Bernoulli **12**(3), 2006, 501–514

Compound Poisson process approximation for locally dependent real-valued random variables via a new coupling inequality

MICHAEL V. BOUTSIKAS

Department of Statistics and Insurance Science, University of Piraeus, 80 Karaoli & Dimitriou Str., 18534 Piraeus, Greece. E-mail: mbouts@unipi.gr

We present a general and quite simple upper bound for the total variation distance d_{TV} between any stochastic process $(X_i)_{i\in\Gamma}$ defined over a countable space Γ, and a compound Poisson process on Γ. This result is sufficient for proving weak convergence for any functional of the process $(X_i)_{i\in\Gamma}$ when the real-valued X_i are rarely non-zero and locally dependent. Our result is established after introducing and employing a generalization of the basic coupling inequality. Finally, two simple examples of application are presented in order to illustrate the applicability of our results.

Keywords: compound Poisson process approximation; coupling inequality; law of small numbers; locally dependent variables; moving sums; rate of convergence; success runs; total variation distance

1. Introduction

Let $(X_i)_{i\in\Gamma}$ be a stochastic process with state space $\mathbb{R}$, where Γ is a countable index set. The main aim of the present work is to provide simple and effective tools for approximating the distribution of any functional of $(X_i)_{i\in\Gamma}$ when the real-valued random variables X_i, $i\in\Gamma$ are locally dependent and rarely differ from zero. This situation appears in numerous applications involving rare and locally dependent events, for example in risk theory, graph theory, extreme value theory, reliability theory, run and scan statistics and biomolecular sequence analysis.

In the simplest case when the X_i are independent, identically distributed (i.i.d.) binary (0–1) random variables, $\sum_i X_i$ follows a binomial distribution, which can be approximated by a Poisson distribution (when $\mathbb{P}(X_i \neq 0) \approx 0$). In the case of dependent X_i, two methods have been mainly used for obtaining Poisson approximation results. The first, initiated by Freedman (1974) and Serfling (1975), concerns sums of dependent indicators $X_1, X_2, \ldots, X_n$ (cf. also Serfling 1978; Serfozo 1986). Typically, this approach offers bounds for the total variation distance between the distribution of the sum of indicators $\sum_i X_i$ and an appropriate Poisson distribution. These bounds are expressed in terms of conditional probabilities of the form $\mathbb{P}(X_i|\mathcal{F}_{i-1})$ assuming that $\{X_i\}$ are adapted to a filtration $\{\mathcal{F}_i\}$. The method exploits coupling techniques, but is also related to martingale theory. For recent developments of this approach we refer to the work of Vellaisamy and Chaudhuri (1999).

1350–7265 © 2006 ISI/BS

The second and most important method for Poisson approximation (for dependent random variables) is based on an adaptation (by Chen 1975) for the Poisson distribution of Stein's technique for normal approximation. This much acclaimed method (referred to as Stein's method for Poisson approximation or the Stein–Chen method) is based on the solution of a difference equation (Stein's equation) but also exploits coupling techniques. It was refined and extended by many authors in various directions and applied to a series of problems and models in diverse research areas. For a complete list of the relevant articles we refer to Barbour *et al.* (1992) and Barbour and Chryssaphinou (2001). In recent years, substantial attention has been drawn to results concerning Poisson process approximation through Stein's method. Refer to Arratia *et al.* (1989) for countable carrier spaces, and to Barbour and Mansson (2002), Chen and Xia (2004) and the references therein for more general carrier spaces.

A third way to obtain compound Poisson approximation results similar to those offered by Stein's method has recently been proposed by Boutsikas and Koutras (2000) for sums of integer-valued associated random variables, Boutsikas and Vaggelatou (2002) for sums of real-valued associated random variables and Boutsikas and Koutras (2001) for sums of locally dependent random variables. These approaches were based on direct probabilistic methods, namely, specific dependence concepts, stochastic orders or coupling techniques.

