

APPROXIMATING THE NUMBER OF SUCCESSES IN INDEPENDENT TRIALS: BINOMIAL VERSUS POISSON¹

BY K. P. CHOI AND AIHUA XIA

National University of Singapore and University of Melbourne

Let I_1, I_2, \dots, I_n be independent Bernoulli random variables with $\mathbb{P}(I_i = 1) = 1 - \mathbb{P}(I_i = 0) = p_i$, $1 \leq i \leq n$, and $W = \sum_{i=1}^n I_i$, $\lambda = \mathbb{E}W = \sum_{i=1}^n p_i$. It is well known that if p_i 's are the same, then W follows a binomial distribution and if p_i 's are small, then the distribution of W , denoted by $\mathcal{L}W$, can be well approximated by the $\text{Poisson}(\lambda)$. Define $r = \lfloor \lambda \rfloor$, the greatest integer $\leq \lambda$, and set $\delta = \lambda - \lfloor \lambda \rfloor$, and κ be the least integer more than or equal to $\max\{\lambda^2/(r-1-(1+\delta)^2), n\}$. In this paper, we prove that, if $r > 1 + (1+\delta)^2$, then

$$d_\kappa < d_{\kappa+1} < d_{\kappa+2} < \dots < d_{TV}(\mathcal{L}W, \text{Poisson}(\lambda)),$$

where d_{TV} denotes the total variation metric and $d_m = d_{TV}(\mathcal{L}W, \text{Bi}(m, \lambda/m))$, $m \geq \kappa$. Hence, in modelling the distribution of the sum of Bernoulli trials, Binomial approximation is generally better than Poisson approximation.

1. Introduction and the main result. Let I_k 's, $1 \leq k \leq n$, denote n independent Bernoulli trials with success probabilities p_k 's. Let $W := \sum_{k=1}^n I_k$ denote the number of successes with mean $\lambda := \mathbb{E}W = \sum_{k=1}^n p_k$. If p_k 's are the same, that is, the trials are identically distributed, then the distribution of W , denoted by $\mathcal{L}W$, has a binomial distribution. However, if p_k 's are different and n is large, then $\mathcal{L}W$ is often difficult to calculate and an approximation is necessary. Binomial, Poisson and normal distributions have often been used to approximate $\mathcal{L}W$ [see Barbour, Holst and Janson (1992) and Volkova (1995)].

Approximating $\mathcal{L}W$ by Poisson has a long history. Indeed, the Poisson distribution $\text{Po}(\lambda)$ was first introduced by Poisson (1837) as the limiting distribution of the binomial distribution $\text{Bi}(m, \lambda/m)$ as $m \rightarrow \infty$. von Bortkewitsch (1898) and Student (1907) were among the earliest to apply Poisson distribution to statistical analysis of rare events. However, the mathematical foundation of why Poisson models could be used to fit the number of rare events was not fully understood until the pioneering work of Chen (1975). Chen adapted Stein's differential method for assessing the accuracy of normal approximation for sums of dependent random variables for Poisson approximations. The resulting Stein–Chen method can be used to work out accurate upper bounds on the total variation distance between

Received June 2001; revised November 2001.

¹Supported in part by ARF Research Grant R-146-000-013-112 from National University of Singapore and an ARC small grant from University of New South Wales.

AMS 2000 subject classifications. Primary 60F05; secondary 60E15.

Key words and phrases. Binomial distribution, Poisson distribution, total variation metric.

a Poisson distribution and the distribution of the sum of dependent 0–1 random variables. In the context of independent Bernoulli trials, Barbour and Hall (1984) use Stein–Chen method to obtain

$$\frac{1}{32} \min\left\{\frac{1}{\lambda}, 1\right\} \sum_{i=1}^n p_i^2 \leq d_{TV}(\mathcal{L}W, \text{Po}(\lambda)) \leq \frac{1 - e^{-\lambda}}{\lambda} \sum_{i=1}^n p_i^2,$$

where d_{TV} is the total variation metric for measuring the distance between two probability measures \mathbf{Q}_1 and \mathbf{Q}_2 on $\mathbf{Z}_+ := \{0, 1, 2, \dots\}$ defined as

$$d_{TV}(\mathbf{Q}_1, \mathbf{Q}_2) = \sup_{A \subset \mathbf{Z}_+} |\mathbf{Q}_1(A) - \mathbf{Q}_2(A)| = \sum_{\mathbf{Q}_1\{i\} > \mathbf{Q}_2\{i\}} [\mathbf{Q}_1\{i\} - \mathbf{Q}_2\{i\}].$$

Hence, the correct order for the distance $d_{TV}(\mathcal{L}W, \text{Po}(\lambda))$ is $\min\{\lambda^{-1}, 1\} \sum_{i=1}^n p_i^2$.

