

BINOMIAL APPROXIMATION FOR DEPENDENT INDICATORS

Spario Y. T. Soon

Nanyang Technological University

Abstract: A binomial approximation theorem for dependent indicators using Stein's method and coupling is proved. The approximating binomial distribution $B(n', p')$ is chosen in such a way that its first moment is equal to that of W and its variance is asymptotically equal to that of W as n' tends to infinity where W is the sum of independent indicators and p' is bounded away from 1. Three examples, one of which concerns two different approximations for the hypergeometric distribution, are given to illustrate applications of the theorem obtained.

Key words and phrases: Binomial distribution, coupling, indicators, Stein's method, total variation distance.

1. Introduction and Statement of Main Result

Stein (1972) introduced a powerful and general method for obtaining an explicit bound for the error in the normal approximation for dependent random variables. This method was adapted and applied to the Poisson approximation by Chen (1974, 1975). Since then, Stein's method has stimulated an area of intensive research in combinatorics, probability and statistics. The method was also extended to the binomial distribution by Stein (1986), Ehm (1991) and Barbour, Holst and Janson (1992), to the compound Poisson distribution by Arratia, Goldstein and Gordon (1990) and Barbour, Chen and Loh (1992), to the multinomial distribution by Loh (1992), to the multivariate normal distribution by Götze (1991), and to the processes by Barbour (1988, 1990), Barbour and Brown (1992). Excellent accounts can be found in Stein (1986) and Barbour, Holst and Janson (1992).

The aim of this paper is to obtain a binomial approximation theorem for dependent indicators using Stein's method and coupling.

Let X_{1n}, \dots, X_{nn} be random indicators with

$$P(X_{in} = 1) = p_{in} = 1 - P(X_{in} = 0),$$

which are not necessarily independent or identical, and let $W_n = \sum_{i=1}^n X_{in}$. To simplify the notation, we write $p_i = p_{in}$, $X_i = X_{in}$ and $W = W_n$.

We are interested in the bound of binomial approximation to the distribution of W , denoted by $\mathcal{L}(W)$, measured in terms of total variation distance. The total variation distance d_{TV} between two probability measures P and Q is defined by

$$d_{TV}(P, Q) = \sup_E |P(E) - Q(E)|,$$

where the supremum is taken over all measurable sets of the real line.

Assuming independence of the indicators X_i , Ehm (1991) gave upper and lower bounds on $d_{TV}(\mathcal{L}(W), B(n, p))$, the total variation distance between $\mathcal{L}(W)$ and the binomial distribution $B(n, p)$, where $p = \sum_{i=1}^n p_i/n (= 1 - q)$, via Stein's method. The bounds are

$$\begin{aligned} (1/124)\{(npq)^{-1} \wedge 1\} \sum_{i=1}^n (p_i - p)^2 &\leq d_{TV}(\mathcal{L}(W), B(n, p)) \\ &\leq (1 - p^{n+1} - q^{n+1})\{(n+1)pq\}^{-1} \sum_{i=1}^n (p_i - p)^2. \end{aligned}$$

This result shows that if the p_i 's are close together, the approximation of $\mathcal{L}(W)$ by $B(n, p)$ is good. However, it may not be good to approximate $\mathcal{L}(W)$ by $B(n, p)$ where some of the p_i 's are far apart. Let us look at one example. Let $p_1 = \dots = p_{[n/2]} = 0$ and $p_{[n/2]+1} = \dots = p_n = 1/2$. The bound for $d_{TV}(\mathcal{L}(W), B(n, p))$ obtained by using Ehm's result is $(1 - 2^{-n})/3$ which is quite large. Therefore, it is natural to ask whether we can improve the situation by approximating $\mathcal{L}(W)$ by another binomial distribution. In this paper we choose $B(n', p')$ to approximate $\mathcal{L}(W)$, where

$$n' = \left[\left(\sum_{i=1}^n p_i \right)^2 / \sum_{i=1}^n p_i^2 + 1/2 \right] \quad \text{and} \quad p' = \sum_{i=1}^n p_i / n'.$$

The bracket $[m]$ represents the integral part of m .