In this paper we present a compound Poisson process approximation result for locally dependent real-valued random variables. More specifically, we present an upper bound for the total variation distance d_{TV} between the law of any stochastic process $(X_i)_{i\in\Gamma}$ defined over a finite or more generally countable space Γ, and an appropriate compound Poisson process on Γ. This bound is small when X_i, $i \in \Gamma$, are locally dependent and $\mathbb{P}(X_i \neq 0) \approx 0$, $i \in \Gamma$. This result was proved by introducing and exploiting a new coupling inequality that can be considered as a generalization of the well-known basic coupling inequality.

It is remarkable that the form of the bound we provide is similar to the bounds offered by Arratia *et al.* (1989) using Stein's method. Their bounds concern the distance between the law of a sequence of indicators X_i and an appropriate Poisson process, and in that sense our result (which concerns real-valued X_i and a compound Poisson process) can be considered as an extension. It is also worth mentioning that the so-called 'magic factor' or 'Stein factor' (a factor that decreases as λ, the mean of the approximating Poisson distribution, increases) that appears in the upper bound of many Poisson approximation results through Stein's method, cannot be present in our bounds since we use the total variation distance and a compound Poisson process (cf. Barbour *et al.* (1992: 203).

2. Preliminaries

A *random element* Ξ is a measurable mapping from a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to a measurable space $(E, \mathcal{A})$. A *coupling* of two random elements Ξ, Ψ from $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$, $(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$ respectively to $(E, \mathcal{A})$ is any random element (Ξ', Ψ') from $(\Omega, \mathcal{F}, \mathbb{P})$ to $(E \times E, \mathcal{A} \otimes \mathcal{A})$ such that $\mathcal{L}(\Xi) = \mathcal{L}(\Xi')$ and $\mathcal{L}(\Psi) = \mathcal{L}(\Psi')$ where, as usual, $\mathcal{L}(\Xi)$ denotes the law of Ξ. Loosely speaking, a coupling of Ξ, Ψ is any 'definition' of Ξ, Ψ in the same probability space. In order to check how 'close' are the laws $\mathcal{L}(\Xi)$, $\mathcal{L}(\Psi)$ of two random elements Ξ, Ψ we shall be using the well-known total variation distance

$$d_{\mathrm{TV}}(\mathcal{L}(\Xi), \mathcal{L}(\Psi)) = \sup_{A\in\mathcal{A}} |\mathbb{P}_1(\Xi \in A) - \mathbb{P}_2(\Psi \in A)| = \sup_{A\in\mathcal{A}} |\mathbb{P}(\Xi' \in A) - \mathbb{P}(\Psi' \in A)|,$$

which may sometimes be too strong for proving convergence of probability measures (requiring 'similarity' of the two measures in *every* event, whereas, for example, vague convergence requires 'similarity' in events with non-zero measure boundaries), but on the other hand it possesses the useful property

$$d_{\mathrm{TV}}(\mathcal{L}(f(\Xi)), \mathcal{L}(f(\Psi))) \leqslant d_{\mathrm{TV}}(\mathcal{L}(\Xi), \mathcal{L}(\Psi)), \tag{1}$$

for *every* measurable f. Therefore, if $\mathcal{L}(\Xi)$ approximates $\mathcal{L}(\Psi)$ with respect to d_{TV} then, with the same accuracy, the law of any functional of Ξ approximates the law of the same functional of Ψ.

A well-known result concerning d_{TV} is the so-called *basic coupling inequality*: for any coupling (Ξ', Ψ') of two random elements Ξ, Ψ,

$$d_{\mathrm{TV}}(\mathcal{L}(\Xi), \mathcal{L}(\Psi)) \leqslant \mathbb{P}(\Xi' \neq \Psi'). \tag{2}$$

The standard way to ensure that the event $[\Xi' = \Psi']$ belongs to the σ-algebra $\mathcal{F}$ is to restrict ourselves to state spaces $(E, \mathcal{B}(E))$ that are Polish (i.e. complete and separable metric spaces) where $\mathcal{B}(E)$ denotes the usual Borel σ-algebra generated by the open sets in E.

Next, we state two well-known preliminary results – see, for example, Serfling (1978), Wang (1986, 1989) or Barbour *et al.* (1992) – that we will need to call upon. In what follows, whenever dependency or independency of some random elements is mentioned, this will immediately imply that these are defined over the same probability space. The following lemma is easily proved by resorting to the basic coupling inequality and the triangle inequality for d_{TV}.

