Adv. Appl. Prob. **34,** 609–625 (2002)
Printed in Northern Ireland

POISSON APPROXIMATION FOR A SUM OF DEPENDENT INDICATORS: AN ALTERNATIVE APPROACH

N. PAPADATOS * ** AND
V. PAPATHANASIOU,* *University of Athens*

Abstract

The random variables $X_1, X_2, \dots, X_n$ are said to be totally negatively dependent (TND) if and only if the random variables X_i and $\sum_{j\neq i} X_j$ are negatively quadrant dependent for all i. Our main result provides, for TND 0–1 indicators $X_1, X_2, \dots, X_n$ with $\mathrm{P}[X_i = 1] = p_i = 1-\mathrm{P}[X_i = 0]$, an upper bound for the total variation distance between $\sum_{i=1}^{n} X_i$ and a Poisson random variable with mean $\lambda \geq \sum_{i=1}^{n} p_i$. An application to a generalized birthday problem is considered and, moreover, some related results concerning the existence of monotone couplings are discussed.

Keywords: Totally negatively dependent indicator; Chen–Stein equation; negatively related indicator; Poisson approximation; total variation distance; birthday problem; monotone coupling

AMS 2000 Subject Classification: Primary 60F05

1. Introduction and a brief history of the Chen–Stein method

One of the most useful topics in applied probability is concerned with the Poisson approximation of a sum of many dependent 0–1 indicators with small means—the generalized Poisson law of small numbers. Perhaps the most accurate results in that direction are obtained by the well-known Chen–Stein method, a method that was firstly obtained by Stein (1972) for the normal approximation of sums of dependent random variables, and later extended to the Poisson approximation by Chen (1975). Since the Chen–Stein method was found to be extremely fruitful, it has become one of the most developed and effective approaches for Poisson approximation in the last 25 years, so that a complete review of related papers seems to be rather difficult. A collection of most results of this kind is contained in Barbour *et al.* (1992). Our main interest, however, focuses on the following classical result, which presents an upper bound for the total variation distance in the very important situation where the indicators $X_1, X_2, \dots, X_n$ are negatively related (NR). For completeness, we firstly give the definition of NR indicators (see Barbour *et al.* (1992, Definition 2.1.1)).

Definition 1.1. Indicator random variables $X_1, X_2, \dots, X_n$ are said to be negatively related (NR) if, for each $i \in \{1, 2, \dots, n\}$, there exist random variables $Y_1, \dots, Y_{i-1}, Y_{i+1}, \dots, Y_n$ (depending on i) such that

(a) $(Y_1, \dots, Y_{i-1}, Y_{i+1}, \dots, Y_n)^\top \stackrel{\mathrm{D}}{=} (X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n)^\top$ given that $X_i = 1$

Received 23 November 2001; revision received 20 June 2002.
* Postal address: Section of Statistics and Operational Research, Department of Mathematics, University of Athens, Panepistemiopolis, 157 84 Athens, Greece.
** Email address: npapadat@cc.uoa.gr
Research partially supported by the research foundation of the University of Athens.

and

(b) with probability 1, $Y_1 \le X_1, \dots, Y_{i-1} \le X_{i-1}, Y_{i+1} \le X_{i+1}, \dots, Y_n \le X_n$.

In the dual case where (b) holds with the inequalities reversed for all i, we say that the indicators $X_1, X_2, \dots, X_n$ are positively related. Since this notion, however, is not related to the results of the present article, it will not be used in the sequel.

The classical Chen–Stein bound for NR indicators is presented in the following theorem (see Barbour *et al.* (1992, Corollary 2.C.2)).

Theorem 1.1. *If the indicators $X_1, X_2, \dots, X_n$ are NR with* $\mathrm{P}[X_i = 1] = p_i = 1 - \mathrm{P}[X_i = 0]$ *and $W = \sum_{i=1}^n X_i$, then*

$$d_{\mathrm{TV}}(W, \mathcal{P}_\mu) \le (1 - \mathrm{e}^{-\mu})\left(1 - \frac{\sigma^2}{\mu}\right), \tag{1.1}$$

where $\mu = \sum_{i=1}^n p_i = \mathrm{E}[W]$, $\sigma^2 = \mathrm{var}[W]$ and $\mathcal{P}_\mu$ denotes a Poisson random variable with mean μ.

We note that here and elsewhere in this article,

$$d_{\mathrm{TV}}(X, Y) = \max\{|\mathrm{P}[X \in A] - \mathrm{P}[Y \in A]|;\ A \subseteq \{0, 1, \dots\}\} \tag{1.2}$$

denotes the total variation distance between the nonnegative integer-valued random variables X and Y, and $\mathcal{P}_\theta$ denotes a Poisson random variable with mean $\theta \ge 0$.

The first step for proving (1.1) hinges on the derivation of a solution $g = g_{\lambda,A}$ for the Chen–Stein difference equation

$$\lambda g(x+1) - x g(x) = \mathbf{1}(x \in A) - \mathrm{P}[\mathcal{P}_\lambda \in A], \qquad x = 0, 1, \dots, \tag{1.3}$$

namely,

$$g_{\lambda,A}(x) = \frac{(x-1)!}{\lambda^x} \sum_{k=0}^{x-1} [\mathbf{1}(k \in A) - \mathrm{P}[\mathcal{P}_\lambda \in A]] \frac{\lambda^k}{k!}, \qquad x = 1, 2, \dots$$

(the value of $g_{\lambda,A}(0)$ is arbitrary and it can be taken to be 0). For this solution, Barbour and Eagleson (1983) obtained the (correct order) bounds

$$\sup_{x,A} |g_{\lambda,A}(x)| \le \min\left\{1, \left(\frac{2}{\mathrm{e}}\right)^{1/2} \lambda^{-1/2}\right\}, \qquad \sup_{x,A} |\Delta g_{\lambda,A}(x)| \le \lambda^{-1}(1 - \mathrm{e}^{-\lambda}), \tag{1.4}$$

where the suprema are taken over $x \in \{0, 1, \dots\}$ and $A \subseteq \{0, 1, \dots\}$, and $\Delta g(x) = g(x+1) - g(x)$ denotes the forward difference of any function $g : \{0, 1, \dots\} \to \mathbb{R}$ (for the first estimate see Barbour *et al.* (1992, Remark 1.1.2)). Since, by (1.3) with $\lambda = \mu$,

$$|\mathrm{P}[W \in A] - \mathrm{P}[\mathcal{P}_\mu \in A]| = |\mathrm{E}[\mu g_{\mu,A}(W+1) - W g_{\mu,A}(W)]|,$$

the remainder of the proof is based on a suitable coupling $\tilde{W} = \sum_{i=1}^n \tilde{X}_i$ of $W = \sum_{i=1}^n X_i$, making X_i and $\tilde{X}_i$ as close as possible for all i. It turns out that such useful couplings exist when the X_i are NR; see Definition 1.1. This is the coupling approach, which was in fact introduced by Serfling (1975); see also Chen (1998). The above coupling approach is used extensively by

Barbour *et al.* (1992), while the local approach to the Chen–Stein method is widely discussed by Arratia *et al.* (1989), (1990).

Our main interest is in extending (1.1) to a wider (than NR) class of indicators. It should be noted that the approach given here is closely related to the Chen–Stein method, because it is based on the solution of the Chen–Stein equation (1.3). Nevertheless, our derivation suggests an alternative coupling method that may be of some interest in other situations as well.

It should also be noted that the connection of the present approach with some previous results (using the so-called w-functions) is strong. In particular, a similar application of the alternative (coupling) method to the normal approximation was given by Cacoullos *et al.* (1994) and extended by Papathanasiou and Utev (1995) to the Poisson case. Neither approach is complete, however, for the following three reasons: (i) the w_X-function (corresponding to the nondegenerate random variable X with finite variance) can be defined only when the support of X is a finite or infinite interval (in the case of a nonnegative integer-valued X, this corresponds to the restriction that the set $\{x : \mathrm{P}[X = x] > 0\}$ is of the form $\{0, 1, \dots, b\}$ with $b \in \{1, 2, \dots\} \cup \{+\infty\}$), (ii) in order to compute the bounds we have to know the w_X-function, which in turn characterizes the distribution of X and (iii) although some results can be obtained when the summands are independent, it seems rather difficult to obtain similar results for the dependent case.

In order to overcome the difficulties of the first two restrictions when X admits a density, Papadatos and Papathanasiou (2001) defined a new random variable X^*, which is similar to the so-called zero-bias transformation independently defined by Goldstein and Reinert (1997). It turns out that this approach is fruitful not only for the classical local central limit theorem (CLT) with i.i.d. summands (Cacoullos *et al.* (2002)), but also for some dependent structures (simple random sampling; Goldstein and Reinert (1997)).