For simplicity, denote $\lambda_i = \sum_{j=1}^n p_j^i$ and $\lambda = \lambda_1$. Also, let $p^* = \lambda_2/\lambda$. Here, n' and p' are chosen in such a way that the binomial distribution $B(n', p')$ has the same expectation as $\mathcal{L}(W)$ and in the case where X_1, \dots, X_n are independent and p^* is bounded away from 1, both distributions will have the same variance asymptotically when n' tends to infinity. (Note that $n' \rightarrow \infty$ implies $n \rightarrow \infty$). To see this, we argue as follows: since $\lambda - p^*/2 \leq n'p^* \leq \lambda + p^*/2$, we have $\lambda/p^* \geq n' - 1/2$. This implies that $p^* = o(\lambda)$ as $n' \rightarrow \infty$, so that $n'p^* \sim \lambda = n'p'$. Thus, under the case where X_1, \dots, X_n are independent and p^* is bounded away from 1,

$$n'p'(1 - p') \sim \lambda(1 - p^*) = \lambda - \lambda_2 = \text{Var}(W).$$

Hence the claim.

Denote $\Gamma = \{1, \dots, n\}$ and set $\Gamma_i = \Gamma - \{i\}$,

$$p_{\max} = \max_{1 \leq i \leq n} p_i, \quad p_{\min} = \min_{1 \leq i \leq n} p_i \quad \text{and} \quad r = p_{\max} - p_{\min}.$$

Also, let $C_{n'p'} = \{1 - (p')^{n'+1} - (q')^{n'+1}\} \{(n' + 1)p'q'\}^{-1}$.

The following theorem is the main result in this paper.

Theorem 1.1. *For each $i \in \Gamma$, let the random indicators $X_i, X_j, J_{ji}, j \in \Gamma_i$ be defined on the same probability space with*

$$\mathcal{L}(J_{ji}; j \in \Gamma_i) = \mathcal{L}(X_j; j \in \Gamma_i | X_i = 1).$$

Let $W_i = \sum_{j \neq i} X_j, W_{ij} = \sum_{l \neq i, j} X_l, V_i = \sum_{j \neq i} J_{ji}, V_{ij}^i = \sum_{l \neq i, j} J_{li}$.

(a) Then

$$\begin{aligned} d_{TV}(\mathcal{L}(W), B(n', p')) &\leq C_{n'p'} \left\{ 2(\lambda_3 - \lambda_2^2/\lambda) + \lambda|p' - p^*| + \sum_i p_i E|W_i - V_i| \right. \\ &\quad \left. + 2\lambda^{-1} \sum_{i,j} |p_i - p_j| p_i p_j E|V_{ij}^i - W_{ij}| \right. \\ &\quad \left. + \lambda^{-1} \sum_{i,j} (p_i - p_j)^2 |\text{Cov}(X_i, X_j)| \right\}. \end{aligned}$$

(b) *If there exists a partition $\Gamma_i = \Gamma_i^+ \cup \Gamma_i^- \cup \Gamma_i^0$ with $J_{ji} \geq X_j$ for $j \in \Gamma_i^+$ and $J_{ji} \leq X_j$ for $j \in \Gamma_i^-$, then for $n \geq 2$,*

$$\begin{aligned} &d_{TV}(\mathcal{L}(W), B(n', p')) \\ &\leq C_{n'p'} \left\{ 2(\lambda_3 - \lambda_2^2/\lambda) + \lambda|p' - p^*| + (1 + 3r) \left(\sum_i \sum_{j \in \Gamma_i^+ \cup \Gamma_i^-} |\text{Cov}(X_i, X_j)| \right. \right. \\ &\quad \left. \left. + \sum_i \sum_{j \in \Gamma_i^0} (E(X_i X_j) + p_i p_j) \right) \right\}. \end{aligned}$$

The possibility of proving Theorem 1.1 was suggested by Barbour, Holst and Janson (1992, p.192). The proof of Theorem 1.1 is given in Section 2. It is similar to the proofs of the Poisson approximation theorems in Barbour, Holst and Janson (1992, pp.21-29).

The following corollaries are immediate consequences of Theorem 1.1.

Corollary 1.1. *Suppose that the random indicators $(X_j; j \in \Gamma)$ are negatively related, that is, $\Gamma_i = \Gamma_i^-$. Then, for $n \geq 2$,*

$$\begin{aligned} &d_{TV}(\mathcal{L}(W), B(n', p')) \\ &\leq C_{n'p'} \left\{ 2(\lambda_3 - \lambda_2^2/\lambda) + \lambda|p' - p^*| - (1 + 3r) \sum_i \sum_{j \neq i} \text{Cov}(X_i, X_j) \right\}. \end{aligned}$$

Corollary 1.2. *Suppose that the random indicators $(X_j; j \in \Gamma)$ are positively related, that is, $\Gamma_i = \Gamma_i^+$. Then, for $n \geq 2$,*

$$d_{TV}(\mathcal{L}(W), B(n', p')) \leq C_{n'p'} \left\{ 2(\lambda_3 - \lambda_2^2/\lambda) + \lambda|p' - p^*| + (1 + 3r) \sum_i \sum_{j \neq i} \text{Cov}(X_i, X_j) \right\}.$$

Since independent indicators are both positively and negatively related, Corollaries 1.1 and 1.2 yield Corollary 1.3.

