The Annals of Applied Probability
2000, Vol. 10, No. 4, 1137–1150

A BOUND FOR THE DISTRIBUTION OF THE SUM OF DISCRETE ASSOCIATED OR NEGATIVELY ASSOCIATED RANDOM VARIABLES[1]

BY MICHAEL V. BOUTSIKAS AND MARKOS V. KOUTRAS

University of Piraeus

Let $X_1, X_2, \dots, X_n$ be a sequence of integer-valued random variables that are either associated or negatively associated. We present a simple upper bound for the distance between the distribution of the sum of X_i and a sum of n independent random variables with the same marginals as X_i. An upper bound useful for establishing a compound Poisson approximation for $\sum_{i=1}^{n} X_i$ is also provided. The new bounds are of the same order as the much acclaimed Stein–Chen bound.

1. Introduction. Let $X_1, X_2, \dots, X_n$ be a sequence of integer-valued random variables (r.v.'s). The need to approximate the distribution of the sum $\sum_{i=1}^{n} X_i$ arises in various disciplines, such as probability theory, statistics, reliability and biology.

The Poisson approximation of the distribution of $\sum_{i=1}^{n} X_i$, known in the statistical literature as *Poisson law of small numbers* when the X_i's are independently distributed binary r.v.'s, has been the subject of continuing theoretical interest for more than one and a half centuries. Throughout statistical history several interesting generalizations and extensions of the classical Poisson law have been brought to light. Recently, the remarkable work by Chen (1975) led to the development of a group of flexible, powerful techniques that can be effectively used to estimate the error in the Poisson, the binomial and the compound Poisson approximations of the sum of dependent indicator r.v.'s. The definitive reference for this method, known as the Stein–Chen method, is the monograph by Barbour, Holst and Janson (1992). For more recent developments on compound Poisson and binomial approximations, the reader may refer to Barbour, Chen and Loh (1992), Roos (1994), Barbour and Utev (1998) and Soon (1996).

The purpose of this paper is to develop simple tools that are useful for approximating the distribution of the sum of integer-valued (not necessarily binary) dependent r.v.'s by the distribution of the sum of independent variables with the same marginals as the original ones. This is accomplished at the expense of restricting the nature of the dependence to that of associated or negatively associated (NA) r.v.'s. As a matter of fact, the main results of this paper are proved under a weaker assumption on the form of the existing

Received April 1999; revised September 1999.

[1]Research supported by the Greek General Secretariat for Research and Technology under Grant PENED 94-No 70/3/2948.

AMS 1991 *subject classifications*. Primary 60E15; secondary 62E17, 60G50.

Key words and phrases. Association, negative association, positive–negative dependence, probability metrics, compound Poisson approximation, rate of convergence.

dependence between X_i's, but we chose to state the results for associated and NA r.v.'s because these classes are widely known, possess a good number of useful properties and have many interesting applications in diverse areas.

The organization of this paper is as follows: Section 2 reviews preliminaries on a number of useful distance metrics (total variation distance, Kolmogorov distance, Wasserstein distance), association, negative association and several other types of positive or negative dependence. In the same section, we present our general result for the approximation of the distribution of the sum of integer-valued associated or NA r.v.'s and specialize to the case of binary r.v.'s thereof, obtaining a simple inequality for the mean of $\prod_{i=1}^{n} X_i$ and an upper bound for generalized binomial approximations. A compound Poisson approximation for $\sum_{i=1}^{n} X_i$ is examined and an application of the general result in a simple sampling scheme is presented. It is worth mentioning that our bounds require the computation of the first and second moments of X_i [namely $\mathbb{E}(X_i)$ and $\mathbb{E}(X_i X_j)$, $i \neq j$], but not the higher ones. Finally, in Section 3, we give the proof of the main theorem along with several auxiliary results that are of independent interest.

2. Notation and statement of results. The total variation and the Wasserstein distance between the distributions of two integer-valued r.v.'s X, Y (with finite expectations) are given by

$$\begin{aligned} d_{\mathrm{TV}}(\mathscr{L}(X), \mathscr{L}(Y)) &= \sup_{A \subseteq \mathbf{Z}} |\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)| \\ &= \tfrac{1}{2} \sum_{j \in \mathbf{Z}} |\mathbb{P}(X = j) - \mathbb{P}(Y = j)| \end{aligned}$$

and

$$d_W(\mathscr{L}(X), \mathscr{L}(Y)) = \sum_{j \in \mathbf{Z}} |\mathbb{P}(X \le j) - \mathbb{P}(Y \le j)| = \inf \mathbb{E}|U - V|, \tag{1}$$

respectively [the infimum ranges over all couplings (U, V) of $\mathscr{L}(X)$ and $\mathscr{L}(Y)$]. A sequence of random variables $\{X_n\}$ converges in distribution to Y if $d_{\mathrm{TV}}(\mathscr{L}(X_n), \mathscr{L}(Y))$ or $d_W(\mathscr{L}(X_n), \mathscr{L}(Y))$ converges to 0. The reverse is always true for d_{TV}, whereas for d_W the additional condition of X_n being uniformly integrable is necessary.

Another metric useful for establishing convergence in distribution is the Kolmogorov distance

$$d_K(\mathscr{L}(X), \mathscr{L}(Y)) = \sup_w |\mathbb{P}(X \le w) - \mathbb{P}(Y \le w)|.$$

Manifestly, $d_K(\mathscr{L}(X), \mathscr{L}(Y)) \le d_{\mathrm{TV}}(\mathscr{L}(X), \mathscr{L}(Y)) \le d_W(\mathscr{L}(X), \mathscr{L}(Y))$.