Furthermore, several generalizations of the alternative method to distributions other than the normal and the Poisson were obtained by Papadatos and Papathanasiou (1995), and a further application to the rate of convergence in the classical CLT (with respect to the total variation distance) was given by Cacoullos *et al.* (1997); moreover, the alternative method for the Poisson approximation was also applied by Majsnerowska (1998) to some specific distributions (such as binomial, negative binomial and hypergeometric), yielding accurate bounds, but it seems that the results depend heavily on the specific distributional assumptions. The above applications (using the w-functions) have the disadvantages described above. Therefore, the novelty of our results is in the unification of the alternative method for the Poisson approximation, and its application to some dependent situations; the bound (3.6) in combination with Theorem 3.1 below shows that the present approach may produce successful bounds.

The paper is organized as follows. In Section 2, we present a general unified upper bound, valid for any nonnegative integer-valued random variable X with finite variance. In fact, this kind of bound has already been studied by Papathanasiou and Utev (1995), Papadatos and Papathanasiou (1995) and Majsnerowska (1998) in some restricted cases (among them, the most general form is given by the last author); however, for our general approach, we need the unified expression given by Theorem 2.1 below, and therefore we provide a brief proof for completeness. Section 3 contains our main results. We introduce a notion of negative dependence (totally negative dependence (TND)) and we show (Theorem 3.1) that NR indicators are always TND, implying that the class of TND indicators is wider than the class of NR indicators. Moreover, under the TND assumption, we show that the bound (2.2) takes a very pleasant form (actually, it is the same as that of Theorem 1.1; see Theorem 3.2 below). In Section 4, we apply the results of Section 3 to a generalized birthday problem. Finally, in Section 5, we present some equivalent conditions for the existence of monotone, weak monotone and very weak monotone couplings of random vectors, implying a strong relationship between the TND property and some existing results.

2. A unified upper bound for Poisson approximation

Let X be a nonnegative integer-valued random variable with finite variance. It will be convenient to denote by $\mathcal{F}_0$ the class of all such random variables. Let also $\mathcal{F}_1 = \{X \in \mathcal{F}_0 : \mathrm{var}[X] > 0\}$ be the subclass of nondegenerate random variables in $\mathcal{F}_0$ and $\mathcal{F}_2 = \{X \in \mathcal{F}_1 : \mathrm{P}[X = x] > 0$ for some x implies that $\mathrm{P}[X = y] > 0$ for all $y = 0, \dots, x\}$ be the subclass of those random variables in $\mathcal{F}_1$ having integer interval supports containing the origin. For the above families of random variables, some useful functions can be defined as follows.

Definition 2.1. Assume that X lies in $\mathcal{F}_0$ and has mean $\mu = \mathrm{E}[X]$, variance $\sigma^2 = \mathrm{var}[X]$ and probability function $p_X(x) = \mathrm{P}[X = x]$, $x \in \{0, 1, \dots\}$.

(a) The function h_X is defined by

$$h_X(x) = \sum_{k=0}^{x} (\mu - k) p_X(k), \qquad x = 0, 1, \dots .$$

(b) If, furthermore, $X \in \mathcal{F}_1$, then the function p_{X^*} is defined by

$$p_{X^*}(x) = \frac{h_X(x)}{\sigma^2}, \qquad x = 0, 1, \dots,$$

with h_X as in (a).

(c) Finally, if X lies in $\mathcal{F}_2$, then the function w_X is defined by

$$w_X(x) = \frac{p_{X^*}(x)}{p_X(x)} \quad \text{for } x \text{ in the support of } X$$

(and it is undefined outside this integer interval support), where p_{X^*} is as in (b).

We note that the above definition (c) for the w-function in the discrete case has been studied by Cacoullos and Papathanasiou (1989), the definition of p_{X^*} in (b) is the discrete analogue of the zero-bias transformation X^* (cf. Goldstein and Reinert (1997), Papadatos and Papathanasiou (2001)) and that the definition of h_X in (a) is presented here only for technical reasons, in order to include the degenerate nonnegative integer-valued random variables.

Using the above terminology, the general Stein-type covariance identity can be restated as follows.

Lemma 2.1. (Cacoullos and Papathanasiou (1989).) *If X lies in $\mathcal{F}_0$, then, with the notation of Definition 2.1,*

$$\mathrm{cov}[X, g(X)] = \sum_{x=0}^{\infty} h_X(x) \Delta g(x), \tag{2.1}$$

for any function $g : \{0, 1, \dots\} \to \mathbb{R}$ for which the series is absolutely convergent.

By (2.1) it follows immediately that the nonnegative unimodal function $h_X(x)$ sums up to $\sigma^2 = \mathrm{var}[X]$, and thus p_{X^*} is a unimodal probability function, defining the discrete zero-bias transformation X^* of X (provided that $X \in \mathcal{F}_1$). Moreover, the covariance identity takes the equivalent forms:

$$\operatorname{cov}[X, g(X)] = \begin{cases} \sigma^2 \operatorname{E}[\Delta g(X^*)] & \text{for } X \in \mathcal{F}_1, \\ \sigma^2 \operatorname{E}[w_X(X)\Delta g(X)] & \text{for } X \in \mathcal{F}_2. \end{cases}$$

In the most interesting case where $X \stackrel{\mathrm{D}}{=} \mathcal{P}_\lambda$ (with $\mu = \sigma^2 = \lambda$), it follows that either $h_X \equiv 0$ (when $\lambda = 0$) or $X^* \stackrel{\mathrm{D}}{=} X$ and $w_X \equiv 1$ (when $\lambda > 0$), and in both cases (2.1) yields the Stein identity for the Poisson law, i.e. $\operatorname{E}[(X-\lambda)g(X)] = \operatorname{cov}[X, g(X)] = \lambda \operatorname{E}[\Delta g(X)]$; this particular identity (which also characterizes the Poisson law) entails the fundamental Chen–Stein equation (1.3).

The unified upper bound for the Poisson approximation is stated in the following theorem.

Theorem 2.1. (Cf. Papathanasiou and Utev (1995), Majsnerowska (1998).) *If $X \in \mathcal{F}_0$, then, under the notation of Definition 2.1,*

$$d_{\mathrm{TV}}(X, \mathcal{P}_\lambda) \le \lambda^{-1}(1 - \mathrm{e}^{-\lambda}) \sum_{x=0}^{\infty} |\lambda p_X(x) - h_X(x)| + \min\left\{1, \frac{(2/\mathrm{e})^{1/2}}{\lambda^{1/2}}\right\}|\mu - \lambda|, \qquad (2.2)$$

where $\lambda^{-1}(1-\mathrm{e}^{-\lambda})$ and $\min\{1, (2/\mathrm{e})^{1/2}\lambda^{-1/2}\}$ is taken to be 1 when $\lambda = 0$.

Proof. For $\lambda = 0$, the upper bound in (2.2) becomes $\mu + \sigma^2 \ge \mu$. In this case, however, $d_{\mathrm{TV}}(X, \mathcal{P}_0) = \mathrm{P}[X \ge 1] \le \mu$, proving the assertion. In any other case, the desired result follows on taking expectations with respect to X in (1.3) and using the identity (2.1) with $g = g_{\lambda,A}$ (a bounded function because of (1.4)) which, taking into account (1.4), yields the desired result (the arguments needed for a detailed proof can be found in Papadatos and Papathanasiou (1995, Theorem 3.1) or Majsnerowska (1998, Theorem 2)).