Lemma 1. *Let* $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ *and* $\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_n$ *be two collections of random vectors* ($\mathbf{X}_i, \mathbf{Y}_i \in \mathbb{R}^{k_i}$, $i = 1, 2, \ldots, n$). *If the couples* $(\mathbf{X}_1, \mathbf{Y}_1), (\mathbf{X}_2, \mathbf{Y}_2), \ldots, (\mathbf{X}_n, \mathbf{Y}_n)$ *are independent then*

$$d_{\mathrm{TV}}(\mathcal{L}(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n), \mathcal{L}(\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_n)) \leqslant \sum_{i=1}^{n} d_{\mathrm{TV}}(\mathcal{L}(\mathbf{X}_i), \mathcal{L}(\mathbf{Y}_i)).$$

As usual, $CP(\lambda, F)$ denotes the distribution of the random sum $\sum_{i=1}^{N} X_i$, where $X_1, X_2, \ldots$ is a sequence of i.i.d. random variables with common distribution F and N is a random variable independent of the X_i following a Poisson distribution with mean λ.

We can now use the above lemma to derive a simple bound for the total variation distance between the joint distribution of a random vector with independent components and a compound Poisson product measure. For the proof of this bound we also use the inequality $d_{\mathrm{TV}}(\mathcal{L}(X), CP(\lambda, F)) \leqslant \mathbb{P}(X \neq 0)^2$, $\lambda = \mathbb{P}(X \neq 0)$, $F(x) = \mathbb{P}(X \leqslant x | X \neq 0)$ which holds for any real-valued random variable X.

Proposition 2. *If* $X_1, X_2, \ldots, X_n$ *are independent real-valued random variables, then*

$$d_{\mathrm{TV}}\left(\mathcal{L}(\mathbf{X}), \prod_{i=1}^{n} CP(\lambda_i, F_i)\right) \leqslant \sum_{i=1}^{n} \mathbb{P}(X_i \neq 0)^2,$$

where $\mathbf{X} = (X_1, X_2, \ldots, X_n)$, $\lambda_i = \mathbb{P}(X_i \neq 0)$ *and* $F_i(x) = \mathbb{P}(X_i \leqslant x | X_i \neq 0)$, $i = 1, 2, \ldots, n$.

The product measure $\prod_{i=1}^{n} CP(\lambda_i, F_i)$ coincides with the distribution $\mathcal{L}(\mathbf{Y})$ of a random vector $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n) \in \mathbb{R}^n$ with independent components where each Y_i follows a $CP(\lambda_i, F_i)$ distribution.

The inequality in Proposition 2, which can be considered a by-product of the basic coupling inequality, can be used to establish compound Poisson process approximation results for sequences of independent random variables. Unfortunately, when we look at cases where the X_i may possibly be dependent, the basic coupling inequality cannot help. In order to obtain similar results for dependent random variables using coupling, it seems reasonable to try first to find an appropriate generalization of the basic coupling inequality. This is the aim of the next section.

3. The generalized coupling inequality

The basic coupling inequality (2) offers a bound for the distance between the laws of two random elements. It would be more flexible, though, to possess a result concerning the change in d_{TV} between the laws of two random elements which occurs when we modify these two elements (e.g. change some of their coordinates). Such a result is offered by the next lemma which, apart from being of independent interest, is the basic ingredient for the establishment of our main result.

Lemma 3. *If* $\Xi_1, \Xi_2, \Psi_1, \Psi_2$ *are four random elements taking values in a Polish space* E, *then*

$$\begin{aligned}|d_{\mathrm{TV}}(\mathcal{L}(\Xi_1), \mathcal{L}(\Xi_2)) - d_{\mathrm{TV}}(\mathcal{L}(\Psi_1), \mathcal{L}(\Psi_2))| \\ \leqslant \mathbb{P}(\Xi_1' \neq \Xi_2', (\Xi_1', \Xi_2') \neq (\Psi_1', \Psi_2')) + \mathbb{P}(\Psi_1' \neq \Psi_2', (\Xi_1', \Xi_2') \neq (\Psi_1', \Psi_2')), \quad (3)\end{aligned}$$

for any coupling $(\Xi_1', \Xi_2', \Psi_1', \Psi_2')$ *of* $\Xi_1, \Xi_2, \Psi_1, \Psi_2$.