A collection of random variables $X_1, X_2, \dots, X_n$ is said to be associated if for every pair of coordinatewise nondecreasing functions f and g,

$$\mathrm{Cov}(f(\mathbf{X}), g(\mathbf{X})) \ge 0, \tag{2}$$

where $\mathbf{X} = (X_1, X_2, \dots, X_n)$. Throughout we assume, without further explicit mention, that functions f and g are such that $\mathrm{Cov}(f(\mathbf{X}), g(\mathbf{X}))$ exists.

The notion of associated r.v.'s was introduced by Esary, Proschan and Walkup (1967), who also developed the fundamental properties of association and indicated several interesting applications of it in order statistics, analysis of variance and so forth. In the special case $n = 2$, the same authors discussed a weaker concept of association offered by assuming that

$$\operatorname{Cov}(f(X_1), g(X_2)) \geq 0 \tag{3}$$

for all nondecreasing functions f and g. Evidently, if X_1 and X_2 are associated, then they are weakly associated, the converse not being always true. Condition (3) is equivalent to

$$\mathbb{P}(X_1 \geq x_1, X_2 \geq x_2) \geq \mathbb{P}(X_1 \geq x_1)\mathbb{P}(X_2 \geq x_2) \tag{4}$$

for all x_1, x_2, which is, in fact, Lehmann's (1966) definition for positive quadrant dependence (PQD). If (4) holds true for all x_1, x_2 with the inequality sign reversed, the r.v.'s X_1, X_2 are called negatively quadrant dependent (NQD).

A collection of r.v.'s $X_1, X_2, \dots, X_n$ is said to be NA if for every pair of disjoint subsets A_1, A_2 of $\{1, 2, \dots, n\}$,

$$\operatorname{Cov}(f(X_i, i \in A_1), g(X_i, i \in A_2)) \leq 0,$$

where f and g are coordinatewise nondecreasing functions of $\{x_i, i \in A_1\}$ and $\{x_i, i \in A_2\}$, respectively.

For a pair of r.v.'s, NQD is equivalent to NA. The definition of NA r.v.'s was introduced by Joag-Dev and Proschan (1983), who also developed their basic properties and discussed several interesting statistical applications.

We are now ready to state our main results. From now on we assume that *all variables involved are integer valued.*

THEOREM 1. *Let $X_1, X_2, \dots, X_n$ be associated or NA random variables with $\mathbb{E}|X_i|$, $\mathbb{E}|X_i X_j| < \infty$, $i, j = 1, 2, \dots, n$, $i \neq j$. If X_i', $i = 1, 2, \dots, n$, are independent random variables such that X_i' is distributed according to the marginal distribution of X_i $[\mathscr{L}(X_i) = \mathscr{L}(X_i')]$, then*

$$d_K\left(\mathscr{L}\left(\sum_{i=1}^n X_i\right), \mathscr{L}\left(\sum_{i=1}^n X_i'\right)\right) \leq \left|\sum_{i<j} \operatorname{Cov}(X_i, X_j)\right|, \tag{5}$$

$$d_W\left(\mathscr{L}\left(\sum_{i=1}^n X_i\right), \mathscr{L}\left(\sum_{i=1}^n X_i'\right)\right) \leq 2\left|\sum_{i<j} \operatorname{Cov}(X_i, X_j)\right|. \tag{6}$$

Moreover, if $X_1, X_2, \dots, X_n$ are associated or NA nonnegative r.v.'s, then

$$\begin{aligned} 0 \leq{}& \mathbb{P}(X_i = 0, i = 1, 2, \dots, n) \\ &- \prod_{i=1}^n \mathbb{P}(X_i = 0) \leq \sum_{i<j} \operatorname{Cov}(X_i, X_j) \end{aligned} \tag{7}$$

and

$$0 \le \prod_{i=1}^{n} \mathbb{P}(X_i = 0) - \mathbb{P}(X_i = 0, i = 1, 2, \ldots, n) \le -\sum_{i<j} \mathrm{Cov}(X_i, X_j), \tag{8}$$

respectively.

The proof is given in Section 3.

In view of the inequality $d_{\mathrm{TV}} \le d_W$, any upper bound established for d_W can also be used for the total variation distance as well.

It is worth mentioning that for associated r.v.'s, the difference between the characteristic functions of $\sum_{i=1}^{n} X_i$ and $\sum_{i=1}^{n} X_i'$ is bounded (above) by the upper bound in (5) multiplied by t^2 (t is the argument of characteristic functions). This is an immediate consequence of Newman and Wright's (1981) result concerning joint and marginal characteristic functions of associated r.v.'s.

Although Theorem 1 was stated for the class of associated or negatively associated r.v.'s, we could relax this condition and assume that for every $i = 2, 3, \ldots, n$, the r.v.'s X_i and $S_{i-1} = \sum_{j=1}^{i-1} X_j$ are PQD and NQD respectively (as a matter of fact, Theorem 1 is proved in Section 3 under this assumption). In the sequel, we refer to these classes of variables by the terms positively and negatively cumulative dependent (PCD and NCD) r.v.'s.