Corollary 1.3. *If the random indicators are independent, then*

$$d_{TV}(\mathcal{L}(W), B(n', p')) \leq C_{n'p'} \{ 2(\lambda_3 - \lambda_2^2/\lambda) + \lambda|p' - p^*| \}.$$

Remarks for Corollary 1.3.

1. The error bound has a smaller absolute constant than that obtained in Theorem 9.E. of Barbour, Holst and Janson (1992, p.190).

2. Note that $\lambda|p' - p^*| = p'|\lambda - n'p^*|$. Thus, if λ^2/λ_2 is an integer, then $n'p^* = \lambda$ and the second term in the braces vanishes.

3. Note that $\lambda|p' - p^*| \leq \lambda p^*/(2n') = \lambda_2/(2n')$ and is of order $o(\lambda_2)$ as $n' \rightarrow \infty$. In addition, $\lambda_3 \leq p_{\max}\lambda_2 = o(\lambda_2)$ if $p_{\max} = o(1)$ as $n' \rightarrow \infty$. Therefore, for the case $p_{\max} = o(1)$ as $n' \rightarrow \infty$, the bound in Corollary 1.3 is of a smaller order than that of the bound on $d_{TV}(\mathcal{L}(W), Po(\lambda))$ obtained by Barbour, Holst and Janson (1992). This has also been observed by Barbour, Holst and Janson (1992, p.190).

4. Applying Corollary 1.3 to the example mentioned earlier in this section, since $p_1 = \dots = p_{[n/2]} = 0$ and $p_{[n/2]+1} = \dots = p_n = 1/2$, we have $d_{TV}(\mathcal{L}(W), B(n', p')) = 0$, where $n' = [(n + 1)/2]$ and $p' = 1/2$. That means, $W \sim B(n', p')$ which is the case.

From Corollary 1.3, we have Corollary 1.4.

Corollary 1.4. *If the random indicators are independent, then*

$$P(W \geq n' + 1) \leq C_{n'p'} \{ 2(\lambda_3 - \lambda_2^2/\lambda) + p'|\lambda - n'p^*| \}.$$

If $p_i = p$, for all $i = 1, \dots, n$, then $\lambda_3 - (\lambda_2^2/\lambda) = 0$, $n' = n$, $\lambda = n'p^*$ and $p' = p$. Thus, Theorem 1.1 yields the following corollary.

Corollary 1.5. *Suppose that the random indicators $(X_j; j \in \Gamma)$ are identical and they are either positively or negatively related, that is, $\Gamma_i = \Gamma_i^+ \cup \Gamma_i^-$. Then, for $n \geq 2$,*

$$d_{TV}(\mathcal{L}(W), B(n', p')) = d_{TV}(\mathcal{L}(W), B(n, p)) \leq C_{np} \sum_i \sum_{j \neq i} |\text{Cov}(X_i, X_j)|.$$

Remarks for Corollary 1.5.

Using the result obtained by Barbour, Holst and Janson (1992) on the Poisson approximation for sums of indicators, we can also obtain a bound for the total variation distance $d_{TV}(\mathcal{L}(W), B(n, p))$ through the Poisson approximation to $\mathcal{L}(W)$ and $B(n, p)$. In particular, when the random indicators X_i are all identical and they are either positively or negatively related, we have

$$\begin{aligned} d_{TV}(\mathcal{L}(W), B(n, p)) &\leq d_{TV}(\mathcal{L}(W), Po(\lambda)) + d_{TV}(Po(\lambda), B(n, p)) \\ &\leq \lambda^{-1}(1 - e^{-\lambda}) \left\{ 2\lambda p + \sum_i \sum_{j \neq i} |\text{Cov}(X_i, X_j)| \right\}. \end{aligned}$$

In this way, an extra term $2\lambda p$ is produced compared to the bound in Corollary 1.5. It is not a good bound when p is large.