Proof. Let $(\Xi_1', \Xi_2', \Psi_1', \Psi_2')$ be a coupling of $\Xi_1, \Xi_2, \Psi_1, \Psi_2$, defined over some probability space $(\Omega, \mathcal{F}, \mathbb{P}$ and taking values in $(E^4, \mathcal{B}(E^4))$. For a fixed $B \in \mathcal{B}(E)$ define the events in $\mathcal{F}$,

$$A_1 = [\Xi_1' \in B], \qquad A_2 = [\Xi_2' \in B], \qquad A_3 = [\Psi_1' \in B], \qquad A_4 = [\Psi_2' \in B].$$

Denoting by A^c the complement of A, it is easy to see that

$$\{\mathbb{P}(A_1) - \mathbb{P}(A_2)\} - \{\mathbb{P}(A_3) - \mathbb{P}(A_4)\}$$
$$= \{\mathbb{P}(A_1 A_2^c) - \mathbb{P}(A_1^c A_2)\} - \{\mathbb{P}(A_3 A_4^c) - \mathbb{P}(A_3^c A_4)\}$$
$$= \{\mathbb{P}(A_1 A_2^c) - \mathbb{P}(A_3 A_4^c)\} - \{\mathbb{P}(A_1^c A_2) - \mathbb{P}(A_3^c A_4)\}$$
$$= \{\mathbb{P}(A_1 A_2^c (A_3 A_4^c)^c) - \mathbb{P}((A_1 A_2^c)^c A_3 A_4^c)\} - \{\mathbb{P}(A_1^c A_2 (A_3^c A_4)^c) - \mathbb{P}((A_1^c A_2)^c A_3^c A_4)\}$$
$$\leqslant \mathbb{P}(A_1 A_2^c (A_3^c \cup A_4)) + \mathbb{P}((A_1 \cup A_2^c) A_3^c A_4)$$
$$= \mathbb{P}(\Xi_1' \in B, \Xi_2' \notin B, (\Psi_1' \notin B \text{ or } \Psi_2' \in B)) + \mathbb{P}((\Xi_1' \in B \text{ or } \Xi_2' \notin B), \Psi_1' \notin B, \Psi_2' \in B),$$

which is bounded from above by $c_1 + c_2$ where

$$c_1 = \mathbb{P}(\Xi_1' \neq \Xi_2', \Xi_1' \neq \Psi_1' \text{ or } \Xi_1' \neq \Xi_2', \Xi_2' \neq \Psi_2') = \mathbb{P}(\Xi_1' \neq \Xi_2', (\Xi_1', \Xi_2') \neq (\Psi_1', \Psi_2')),$$
$$c_2 = \mathbb{P}(\Psi_1' \neq \Psi_2', \Psi_1' \neq \Xi_1' \text{ or } \Psi_1' \neq \Psi_2', \Psi_2' \neq \Xi_2') = \mathbb{P}(\Psi_1' \neq \Psi_2', (\Xi_1', \Xi_2') \neq (\Psi_1', \Psi_2')).$$

Interchanging Ξ_1 with Ψ_1 and Ξ_2 with Ψ_2, we obtain

$$\{\mathbb{P}(A_3) - \mathbb{P}(A_4)\} - \{\mathbb{P}(A_1) - \mathbb{P}(A_2)\} \leqslant c_2 + c_1,$$

and thus, $|\{\mathbb{P}(A_1) - \mathbb{P}(A_2)\} - \{\mathbb{P}(A_3) - \mathbb{P}(A_4)\}| \leqslant c_1 + c_2$. Since $||a| - |b|| \leqslant |a + b|$, for every $a, b \in \mathbb{R}$, we conclude that

$$||\mathbb{P}(A_1) - \mathbb{P}(A_2)| - |\mathbb{P}(A_3) - \mathbb{P}(A_4)|| \leqslant c_1 + c_2.$$