The fact that the class PCD is wider than the class of associated r.v.'s is easily verified by considering the coordinatewise nondecreasing functions $f_{i,w}(\mathbf{x}) = I_{[w,\infty)}(\sum_{j=1}^{i-1} x_j)$ and $g_{i,y}(\mathbf{x}) = I_{[y,\infty)}(x_i)$ for fixed w, y and i. If $X_1, X_2, \ldots, X_n$ are associated, then $\mathrm{Cov}(f_{i,w}(\mathbf{X}), g_{i,y}(\mathbf{X})) \ge 0$ and whereas

$$\mathrm{Cov}(f_{i,w}(\mathbf{X}), g_{i,y}(\mathbf{X})) = \mathbb{P}\left(\sum_{j=1}^{i-1} X_j \ge w, X_i \ge y\right) - \mathbb{P}\left(\sum_{j=1}^{i-1} X_j \ge w\right)\mathbb{P}(X_i \ge y),$$

we conclude that $X_i, S_{i-1} = \sum_{j=1}^{i-1} X_j$ are PQD; therefore $X_1, X_2, \ldots, X_n$ are PCD. By the same reasoning (note that $f_{i,w}$ and $g_{i,y}$ are defined over disjoint subsets of $\{x_1, x_2, \ldots, x_n\}$), we may easily verify that any set of NA r.v.'s is NCD.

Let us now restrict ourselves to associated binary random variables. A straightforward application of (7) for $1 - X_1, 1 - X_2, \ldots, 1 - X_n$ (which are also associated) reveals the following result.

THEOREM 2. *If $X_1, X_2, \ldots, X_n$ are binary associated r.v.'s, then*

$$0 \le \mathbb{E}\left(\prod_{i=1}^{n} X_i\right) - \prod_{i=1}^{n} \mathbb{E}(X_i) \le \sum_{i<j} \mathrm{Cov}(X_i, X_j). \tag{9}$$

The left inequality gives a well known result established by Esary, Proschan and Walkup (1967) that has been effectively used for the development of reliability bounds of coherent structures [see, e.g., Barlow and Proschan (1981)].

It is of interest to notice that in the case of binary NA random variables $X_1, X_2, \ldots, X_n$, the inequality of Theorem 2 reads

$$0 \le \prod_{i=1}^{n} \mathbb{E}(X_i) - \mathbb{E}\left(\prod_{i=1}^{n} X_i\right) \le -\sum_{i<j} \operatorname{Cov}(X_i, X_j).$$

The aforementioned formulas are also valid under the weaker assumption of PCD and NCD, respectively.

It is also noteworthy that Theorem 1 immediately concludes that if $X_1, X_2, \ldots, X_n$ are binary associated or NA r.v.'s, then

$$d_K\left(\mathscr{L}\left(\sum_{i=1}^{n} X_i\right), \mathrm{GB}\right) \le \left|\sum_{i<j}(\mathbb{P}(X_i = 1, X_j = 1) - \mathbb{P}(X_i = 1)\mathbb{P}(X_j = 1))\right|, \tag{10}$$

where $\mathrm{GB} = \mathscr{L}(\sum_{i=1}^{n} X_i')$ denotes the generalized binomial distribution.

Let us next turn our attention to the problem of approximating the distribution of the sum of associated or NA r.v.'s by a compound Poisson distribution.

THEOREM 3. *Let $X_1, X_2, \ldots, X_n$ be nonnegative associated or NA r.v.'s with $\mathbb{E}(X_i)$, $\mathbb{E}(X_iX_j) < \infty$ for $i, j = 1, 2, \ldots, n$, $i \ne j$. Then*

$$d_K\left(\mathscr{L}\left(\sum_{i=1}^{n} X_i\right), \mathrm{CP}(\lambda_1, \lambda_2, \ldots)\right) \le \left|\sum_{i<j} \operatorname{Cov}(X_i, X_j)\right| + \tfrac{1}{2}\sum_{i=1}^{n} \mathbb{E}(X_i)^2,$$

$$d_W\left(\mathscr{L}\left(\sum_{i=1}^{n} X_i\right), \mathrm{CP}(\lambda_1, \lambda_2, \ldots)\right) \le 2\left|\sum_{i<j} \operatorname{Cov}(X_i, X_j)\right| + \sum_{i=1}^{n} \mathbb{E}(X_i)^2,$$

where $\lambda_j = \sum_{i=1}^{n} \mathbb{P}(X_i = j)$ and $\mathrm{CP}(\lambda_1, \lambda_2, \ldots)$ denotes a compound Poisson distribution with probability generating function $\exp(-\lambda(1 - \sum_{j\ge1}(\lambda_j/\lambda)z^j))$, $\lambda = \sum_{j\ge1} \lambda_j$.

Loosely speaking, Theorem 3 (the proof of which is given in Section 3) states that if the nonnegative associated or NA r.v.'s X_i, $i = 1, 2, \ldots, n$, are almost uncorrelated [i.e., $\sum_{i<j} \operatorname{Cov}(X_i, X_j)$ is negligible] and their distributions are concentrated around zero [e.g., $\mathbb{P}(X_i > 0) \le \mathbb{E}(X_i) = o(n^{-1/2})$], then their sum $\sum_{i=1}^{n} X_i$ can be approximated by an appropriate compound Poisson distribution. Needless to say, Theorem 3 is also valid for the case of PCD and NCD r.v.'s. Theorem 3 can be used to handle almost all cases where a Poisson or compound Poisson Stein–Chen approximation has been applied, provided that the variables involved are associated or NA (e.g., overlapping success runs, etc.). To this end, it suffices to define X_i as the size of clump clustered at the point where the ith occurrence of the event we are interested in took place. We mention that to attack such problems by the Stein–Chen method, a typical declumping technique should first be established [e.g., see Arratia, Goldstein and Gordon (1990)].