A direct setup of Stein identity for estimating binomial approximation error to $\mathcal{L}W$ is given in Ehm (1991) for $m = n$, $p = \lambda/n$ and $q = 1 - p$, and according to Ehm's Lemma 2, the correct order for the distance $d_{TV}(\mathcal{L}W, \text{Bi}(n, p))$ is

$$\min\{1, (npq)^{-1}\} \sum_{i=1}^n (p_i - p)^2.$$

For more investigations on binomial approximation, see Barbour, Holst and Janson (1992), Soon (1996) and Roos (2000).

These bounds, though informative, do not quite compare the binomial and Poisson approximations to $\mathcal{L}W$. For a given W , Deheuvels and Pfeifer (1986) determined asymptotically those Poisson distributions $\text{Po}(\nu)$ which minimize the total variation distance of Poisson approximations to $\mathcal{L}W$. Our main interest in this paper is to compare the binomial and Poisson approximations with the same mean λ . More precisely, if one wants to compute $\mathbb{P}(W \in A)$, should one use a binomial or a Poisson? Obviously, the answer depends on the form of set A . For sets A of certain intervals, it is well understood which model should be used due to an elegant result of Hoeffding [(1956) Theorem 4] which states the following.

THEOREM A. *If $\mathbb{E}W = \lambda$, and a is an integer, then*

$$\begin{aligned} \mathbb{P}(W \leq a) &\leq \mathbb{P}\left(\text{Bi}\left(n, \frac{\lambda}{n}\right) \leq a\right) && \text{if } 0 \leq a \leq \lambda - 1, \\ \mathbb{P}(W \leq a) &\geq \mathbb{P}\left(\text{Bi}\left(n, \frac{\lambda}{n}\right) \leq a\right) && \text{if } \lambda \leq a \leq n. \end{aligned}$$

To avoid introducing more notation, we generally use distributions [here, e.g., $\text{Bi}(n, \frac{\lambda}{n})$] as random variables having the corresponding distributions.

Combining this result and an observation by Anderson and Samuels (1967), Theorem 2.1 and Corollary 2.1, we have for integer a ,

if $0 \leq a \leq \lambda - 1$,

$$(1.1) \quad \begin{aligned} \mathbb{P}(W \leq a) &\leq \mathbb{P}\left(\text{Bi}\left(n, \frac{\lambda}{n}\right) \leq a\right) \\ &< \mathbb{P}\left(\text{Bi}\left(n+1, \frac{\lambda}{n+1}\right) \leq a\right) < \dots < \mathbb{P}(\text{Po}(\lambda) \leq a) \end{aligned}$$

and if $a \geq \lambda + 1$,

$$(1.2) \quad \begin{aligned} \mathbb{P}(W \geq a) &\leq \mathbb{P}\left(\text{Bi}\left(n, \frac{\lambda}{n}\right) \geq a\right) \\ &< \mathbb{P}\left(\text{Bi}\left(n+1, \frac{\lambda}{n+1}\right) \geq a\right) < \dots < \mathbb{P}(\text{Po}(\lambda) \geq a) \end{aligned}$$

showing that the binomial model is preferred over the Poisson model for approximating $\mathcal{L}W$ for such sets A .

What happens if the set A is not of such intervals? We are then led to compare $d_{TV}(\mathcal{L}W, \text{Bi}(n, \frac{\lambda}{n}))$ and $d_{TV}(\mathcal{L}W, \text{Po}(\lambda))$. Our main result (Theorem 1.1) states that, under some mild conditions on λ , binomial approximation is better than Poisson approximation in the sense of total variation. Moreover, there is a monotone analogue of (1.1) or (1.2). This result is also motivated by the following considerations.