Hence, for any $B \in \mathcal{B}(E)$,

$$|\mathbb{P}(\Xi_1' \in B) - \mathbb{P}(\Xi_2' \in B)| \leqslant |\mathbb{P}(\Psi_1' \in B) - \mathbb{P}(\Psi_2' \in B)| + c_1 + c_2,$$

and

$$|\mathbb{P}(\Psi_1' \in B) - \mathbb{P}(\Psi_2' \in B)| \leqslant |\mathbb{P}(\Xi_1' \in B) - \mathbb{P}(\Xi_2' \in B)| + c_1 + c_2.$$

Considering the supremum with respect to B on both sides of the above inequalities, we deduce that

$$d_{\mathrm{TV}}(\mathcal{L}(\Xi_1), \mathcal{L}(\Xi_2)) \leqslant d_{\mathrm{TV}}(\mathcal{L}(\Psi_1), \mathcal{L}(\Psi_2)) + c_1 + c_2,$$

and

$$d_{\mathrm{TV}}(\mathcal{L}(\Psi_1), \mathcal{L}(\Psi_2)) \leqslant d_{\mathrm{TV}}(\mathcal{L}(\Xi_1), \mathcal{L}(\Xi_2)) + c_1 + c_2,$$

which completes the proof. □

It is easy to see that Lemma 3 can be considered as a generalization of the basic coupling inequality (2). Indeed, if (Ξ_1', Ξ_2') is a coupling of some random elements Ξ_1, Ξ_2, then the inequality in Lemma 3 (considering the coupling $(\Xi_1', \Xi_2', \Xi_1', \Xi_1')$ of $\Xi_1, \Xi_2, \Xi_1, \Xi_1$) leads to

$$d_{\mathrm{TV}}(\mathcal{L}(\Xi_1), \mathcal{L}(\Xi_2)) = |d_{\mathrm{TV}}(\mathcal{L}(\Xi_1), \mathcal{L}(\Xi_2)) - d_{\mathrm{TV}}(\mathcal{L}(\Xi_1), \mathcal{L}(\Xi_1))| \leqslant \mathbb{P}(\Xi_1' \neq \Xi_2').$$

Similar to the basic coupling inequality, (3) can also be used for bounding the d_{TV} between the laws of Ξ_1 and Ξ_2. This can be accomplished by choosing appropriate auxiliary elements Ψ_1, Ψ_2 (more precisely, an appropriate coupling of Ξ_1, Ξ_2, Ψ_1, Ψ_2) so that the upper bound of (3) is small and $d_{\mathrm{TV}}(\mathcal{L}(\Psi_1), \mathcal{L}(\Psi_2))$ can easily be calculated or upper-bounded. This approach offers increased flexibility in the bounding procedure, due to the presence of two additional random elements Ψ_1, Ψ_2 to play with.

The proof of Lemma 1 was based on the triangle inequality and the relation

$$d_{\mathrm{TV}}(\mathcal{L}(\mathbf{X}, \mathbf{Z}), \mathcal{L}(\mathbf{Y}, \mathbf{Z})) = d_{\mathrm{TV}}(\mathcal{L}(\mathbf{X}), \mathcal{L}(\mathbf{Y})), \qquad \mathbf{X}, \mathbf{Y} \in \mathbb{R}^k, \mathbf{Z} \in \mathbb{R}^r, \tag{4}$$

which holds when the random vector $\mathbf{Z}$ is independent of $\mathbf{X}$ and $\mathbf{Y}$ (actually only the $\leqslant$ part of the above relation is needed), which in turn can be proved using the basic coupling inequality (2). Therefore, we cannot use this relation when the random variables involved are possibly dependent. It would thus be very convenient to possess an analogous result for $d_{\mathrm{TV}}(\mathcal{L}(\mathbf{X}, \mathbf{Z}), \mathcal{L}(\mathbf{Y}, \mathbf{Z}))$ that holds even when $\mathbf{Z}$ is dependent on $\mathbf{X}, \mathbf{Y}$. The following corollary of Lemma 3 offers such a result which is quite flexible since it involves an arbitrarily chosen random vector $\mathbf{Z}'$.