2. Proof of Theorem 1.1

Let $Z \sim B(n', p')$, where $n' = \lceil (\lambda^2/\lambda_2) + 1/2 \rceil$ and $p' = \lambda/n'$. Let A be a subset of integers and $A^* = A \cap \{0, 1, \dots, n'\}$. We have

$$\begin{aligned} d_{TV}(\mathcal{L}(W), \mathcal{L}(Z)) &= \sup_A (P(Z \in A) - P(W \in A)) \\ &= \sup_{A^*} (P(Z \in A^*) - P(W \in A^*)). \end{aligned}$$

The last equality is due to $P(Z > n') = 0$. In light of the above remark, we just need to bound $|E(I(W \in A^*) - P(Z \in A^*))|$. By following Barbour, Holst and Janson (1992, p.188), let $f = f_{A^*} : Z^+ \cup \{0, -1\} \rightarrow R$ satisfy the following equations

$$\begin{aligned} p'(n' - x)f(x) - q'xf(x - 1) &= I(x \in A^*) - P(Z \in A^*), \quad \text{for } 0 \leq x \leq n', \\ f(-1) &= f(0), \quad f(x) = f(n'), \quad \text{for } x \geq n'. \end{aligned}$$

Letting $\Delta f(x) = f(x) - f(x - 1)$, Ehm (1991) proved that

$$\|\Delta f\|_\infty \leq C_{n'p'}. \tag{1}$$

We shall obtain an upper bound for $|P(W \in A^*) - P(Z \in A^*)|$ involving Δf . For simplicity, let $W_i = \sum_{j \neq i} X_j$ and $W_{ij} = \sum_{l \neq i, j} X_l$. Now,

$$\begin{aligned} E(I(W \in A^*) - P(Z \in A^*)) &= E(p'(n' - W)f(W) - q'Wf(W - 1)) \\ &= E(\lambda f(W) - Wf(W - 1) - p'W\Delta f(W)). \end{aligned} \tag{2}$$

We have $p'E(W\Delta f(W)) = \sum_{i=1}^n p'p_i E(\Delta f(W_i + 1)|X_i = 1)$ and

$$\begin{aligned} &E(\lambda f(W) - Wf(W - 1)) \\ &= \sum_{i=1}^n p_i E(f(W)) - \sum_{i=1}^n E(X_i f(W - 1)) \\ &= \sum_{i=1}^n p_i^2 E(\Delta f(W_i + 1)|X_i = 1) + \sum_{i=1}^n p_i q_i (E(f(W_i)|X_i = 0) - E(f(W_i)|X_i = 1)). \end{aligned}$$

Thus, the right hand side of (2) can be rewritten as

$$\sum_{i=1}^n p_i(p_i - p^*)E(\Delta f(W_i + 1)|X_i = 1) + \sum_{i=1}^n p_i(p^* - p')E(\Delta f(W_i + 1)|X_i = 1) + \sum_{i=1}^n p_i q_i (E(f(W_i)|X_i = 0) - E(f(W_i)|X_i = 1)). \tag{3}$$

Therefore, to obtain a bound on (2), we need to bound the terms in (3).

Let us first look at the last term of (3). Since

$$q_i(P(W_i = k|X_i = 0) - P(W_i = k|X_i = 1)) = P(W_i = k) - P(W_i = k|X_i = 1),$$

the last term of (3) is equal to

$$\sum_{i=1}^n p_i \{E(f(W_i)) - E(f(W_i)|X_i = 1)\}. \tag{4}$$

Following Barbour, Holst and Janson (1992, pp.23-24), let

$$\mathcal{L}(J_{ji}; j \in \Gamma) = \mathcal{L}(X_j; j \in \Gamma|X_i = 1),$$

where $\Gamma = \{1, \dots, n\}$, and set $V_i = \sum_{j \neq i} J_{ji}$. Then, the expression (4) and also the last term of (3) is bounded by

$$\|\Delta f\|_\infty \sum_{i=1}^n p_i E|W_i - V_i|. \tag{5}$$

Next, the second term of (3) is bounded by $\|\Delta f\|_\infty \lambda |p' - p^*|$ which is equal to

$$\|\Delta f\|_\infty p' |\lambda - n' p^*|. \tag{6}$$

Finally, let us look at the first term of (3).

$$\sum_{i=1}^n p_i(p_i - p^*)E(\Delta f(W_i + 1)|X_i = 1) = \sum_{k=1}^n \Delta f(k) \sum_{i=1}^n (p_i - p^*)P(X_i = 1, W = k). \tag{7}$$