The above unified upper bound can be rewritten in a simpler form when X satisfies some additional assumptions. In particular, if $X \in \mathcal{F}_1$ and $\sigma^2 = \lambda$, then (2.2) takes the form

$$\begin{aligned} d_{\mathrm{TV}}(X, \mathcal{P}_{\sigma^2}) &\le (1 - \mathrm{e}^{-\sigma^2}) \sum_{x=0}^{\infty} |p_X(x) - p_{X^*}(x)| + \min\left\{1, \frac{(2/\mathrm{e})^{1/2}}{\sigma}\right\}|\mu - \sigma^2| \\ &= 2(1 - \mathrm{e}^{-\sigma^2}) d_{\mathrm{TV}}(X, X^*) + \min\left\{1, \frac{(2/\mathrm{e})^{1/2}}{\sigma}\right\}|\mu - \sigma^2|, \end{aligned}$$

showing that the total variation distance between X and its zero-bias transformation X^* provides an upper estimate for the total variation distance between X and $\mathcal{P}_{\sigma^2}$ (provided that μ and σ^2 are close to each other); a similar result holds for the normal approximation. On the other hand, if $X \in \mathcal{F}_2$, then (2.2) yields the known upper bound (Papadatos and Papathanasiou (1995), Majsnerowska (1998))

$$d_{\mathrm{TV}}(X, \mathcal{P}_\lambda) \le \lambda^{-1}(1 - \mathrm{e}^{-\lambda}) \operatorname{E}|\lambda - \sigma^2 w_X(X)| + \min\left\{1, \frac{(2/\mathrm{e})^{1/2}}{\lambda^{1/2}}\right\}|\mu - \lambda|.$$

Our experience with the above kind of bounds leads us to the following: although the estimates are often accurate, the bounds are difficult to use in most interesting situations (with the possible exception of the case where X is a sum of independent random variables; cf. Papathanasiou and Utev (1995)), because we have to evaluate the function h_X (equivalently, the distribution of X^* or the w_X-function), which usually happens to be intractable. Therefore, except in some special cases (where the distribution of X is completely known) the bound (2.2) is only of theoretical interest. The purpose of the next section (which contains our main results) is to carry the information contained in the bound (2.2) to a particular interesting case where the random variable X presents a sum of totally negatively dependent indicators (see Definition 3.1 below).

3. A bound for the sum of totally negatively dependent indicators

In the sequel we shall use the following notion.

Definition 3.1. Random variables $X_1, X_2, \dots, X_n$ $(n \geq 2)$ are said to be totally negatively dependent (TND) if, for all $i = 1, 2, \dots, n$, the random variables X_i and $X^{(i)} = \sum_{j \neq i} X_j$ are negatively quadrant dependent (NQD).

We see that the TND property is defined via the NQD property, introduced by Lehmann (1966). Specifically, the random variables X, Y are defined to be NQD if

$$\operatorname{cov}[f(X), g(Y)] \leq 0 \tag{3.1}$$

for all nondecreasing functions $f, g : \mathbb{R} \to \mathbb{R}$ for which the covariance is finite (note that it suffices to test (3.1) only for nondecreasing indicators $f, g : \mathbb{R} \to \{0, 1\}$). Obviously, the NQD and TND properties coincide when $n = 2$; on the other hand, independent random variables are TND for all $n \geq 2$ (thus, it will be convenient to define a single random variable X_1 as TND).

It should be noted that a notion similar to TND was introduced recently by Boutsikas and Koutras (2000). Specifically, they defined the random variables $X_1, X_2, \dots, X_n$ to be negatively cumulative dependent (NCD) if, for all $i \geq 2$, the random variables X_i and $S_{i-1} = \sum_{j<i} X_j$ are NQD. Although the definitions of TND and NCD do coincide for $n = 2$, it will be shown that for $n \geq 3$ there exist TND 0–1 indicators that are not NCD (see Remark 3.1 below), implying that the NCD class does not contain the TND class. It is more important, however, that NR indicators are always TND, showing that the class of TND indicators is wider than that of NR indicators.

Theorem 3.1. (a) *If the* 0–1 *indicators* $X_1, X_2, \dots, X_n$ *are NR then they are TND.*

(b) *For* $n \geq 3$, *there exist* n *TND* 0–1 *indicators that are not NR.*

(c) *For exchangeable* 0–1 *indicators, the TND and NR properties are equivalent.*

Proof. (a) We firstly show that X_1 and $X^{(1)} = X_2 + \cdots + X_n$ are NQD. To this end, it is enough to test (3.1) with $X = X_1$ and $Y = X^{(1)}$ for nondecreasing $f, g : \mathbb{R} \to \{0, 1\}$. Since $X_1 \in \{0, 1\}$ with probability 1, the case $f(0) = f(1)$ trivially yields equality in (3.1). Thus, it suffices to show that, for all nondecreasing $g : \mathbb{R} \to \{0, 1\}$,

$$\operatorname{cov}[X_1, g(X^{(1)})] = \mathrm{E}\{X_1[\mathrm{E}[g(X^{(1)}) \mid X_1] - \mathrm{E}\, g(X^{(1)})]\} = \mathrm{E}[h(X_1)] \leq 0, \tag{3.2}$$

where

$$h(x) = x[\mathrm{E}[g(X_2 + \cdots + X_n) \mid X_1 = x] - \mathrm{E}\, g(X_2 + \cdots + X_n)].$$

Obviously $h(0) = 0$. On the other hand, by the NR property of $X_1, X_2, \dots, X_n$ it follows that there exist random variables $Y_2, \dots, Y_n$ satisfying (a) and (b) of Definition 1.1 (with $i = 1$). This shows that

$$h(1) = \mathrm{E}[g(X^{(1)}) \mid X_1 = 1] - \mathrm{E}\, g(X^{(1)}) = \mathrm{E}[g(Y_2 + \cdots + Y_n) - g(X^{(1)})] \leq 0$$

with probability 1, where the equality follows from Definition 1.1(a) and the inequality from (b) and the fact that g is nondecreasing. This shows that $h(X_1) \leq 0$ with probability 1, and thus (3.2) follows. The general case (that X_i and $X^{(i)}$ are NQD) can be proved similarly.

(b) Consider the indicators X_1, X_2, X_3, uniformly distributed over

$$S = \{(0,0,1), (1,1,0), (1,0,1), (0,1,1), (1,1,1)\}.$$

It is easily seen that X_1, X_2, X_3 are TND. Since, however, $\operatorname{cov}[X_1, X_2] = \frac{1}{25} > 0$, they are not NR. This proves the assertion for $n = 3$, while for any $n > 3$ it suffices to consider the indicators $X_1, X_2, X_3, 1, \ldots, 1$, with X_1, X_2, X_3 as above.

(c) Because of (a), it suffices to prove that if the 0–1 TND indicators $X_1, X_2, \ldots, X_n$ ($n \geq 2$) are exchangeable, then they are NR. Without loss of generality, we may assume that $\mathrm{P}[X_i = 1] > 0$ for all i. Using a result of Strassen (1965) (see also Preston (1974, Proposition 2), Barbour *et al.* (1992, Theorem 2.D)), it follows that the NR property is equivalent to the fact that, for all $i = 1, \ldots, n$,

$$\mathrm{E}[\phi(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n) \mid X_i = 1] \leq \mathrm{E}[\phi(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)]$$

for any component-wise nondecreasing test function $\phi : \mathbb{R}^{n-1} \to \{0, 1\}$. By the exchangeability of $X_1, X_2, \ldots, X_n$, it follows that, for any such ϕ, there exists a nondecreasing function $g : \mathbb{R} \to \{0, 1\}$ such that

$$\phi(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n) = g(X^{(i)}) \quad \text{with probability 1},$$

where $X^{(i)} = X_1 + \cdots + X_{i-1} + X_{i+1} + \cdots + X_n$. Therefore, assuming $X_1, X_2, \ldots, X_n$ to be TND, it follows that

$$\begin{aligned}
\mathrm{E}[\phi(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n) \mid X_i = 1] &= \mathrm{E}[g(X^{(i)}) \mid X_i = 1] \\
&\leq \mathrm{E}[g(X^{(i)})] \\
&= \mathrm{E}[\phi(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)]
\end{aligned}$$

(the inequality is a consequence of the NQD property of X_i and $X^{(i)}$ and the fact that g is nondecreasing), proving the assertion.

Remark 3.1. It can easily be verified that, for $n = 2$, the NR and TND notions do coincide for 0–1 indicators X_1, X_2 (they also coincide with the NCD and NQD notions, stated above). For $n \geq 3$, however, there exist TND 0–1 indicators that are not NCD, e.g. the TND indicators $X_1, X_2, X_3, 1, \ldots, 1$, with X_1, X_2, X_3 defined in the proof of Theorem 3.1(b), are not NCD because X_2 and $S_1 = X_1$ are positively correlated and thus they are not NQD (nevertheless, $X_3, X_2, X_1, 1, \ldots, 1$ are NCD). As a conclusion, the main difference between TND and NCD notions is that the former is symmetric in its arguments, in contrast to the latter.

We also note that negatively associated random variables $X_1, X_2, \ldots, X_n$ are TND, as can be readily shown from the definition due to Joag-Dev and Proschan (1983), stated below for completeness.