Corollary 4. *For any random vectors* $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^k$ *and* $\mathbf{Z}$, $\mathbf{Z}' \in \mathbb{R}^r$ *defined on the same probability space, we have that*

$$|d_{\mathrm{TV}}(\mathcal{L}(\mathbf{Z}, \mathbf{X}), \mathcal{L}(\mathbf{Z}, \mathbf{Y})) - d_{\mathrm{TV}}(\mathcal{L}(\mathbf{Z}', \mathbf{X}), \mathcal{L}(\mathbf{Z}', \mathbf{Y}))| \leqslant 2\mathbb{P}(\mathbf{X} \neq \mathbf{Y}, \mathbf{Z} \neq \mathbf{Z}'). \tag{5}$$

Proof. A direct application of (3) reveals that

$$\begin{aligned} &|d_{\mathrm{TV}}(\mathcal{L}(\mathbf{Z}, \mathbf{X}), \mathcal{L}(\mathbf{Z}, \mathbf{Y})) - d_{\mathrm{TV}}(\mathcal{L}(\mathbf{Z}', \mathbf{X}), \mathcal{L}(\mathbf{Z}', \mathbf{Y}))| \\ &\quad \leqslant \mathbb{P}((\mathbf{Z}, \mathbf{X}) \neq (\mathbf{Z}, \mathbf{Y}), ((\mathbf{Z}, \mathbf{X}), (\mathbf{Z}, \mathbf{Y})) \neq ((\mathbf{Z}', \mathbf{X}), (\mathbf{Z}', \mathbf{Y}))) \\ &\qquad + \mathbb{P}((\mathbf{Z}', \mathbf{X}) \neq (\mathbf{Z}', \mathbf{Y}), ((\mathbf{Z}, \mathbf{X}), (\mathbf{Z}, \mathbf{Y})) \neq ((\mathbf{Z}', \mathbf{X}), (\mathbf{Z}', \mathbf{Y}))) \\ &\quad = \mathbb{P}(\mathbf{X} \neq \mathbf{Y}, \mathbf{Z} \neq \mathbf{Z}') + \mathbb{P}(\mathbf{X} \neq \mathbf{Y}, \mathbf{Z} \neq \mathbf{Z}'). \end{aligned}$$

□

In the following section we shall exploit the above inequality in order to obtain a compound Poisson process result for locally dependent sequences.

4. Compound Poisson process approximation using the generalized coupling inequality

Consider a collection of real-valued random variables X_i, $i \in \Gamma_n = \{1, 2, \ldots, n\}$, and assume that for every X_i there exist a set of indices $B_i \subset \Gamma_n - \{i\}$ such that X_i is independent of or weakly dependent on X_j, $j \in B_i^c \equiv (\Gamma_n - \{i\}) - B_i$. Assume also that the sets B_i, $i \in \Gamma_n$, satisfy the reflexivity condition $j \in B_i \Leftrightarrow i \in B_j$. The set B_i can be considered as the *neighbourhood of strong dependence of* X_i and therefore the set $B_i \cap \Gamma_i$ is the *left neighbourhood of strong dependence of* X_i.

Note that the sets B_i, $i \in \Gamma_n$, can be chosen arbitrarily but the next theorem offers better (smaller) bounds when these sets are chosen so that every X_i is independent of or weakly dependent on all X_j outside its neighbourhood.

Let $X_i^\perp$, $i \in \Gamma_n$, be a sequence of independent random variables (also independent of X_i, $i \in \Gamma_n$) with the same marginal distributions as X_i, $i \in \Gamma_n$ (i.e. $\mathcal{L}(X_i) = \mathcal{L}(X_i^\perp)$, $i \in \Gamma_n$).