Thus, to get a good bound on the first term of (3), we express the sum

$$\sum_{i=1}^n (p_i - p^*)P(X_i = 1, W = k)$$

in terms of the positive terms $(p_i - p_j)^2$. In fact,

$$\begin{aligned} & \sum_{i=1}^n (p_i - p^*)P(X_i = 1, W = k) \\ &= (2\lambda)^{-1} \left\{ \sum_{i,j} (p_i - p_j)^2 p_i P(W_{ij} = k - 1) - \sum_{i,j} (p_i - p_j)^2 P(X_i = 1, W = k) \right. \\ & \quad \left. + \sum_{i,j} (p_i - p_j) p_i \text{Cov}(X_i - X_j, I_{(W_{ij}=k-1)}) \right\}. \tag{8} \end{aligned}$$

This is obtained by applying the fact that

$$\begin{aligned} & P(X_i = 1, W = k) - P(X_j = 1, W = k) \\ &= (p_i - p_j)P(W_{ij} = k - 1) + \text{Cov}(X_i - X_j, I_{(W_{ij}=k-1)}), \end{aligned}$$

where $W_{ij} = \sum_{l \neq i,j} X_l$. Next, express the sums on the right hand side of (8) in terms of $(p_i - p_j)p_i p_j$. For simplicity, let $t_{ij} = P(X_i = 1, X_j = 0, W_{ij} = k - 1)$. Since

$$\begin{aligned} & p_i P(W_{ij} = k - 1) \\ &= P(X_i = 1, W_{ij} = k - 1) - \text{Cov}(X_i, I_{(W_{ij}=k-1)}) \\ &= t_{ij} + (p_i p_j + \text{Cov}(X_i, X_j))P(W = k + 1 | X_i = 1 = X_j) - \text{Cov}(X_i, I_{(W_{ij}=k-1)}) \end{aligned}$$

and

$$\begin{aligned} P(X_i = 1, W = k) &= P(X_i = 1, X_j = 0, W = k) + P(X_i = 1, X_j = 1, W = k) \\ &= t_{ij} + (p_i p_j + \text{Cov}(X_i, X_j))P(W = k | X_i = 1 = X_j), \end{aligned}$$

the right hand side of (8) is equal to

$$\begin{aligned} (2\lambda)^{-1} \left\{ \sum_{i,j} (p_i - p_j)^2 p_i p_j P_{ij}(k) + \sum_{i,j} (p_i - p_j)^2 \text{Cov}(X_i, X_j) P_{ij}(k) \right. \\ \left. + 2 \sum_{i,j} (p_i - p_j) p_j \text{Cov}(X_i, I_{(W_{ij}=k-1)}) \right\}, \end{aligned} \tag{9}$$

where $P_{ij}(k) = P(W = k + 1 | X_i = 1 = X_j) - P(W = k | X_i = 1 = X_j)$. Thus, combining (7), (8) and (9), the first term of (3) is bounded by

$$\begin{aligned} (2\lambda)^{-1} \left\{ \left| \sum_k \Delta f(k) \sum_{i,j} (p_i - p_j)^2 p_i p_j P_{ij}(k) \right| \right. \\ \left. + \left| \sum_k \Delta f(k) \sum_{i,j} (p_i - p_j)^2 \text{Cov}(X_i, X_j) P_{ij}(k) \right| \right. \\ \left. + 2 \left| \sum_k \Delta f(k) \sum_{i,j} (p_i - p_j) p_j \text{Cov}(X_i, I_{(W_{ij}=k-1)}) \right| \right\}. \end{aligned} \tag{10}$$

Observe that the first term on the right hand side of (10) is bounded by

$$2 \|\Delta f\|_\infty \sum_{i,j} (p_i - p_j)^2 p_i p_j = 4 \|\Delta f\|_\infty \left(\sum_i p_i^3 \lambda - \left(\sum_i p_i^2 \right)^2 \right),$$

the second term of (10) is bounded by

$$2 \|\Delta f\|_\infty \sum_{i,j} (p_i - p_j)^2 |\text{Cov}(X_i, X_j)|,$$

and the last term of (10) is equal to

$$2 \left| \sum_{i,j} (p_i - p_j) p_i p_j (E(\Delta f(V_{ij}^i + 1)) - E(\Delta f(W_{ij} + 1))) \right|,$$

where $V_{ij}^i = \sum_{l \neq i,j} J_{li}$, which is bounded by $4 \|\Delta f\|_\infty \sum_{i,j} |p_i - p_j| p_i p_j E|V_{ij}^i - W_{ij}|$. Thus, we conclude that (10) is bounded by

$$2 \|\Delta f\|_\infty \left(\sum_i p_i^3 - \lambda^{-1} \left(\sum_i p_i^2 \right)^2 \right) + \lambda^{-1} \|\Delta f\|_\infty \sum_{i,j} (p_i - p_j)^2 |\text{Cov}(X_i, X_j)| + 2\lambda^{-1} \|\Delta f\|_\infty \sum_{i,j} |p_i - p_j| p_i p_j E|V_{ij}^i - W_{ij}|. \tag{11}$$