Definition 3.2. The random variables $X_1, X_2, \ldots, X_n$ are said to be negatively associated (NA) if, for all nonempty subsets A, B of $\{1, \ldots, n\}$ with $AB = \varnothing$ and $A \cup B = \{1, \ldots, n\}$ and all component-wise nondecreasing functions $f : \mathbb{R}^{|A|} \to \mathbb{R}$ and $g : \mathbb{R}^{|B|} \to \mathbb{R}$ (where $|C|$ denotes the cardinal number of a set C),

$$\operatorname{cov}[f(X_i; i \in A), g(X_j; j \in B)] \le 0, \tag{3.3}$$

provided that the covariance is finite.

Once again, it suffices to verify (3.3) only for nondecreasing 0–1 indicators f, g. As a general observation, the notions NA, NQD, TND and NCD coincide for $n = 2$ (and they are equivalent to NR, when we consider the 0–1 indicators X_1, X_2). However, for $n \ge 3$, the strictly restricted class is that of NA.

We next prove the following basic lemma satisfied by TND indicators.

Lemma 3.1. *Assume that $X_1, X_2, \dots, X_n$ are TND 0–1 indicators with $\mathrm{P}[X_1 = 1] = p_i = 1 - \mathrm{P}[X_i = 0]$, $i = 1, 2, \dots, n$, and let $W = \sum_{i=1}^n X_i$ and $\mu = \mathrm{E}[W]$. Then, for any nondecreasing function g,*

$$\mathrm{E}[Wg(W)] \le \mu\, \mathrm{E}[g(1 + W)]. \tag{3.4}$$

Proof. First observe that W has bounded support, and thus the expectations in (3.4) are finite. Since, by the assumptions, the random variables X_i and $X^{(i)} = W - X_i$ are NQD for all i, taking into account the fact that $W \le 1 + X^{(i)} \le 1 + W$ with probability 1, we have

$$\mathrm{E}[X_i g(W)] \le \mathrm{E}[X_i g(1 + X^{(i)})] \le p_i\, \mathrm{E}[g(1 + X^{(i)})] \le p_i\, \mathrm{E}[g(1 + W)],$$

where the first and third inequalities follow because g is nondecreasing and the second because X_i and $X^{(i)}$ are NQD. Adding for $i = 1, \dots, n$, we get (3.4).

Corollary 3.1. *Under the terminology of Definition 2.1 and the assumptions of Lemma 3.1,*

$$h_W(x) \le \mu p_W(x), \qquad x = 0, 1, \dots, \tag{3.5}$$

where p_W is the probability function of W (which lies in $\mathcal{F}_0$) and h_W is given by Definition 2.1(a).

Proof. For all $x \in \{0, 1, \dots\}$, we have

$$\begin{aligned} h_W(x) &= \sum_{k=x+1}^{\infty} (k - \mu) p_W(k) \\ &= \mathrm{E}[(W - \mu)\mathbf{1}(W \ge x + 1)] \\ &= \mathrm{E}[W\mathbf{1}(W \ge x + 1)] - \mu\, \mathrm{P}[W \ge x + 1] \\ &\le \mu\, \mathrm{P}[W + 1 \ge x + 1] - \mu\, \mathrm{P}[W \ge x + 1] \\ &= \mu\, \mathrm{P}[W = x], \end{aligned}$$

where for the inequality we applied (3.4) to the nondecreasing function $g(w) = \mathbf{1}(w \ge x + 1)$.

We are now in a position to state and prove our main result for TND 0–1 indicators.

Theorem 3.2. *Let $X_1, X_2, \dots, X_n$ be TND 0–1 indicators and set $W = \sum_{i=1}^n X_i$, $\mu = \mathrm{E}[W]$ and $\sigma^2 = \operatorname{var}[W]$. Then, for any $\lambda \ge \mu$,*

$$d_{\mathrm{TV}}(W, \mathcal{P}_\lambda) \le (1 - \mathrm{e}^{-\lambda})\left(1 - \frac{\sigma^2}{\lambda}\right) + \min\left\{1, \frac{(2/\mathrm{e})^{1/2}}{\lambda^{1/2}}\right\}(\lambda - \mu), \tag{3.6}$$

where for $\lambda = \mu = 0$ the upper bound should be taken to be 0.

Proof. From (2.2) we have that, for all $\lambda \geq 0$,

$$d_{\mathrm{TV}}(W, \mathcal{P}_\lambda) \leq \lambda^{-1}(1-\mathrm{e}^{-\lambda}) \sum_{x=0}^{\infty} |\lambda p_W(x) - h_W(x)| + \min\left\{1, \frac{(2/\mathrm{e})^{1/2}}{\lambda^{1/2}}\right\}|\lambda - \mu|.$$

If we take $\lambda \geq \mu$, then $h_W \leq \lambda p_W$ by (3.5), all the absolute value bars can be removed from the above expression, and (3.6) follows since $\sum_{x=0}^{\infty} h_W(x) = \mathrm{var}[W] = \sigma^2$ from the covariance identity (2.1).

Remark 3.2. (a) For $\lambda = \mu$, the upper bound in (3.6) is the same as that of Theorem 1.1; therefore, as was already noted, Theorem 3.2 provides an extension of the classical Chen–Stein bound to a wider class of 0–1 indicators—the TND class.

(b) Although, under the TND assumption, some pairs X_i, X_j may be positively correlated (see the proof of Theorem 3.1(b)), the bound (3.6) shows that the inequality $\mathrm{E}[W] \geq \mathrm{var}[W]$ (with strict inequality unless $X_1 = \cdots = X_n = 0$) holds for TND 0–1 indicators.

(c) Since for all μ and σ^2 satisfying $0 < \sigma^2 < \mu$, the three functions $1 - \mathrm{e}^{-\lambda}$, $1 - \sigma^2/\lambda$ and $\min\{1, (2/\mathrm{e})^{1/2}/\lambda^{1/2}\}(\lambda - \mu)$ are strictly increasing in $\lambda \geq \mu$ and positive for $\lambda \in (\mu, \infty)$, it follows that the right-hand side of (3.6) is strictly increasing in $\lambda \geq \mu$. Thus, the best estimate in (3.6) is given for $\lambda = \mu$, in which case the bound simplifies to (1.1). In most applications, however, the distribution of $W = W_n$ depends on the sample size n (and its mean $\mu = \mu_n$ and variance $\sigma^2 = \sigma_n^2$ are functions of n), so that it is more natural to consider the distance between W_n and its constant Poisson limit $\mathcal{P}_\lambda$ (provided that $\mu_n \uparrow \lambda$ and $d_{\mathrm{TV}}(W_n, \mathcal{P}_\lambda) \to 0$ as $n \to \infty$; see Section 4) rather than considering the moving 'limit' $\mathcal{P}_{\mu_n}$. Therefore, if we are interested in bounding the distance between W and $\mathcal{P}_\lambda$ (with $\mathrm{E}[W] = \mu < \lambda$), it is easily verified that the upper estimate (3.6) is strictly better than the corresponding one which is commonly used in similar situations, namely, the estimate obtained by using first the triangular inequality $d_{\mathrm{TV}}(W, \mathcal{P}_\lambda) \leq d_{\mathrm{TV}}(W, \mathcal{P}_\mu) + d_{\mathrm{TV}}(\mathcal{P}_\mu, \mathcal{P}_\lambda)$ and next the estimates (3.6) with $\lambda = \mu$ and

$$d_{\mathrm{TV}}(\mathcal{P}_\mu, \mathcal{P}_\lambda) = d_{\mathrm{TV}}(\mathcal{P}_\lambda, \mathcal{P}_\mu) \leq (1 - \mathrm{e}^{-\lambda} + \min\{\lambda, (2\lambda/\mathrm{e})^{1/2}\})\Big(1 - \frac{\mu}{\lambda}\Big) \tag{3.7}$$

(which is the minimum of the two estimates obtained from Theorem 2.1 on taking first $X = \mathcal{P}_\mu$ and next interchanging the roles of μ and λ). For this reason, we presented the wide family of bounds (3.6) ($\lambda \geq \mu$) instead of giving the single estimate (1.1).

4. Application to a generalized birthday problem

Assume that, for some fixed $c \in \{2, 3, \dots\}$, $m_n = \lfloor tn^{1-1/c} \rfloor$ balls (persons) are randomly placed into n urns (the possible, equally likely, dates of birth; usually $n = 365$), where $t > 0$ is a constant (which is considered for practical reasons, i.e. in order to ensure that, for all fixed n and c, any given number of m balls can be expressed in the above form $m = m_n$) and $\lfloor x \rfloor$ denotes the integer part of x. If W_n presents the number of urns with at least c balls, then Henze (1998) proved that, as $n \to \infty$, W_n converges to $\mathcal{P}_\lambda$ with $\lambda = t^c/c!$. As an immediate application for $T_{n,c}^{(k)}$, the number of balls needed until for the first time exactly k urns contain at least c balls, Henze (1998) showed that $n^{-(c-1)}(T_{n,c}^{(k)})^c/c!$ converges to a limiting Erlang distribution, as $n \to \infty$, with parameter k (standard exponential for $k = 1$).