We can easily verify that the approximating compound Poisson distribution $\mathrm{CP}(\lambda_1, \lambda_2, \ldots)$ coincides with the distribution of the sum $\sum_{i=1}^{N} Y_i$, where N is a Poisson r.v. with parameter $\lambda = \sum_{i=1}^{n} \mathbb{P}(X_i > 0)$ and Y_i are independent r.v.'s with distribution function $\mathbb{P}(Y_i \le y) = (1/\lambda) \sum_{i=1}^{n} \mathbb{P}(0 < X_i \le y)$ [the latter can also be viewed as a mixture of the conditional distributions of X_i given that $X_i > 0$ with weights $\mathbb{P}(X_i > 0)$, $i = 1, 2, \ldots, n$].

Recently, the powerful Stein–Chen method has become very popular for studying the Poisson approximation of sums of dependent binary variables [see, e.g., Barbour, Holst and Janson (1992)]. Attempting a comparison between the Stein–Chen method and the results reported in this section, we may state that both approaches involve first and second moments and offer error estimates on the same order. The Stein–Chen method applies to binary variables only, whereas Theorem 1 is valid for any set of integer-valued variables; on the other hand, Theorem 1 applies only to associated or NA r.v.'s, whereas the Stein–Chen formula is valid for any set of indicator variables. Finally, the Stein–Chen method offers an estimate of the error incurred when the distribution of a sum of dependent r.v.'s is approximated by a Poisson, compound Poisson or binomial distribution [cf. Arratia, Goldstein and Gordon (1990), Barbour, Holst and Janson (1992), Roos (1994), Barbour and Utev (1998) and Soon (1996)], whereas Theorem 1 provides an estimate of the error incurred when approximating the distribution of a sum of dependent r.v.'s by the convolution of (independent) variables following the marginal distributions of the original set.

In closing we illustrate very briefly how the outcome of Theorem 1 can be used in a simple sampling problem. Suppose a finite population consists of N integer values, $y_1, y_2, \ldots, y_N$. Let $Y'_1, Y'_2, \ldots, Y'_n$ and $Y_1, Y_2, \ldots, Y_n$, be two random samples obtained from this population with and without replacement, respectively. The random variables $Y_1, Y_2, \ldots, Y_n$ are, as Joag-Dev and Proschan (1983) indicated, NA. The population total $y = y_1 + y_2 + \cdots + y_N$ can then be estimated either by $Y' = (N/n) \sum_{i=1}^{n} Y'_i$ or by $Y = (N/n) \sum_{i=1}^{n} Y_i$ [e.g., see Cochran (1977)]. Clearly, the first estimator has larger variance than the second, but because it is expressed in terms of independent r.v.'s, it has more tractable distribution than Y. Were we interested in estimating the discrepancy between the distributions of Y' and Y, we could make use of Theorem 1 to get

$$\begin{aligned} d_K(\mathscr{L}(Y), \mathscr{L}(Y')) &\le -\sum_{i<j} \mathrm{Cov}(Y_i, Y_j) \\ &= \tfrac{1}{2}\left(\mathrm{Var}\left(\sum_{i=1}^{n} Y'_i\right) - \mathrm{Var}\left(\sum_{i=1}^{n} Y_i\right)\right). \end{aligned}$$

Denoting now by $\sigma^2 = \sum_{j=1}^{N}(y_i - y/N)^2/N$ the variance of Y_i in the finite population and recalling the well known formulas $\mathrm{Var}(\sum_{i=1}^{n} Y_i/n) = \sigma^2(N - n)/(n(N-1))$ and $\mathrm{Var}(\sum_{i=1}^{n} Y_i'/n) = \sigma^2/n$, we get

$$d_K(\mathscr{L}(Y), \mathscr{L}(Y')) \le \frac{n(n-1)}{2(N-1)}\sigma^2.$$

Clearly, the RHS of the last inequality is also a bound for $d_K(\mathscr{L}(\sum_{i=1}^{n} Y_i), \mathscr{L}(\sum_{i=1}^{n} Y_i'))$ and for the distance between the distributions of the mean estimators $(1/n)\sum_{i=1}^{n} Y_i$ and $(1/n)\sum_{i=1}^{n} Y_i'$.

If the population consisted of m 1's and $N - m$ 0's, the r.v.'s $\sum_{i=1}^{n} Y_i'$ and $\sum_{i=1}^{n} Y_i$ would follow the binomial distribution with parameters n, $p = m/N$ and a hypergeometrical distribution with parameters n, m and N, respectively. In this case, we have $\sigma^2 = m/N(1 - m/N)$ and, therefore, the distance between the aforementioned distributions is bounded by $(n(n-1))/(2(N-1))m/N(1-m/N)$ [or $(n(n-1))/(8(N-1))$ when m is unknown]. Needless to say, an upper bound for the total variation or Wasserstein distance is twice the aforementioned value. For comparable results through the Stein–Chen method, see Barbour, Holst and Janson (1992) and Soon (1996).

Additional applications that pertain to reliability bounds and approximations of runs, scans and urn model distributions will be reported in subsequent works.

3. Proof of the main results. Before starting the proof process of our main results, we note that if X, Y are integer-valued r.v.'s with finite $\mathbb{E}|X|$, $\mathbb{E}|Y|$, $\mathbb{E}|XY|$, then

$$\begin{aligned} \mathrm{Cov}(X, Y) &= \sum_{(x,y)\in\mathbf{Z}^2} \mathrm{Cov}(I_{[X\ge x]}, I_{[Y\ge y]}) \\ &= \sum_{(x,y)\in\mathbf{Z}^2} \big[\mathbb{P}(X \ge x, Y \ge y) - \mathbb{P}(X \ge x)\mathbb{P}(Y \ge y)\big]. \end{aligned} \tag{11}$$

This is the discrete analogue of an identity presented in Lehmann (1966). The following lemma is an immediate consequence of (4) and (11).