Combining (2), (3), (5), (6), (10) and (11), we have

$$\begin{aligned} & |E(I(W \in A^*) - P(Z \in A^*))| \\ & \leq \|\Delta f\|_\infty \left(2 \left(\sum_i p_i^3 - \lambda^{-1} \left(\sum_i p_i^2 \right)^2 \right) + p' |\lambda - n' p^*| \right) \\ & \quad + \|\Delta f\|_\infty \lambda^{-1} \sum_{i,j} (p_i - p_j)^2 |\text{Cov}(X_i, X_j)| \\ & \quad + \|\Delta f\|_\infty \sum_{i=1}^n p_i E|W_i - V_i| + 2(\lambda)^{-1} \sum_{i,j} |p_i - p_j| p_i p_j E|V_{ij}^i - W_{ij}|. \end{aligned}$$

Utilizing (1), (a) of Theorem 1.1 is proved.

To prove (b) of Theorem 1.1, observe that for $j \in \Gamma_i^+$:

$$0 \leq p_i E|X_j - J_{ji}| = p_i E(J_{ji} - X_j) = E(X_i X_j) - p_i p_j = \text{Cov}(X_i, X_j).$$

Similarly for $j \in \Gamma_i^-$, $0 \leq p_i E|X_j - J_{ji}| = -\text{Cov}(X_i, X_j)$. Thus,

$$\begin{aligned} p_i E|W_i - V_i| &= p_i E \left| \sum_{j \in \Gamma_i^+} (X_j - J_{ji}) + \sum_{j \in \Gamma_i^-} (X_j - J_{ji}) + \sum_{j \in \Gamma_i^0} (X_j - J_{ji}) \right| \\ &\leq \sum_{j \in \Gamma_i^+ \cup \Gamma_i^-} |\text{Cov}(X_i, X_j)| + \sum_{j \in \Gamma_i^0} (p_i p_j + E(X_i X_j)). \end{aligned}$$

Similarly,

$$p_i E|V_{ij}^i - W_{ij}| \leq \sum_{l \in \Gamma_i^+ \cup \Gamma_i^-, l \neq j} |\text{Cov}(X_i, X_l)| + \sum_{l \in \Gamma_i^0, l \neq j} (p_i p_l + E(X_i X_l)).$$

In addition, we note that

$$\begin{aligned} & \lambda^{-1} \sum_{i,j} |p_i - p_j| p_j \sum_{l \in \Gamma_i^+ \cup \Gamma_i^-, l \neq j} |\text{Cov}(X_l, X_i)| \\ &= \lambda^{-1} \sum_j p_j \sum_i |p_i - p_j| \sum_{l \in \Gamma_i^+ \cup \Gamma_i^-, l \neq j} |\text{Cov}(X_l, X_i)| \leq r \sum_i \sum_{l \in \Gamma_i^+ \cup \Gamma_i^-} |\text{Cov}(X_l, X_i)|, \end{aligned}$$

where $r = p_{\max} - p_{\min}$. Thus, Theorem 1.1(b) follows from Theorem 1.1(a).

3. Some Applications

When applying Theorem 1.1 in a specific situation it is necessary either to construct an efficient coupling defining the partition of Γ_i and the J 's or to know the existence of a suitable one. Consider the following examples.

Example 1. (Hypergeometric distribution)

Suppose a random sample of size n is taken from a finite population containing N elements of two different types, of which m are of type A and $N - m (\neq 0)$ are of type B. Let the random variable W be the number of type A elements in the sample. Clearly W has the hypergeometric distribution $H(N, n, m)$ where, for $\max(0, n + m - N) \leq k \leq \min(n, m)$,