Since the above generalized birthday problem, however, lies in the extremely useful category of occupancy models, more general results can be found in several previous works such as Arratia *et al.* (1989, Example 2) and Kolchin *et al.* (1978, Chapter III, Section 3, Theorem 1).

Moreover, in Barbour *et al.* (1992, Chapter 6, Corollary 6.C.1) it is shown that, in order to have a Poisson limit for W_n, it suffices to choose $m_n = t_n n^{1-1/c}$ where $t_n \to t > 0$ as $n \to \infty$ (observe that Henze's choice satisfies this condition), in which case

$$d_{\mathrm{TV}}(W_n, \mathcal{P}_{\mu_n}) = O(n^{-(1-1/c)}) \quad \text{as } n \to \infty, \tag{4.1}$$

where $\mu_n = \mathrm{E}[W_n]$ (in fact, the order $n^{-(1-1/c)}$ was shown to hold for the much wider class of urns with unequal uniformly small probabilities satisfying some regularity conditions).

It seems, however, that all the preceding results about the rate of convergence are concerned with $d_{\mathrm{TV}}(W_n, \mathcal{P}_{\mu_n})$ rather than the more natural $d_{\mathrm{TV}}(W_n, \mathcal{P}_\lambda)$, with $\lambda = t^c/c!$; thus, a convergence rate for the latter may be of some interest. With the help of Theorem 3.2 we can easily show the following result.

Theorem 4.1. *For fixed $c \in \{2, 3, \dots\}$ and $t > 0$, let $m_n = \lfloor t n^{1-1/c} \rfloor$ be the number of balls randomly placed into n equally likely urns and W_n be the number of urns that contain at least c balls. Then we have*

$$d_{\mathrm{TV}}(W_n, \mathcal{P}_\lambda) \le O(n^{-1/c}) \quad \textit{as } n \to \infty, \tag{4.2}$$

where $\lambda = t^c/c!$.

For the proof of Theorem 4.1, we shall make use of the following, very particular, cases of the well-known identities expressing the tail probabilities of a binomial and a trinomial distribution in terms of incomplete beta and Dirichlet integrals respectively (see Olkin and Sobel (1965), Henze (1998)).

Lemma 4.1. *If Y_1, Y_2 follow a trinomial distribution with m trials and probabilities $1/n$ and $1/n$ $(n > 2)$, then, for any $c \in \{1, \dots, \lfloor m/2 \rfloor\}$,*

$$\mathrm{P}[Y_1 \ge c] = \frac{m!}{c!\,(m-c)!} \int_0^{1/n} (1-u)^{m-c}\, \mathrm{d}u^c, \tag{4.3}$$

$$\mathrm{P}[Y_1 \ge c, Y_2 \ge c] = \frac{m!}{(c!)^2 (m-2c)!} \int_0^{1/n} \int_0^{1/n} (1-u-v)^{m-2c}\, \mathrm{d}u^c\, \mathrm{d}v^c. \tag{4.4}$$

Proof of Theorem 4.1. Since $n, m_n \to \infty$, we may assume that $n \ge 3$ and $m_n \ge 2c$. Let $Y_i = Y_i^{(n)}$, $i = 1, \dots, n$, be the number of balls contained in the urn i, and observe that

$$W_n = \sum_{i=1}^n \mathbf{1}(Y_i \ge c) = \sum_{i=1}^n X_i,$$

where the $X_i = X_i^{(n)} = \mathbf{1}(Y_i \ge c)$ are 0–1 indicators which are nondecreasing functions of the NA random variables $Y_1, Y_2, \dots, Y_n$ (because every multinomial random vector consists of NA random variables). Since each X_i is a function of Y_i alone, it follows that $X_1, X_2, \dots, X_n$ are also NA, and thus NR and TND. Therefore, both Theorems 1.1 and 3.2 could equally be applied to W_n. However, in the remainder of the proof we shall use Theorem 3.2, because we are interested in the constant limit $\mathcal{P}_\lambda$ (see Remark 3.2(c)). By the exchangeability of $X_1, X_2, \dots, X_n$, it follows that

$$\mu_n = \mathrm{E}[W_n] = n\,\mathrm{E}[X_1] = n\,\mathrm{P}[X_1 = 1] = n\,\mathrm{P}[Y_1 \ge c] = np_n$$

with

$$p_n = \mathrm{P}[Y_1 \geq c] = \sum_{j=c}^{m_n} \frac{m_n!}{j!\,(m_n - j)!} \left(\frac{1}{n}\right)^j \left(1 - \frac{1}{n}\right)^{m_n - j},$$

and thus by (4.3) we get

$$\mu_n = np_n = \frac{n(m_n!)}{c!\,(m_n - c)!} \int_0^{1/n} (1-u)^{m_n - c}\,\mathrm{d}u^c. \tag{4.5}$$

On the other hand,

$$\begin{aligned}\sigma_n^2 = \mathrm{var}[W_n] &= n\,\mathrm{var}[X_1] + n(n-1)\,\mathrm{cov}[X_1, X_2]\\ &= np_n(1-p_n) + n(n-1)(q_n - p_n^2)\\ &= \mu_n - \mu_n^2 + n(n-1)q_n\end{aligned}$$

with

$$\begin{aligned}q_n &= \mathrm{E}[X_1 X_2] = \mathrm{P}[X_1 = 1, X_2 = 1]\\ &= \mathrm{P}[Y_1 \geq c, Y_2 \geq c]\\ &= \sum_{j=c}^{m_n - c} \sum_{k=c}^{m_n - j} \frac{m_n!}{j!\,k!\,(m_n - j - k)!} \left(\frac{1}{n}\right)^{j+k} \left(1 - \frac{2}{n}\right)^{m_n - j - k};\end{aligned}$$

therefore, (4.4) yields

$$\sigma_n^2 - \mu_n + \mu_n^2 = \frac{n(n-1)(m_n!)}{(c!)^2(m_n - 2c)!} \int_0^{1/n}\!\!\int_0^{1/n} (1-u-v)^{m_n - 2c}\,\mathrm{d}u^c\,\mathrm{d}v^c, \tag{4.6}$$

with μ_n given by (4.5). Using the obvious inequalities $m_n \leq tn^{1-1/c}$, $m_n!/(m_n - c)! < m_n^c \leq t^c n^{c-1}$ and $(1-u)^{m_n - c} < 1$, we conclude from (4.5) that

$$\mu_n < \frac{t^c}{c!} = \lambda, \tag{4.7}$$

which shows that we can apply Theorem 3.2 for this particular value of λ. On the other hand, since $\sigma_n^2 < \mu_n < \lambda$ by the TND property of $X_1, X_2, \ldots, X_n$ (see Remark 3.2(b)), the bound (3.6) yields the estimate

$$d_{\mathrm{TV}}(W_n, \mathcal{P}_\lambda) \leq B_\lambda(\lambda - \sigma_n^2) + A_\lambda(\lambda - \mu_n) = O(\lambda - \sigma_n^2) \quad \text{as } n \to \infty, \tag{4.8}$$

where A_λ, B_λ are positive constants depending only on $\lambda = t^c/c! > 0$ (in fact, $B_\lambda = \lambda^{-1}(1 - \mathrm{e}^{-\lambda})$ and $A_\lambda = \min\{1, (2/\mathrm{e})^{1/2}/\lambda^{1/2}\}$ are given in (1.4)). Therefore, for the proof of (4.2), it suffices to verify that

$$\lambda - \sigma_n^2 \leq O(n^{-1/c}) \quad \text{as } n \to \infty. \tag{4.9}$$