LEMMA 4. *If X and Y are PQD r.v.'s, then*

$$0 \le \sum_{(x,y)\in A} [\mathbb{P}(X \ge x, Y \ge y) - \mathbb{P}(X \ge x)\mathbb{P}(Y \ge y)] \le \mathrm{Cov}(X, Y)$$

for any subset A of $\mathbf{Z}^2$.

We now proceed to another lemma, which will play a crucial role in the development of our main result.

LEMMA 5. *If X and Y are PQD r.v.'s, then*:

(a) $\big|\sum_{w\in A}(\mathbb{P}(X+Y \ge w) - \mathbb{P}(X'+Y' \ge w))\big| \le \mathrm{Cov}(X, Y)$ *for every* $A \subseteq \mathbf{Z}$,
(b) $d_K(\mathscr{L}(X+Y), \mathscr{L}(X'+Y')) \le \mathrm{Cov}(X, Y)$,
(c) $d_W(\mathscr{L}(X+Y), \mathscr{L}(X'+Y')) \le 2\,\mathrm{Cov}(X, Y)$,

where X' and Y' are independent random variables such that $\mathscr{L}(X)=\mathscr{L}(X')$ and $\mathscr{L}(Y)=\mathscr{L}(Y')$.

PROOF. (a) We first express $\mathbb{P}(X+Y\geq w)$ in terms of probabilities on lower bounded quadrants as

$$\mathbb{P}(X+Y\geq w)=\sum_{y=-\infty}^{\infty}(\mathbb{P}(X\geq w-y, Y\geq y)-\mathbb{P}(X\geq w-y, Y\geq y+1)).$$

Whereas X' and Y' are independent and follow the same distribution with X and Y, respectively, we may write, by the same token,

$$\begin{aligned}\mathbb{P}(X'+Y'\geq w)=\sum_{y\in\mathbf{Z}}(&\mathbb{P}(X\geq w-y)\mathbb{P}(Y\geq y)\\&-\mathbb{P}(X\geq w-y)\mathbb{P}(Y\geq y+1)).\end{aligned}$$

A combined use of these expressions yields

$$\begin{aligned}&\sum_{w\in A}(\mathbb{P}(X+Y\geq w)-\mathbb{P}(X'+Y'\geq w))\\&\quad=\sum_{w\in A}\sum_{y\in\mathbf{Z}}(\mathbb{P}(X\geq w-y, Y\geq y)-\mathbb{P}(X\geq w-y)\mathbb{P}(Y\geq y))\\&\qquad-\sum_{w\in A}\sum_{y\in\mathbf{Z}}(\mathbb{P}(X\geq w-y, Y\geq y+1)-\mathbb{P}(X\geq w-y)\mathbb{P}(Y\geq y+1)).\end{aligned}$$

The proof of part (a) is now easily completed by observing that in view of Lemma 4, each of the last two double sums is bounded below by 0 and above by $\mathrm{Cov}(X,Y)$.

(b) Follows readily from part (a) on choosing $A=\{w\}$, $w\in\mathbf{Z}$.

(c) Let I_w denote the sign of the difference $\mathbb{P}(X+Y\geq w)-\mathbb{P}(X'+Y'\geq w)$. By the definition of the Wasserstein distance, we have

$$\begin{aligned}&d_W(\mathscr{L}(X+Y),\mathscr{L}(X'+Y'))\\&\quad=\sum_{w\in\mathbf{Z}:I_w=1}(\mathbb{P}(X+Y\geq w)-\mathbb{P}(X'+Y'\geq w))\\&\qquad-\sum_{w\in\mathbf{Z}:I_w=-1}(\mathbb{P}(X+Y\geq w)-\mathbb{P}(X'+Y'\geq w))\end{aligned}$$

and applying (a) twice, we gain the desired inequality. □

It is also of interest to note that should X and Y be NQD, following an exact parallel to the procedure used before, we may gain the inequalities of Lemma 5 with $\mathrm{Cov}(X,Y)$ replaced by $-\mathrm{Cov}(X,Y)\geq 0$.

The following simple lemma also proves useful for the development of our main result.

LEMMA 6. *If X and Y are PQD and Z is independent of X and Y, then $X+Z, Y$ are also PQD.*

PROOF. The proof results immediately by observing that for every w, y, we have

$$\begin{aligned}\mathbb{P}(X+Z\ge w, Y\ge y) &= \sum_z \mathbb{P}(X\ge w-z, Y\ge y)\mathbb{P}(Z=z)\\ &\ge \sum_z \mathbb{P}(X\ge w-z)\mathbb{P}(Y\ge y)\mathbb{P}(Z=z)\\ &= \mathbb{P}(X+Z\ge w)\mathbb{P}(Y\ge y). \qquad\square\end{aligned}$$

Similar reasoning assures that Lemma 6 is also valid for NQD r.v.'s.

We are now in possession of the "machinery" needed to prove Theorem 1.

PROOF OF THEOREM 1. The proof will be established under the assumption that $X_1, X_2, \dots, X_n$ are PCD. Trivial adjustments in the arguments involved therein (e.g., replacing Cov by $-$Cov, etc.) yield an analogous proof for the case of NCD r.v.'s. Whereas the class of associated and NA r.v.'s is a subclass of PCD and NCD respectively, the proof covers these classes as well.