As the mean of a Poisson distribution is the same as its variance, if the data of “rare events” have the sample mean close to the sample variance, then the Poisson fitting is likely a good choice, as evidenced in von Bortkewitsch (1898) and Student (1907). However, when it is found that the variance of the counts is substantially larger than the mean, the negative binomial distribution is sometimes instead considered as a model [see Bliss and Fisher (1953) or Hand, Daly, McConway and Ostrowski (1996), pages 173–176, for many sets of data of this kind]. The underlying structure of getting such data suggests that the data exhibit a kind of positive dependence and it is well known that compound Poisson distribution can take into consideration of the dependence of the data. As a member of compound Poisson distribution family, negative binomial can play a better role than Poisson.

On the other hand, if the variance of the counts is much smaller than the mean, then a binomial could match the variance of the data better than that of a Poisson and hence could fit the data better than a Poisson. The theorem below confirms that for the counts of *independent* Bernoulli trials, binomial is indeed a more favored model than Poisson.

THEOREM 1.1. *Let $W = \sum_{k=1}^n I_k$, where I_k 's are independent Bernoulli random variables such that $\mathbb{P}(I_k = 1) = 1 - \mathbb{P}(I_k = 0) = p_k$, for $1 \leq k \leq n$. Let $d_m = d_{TV}(\mathcal{L}W, \text{Bi}(m, \lambda/m))$, $r = \lfloor \lambda \rfloor$, the integer part of λ , $\delta = \lambda - \lfloor \lambda \rfloor$ and κ the least integer more than or equal to $\max\{n, \lambda^2/(r - 1 - (1 + \delta)^2)\}$. If $r > 1 + (1 + \delta)^2$, then*

$$d_\kappa < d_{\kappa+1} < d_{\kappa+2} < \dots < d_{TV}(\mathcal{L}W, \text{Po}(\lambda)).$$

In particular, if $\lambda \geq 6$ and $n \geq (r + 1)^2 / (r - 5)$, then

$$(1.3) \quad d_{TV}\left(\mathcal{L}W, \text{Bi}\left(n, \frac{\lambda}{n}\right)\right) < d_{TV}\left(\mathcal{L}W, \text{Bi}\left(n + 1, \frac{\lambda}{n + 1}\right)\right) < \dots < d_{TV}(\mathcal{L}W, \text{Po}(\lambda)).$$

And these inequalities are strict.

REMARK 1.1. When $6 \leq \lambda < 7$, (1.3) holds for $n \geq 49$, which is hardly restrictive in applications.

REMARK 1.2. Theorem 1.1 says that d_m is increasing for $m \geq \kappa$ if $r > 1 + (1 + \delta)^2$. We believe the restrictions of $r > 1 + (1 + \delta)^2$ and $m \geq \kappa$ could be slightly relaxed. In fact, the idea of the moment estimates suggests that the best binomial distributions $\text{Bi}(m, p)$ to fit the distribution of W should be those with $mp \approx \mathbb{E}W$ and $mp(1 - p) \approx \text{Var}(W)$, giving $m \approx \frac{\lambda^2}{\sum_{i=1}^n p_i^2}$. This leads to the following conjecture.

CONJECTURE. Let m_0 be the integer part of $\frac{\lambda^2}{\sum_{i=1}^n p_i^2}$, then d_m is decreasing for $m \leq m_0$ and increasing for $m > m_0$.

EXAMPLE. Suppose

$$\begin{aligned} X_1 &\sim \text{Binomial}(70, 0.1), & X_2 &\sim \text{Binomial}(9, 1/3), \\ X_3 &\sim \text{Binomial}(2, 1/\sqrt{2}) \end{aligned}$$

are independent random variables and let $W = X_1 + X_2 + X_3$. It has been shown in Brown and Xia (2001) that binomial approximation to $\mathcal{L}W$ is much better