$$P(W = k) = \binom{m}{k} \binom{N - m}{n - k} \binom{N}{n}^{-1}.$$

Barbour, Holst and Janson (1992, p.113) remarked that the distribution of W can be approximated by a binomial distribution via Stein's method together with the coupling technique. In this example, we approximate $\mathcal{L}(W)$ by two binomial distributions using different representations of W . Firstly, W can be represented as $W = \sum_{i=1}^n X_i$ where $X_i = 1$ if the i th element in the sample is of type A and $X_i = 0$ otherwise. We then have $P(X_i = 1) = mN^{-1} = p$, $\lambda = E(W) = nmN^{-1}$, $E(X_i X_j) = P(X_i = 1)P(X_j = 1 | X_i = 1) = mN^{-1}(m - 1)(N - 1)^{-1}$. Consider the following coupling. If $X_k = 1$, set $J_{ik} = X_i$. If $X_k = 0$, interchange the k th element in the sample with a randomly chosen element of type A and then, for $i \neq k$, set $J_{ik} = 1$ if the i th element in the sample is of type A, $J_{ik} = 0$ otherwise. By construction, this coupling satisfies the hypothesis of Corollary 1.5. We have

$$\begin{aligned} & d_{TV}(\mathcal{L}(W), B(n, p)) \\ & \leq C_{np}n(n - 1)p(mN^{-1} - (m - 1)(N - 1)^{-1}) \\ & < q^{-1}(1 - p^{n+1} - q^{n+1})(N - m)(n - 1)(N(N - 1))^{-1} \\ & < (n - 1)(N - 1)^{-1} = o(1), \text{ if } n = o(N). \end{aligned}$$

On the other hand, W can be represented as $W = \sum_{i=1}^m Y_i$ where $Y_i = 1$ if the i th element of type A is in the sample and $Y_i = 0$ otherwise. We have

$$P(Y_i = 1) = nN^{-1} = p', \quad \lambda = E(W) = (nm)N^{-1}, \quad E(Y_i Y_j) = n(n - 1)(N(N - 1))^{-1}.$$

Consider the following coupling. If $Y_k = 1$, set $J_{ik} = Y_i$. If $Y_k = 0$, interchange the k th element of type A with a randomly chosen element in the sample and

then, for $i \neq k$, set $J_{ik} = 1$ if the i th element of type A is in the sample, $J_{ik} = 0$ otherwise. As before, this coupling satisfies the hypothesis of Corollary 1.5. Therefore, we have

$$\begin{aligned} & d_{TV}(\mathcal{L}(W), B(m, p')) \\ & \leq C_{mp'} m(m-1) p' (nN^{-1} - (n-1)(N-1)^{-1}) \\ & < (q')^{-1} (1 - (p')^{m+1} - (q')^{m+1}) (N-n)(N-1)^{-1} (m-1) N^{-1} \\ & < (m-1)(N-1)^{-1} = o(1), \quad \text{if } m = o(N). \end{aligned}$$

Hence, the hypergeometric distribution $H(N, n, m)$ is well approximated by the binomial distributions $B(n, m/N)$ or $B(m, n/N)$ depending on whether $n = o(N)$ or $m = o(N)$ respectively.

Example 2. (Random graphs problem)

Consider a complete graph K_n with n vertices and $n!/[(n-2)!2!]$ edges. Delete each edge with probability $1-p$ ($=q$) independently of the other edges. Thus, we get the random graph $K_{n,p}$. Let W be the number of vertices i whose degree D_i in $K_{n,p}$ is greater than or equal to m .

Following Barbour, Holst and Janson (1992, p. 97), let $X_i = I(D_i \geq m)$ and $W = \sum_{i=1}^n X_i$. Since $\{X_i, 1 \leq i \leq n\}$ are increasing functions of the independent edge indicators, they are positively related. Thus, by Corollary 1.5,

$$\begin{aligned} d_{TV}(\mathcal{L}(W), B(n, \pi_1)) & \leq C_{n\pi_1} \sum_i \sum_{j \neq i} \text{Cov}(X_i, X_j) \\ & \leq \sum_i \sum_{j \neq i} \text{Cov}(X_i, X_j) = n(n-1) p q p_{m-1}^2. \end{aligned}$$

Here, $\pi_1 = \pi_1(m) = P(Z_{n-1} \geq m)$, $p_{m-1} = P(Z_{n-2} = m-1)$ where $Z_l \sim B(l, p)$, $l = n-2$ or $n-1$. (Refer to Barbour, Holst and Janson (1992, pp. 97-98) for the calculation of $\text{Cov}(X_i, X_j)$).

Note that if $m = 1$, $d_{TV}(\mathcal{L}(W), B(n, \pi_1)) \leq np(n-1)q^{2n-3} < n^2 p e^{-2np}$. For any $\epsilon > 0$, if $p = p(n)$ is such that $np > (1 + \epsilon) \ln n$ for sufficiently large n , then $n^2 p e^{-2np} = o(1/n^\epsilon)$ as $n \rightarrow \infty$. In this case, we have a binomial approximation but since $\text{Var}(W) \sim 2E(W)$, the Poisson approximation cannot be good. (Barbour, Holst and Janson (1992, p.98).)