Let $X_1, X_2, \dots, X_n$ be PCD r.v.'s and define $S_i = \sum_{j=1}^{i} X_j$, $i = 1, 2, \dots, n$. By the definition of PCD, the r.v.'s S_{n-1} and X_n are PQD, and choosing X'_n so that $\mathscr{L}(X_n) = \mathscr{L}(X'_n)$ and X'_n is independent of $X_1, X_2, \dots, X_{n-1}$, allows us to apply Lemma 5(b) to get

$$d_K(\mathscr{L}(S_{n-1}+X_n), \mathscr{L}(S_{n-1}+X'_n)) \le \mathrm{Cov}(S_{n-1}, X_n).$$

Whereas X'_n is independent of S_{n-2}, X_{n-1} and S_{n-2}, X_{n-1} are PQD, Lemma 6 assures that $S_{n-2}+X'_n, X_{n-1}$ are also PQD. Choosing now X'_{n-1} so that $\mathscr{L}(X_{n-1}) = \mathscr{L}(X'_{n-1})$ and X'_{n-1} is independent of $X_1, X_2, \dots, X_{n-2}, X'_n$, we may write, by virtue of Lemma 5(b),

$$\begin{aligned}&d_K(\mathscr{L}((S_{n-2}+X'_n)+X_{n-1}), \mathscr{L}((S_{n-2}+X'_n)+X'_{n-1}))\\ &\quad\le \mathrm{Cov}(S_{n-2}+X'_n, X_{n-1}).\end{aligned}$$

Invoking the triangle inequality, we have

$$\begin{aligned}&d_K(\mathscr{L}(S_n), \mathscr{L}(S_{n-2}+X'_{n-1}+X'_n))\\ &\quad\le \mathrm{Cov}(S_{n-2}+X'_n, X_{n-1}) + \mathrm{Cov}(S_{n-1}, X_n),\end{aligned}$$

which, after taking into account that $\mathrm{Cov}(X'_n, X_{n-1}) = 0$, simplifies to

$$\begin{aligned}&d_K\left(\mathscr{L}\left(\sum_{i=1}^{n} X_i\right), \mathscr{L}\left(\sum_{i=1}^{n-2} X_i + X'_{n-1} + X'_n\right)\right)\\ &\quad\le \mathrm{Cov}(S_{n-2}, X_{n-1}) + \mathrm{Cov}(S_{n-1}, X_n).\end{aligned}$$

It is now clear that by induction, we are led to

$$d_K\left(\mathscr{L}\left(\sum_{i=1}^{n} X_i\right), \mathscr{L}\left(\sum_{i=1}^{n} X'_i\right)\right) \le \sum_{i=2}^{n} \mathrm{Cov}(S_{i-1}, X_i) = \sum_{i=2}^{n}\sum_{j=1}^{i-1} \mathrm{Cov}(X_j, X_i)$$

and the proof of (5) is done.

For the proof of (6), we employ exactly the same reasoning and the respective inequality for d_W as mentioned in Lemma 5(c).

Let us now proceed to the proof of (7). It can be easily verified that for any pair of nonnegative PQD r.v.'s X, Y, we have $\mathbb{P}(X = 0, Y = 0) \geq \mathbb{P}(X = 0)$ $\mathbb{P}(Y = 0)$. Now, if $X_1, X_2, \ldots, X_n$ are nonnegative PCD r.v.'s, then S_{n-1}, X_n are PQD and, therefore,

$$\mathbb{P}(S_n = 0) = \mathbb{P}(S_{n-1} = 0, X_n = 0) \geq \mathbb{P}(S_{n-1} = 0)\mathbb{P}(X_n = 0).$$

By the same token, $\mathbb{P}(S_{n-1} = 0) \geq \mathbb{P}(S_{n-2} = 0)\mathbb{P}(X_{n-1} = 0)$, and working by induction we get the LHS inequality of (7). For the proof of the RHS, observe that

$$\begin{aligned} 0 &\leq \mathbb{P}(X_i = 0, 1 \leq i \leq n) - \prod_{i=1}^{n} \mathbb{P}(X_i = 0) \\ &= \mathbb{P}\left(\sum_{i=1}^{n} X_i = 0\right) - \mathbb{P}\left(\sum_{i=1}^{n} X_i' = 0\right) \\ &\leq d_K\left(\mathscr{L}\left(\sum_{i=1}^{n} X_i\right), \mathscr{L}\left(\sum_{i=1}^{n} X_i'\right)\right) \leq \sum_{i<j} \operatorname{Cov}(X_i, X_j). \end{aligned}$$

□

The next two lemmas, apart from being instrumental in constructing the proof of Theorem 3, contain results that are of independent interest. In the sequel, $\mathrm{Po}(\lambda)$ denotes the Poisson distribution with mean λ.

LEMMA 7. *If* $X_1, X_2, \ldots, X_n$ *are independent binary variables with* $\mathbb{P}(X_i = 1) = p_i$, *then*

$$d_W\left(\mathscr{L}\left(\sum_{i=1}^{n} X_i\right), \mathrm{Po}\left(\sum_{i=1}^{n} p_i\right)\right) \leq \sum_{i=1}^{n} p_i^2.$$