TABLE 1

m	48	49	50	51	52	53	54	55
d_m	0.48006	0.48072	0.53914	0.60633	0.69992	0.80139	0.90084	1.0055
m	56	57	58	59	60	65	70	75
d_m	1.1080	1.2064	1.3099	1.4103	1.5070	1.9409	2.3065	2.6187
m	80	81	85	90	95	100	200	300
d_m	2.8886	2.9382	3.1241	3.3314	3.5153	3.6795	5.1831	5.6626
m	400	500	600	700	800	900	1000	1250
d_m	5.8984	6.0388	6.1318	6.1980	6.2475	6.2860	6.3167	6.3719
m	1500	1750	2000	2250	2500	2750	3000	3250
d_m	6.4086	6.4348	6.4544	6.4696	6.4818	6.4918	6.5001	6.5071
m	3500	3750	4000	5000	6000	7000	8000	∞
d_m	6.5131	6.5183	6.5229	6.5365	6.5457	6.5522	6.5570	6.5912

than Poisson approximation. In fact, if we choose $m = 48$ and $p = 0.237796$ so that $mp = \mathbb{E}(W)$ and $mp(1 - p) \approx \text{Var}(W)$, then $d_{48} = 0.48\%$ and $d_\infty = d_{TV}(\mathcal{L}W, \text{Po}(\mathbb{E}(W))) = 6.6\%$. More details of approximation errors (in percentage) for other binomial distributions are provided in Table 1. Note that, in this case, $\kappa = n = 81$ and one can easily verify that the values of the table are in agreement with the conjecture.

2. Proofs. For this section, $m \geq n$. To simplify our notation, we write

$$\text{Bi}_m(j) := \text{Bi}(m, \lambda/m)\{j\}$$

and, for each $A \subset \mathbf{Z}_+$,

$$\text{Bi}_m(A) := \sum_{j \in A} \text{Bi}_m(j).$$

Let

$$A_m(j) = \mathbb{P}(W = j) / \text{Bi}_m(j), \quad 0 \leq j \leq m.$$

We need the following results from Hoeffding [(1956), Corollary 2.1 and Theorem 5].

LEMMA 2.1. *For any function $g : \mathbf{Z}_+ \rightarrow \mathbf{R}$, the maximum and minimum of*

$$f(p_1, \dots, p_n) := \mathbb{E}g(W),$$

as a function of p_1, \dots, p_n with $p_1 + \dots + p_n = \lambda$, are attained at points whose coordinates take on, at most, three different values, only one of which is distinct from 0 and 1.

LEMMA 2.2. *For integers a and b such that $0 \leq a \leq \lambda \leq b \leq n$,*

$$\mathbb{P}(a \leq \text{Bi}_m \leq b) \leq \mathbb{P}(a \leq W \leq b)$$

for $m \geq n$.

REMARK 2.1. As in (1.2), for $m \geq n$, and integers a and b such that $0 \leq a \leq \lambda \leq b$, we have

$$\mathbb{P}(a \leq \text{Bi}_{m+1} \leq b) < \mathbb{P}(a \leq \text{Bi}_m \leq b).$$

We need the following lemmas to prove Theorem 1.1.

LEMMA 2.3. *Let $m \geq n$ and $r = \lfloor \lambda \rfloor$ as above, then*

$$(2.1) \quad A_m(r - 2) \leq A_m(r + 1),$$

$$(2.2) \quad A_m(r) \geq A_m(r + 3).$$

Moreover, if $\delta > 0.5$, then $A_m(r - 1) \leq A_m(r + 1)$, and if $\delta \leq 0.5$, then $A_m(r) \geq A_m(r + 2)$.

PROOF. We will prove (2.1) only as the other claims can be proved in the same way. Let

$$(2.3) \quad a_m = \frac{\text{Bi}_m(r - 2)}{\text{Bi}_m(r + 1)} = \frac{(r - 1)r(r + 1)(m - \lambda)^3}{(m - r + 2)(m - r + 1)(m - r)\lambda^3}.$$

Note that a_m is dependent on λ . Then (2.1) is equivalent to

$$\mathbb{P}(W = r - 2) - a_m \mathbb{P}(W = r + 1) \leq 0.$$

Define

$$g(w) = \begin{cases} 1, & \text{for } w = r - 2, \\ -a_m, & \text{for } w = r + 1, \\ 0, & \text{otherwise.} \end{cases}$$