Observe that, by (4.7),

$$\begin{aligned}\lambda - \sigma_n^2 &= (\lambda - \mu_n) - (\lambda^2 - \mu_n^2) + (\lambda^2 - (\sigma_n^2 - \mu_n + \mu_n^2))\\ &\leq (\lambda - \mu_n) + (\lambda^2 - (\sigma_n^2 - \mu_n + \mu_n^2)),\end{aligned}$$

which implies that, for (4.9) to hold, it is sufficient to verify that both $\lambda - \mu_n$ and $\lambda^2 - (\sigma_n^2 - \mu_n + \mu_n^2)$ converge to 0 at least as fast as $n^{-1/c}$. To this end, we have to obtain accurate lower bounds for μ_n and $\sigma_n^2 - \mu_n + \mu_n^2$. Using first the obvious inequalities $m_n!/(m_n - c)! \geq (m_n - c + 1)^c \geq (tn^{1-1/c} - c)^c$ (assuming that $n > (c/t)^{c/(c-1)}$) and

$$(1-u)^{m_n - c} > \left(1 - \frac{1}{n}\right)^{m_n - c} \geq \left(1 - \frac{1}{n}\right)^{tn^{1-1/c}} \qquad \text{for } 0 < u < \frac{1}{n},$$

we get from the integral expansion (4.5) the inequality

$$\mu_n > \lambda\left(1 - \frac{c}{tn^{1-1/c}}\right)^c \left(1 - \frac{1}{n}\right)^{tn^{1-1/c}} = \lambda(1 - O(n^{-1/c})),$$

which implies that $\lambda - \mu_n \leq O(n^{-1/c})$. For the other term, we use the inequalities $m_n!/(m_n - 2c)! \geq (tn^{1-1/c} - 2c)^{2c}$ (provided that $n > (2c/t)^{c/(c-1)}$) and

$$(1-u-v)^{m_n - 2c} > \left(1 - \frac{2}{n}\right)^{m_n - 2c} \geq \left(1 - \frac{2}{n}\right)^{tn^{1-1/c}} \qquad \text{for } 0 < u, v < \frac{1}{n},$$

which, combined with (4.6), yield the lower bound

$$\sigma_n^2 - \mu_n + \mu_n^2 > \lambda^2\left(1 - \frac{1}{n}\right)\left(1 - \frac{2c}{tn^{1-1/c}}\right)^{2c}\left(1 - \frac{2}{n}\right)^{tn^{1-1/c}} = \lambda^2(1 - O(n^{-1/c}));$$

this shows that $\lambda^2 - (\sigma_n^2 - \mu_n + \mu_n^2) \leq O(n^{-1/c})$, and the proof is complete.

Remark 4.1. (a) The rate $n^{-(1-1/c)}$ given in (4.1) is strictly better than that suggested by (4.2), namely $n^{-1/c}$, except if $c = 2$. The latter rate, however, cannot be improved, for the following reason: working as in the proof of Theorem 4.1, it can be shown that

$$\lambda - \mu_n = O(n^{-1/c}).$$

Indeed, using the obvious inequalities $m_n!/(m_n - c)! \leq t^c n^{c-1}$ and

$$(1-u)^{m_n - c} \leq 1 - (m_n - c)u + \tfrac{1}{2}(m_n - c)(m_n - c - 1)u^2, \qquad 0 < u < 1,$$

in the integral expression (4.5), we have

$$\mu_n \leq \lambda\left(1 - \frac{c}{c+1}(m_n - c)n^{-1} + O(n^{-2/c})\right) = \lambda(1 - O(n^{-1/c}));$$

this shows that $\lambda - \mu_n \geq O(n^{-1/c})$, while the opposite inequality was shown in the proof of the theorem. Therefore, since $d_{\mathrm{TV}}(\mathcal{P}_{\mu_n}, \mathcal{P}_\lambda) = O(\lambda - \mu_n)$, which follows from (3.7) and the fact that

$$d_{\mathrm{TV}}(\mathcal{P}_{\mu_n}, \mathcal{P}_\lambda) \geq \mathrm{P}[\mathcal{P}_{\mu_n} = 0] - \mathrm{P}[\mathcal{P}_\lambda = 0] = \mathrm{e}^{-\mu_n} - \mathrm{e}^{-\lambda} = O(\lambda - \mu_n),$$

we conclude from (4.1) and the triangular inequality that

$$d_{\mathrm{TV}}(W_n, \mathcal{P}_\lambda) \geq d_{\mathrm{TV}}(\mathcal{P}_{\mu_n}, \mathcal{P}_\lambda) - d_{\mathrm{TV}}(W_n, \mathcal{P}_{\mu_n}) = O(n^{-1/c}) - O(n^{-(1-1/c)}) = O(n^{-1/c});$$

thus, the correct order of magnitude for the approximation is exactly $n^{-1/c}$.

(b) Comparing with other techniques, it seems that the method for proving Theorem 4.1 is quite simple; this suggests that the assertion (4.2) may be extended to the case of unequal urn probabilities $p_i^{(n)}$. However, more restricted conditions (than those given in Corollary 6.C.1 of Barbour *et al.* (1992)) may be required for $p_i^{(n)}$, in order to ensure that $\mu_n \uparrow \lambda$; a condition needed in applying Theorem 3.2.

5. On the existence of monotone couplings

In Barbour *et al.* (1992, Remark 2.1.4) it is stated that the assertion of Theorem 1.1 remains valid for a wider (than NR) class of indicators, namely the weakly negatively related indicators given by the following definition.

Definition 5.1. Indicator random variables $X_1, X_2, \ldots, X_n$ are said to be weakly negatively related (WNR) if, for each $i \in \{1, 2, \ldots, n\}$, there exist random variables $Y_1, \ldots, Y_{i-1}, Y_{i+1}, \ldots, Y_n$ (depending on i) satisfying Definition 1.1(a) and

(b′) with probability 1, $\sum_{j \neq i} Y_j \leq \sum_{j \neq i} X_j$.

Clearly, NR indicators are always WNR, while the converse is false for $n \geq 3$ (see the proof of Theorem 3.1(b)). The following definition seems to give an even weaker negative relationship among indicators.

Definition 5.2. The 0–1 indicators $X_1, X_2, \ldots, X_n$ are said to be very weakly negatively related (VWNR) if, for each $i \in \{1, 2, \ldots, n\}$, there exists a random variable Y (depending on i) such that

(a′) $Y \stackrel{\mathrm{D}}{=} (\sum_{j \neq i} X_j)$ given that $X_i = 1$, and

(b″) with probability 1, $Y \leq \sum_{j \neq i} X_j$.

From the above definitions it becomes clear that WNR indicators are always VWNR (take $Y = \sum_{j \neq i} Y_j$), while the situation concerning the converse is not very clear at this point. Another natural question arising from the above definitions is the interrelation between the TND and NR or WNR or VWNR classes of random variables (we have already seen in Theorem 3.1 that the TND class is strictly wider than the NR class of indicators). In order to see exactly what happens between the above classes, it will be convenient to use the following three notions of a monotone coupling.

Definition 5.3. Let $\boldsymbol{X} = (X_1, \ldots, X_n)$ and $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$ be two random vectors taking values in $\mathbb{R}^n$ with distributions $\mathcal{L}(\boldsymbol{X})$ and $\mathcal{L}(\boldsymbol{Y})$ respectively.

(i) We say that there exists a monotone coupling for $\boldsymbol{X}$ and $\boldsymbol{Y}$ (or for $\mathcal{L}(\boldsymbol{X})$ and $\mathcal{L}(\boldsymbol{Y})$) if there is a random vector $(U_1, \ldots, U_n; V_1, \ldots, V_n) = (\boldsymbol{U}; \boldsymbol{V})$ in $\mathbb{R}^{2n}$ such that $\mathcal{L}(\boldsymbol{U}) = \mathcal{L}(\boldsymbol{X})$, $\mathcal{L}(\boldsymbol{V}) = \mathcal{L}(\boldsymbol{Y})$ and $\mathrm{P}[U_i \leq V_i,\ i = 1, 2, \ldots, n] = 1$. The existence of a monotone coupling will be denoted by $\mathrm{MC}(\mathcal{L}(\boldsymbol{X}), \mathcal{L}(\boldsymbol{Y}))$.

(ii) We say that there exists a weak monotone coupling for $\boldsymbol{X}$ and $\boldsymbol{Y}$ if there is a random vector $(U_1, \ldots, U_n; V_1, \ldots, V_n) = (\boldsymbol{U}; \boldsymbol{V})$ in $\mathbb{R}^{2n}$ such that $\mathcal{L}(\boldsymbol{U}) = \mathcal{L}(\boldsymbol{X})$, $\mathcal{L}(\boldsymbol{V}) = \mathcal{L}(\boldsymbol{Y})$ and $\mathrm{P}[\sum_i U_i \leq \sum_i V_i] = 1$. The existence of a weak monotone coupling will be denoted as $\mathrm{WMC}(\mathcal{L}(\boldsymbol{X}), \mathcal{L}(\boldsymbol{Y}))$.