PROOF. Let $Y_1, Y_2, \ldots, Y_n$ be independent r.v.'s such that $\mathscr{L}(Y_i) = \mathrm{Po}(p_i)$. Then, recalling the subadditivity property of d_W for independent summands, we get

$$\begin{aligned} d_W\left(\mathscr{L}\left(\sum_{i=1}^{n} X_i\right), \mathscr{L}\left(\sum_{i=1}^{n} Y_i\right)\right) &\leq \sum_{i=1}^{n} d_W(\mathscr{L}(X_i), \mathscr{L}(Y_i)) \\ &= \sum_{i=1}^{n} \sum_{x=0}^{\infty} |\mathbb{P}(X_i > x) - \mathbb{P}(Y_i > x)| \\ &= \sum_{i=1}^{n} \left(|1 - p_i - e^{-p_i}| + \mathbb{E}(Y_i) - \mathbb{P}(Y_i > 0)\right) \\ &= 2\sum_{i=1}^{n}(e^{-p_i} - 1 + p_i) \leq \sum_{i=1}^{n} p_i^2 \end{aligned}$$

and the proof is complete. □

Whereas $d_K \le d_{\mathrm{TV}} \le d_W$, Lemma 7 implies the well known fact that the total variation distance between a generalized binomial distribution with parameters $p_1, p_2, \ldots, p_n$ and a Poisson distribution with mean $\sum_{i=1}^n p_i$ is bounded from above by the quantity $\sum_{i=1}^n p_i^2$. For a similar result, see Xia (1997).

LEMMA 8. *If $X_i, i = 1, 2, \ldots, n$, are independent nonnegative integer-valued r.v.'s with $\mathbb{E}(X_i) < \infty$, then*

$$d_W\left(\mathscr{L}\left(\sum_{i=1}^n X_i\right), \mathrm{CP}(\lambda_1, \lambda_2, \ldots)\right) \le \sum_{i=1}^n \mathbb{E}(X_i)^2, \tag{12}$$

where $\lambda_j = \sum_{i=1}^n \mathbb{P}(X_i = j)$ and $\mathrm{CP}(\lambda_1, \lambda_2, \ldots)$ denotes a compound Poisson distribution with probability generating function $\exp(-\lambda(1 - \sum_{j\ge1}(\lambda_j/\lambda)z^j))$, $\lambda = \sum_{j\ge1}\lambda_j$.

PROOF. Assume first that X_i's have finite support $\{0, 1, \ldots, m\}$ and express $\sum_{i=1}^n X_i$ as a sum of NA r.v.'s as

$$\sum_{i=1}^n X_i = \sum_{i=1}^n \sum_{j=1}^m jI_{[X_i=j]}$$

[the multinomially distributed r.v.'s $I_{[X_i=j]}$, $j = 1, 2, \ldots, m$, are NA; see Joag-Dev and Proschan(1983) and, therefore, the r.v's $jI_{[X_i=j]}$, $j = 1, \ldots, m$, $i = 1, 2, \ldots, n$, are also NA]. Denoting by I'_{ij}, $j = 1, 2, \ldots, m$, $i = 1, 2, \ldots, n$, a set of independent binary r.v.'s with success probabilities $\mathbb{P}(I'_{ij} = 1) = 1 - \mathbb{P}(I'_{ij} = 0) = \mathbb{P}(X_i = j)$ and employing Theorem 1, we deduce

$$\begin{aligned}
&d_W\left(\mathscr{L}\left(\sum_{i=1}^n \sum_{j=1}^m jI_{[X_i=j]}\right), \mathscr{L}\left(\sum_{i=1}^n \sum_{j=1}^m jI'_{ij}\right)\right)\\
&\quad\le \sum_{i=1}^n d_W\left(\mathscr{L}\left(\sum_{j=1}^m jI_{[X_i=j]}\right), \mathscr{L}\left(\sum_{j=1}^m jI'_{ij}\right)\right)\\
&\quad\le -2\sum_{i=1}^n \sum_{k<j} \mathrm{Cov}(jI_{[X_i=j]}, kI_{[X_i=k]})\\
&\quad= 2\sum_{i=1}^n \sum_{k<j} jk\mathbb{P}(X_i = j)\mathbb{P}(X_i = k)\\
&\quad= \sum_{i=1}^n \mathbb{E}(X_i)^2 - \sum_{i=1}^n \sum_{j=1}^m j^2\mathbb{P}(X_i = j)^2.
\end{aligned}$$

Let now W_j, $j = 1, 2, \ldots, m$, be independent Poisson r.v.'s such that $\lambda_j = \mathbb{E}(W_j) = \sum_{i=1}^{n} \mathbb{P}(X_i = j)$. Employing Lemma 7, we get

$$\begin{aligned}
d_W&\biggl(\mathscr{L}\biggl(\sum_{j=1}^{m} j \sum_{i=1}^{n} I'_{ij}\biggr), \mathscr{L}\biggl(\sum_{j=1}^{m} jW_j\biggr)\biggr)\\
&\le \sum_{j=1}^{m} d_W\biggl(\mathscr{L}\biggl(j\sum_{i=1}^{n} I'_{ij}\biggr), \mathscr{L}(jW_j)\biggr)\\
&= \sum_{j=1}^{m} j d_W\biggl(\mathscr{L}\biggl(\sum_{i=1}^{n} I'_{ij}\biggr), \mathrm{Po}\biggl(\sum_{i=1}^{n} \mathbb{P}(X_i = j)\biggr)\biggr)\\
&\le \sum_{j=1}^{m} j \sum_{i=1}^{n} \mathbb{P}(X_i = j)^2.
\end{aligned}$$