Example 3. (The classical occupancy problem)

Let n balls be thrown independently of each other into r boxes, with probability $1/r$ falling into the k th box. The random variable W is equal to the number of empty boxes and can be represented as $W = \sum_{i=1}^r X_i$ where $X_k = 1$ if the k th box is empty and $X_k = 0$ otherwise. We then have $P(X_k = 1) = (1 - 1/r)^n = p$, $\lambda = E(W) = r(1 - 1/r)^n$, $E(X_i X_j) = P(X_i = 1 = X_j) = (1 - 2/r)^n$. Following Barbour, Holst and Janson (1992, p.123), we introduce the following

coupling. Throw each of the balls which have fallen into the k th box independently into one of the other boxes, in such a way that the probability of a ball falling into box i ($i \neq k$) is $1/(n-1)$. Then let $J_{ik} = 1$ if box i is empty, $J_{ik} = 0$ otherwise, and $X_{ik} = X_i$. Evidently, $J_{ik} \leq X_{ik}$ for $i \neq k$, and for each k , $\mathcal{L}(J_{1k}, \dots, J_{nk}) = \mathcal{L}(X_1, \dots, X_n | X_k = 1)$. By Corollary 1.5., we have

$$\begin{aligned} d_{TV}(\mathcal{L}(W), B(n', p')) &= d_{TV}(\mathcal{L}(W), B(r, p)) \\ &\leq C_{rp} \sum_i \sum_{j \neq i} |\text{Cov}(X_i, X_j)| \leq C_{rp} \sum_i \sum_{j \neq i} ((1 - 1/r)^{2n} - (1 - 2/r)^n) \\ &\leq np(1 - q^{r+1})((r+1)q)^{-1}. \end{aligned}$$

If $n = ra_r$, then

$$d_{TV}(\mathcal{L}(W), B(r, p)) \leq (r/(r+1))(a_r/(e^{a_r} - 1))(1 - (1 - e^{-a_r})^{r+1})$$

giving a binomial approximation throughout the range $a_r \rightarrow \infty$.

Acknowledgement

This paper is based on a part of my Ph.D. dissertation. I am grateful to my advisor, Professor Louis H. Y. Chen, for his inspiring guidance and invaluable advice.

References

- Arratia, R., Goldstein, L. and Gordon, L. (1990). Poisson approximation and the Chen-Stein method (with discussion). *Statist. Sci.* **5**, 403-424.
- Barbour, A. D. (1988). Stein's method and Poisson process convergence. *J. Appl. Probab.* Special Volume **25A**, 175-184.
- Barbour, A. D. (1990). Stein's method for diffusion approximations. *Probab. Theory Related Fields* **84**, 297-322.
- Barbour, A. D. and Brown, T. C. (1992). Stein's method and point process approximation. *Stochastic Processes and Their Applications* **43**, 9-31.
- Barbour, A. D., Chen, L. H. Y. and Loh, W.-L. (1992). Compound Poisson approximation for nonnegative random variables via Stein's method. *Ann. Probab.* **20**, 1843-1866.
- Barbour, A. D., Holst, L. and Janson, S. (1992). *Poisson Approximation*. *Oxford Studies in Probability* **2**. Clarendon Press, Oxford.
- Chen, L. H. Y. (1974). On the convergence of Poisson binomial to Poisson distributions. *Ann. Probab.* **2**, 178-180.
- Chen, L. H. Y. (1975). Poisson approximation for dependent trials. *Ann. Probab.* **3**, 534-545.
- Ehm, W. (1991). Binomial approximation to the Poisson binomial distribution. *Statist. & Probab. Lett.* **11**, 7-16.
- Götze, F. (1991). On the rate of convergence in the multivariate CLT. *Ann. Probab.* **19**, 724-739.
- Loh, W.-L. (1992). Stein's method and multinomial approximation. *Ann. Appl. Probab.* **2**, 536-554.

- Stein, C. M. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **2**, 583-602. Univ. California Press, Berkeley, CA.
- Stein, C. M. (1986). *Approximate Computation of Expectations*. IMS, Hayward, CA.

Division of Mathematics, National Institute of Education, Nanyang Technological University, 469, Bukit Timah Road, S'1025, Singapore.

(Received December 1994; accepted July 1995)