Theorem 5. *If X_i, $i \in \Gamma_n$, is a collection of real-valued random variables and B_i, $i \in \Gamma_n$, denotes their neighbourhoods of strong dependence, then*

$$d_{\mathrm{TV}}(\mathcal{L}(X_1, \ldots, X_n), \mathcal{L}(X_1^\perp, \ldots, X_n^\perp)) \leqslant 2\sum_{i=2}^{n} \mathbb{P}\big((X_b)_{b \in B_i \cap \Gamma_i} \neq \mathbf{0}, X_i \neq X_i^\perp\big) + c_{\mathbf{X}}(B), \quad (6)$$

where

$$c_{\mathbf{X}}(B) = \sum_{i=2}^{n} d_{\mathrm{TV}}(\mathcal{L}((X_b)_{b \in B_i^c \cap \Gamma_i}, X_i), \mathcal{L}((X_b)_{b \in B_i^c \cap \Gamma_i}, X_i^\perp)). \quad (7)$$

Proof. Define the random vectors

$$\mathbf{Z}_i = (X_1, \ldots, X_{i-1}), \qquad \mathbf{Z}_i{}' = (X_1 \cdot I(1 \notin B_i), X_2 \cdot I(2 \notin B_i), \ldots, X_{i-1} \cdot I(i-1 \notin B_i)),$$

where $I(A) = 1$ if relation A holds and $I(A) = 0$ otherwise, that is, $\mathbf{Z}_i{}'$ emerges by replacing the 'neighbours' of X_i in $\mathbf{Z}_i$ with zeros. Applying Corollary 4, we obtain

$$\begin{aligned}|d_{\mathrm{TV}}(\mathcal{L}(\mathbf{Z}_i, X_i), \mathcal{L}(\mathbf{Z}_i, X_i^\perp)) - d_{\mathrm{TV}}(\mathcal{L}(\mathbf{Z}_i{}', X_i), \mathcal{L}(\mathbf{Z}_i{}', X_i^\perp))| &\leqslant 2\mathbb{P}(\mathbf{Z}_i \neq \mathbf{Z}_i{}', X_i \neq X_i^\perp)\\ &= 2\mathbb{P}((X_b)_{b \in B_i \cap \Gamma_i} \neq \mathbf{0}, X_i \neq X_i^\perp),\end{aligned}$$

where $\mathbf{0} = (0, 0, \ldots, 0)$. Note now that

$$d_{\mathrm{TV}}(\mathcal{L}(\mathbf{Z}_i{}', X_i), \mathcal{L}(\mathbf{Z}_i{}', X_i^\perp)) = d_{\mathrm{TV}}(\mathcal{L}((X_b)_{b \in B_i^c \cap \Gamma_i}, X_i), \mathcal{L}((X_b)_{b \in B_i^c \cap \Gamma_i}, X_i^\perp)),$$

where the right-hand side is just the left-hand side without the coordinates of $\mathbf{Z}_i{}'$ that are equal to 0 (i.e. all the coordinates with index $j \in B_i \cap \Gamma_i$). This follows from the general equality, $d_{\mathrm{TV}}(\mathcal{L}(\mathbf{X}, \mathbf{0}), \mathcal{L}(\mathbf{Y}, \mathbf{0})) = d_{\mathrm{TV}}(\mathcal{L}(\mathbf{X}), \mathcal{L}(\mathbf{Y}))$ that holds for every $\mathbf{X}$, $\mathbf{Y}$, which can be considered as a special case of (4).

Moreover, using (4) again, we obtain

$$\begin{aligned}d_{\mathrm{TV}}(\mathcal{L}(\mathbf{Z}_i, X_i), \mathcal{L}(\mathbf{Z}_i, X_i^\perp)) &= d_{\mathrm{TV}}(\mathcal{L}(X_1, \ldots, X_i), \mathcal{L}(X_1, \ldots, X_{i-1}, X_i^\perp))\\ &= d_{\mathrm{TV}}(\mathcal{L}(X_1, \ldots, X_i, X, \ldots, X_n^\perp), \mathcal{L}(X_1, \ldots, X_{i-1}, X_i^\perp, \ldots, X_n^\perp)).\end{aligned}$$

Therefore, for $i = 2, \ldots, n$,

$$\begin{aligned}&d_{\mathrm{TV}}(\mathcal{L}(X_1, \ldots, X_i, X_{i+1}^\perp, \ldots, X_n^\perp), \mathcal{L}(X_1, \ldots, X_{i-1}, X_i^\perp, \ldots, X_n^\perp))\\ &\quad\leqslant 2\mathbb{P}\big((X_b)_{b \in B_i \cap \Gamma_i} \neq \mathbf{0}, X_i \neq X_i^\perp\big) + d_{\mathrm{TV}}(\mathcal{L}((X_b)_{b \in B_i^c \cap \Gamma_i}, X_i), \mathcal{L}((X_b)_{b \in B_i^c \cap \Gamma_i}, X_i^\perp)). \quad (8)\end{aligned}$$