The proof is easily completed if we combine the preceding inequalities to deduce (by virtue of the triangle inequality)

$$\begin{aligned}
d_W&\biggl(\mathscr{L}\biggl(\sum_{i=1}^{n} X_i\biggr), \mathscr{L}\biggl(\sum_{j=1}^{m} jW_j\biggr)\biggr)\\
&\le \sum_{i=1}^{n} \mathbb{E}(X_i)^2 - \sum_{i=1}^{n}\sum_{j=1}^{m} (j^2 - j)\mathbb{P}(X_i = j)^2 \le \sum_{i=1}^{n} \mathbb{E}(X_i)^2
\end{aligned} \tag{13}$$

and observe that the probability generating function of $\sum_{j=1}^{m} jW_j$ is given by

$$\begin{aligned}
\mathbb{E}\bigl(z^{\sum_{j=1}^{m} jW_j}\bigr) &= \prod_{j=1}^{m} \mathbb{E}(z^{jW_j}) = \prod_{j=1}^{m} \exp(-\lambda_j(1 - z^j))\\
&= \exp\biggl(-\lambda\biggl(1 - \sum_{j=1}^{m} \frac{\lambda_j}{\lambda} z^j\biggr)\biggr).
\end{aligned}$$

The extension of inequality (13) to the case of r.v.'s with infinite support ($X_i \in \mathbf{Z}^+$) can be easily achieved by considering first the truncated r.v.'s $Y_{i,m} = X_i I_{[X_i \le m]}$ and a sequence W_j, $j = 1, 2, \ldots,$ of independent Poisson r.v.'s such that $\lambda_j = \mathbb{E}(W_j) = \sum_{i=1}^{n} \mathbb{P}(X_i = j)$. Recalling (1), we may state that for $m = 1, 2, \ldots,$

$$\begin{aligned}
d_W&\biggl(\mathscr{L}\biggl(\sum_{i=1}^{n} X_i\biggr), \mathrm{CP}(\lambda_1, \lambda_2, \ldots)\biggr)\\
&\le d_W\biggl(\mathscr{L}\biggl(\sum_{i=1}^{n} X_i\biggr), \mathscr{L}\biggl(\sum_{i=1}^{n} Y_{i,m}\biggr)\biggr) + \sum_{i=1}^{n} \mathbb{E}(Y_{i,m})^2\\
&\quad + d_W\biggl(\mathscr{L}\biggl(\sum_{j=1}^{m} jW_j\biggr), \mathscr{L}\biggl(\sum_{j=1}^{\infty} jW_j\biggr)\biggr)
\end{aligned}$$

$$\leq \mathbb{E}\left|\sum_{i=1}^{n}(X_i - Y_{i,m})\right| + \sum_{i=1}^{n}\mathbb{E}(Y_{i,m})^2 + \mathbb{E}\left|\sum_{j=m+1}^{\infty} jW_j\right|$$

$$= 2\sum_{i=1}^{n}\mathbb{E}(X_i I_{[X_i>m]}) + \sum_{i=1}^{n}\mathbb{E}(Y_{i,m})^2$$

and, because $Y_{i,m} \to X_i$, $X_i I_{[X_i>m]} \to 0$ (as $m \to \infty$) a.s. and $0 \leq Y_{i,m}$, $X_i I_{[X_i>m]} \leq X_i$, the bounded convergence theorem completes the proof. $\square$

PROOF OF THEOREM 3. If X_i' are independent r.v.'s such that $\mathscr{L}(X_i') = \mathscr{L}(X_i)$, we may write

$$d_W\left(\mathscr{L}\left(\sum_{i=1}^{n} X_i\right), \mathrm{CP}(\lambda_1, \lambda_2, \ldots)\right)$$
$$\leq d_W\left(\mathscr{L}\left(\sum_{i=1}^{n} X_i\right), \mathscr{L}\left(\sum_{i=1}^{n} X_i'\right)\right) + d_W\left(\mathscr{L}\left(\sum_{i=1}^{n} X_i'\right), \mathrm{CP}(\lambda_1, \lambda_2, \ldots)\right)$$

and the proof results immediately by employing Theorem 1 and Lemma 8. The proof of the upper bound for d_K is captured by reasoning similar to that used earlier for d_W. $\square$

Acknowledgment. This work was done while the authors were with the University of Athens, Greece.

REFERENCES

ARRATIA, R., GOLDSTEIN, L. and GORDON, L. (1990). Poisson approximation and the Chen–Stein method. *Statist. Sci.* **5** 403–434.

BARBOUR, A. D., and UTEV, S. (1998). Solving the Stein equation in compound Poisson approximation. *Adv. Appl. Probab.* **30** 449–475.

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*. Clarendon Press, Oxford.

BARLOW, R. E. and PROSCHAN, F. (1981). *Statistical Theory of Reliability and Life Testing*. To Begin With, Silver Spring, MD.

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

COCHRAN, W. G. (1977). *Sampling Techniques*. Wiley, New York.

ESARY, J. D., PROSCHAN, F. and WALKUP, D. (1967). Association of random variables with applications. *Ann. Math. Statist.* **38** 1466–1474.

JOAG-DEV, K. and PROSCHAN, F. (1983). Negative association of random variables with applications. *Ann. Statist.* **11** 286–295.

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

NEWMAN, C. M. and WRIGHT, A. L. (1981). An invariance principle for certain dependent sequences. *Ann. Probab.* **9** 671–675.

ROOS, M. (1994). Stein's method for compound Poisson approximation: The local approach. *Ann. Appl. Probab.* **4** 1177–1187.

SOON, S. Y. T. (1996). Binomial approximation for dependent indicators. *Statist. Sinica* **6** 703–714.

XIA, A. (1997). On using the first difference in the Stein–Chen method. *Ann. Appl. Probab.* **7** 899–916.

DEPARTMENT OF STATISTICS
AND INSURANCE SCIENCE
UNIVERSITY OF PIRAEUS
80 KARAOLI & DIMITRIOU STR.
18534 PIRAEUS
GREECE
E-MAIL: mvbouts@hotmail.com
mkoutras@unipi.gr