(iii) We say that there exists a very weak monotone coupling for $\boldsymbol{X}$ and $\boldsymbol{Y}$ if there is a random vector $(U; V)$ in $\mathbb{R}^2$ such that $\mathcal{L}(U) = \mathcal{L}(\sum_i X_i)$, $\mathcal{L}(V) = \mathcal{L}(\sum_i Y_i)$ and $\mathrm{P}[U \leq V] = 1$. The existence of a very weak monotone coupling will be denoted as $\mathrm{VWMC}(\mathcal{L}(\boldsymbol{X}), \mathcal{L}(\boldsymbol{Y}))$.

From the above definitions it is clear that a monotone coupling implies the existence of a weak monotone one, and the existence of a weak monotone coupling implies the existence of a very weak monotone one. Moreover,

$$\mathrm{VWMC}(\mathcal{L}(\boldsymbol{X}), \mathcal{L}(\boldsymbol{Y})) \quad \text{if and only if} \quad \mathrm{MC}\bigg(\mathcal{L}\bigg(\sum_i X_i\bigg), \mathcal{L}\bigg(\sum_i Y_i\bigg)\bigg). \tag{5.1}$$

Therefore, if $X_1, X_2, \ldots, X_n$ are 0–1 indicators, then the alternative definitions become

$$\mathrm{NR} \equiv \mathrm{MC}(\mathcal{L}(\boldsymbol{X}_i \mid X_i = 1), \mathcal{L}(\boldsymbol{X}_i)) \quad \text{for all } i, \tag{5.2}$$

$$\mathrm{WNR} \equiv \mathrm{WMC}(\mathcal{L}(\boldsymbol{X}_i \mid X_i = 1), \mathcal{L}(\boldsymbol{X}_i)) \quad \text{for all } i, \tag{5.3}$$

$$\mathrm{VWNR} \equiv \mathrm{VWMC}(\mathcal{L}(\boldsymbol{X}_i \mid X_i = 1), \mathcal{L}(\boldsymbol{X}_i)) \quad \text{for all } i \tag{5.4}$$

(where $\boldsymbol{X}_i = (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$). Thus, using Strassen's (1965) results on monotone couplings, the following lemma can be easily established.

Lemma 5.1. *The* 0–1 *indicators $X_1, \ldots, X_n$ are TND if and only if they are VWNR.*

Proof. Without loss of generality assume that $\mathrm{P}[X_i = 1] > 0$ for all i. Write $\boldsymbol{X}_i = (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$ and $X^{(i)} = \sum_{j \neq i} X_j$ and observe that $\mathrm{E}[X_i g(X^{(i)})] = \mathrm{P}[X_i = 1]\,\mathrm{E}[g(X^{(i)}) \mid X_i = 1]$ for all nondecreasing test functions $g : \mathbb{R} \to \{0, 1\}$; thus, the TND property is equivalent to the fact that

$$\mathrm{E}[g(X^{(i)}) \mid X_i = 1] \le \mathrm{E}[g(X^{(i)})] \quad \text{for all such } g \text{ and for all } i.$$

By (5.4), the VWNR property is equivalent to $\mathrm{VWMC}(\mathcal{L}(\boldsymbol{X}_i \mid X_i = 1), \mathcal{L}(\boldsymbol{X}_i))$ for all i, which, in turn, is equivalent by (5.1) to $\mathrm{MC}(\mathcal{L}(X^{(i)} \mid X_i = 1), \mathcal{L}(X^{(i)}))$ for all i, and the assertion follows by Strassen's results on monotone couplings (see also Barbour *et al.* (1992, Theorem 2.D)).

Since Barbour *et al.* (1992, Remark 2.1.4) state that (3.6) holds for WNR 0–1 indicators, while in the present article it is proved that this holds for TND indicators (i.e. for VWNR indicators because of Lemma 5.1), it would be desirable to show that the TND (VWNR) class is strictly larger than the WNR class of indicators. This is not true, however, because of the following lemma.

Lemma 5.2. *The* 0–1 *indicators $X_1, X_2, \ldots, X_n$ are VWNR (TND) if and only if they are WNR.*

Thus, surprisingly enough, the notions of TND, WNR and VWNR do coincide. The proof of this fact is a consequence of (5.3), (5.4) and the following very general result, which may be of some independent interest.

Theorem 5.1. (a) *For any random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$ with values in $\mathbb{R}^n$, the properties $WMC(\mathcal{L}(\boldsymbol{X}), \mathcal{L}(\boldsymbol{Y}))$ and $VWMC(\mathcal{L}(\boldsymbol{X}), \mathcal{L}(\boldsymbol{Y}))$ are equivalent. In other words, the existence of a very weak monotone coupling implies the existence of a weak monotone one.*

(b) *Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ and $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_m)$ be two arbitrary random vectors with values in $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively. If the random vector $(U; V)$ in $\mathbb{R}^2$ satisfies*

$$\mathcal{L}(U) = \mathcal{L}\bigg(\sum_{i=1}^n X_i\bigg) \quad \text{and} \quad \mathcal{L}(V) = \mathcal{L}\bigg(\sum_{j=1}^m Y_j\bigg),$$

then there exists a random vector $(\boldsymbol{U};\boldsymbol{V}) = (U_1,\dots,U_n;V_1,\dots,V_m)$ *with values in* $\mathbb{R}^{n+m}$ *such that*

$$\mathcal{L}(\boldsymbol{U}) = \mathcal{L}(\boldsymbol{X}), \qquad \mathcal{L}(\boldsymbol{V}) = \mathcal{L}(\boldsymbol{Y}) \quad \textit{and} \quad \mathcal{L}\biggl(\sum_{i=1}^{n} U_i;\sum_{j=1}^{m} V_j\biggr) = \mathcal{L}(U;V).$$

Proof. We shall only show (b) which trivially implies (a). Denote by $\mathcal{K}$ the probability distributions of the random vectors $(\boldsymbol{U};\boldsymbol{V})$ in $\mathbb{R}^{n+m}$ such that

$$\mathcal{L}\biggl(\sum_{i=1}^{n} U_i;\sum_{j=1}^{m} V_j\biggr) = \mathcal{L}(U;V).$$

It follows that $\mathcal{K}$ is a nonempty convex subset of all probability measures in $\mathbb{R}^{n+m}$, which is closed with respect to the topology generated by the weak convergence of probability measures in $\mathbb{R}^{n+m}$. By Theorem 7 of Strassen (1965), there exists an element of $\mathcal{K}$ satisfying the assertion if and only if, for all bounded continuous functions $f:\mathbb{R}^n\to\mathbb{R}$ and $g:\mathbb{R}^m\to\mathbb{R}$,

$$\mathrm{E}[f(\boldsymbol{X})] + \mathrm{E}[g(\boldsymbol{Y})] \le \sup_{\mathcal{K}}\{\mathrm{E}[f(\boldsymbol{U})] + \mathrm{E}[g(\boldsymbol{V})]\}, \tag{5.5}$$

where the supremum is taken over the random vectors $(\boldsymbol{U};\boldsymbol{V})$ for which $\mathcal{L}(\boldsymbol{U};\boldsymbol{V})\in\mathcal{K}$. Fix two functions f and g as above and $\varepsilon>0$. Without loss of generality, we may assume that $0\le f(\boldsymbol{x})\le 1$ and $0\le g(\boldsymbol{y})\le 1$ for all $(\boldsymbol{x};\boldsymbol{y})\in\mathbb{R}^{n+m}$. Let $M=M(\varepsilon)>0$ be large enough so that

$$\begin{aligned}
\mathrm{P}[\max|X_i|>M] &< \varepsilon,\\
\mathrm{P}[\max|Y_j|>M] &< \varepsilon,\\
\mathrm{P}\biggl[\max\biggl\{\frac{|U|}{n},\frac{|V|}{m}\biggr\}>M\biggr] &< \varepsilon,
\end{aligned}$$

and define the functions

$$\begin{aligned}
f_M(\boldsymbol{x}) &= f(\boldsymbol{x})\mathbf{1}(\max|x_i|\le M),\\
g_M(\boldsymbol{y}) &= g(\boldsymbol{y})\mathbf{1}(\max|y_j|\le M).
\end{aligned}$$

Since $|f(\boldsymbol{x})-f_M(\boldsymbol{x})|\le\mathbf{1}(\max|x_i|>M)$ and similarly for $|g-g_M|$, it follows that

$$\mathrm{E}[f(\boldsymbol{X})]\le\mathrm{E}[f_M(\boldsymbol{X})]+\varepsilon, \qquad \mathrm{E}[g(\boldsymbol{Y})]\le\mathrm{E}[g_M(\boldsymbol{Y})]+\varepsilon.$$