It follows from Lemma 2.1 that $\mathbb{E}g(W)$ is maximized at one of the shifted binomial distributions with the same mean:

$$\begin{aligned} & \mathbb{P}(W = r - 2) - a_m \mathbb{P}(W = r + 1) \\ &= \mathbb{E}g(W) \leq \max_{s, k \in \mathbf{Z}_+, s \leq \lambda \leq s + k \leq m} \mathbb{E}g(s + \text{Bi}(k, (\lambda - s)/k)). \end{aligned}$$

Hence, it suffices to show

$$\begin{aligned} & \mathbb{P}(s + \text{Bi}(k, (\lambda - s)/k) = r - 2) \\ & \leq a_m \mathbb{P}(s + \text{Bi}(k, (\lambda - s)/k) = r + 1), \quad s \leq r - 1, \end{aligned}$$

which is equivalent to, after simple expansion, much as for (2.3),

$$(2.4) \quad b_{s,k} := \frac{(r - s + 1)(r - s)(r - s - 1)(k - \lambda + s)^3}{(k - r + s + 2)(k - r + s + 1)(k - r + s)(\lambda - s)^3} \leq a_m,$$

for all $s, k \in \mathbf{Z}_+$ with $s \leq r - 1$ and $s + k \leq m$. However,

$$(k - \lambda + s)^3 / [(k - r + s + 2)(k - r + s + 1)(k - r + s)]$$

is an increasing function of k , (2.4) is ensured by $b_{s,m-s} \leq a_m$, that is,

$$\frac{(r - s + 1)(r - s)(r - s - 1)}{(\lambda - s)^3} \leq \frac{(r + 1)r(r - 1)}{\lambda^3}.$$

The last inequality is obvious because $(r - s + 1)(r - s)(r - s - 1)/(\lambda - s)^3$ is a decreasing function of s in $[0, r - 1]$. This completes the proof. \square

For $\frac{\lambda}{m+1} \leq p \leq \frac{\lambda}{m}$, let

$$K_m(p) = \text{Bi}(m, p) + I(p),$$

where $I(p)$ is an indicator random variable with $\mathbb{P}(I(p) = 1) = 1 - \mathbb{P}(I(p) = 0) = \lambda - mp \in [0, 1]$, and $I(p)$ is independent of $\text{Bi}(m, p)$. Then, we have $K_m(\lambda/m) \sim \text{Bi}_m$ and $K_m(\lambda/(m + 1)) \sim \text{Bi}_{m+1}$.

LEMMA 2.4. For each $j \in \mathbf{Z}_+$ and $\frac{\lambda}{m+1} < p < \frac{\lambda}{m}$, we have

$$(2.5) \quad \frac{\partial \mathbb{P}(K_m(p) = j)}{\partial p} = \frac{\text{Bi}(m, p)\{j\}[\lambda - (m + 1)p]}{(m + 1 - j)p^2(1 - p)} [(mp - j)^2 + mp^2 - j],$$

$$(2.6) \quad \frac{\partial \mathbb{P}(K_m(p) \leq j)}{\partial p} = \frac{\text{Bi}(m, p)\{j\}}{p(1 - p)} [\lambda - (m + 1)p](mp - j).$$

PROOF. Clearly, for each $i \in \mathbf{Z}_+$, we have

$$\frac{\partial \text{Bi}(m, p)\{i\}}{\partial p} = m[\text{Bi}(m - 1, p)\{i - 1\} - \text{Bi}(m - 1, p)\{i\}],$$

and so

$$\frac{\partial \text{Bi}(m, p)([0, i])}{\partial p} = -m \text{Bi}(m - 1, p)\{i\}.$$

Now,

$$\mathbb{P}(K_m(p) \leq j) = \text{Bi}(m, p)([0, j])(1 - \lambda + mp) + \text{Bi}(m, p)([0, j - 1])(\lambda - mp),$$

which implies

$$\begin{aligned} \frac{\partial \mathbb{P}(K_m(p) \leq j)}{\partial p} &= -m[\text{Bi}(m - 1, p)\{j\}(1 - \lambda + mp) \\ &\quad + \text{Bi}(m - 1, p)\{j - 1\}(\lambda - mp) - \text{Bi}(m, p)\{j\}] \\ &= \frac{\text{Bi}(m, p)\{j\}}{p(1 - p)} [\lambda - (m + 1)p](mp - j). \end{aligned}$$

