciences at the National University of Singapore, whose support is gratefully acknowledged. We also thank the referee for a number of helpful comments, and for drawing our attention to the work of Mirakhmedov.

References

A. D. Barbour. Notes on poisson approximation (2009). Tutorial notes, Institute for Mathematical Sciences, National University of Singapore www.ims.nus.edu.sg/Programs/stein09/files/AndrewBarbour_tut.pdf.

A. D. Barbour and V. Čekanavičius. Total variation asymptotics for sums of independent integer random variables. *Ann. Probab.* **30** (2), 509–545 (2002). MR1905850.

A. D. Barbour and A. V. Gnedin. Small counts in the infinite occupancy scheme. *Electron. J. Probab.* **14**, no. 13, 365–384 (2009). MR2480545.

M. Dutko. Central limit theorems for infinite urn models. *Ann. Probab.* **17** (3), 1255–1263 (1989). MR1009456.

A. Gnedin, B. Hansen and J. Pitman. Notes on the occupancy problem with infinitely many boxes: general asymptotics and power laws. *Probab. Surv.* **4**, 146–171 (electronic) (2007). doi:10.1214/07-PS092. MR2318403.

H.-K. Hwang and S. Janson. Local limit theorems for finite and infinite urn models. *Ann. Probab.* **36** (3), 992–1022 (2008). doi:10.1214/07-AOP350. MR2408581.

S. Karlin. Central limit theorems for certain infinite urn schemes. *J. Math. Mech.* **17**, 373–401 (1967). MR0216548.

L. Le Cam. An approximation theorem for the Poisson binomial distribution. *Pacific J. Math.* **10**, 1181–1197 (1960). MR0142174.

R. Michel. An improved error bound for the compound poisson approximation of a nearly homogeneous portfolio. *ASTIN Bulletin* **17** (2), 165 – 169 (1987).

Sh. A. Mirakhmedov. Randomized decomposable statistics in a generalized allocation scheme over a countable set of cells. *Diskretnaya Matematika* **1** (4), 46–62 (1989). MR1041684.

Sh. A. Mirakhmedov. Randomized decomposable statistics in the scheme of independent allocating particles into boxes. *Discrete Mathematics and Applications* **2** (1), 91–108 (1992).

A. Röllin. Approximation of sums of conditionally independent variables by the translated Poisson distribution. *Bernoulli* **11** (6), 1115–1128 (2005). doi:10.3150/bj/1137421642. MR2189083.