Since f_M and g_M have a compact support, we may construct the random vector $(\boldsymbol{U};\boldsymbol{V}) = (\boldsymbol{U}(U;V);\boldsymbol{V}(U;V))$ defined by

$$(\boldsymbol{U};\boldsymbol{V}) = \begin{cases} (u_1^*(U),\dots,u_n^*(U);v_1^*(V),\dots,v_m^*(V)) & \text{if } |U|\le nM \text{ and } |V|\le mM,\\ (U/n,\dots,U/n;V/m,\dots,V/m) & \text{otherwise,}\end{cases}$$

where $\boldsymbol{u}^*(u) = (u_1^*(u),\dots,u_n^*(u))$ and $\boldsymbol{v}^*(v) = (v_1^*(v),\dots,v_m^*(v))$ are, respectively, any maximizing points for f_M and g_M over the compact sets $A_{u,M}=\{\boldsymbol{u}:\max|u_i|\le M,\ u_1+\cdots+u_n=u\}$ and $B_{v,M}=\{\boldsymbol{v}:\max|v_j|\le M,\ v_1+\cdots+v_m=v\}$, that is,

$$f_M(\boldsymbol{u}^*(u)) = \max\{f_M(\boldsymbol{u}):\boldsymbol{u}\in A_{u,M}\} \quad \text{for } |u|\le nM,$$

and similarly for g_M. By construction, $\mathcal{L}(\boldsymbol{U}; \boldsymbol{V}) \in \mathcal{K}$. On the other hand, by the definition of $(\boldsymbol{U}; \boldsymbol{V})$ and the fact that $\mathcal{L}(U) = \mathcal{L}(\sum_i X_i)$, we have

$$\begin{aligned}
\mathrm{E}[f_M(\boldsymbol{X})] &= \mathrm{E}\bigg\{\mathrm{E}\bigg[f_M(\boldsymbol{X}) \,\bigg|\, \sum_i X_i\bigg]\bigg\} \\
&\le \mathrm{E}\bigg\{\mathbf{1}\bigg(\bigg|\sum_i X_i\bigg| \le nM\bigg)\mathrm{E}\bigg[f_M(\boldsymbol{X}) \,\bigg|\, \sum_i X_i\bigg]\bigg\} + \mathrm{P}\bigg[\bigg|\sum_i X_i\bigg| > nM\bigg] \\
&\le \mathrm{E}\bigg[\mathbf{1}\bigg(\bigg|\sum_i X_i\bigg| \le nM\bigg) f_M\bigg(\boldsymbol{u}^*\bigg(\sum_i X_i\bigg)\bigg)\bigg] + \mathrm{P}\bigg[\bigg|\sum_i X_i\bigg| > nM\bigg] \\
&= \mathrm{E}[\mathbf{1}(|U| \le nM) f_M(\boldsymbol{u}^*(U))] + \mathrm{P}[|U| > nM] \\
&\le \mathrm{E}[\mathbf{1}(|U| \le nM)\mathbf{1}(|V| \le mM) f_M(\boldsymbol{u}^*(U))] + \mathrm{P}[|U| > nM] + \mathrm{P}[|V| > mM] \\
&= \mathrm{E}[\mathbf{1}(|U| \le nM)\mathbf{1}(|V| \le mM) f(\boldsymbol{u}^*(U))] + \mathrm{P}[|U| > nM] + \mathrm{P}[|V| > mM] \\
&= \mathrm{E}[\mathbf{1}(|U| \le nM)\mathbf{1}(|V| \le mM) f(\boldsymbol{U})] + \mathrm{P}[|U| > nM] + \mathrm{P}[|V| > mM] \\
&\le \mathrm{E}[f(\boldsymbol{U})] + 2\,\mathrm{P}[\max\{|U|/n, |V|/m\} > M] \\
&\le \mathrm{E}[f(\boldsymbol{U})] + 2\varepsilon.
\end{aligned}$$

By the same arguments it can be shown that $\mathrm{E}[g_M(\boldsymbol{Y})] \le \mathrm{E}[g(\boldsymbol{V})] + 2\varepsilon$ and, therefore,

$$\mathrm{E}[f(\boldsymbol{X})] + \mathrm{E}[g(\boldsymbol{Y})] \le \mathrm{E}[f(\boldsymbol{U})] + \mathrm{E}[g(\boldsymbol{V})] + 6\varepsilon.$$

Since ε is arbitrary and $\mathcal{L}(\boldsymbol{U}; \boldsymbol{V}) \in \mathcal{K}$, (5.5) holds and the proof is complete.

References

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

Arratia, R., Goldstein, L. and Gordon, L. (1990). Poisson approximation and the Chen–Stein method. *Statist. Sci.* **5,** 403–434.

Barbour, A. D. and Eagleson, G. K. (1983). Poisson approximation for some statistics based on exchangeable trials. *Adv. Appl. Prob.* **15,** 585–600.

Barbour, A. D., Holst, L. and Janson, S. (1992). *Poisson Approximation* (Oxford Studies Prob. **2**). Oxford University Press.

Boutsikas, M. V. and Koutras, M. V. (2000). A bound for the distribution of the sum of discrete associated or negatively associated random variables. *Ann. Appl. Prob.* **10,** 1137–1150.

Cacoullos, T. and Papathanasiou, V. (1989). Characterizations of distributions by variance bounds. *Statist. Prob. Lett.* **7,** 351–356.

Cacoullos, T., Papadatos, N. and Papathanasiou, V. (1997). Variance inequalities for covariance kernels and applications to central limit theorems. *Theory Prob. Appl.* **42,** 195–201.

Cacoullos, T., Papadatos, N. and Papathanasiou, V. (2002). An application of a density transform and the local limit theorem. *Theory Prob. Appl.* **46,** 803–810.

Cacoullos, T., Papathanasiou, V. and Utev, S. (1994). Variational inequalities with examples and an application to the central limit theorem. *Ann. Prob.* **22,** 1607–1618.

Chen, L. H. Y. (1975). Poisson approximation for dependent trials. *Ann. Prob.* **3,** 534–545.

Chen, L. H. Y. (1998). Stein's method: some perspectives with applications. In *Probability Towards 2000*, eds L. Accardi and C. C. Heyde (Lecture Notes Statist. **128**), Springer, New York, pp. 97–122.

Goldstein, L. and Reinert, G. (1997). Stein's method and the zero bias transformation with application to simple random sampling. *Ann. Appl. Prob.* **7,** 935–952.

Henze, N. (1998). A Poisson limit law for a generalized birthday problem. *Statist. Prob. Lett.* **39,** 333–336.

Joag-Dev, K. and Proschan, F. (1983). Negative association of random variables, with applications. *Ann. Statist.* **11,** 286–295.

Kolchin, V. F., Sevast´yanov, B. A. and Chistyakov, V. P. (1978). *Random Allocations.* Winston, Washington, DC.

Lehmann, E. L. (1966). Some concepts of dependence. *Ann. Math. Statist.* **37,** 1137–1153.

Majsnerowska, M. (1998). A note on Poisson approximation by w-functions. *Appl. Math. (Warsaw)* **25,** 387–392.

Olkin, I. and Sobel, M. (1965). Integral expressions for tail probabilities of the multinomial and the negative multinomial distribution. *Biometrika* **52,** 167–179.

Papadatos, N. and Papathanasiou, V. (1995). Distance in variation between two arbitrary distributions via the associated w-functions. *Theory Prob. Appl.* **40,** 685–694.

Papadatos, N. and Papathanasiou, V. (2001). Unified variance bounds and a Stein-type identity. In *Probability and Statistical Models with Applications*, eds Ch. A. Charalambides, M. V. Koutras and N. Balakrishnan, Chapman and Hall/CRC, New York, pp. 87–100.

Papathanasiou, V. and Utev, S. A. (1995). Integro-differential inequalities and the Poisson approximation. *Siberian Adv. Math.* **5,** 138–150.

Preston, C. J. (1974). A generalization of the FKG inequalities. *Commun. Math. Phys.* **36,** 233–241.

Serfling, R. J. (1975). A general Poisson approximation theorem. *Ann. Prob.* **3,** 726–731.

Stein, C. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proc. 6th Berkeley Symp. Math. Statist. Prob.*, Vol. II, University of California Press, Berkeley, pp. 583–602.

Strassen, V. (1965). The existence of probability measures with given marginals. *Ann. Math. Statist.* **36,** 423–439.