Finally, (2.5) follows from (2.6) by taking difference and some algebraic rearrangement. \square

LEMMA 2.5. Suppose that $r > 1 + (1 + \delta)^2$ and $m \geq \lambda^2/[r - 1 - (1 + \delta)^2]$. If $r - 1 \leq j \leq r + 1$ or $j = r + 2$ and $\delta > 0.5$, then

$$\text{Bi}_{m+1}(j) < \text{Bi}_m(j).$$

PROOF. Let

$$h(p) = (mp - j)^2 + mp^2 - j = m(m + 1)p^2 - 2mpj + j(j - 1).$$

Since

$$m \geq \frac{\lambda^2}{r - 1 - (1 + \delta)^2} \geq \frac{j(j - 1)}{j - (\lambda - j)^2}$$

for $r - 1 \leq j \leq r + 1$ or $j = r + 2$ and $\delta > 0.5$, it follows that

$$h\left(\frac{\lambda}{m+1}\right) = \frac{m}{m+1}[(\lambda - j)^2 + j(j - 1)/m - j] \leq 0.$$

On the other hand, by the fact that

$$m \geq \frac{\lambda^2}{r - 1 - (1 + \delta)^2} \geq \frac{\lambda^2}{j - (\lambda - j)^2}$$

for $r - 1 \leq j \leq r + 2$, we have

$$h\left(\frac{\lambda}{m}\right) = (\lambda - j)^2 + \frac{\lambda^2}{m} - j \leq 0.$$

Hence $h(p) \leq 0$ for all $\frac{\lambda}{m+1} < p < \frac{\lambda}{m}$, giving

$$\frac{\partial \mathbb{P}(K_m(p) = j)}{\partial p} \geq 0, \quad \frac{\lambda}{m+1} < p < \frac{\lambda}{m},$$

and consequently,

$$\text{Bi}_{m+1}(j) = \mathbb{P}(K_m(\lambda/(m+1)) = j) < \mathbb{P}(K_m(\lambda/m) = j) = \text{Bi}_m(j),$$

for $r - 1 \leq j \leq r + 1$ or $j = r + 2$ and $\delta > 0.5$. \square

PROOF OF THEOREM 1.1. We first note that $\lambda^2/(r - 1 - (1 + \delta)^2)$ is a non-decreasing function of δ ; (1.3) follows from the first assertion of Theorem 1.1.

Next, for each $m \geq \kappa$, define

$$l_m = \min\{j : \mathbb{P}(W = j) \geq \text{Bi}_m(j)\},$$

$$r_m = \max\{j : \mathbb{P}(W = j) \geq \text{Bi}_m(j)\}$$

and set

$$L_m = [0, l_m - 1], \quad M_m = [l_m, r_m] \quad \text{and} \quad R_m = [r_m + 1, \infty).$$

Since $A_m(j)/A_m(j - 1)$ is decreasing in $1 \leq j \leq m$ [see the inequality (5) of Samuels (1965)], we have

$$\mathbb{P}(W = j) \begin{cases} \geq \text{Bi}_m(j), & \text{if } l_m \leq j \leq r_m, \\ < \text{Bi}_m(j), & \text{if } 0 \leq j < l_m \text{ or } r_m < j \leq m. \end{cases}$$

Therefore, we have

$$(2.7) \quad d_m = \mathbb{P}(W \in M_m) - \text{Bi}_m(M_m).$$

Fix $m \geq \kappa$. Letting $a = r$, $b = r + 1$ in Lemma 2.2, we have

$$\mathbb{P}(r \leq W \leq r + 1) \geq \text{Bi}_m([r, r + 1]).$$

There are three cases to consider:

$$(I) \mathbb{P}(W = r) \geq \text{Bi}_m(r) \text{ and } \mathbb{P}(W = r + 1) \geq \text{Bi}_m(r + 1),$$

$$(II) \mathbb{P}(W = r) > \text{Bi}_m(r) \text{ but } \mathbb{P}(W = r + 1) < \text{Bi}_m(r + 1)$$

and

$$(III) \mathbb{P}(W = r) < \text{Bi}_m(r) \text{ and } \mathbb{P}(W = r + 1) > \text{Bi}_m(r + 1).$$

To complete the proof, it is sufficient to show that $d_m < d_{m+1}$ for the three cases.

Case I. We have, under this case, $l_m \leq r$ and $r + 1 \leq r_m$ and hence

$$\begin{aligned} d_m &= \mathbb{P}(W \in M_m) - \text{Bi}_m(M_m) \\ &< \mathbb{P}(W \in M_m) - \text{Bi}_{m+1}(M_m) \quad (\text{by Remark 2.1}) \\ &\leq d_{m+1}. \end{aligned}$$

Case II. Under this situation, we have $l_m \leq r$ and $r_m = r$. Lemma 2.3 shows that $l_m = r$ or $r - 1$. Now, it follows from (2.7) and Lemma 2.5 that

$$d_m = \mathbb{P}(W \in M_m) - \text{Bi}_m(M_m) < \mathbb{P}(W \in M_m) - \text{Bi}_{m+1}(M_m) \leq d_{m+1}.$$

Case III. The proof is essentially the same as in Case II. Here Lemma 2.3 implies that $l_m = r + 1$ and $r_m = r + 1$ or $r + 2$ and $\delta > 0.5$. The proof is thus completed by applying (2.7) and Lemma 2.5 again. \square

REFERENCES

- ANDERSON, T. W. and SAMUELS, S. (1967). Some inequalities among binomial and Poisson probabilities. *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* **1** 1–12. Univ. California Press, Berkeley.
- BARBOUR, A. D. and HALL, P. (1984). On the rate of Poisson convergence. *Math. Proc. Cambridge Philos. Soc.* **95** 473–480.
- BARBOUR, A. D., HOLST, L. and JANSON, S. (1992). *Poisson Approximation*. Oxford Univ. Press, Oxford.
- BLISS, C. and FISHER, R. A. (1953). Fitting the negative binomial distribution to biological data. *Biometrics* **9** 174–200.
- BROWN, T. C. and XIA, A. (2001). Stein's method and birth–death processes. *Ann. Probab.* **29** 1373–1403.
- CHEN, L. H. Y. (1975). Poisson approximation for dependent trials. *Ann. Probab.* **3** 534–545.
- DEHEUVELS, P. and PFEIFER, D. (1986). A semigroup approach to Poisson approximation. *Ann. Probab.* **14** 663–676.
- EHM, W. (1991). Binomial approximation to the Poisson binomial distribution. *Statist. Probab. Lett.* **11** 7–16.
- HAND, D. J., DALY, F., MCCONWAY, K. J. and OSTROWSKI, E. (1996). *A Handbook of Small Data Sets*. Chapman and Hall, London.
- HOEFFDING, W. (1956). On the distribution of the number of successes in independent trials. *Ann. Math. Statist.* **27** 713–721.
- POISSON, S. D. (1837). *Recherches sur la probabilité des jugements en matière criminelle et en matière civile, précédées des règles générales du calcul des probabilités*. Bachelier, Paris.

- ROOS, B. (2000). Binomial approximation to the Poisson binomial distribution: The Krawtchouk expansion. *Theory Probab. Appl.* **45** 258–272.
- SAMUELS, S. M. (1965). On the number of successes in independent trials. *Ann. Math. Statist.* **36** 1272–1278.
- SOON, S. Y. T. (1996). Binomial approximation for dependent indicators. *Statist. Sinica* **6** 703–714.
- STUDENT (1907). On the error of counting with a haemocytometer. *Biometrika* **5** 351–360.
- VOLKOVA, A. YU. (1995). A refinement of the central limit theorem for sums of independent random indicators. *Theory Probab. Appl.* **40** 791–794.
- VON BORTKEWITSCH, L. (1898). *Das Gesetz der kleinen Zahlen*. Teubner, Leipzig.

DEPARTMENT OF MATHEMATICS
NATIONAL UNIVERSITY OF SINGAPORE
2 SCIENCE DRIVE 2
SINGAPORE 117543
REPUBLIC OF SINGAPORE
E-MAIL: matckp@nus.edu.sg

DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF MELBOURNE
VIC 3010
AUSTRALIA
E-MAIL: xia@ms.unimelb.edu.au