Finally, from the triangle inequality we conclude that

$$d_{\mathrm{TV}}(\mathcal{L}(\mathbf{X}), \mathcal{L}(\mathbf{X}^{\perp})) \leqslant \sum_{i=1}^{n} d_{\mathrm{TV}}(\mathcal{L}(X_1, \ldots, X_i, X_{i+1}^{\perp}, \ldots, X_n^{\perp}), \mathcal{L}(X_1, \ldots, X_{i-1}, X_i^{\perp}, \ldots, X_n^{\perp})),$$

which, combined with (8), leads to (6) (from Lemma 1, the first term of the above sum equals 0). □

From the above proof we understand that $\mathbb{P}\big((X_b)_{b\in B_i\cap\Gamma_i} \neq \mathbf{0}, X_i \neq X_i^{\perp}\big) = 0$ when some $B_i \cap \Gamma_i = \varnothing$. Now, in order to minimize the upper bound, it is essential that the X_i are rarely non-zero. This situation immediately calls for a compound Poisson approximation result. Specifically, we state the following theorem.

Theorem 6. *If X_i, $i \in \Gamma_n$, is a collection of real-valued random variables and B_i, $i \in \Gamma_n$, denotes their neighbourhoods of strong dependence, then*

$$d_{\mathrm{TV}}\left(\mathcal{L}(\mathbf{X}), \prod_{i=1}^{n} CP(\lambda_i, F_i)\right) \leqslant 2\sum_{i=2}^{n} \mathbb{P}\big((X_b)_{b\in B_i\cap\Gamma_i} \neq \mathbf{0}, X_i \neq X_i^{\perp}\big) + \sum_{i=1}^{n} \mathbb{P}(X_i \neq 0)^2 + c_{\mathbf{X}}(B),$$

where $\mathbf{X} = (X_1, X_2, \ldots, X_n)$, $\lambda_i = \mathbb{P}(X_i \neq 0)$, $F_i(x) = \mathbb{P}(X_i \leqslant x | X_i \neq 0)$ *and* $c_{\mathbf{X}}(B)$ *is given by* (7).

Proof. From Proposition 2 we obtain that

$$d_{\mathrm{TV}}\left(\mathcal{L}(\mathbf{X}^{\perp}), \prod_{i=1}^{n} CP(\lambda_i, F_i)\right) \leqslant \sum_{i=1}^{n} \mathbb{P}(X_i \neq 0)^2,$$

where $\mathbf{X}^{\perp} = (X_1^{\perp}, X_2^{\perp}, \ldots, X_n^{\perp})$, $\lambda_i = \mathbb{P}(X_i \neq 0)$ and $F_i(x) = \mathbb{P}(X_i \leqslant x | X_i \neq 0)$. The proof is now easily completed by involving the triangle inequality and Theorem 5. □

Given that

$$[(X_b)_{b\in B_i\cap\Gamma_i} \neq \mathbf{0}] \cap [X_i \neq X_i^{\perp}] \subset \bigcup_{b\in B_i\cap\Gamma_i} [X_b \neq 0] \cap ([X_i \neq 0] \cup [X_i^{\perp} \neq 0]),$$

we obtain the next corollary which offers a slightly worse but computationally more convenient bound.

Corollary 7. *If X_i, $i \in \Gamma_n$, is a collection of real-valued random variables and B_i, $i \in \Gamma_n$, denotes their neighbourhoods of strong dependence, then*

$$d_{\mathrm{TV}}\left(\mathcal{L}(\mathbf{X}), \prod_{i=1}^{n} CP(\lambda_i, F_i)\right) \leqslant UB \equiv \sum_{i=1}^{n} \sum_{b\in B_i} (\mathbb{P}(X_b \neq 0, X_i \neq 0) + \mathbb{P}(X_b \neq 0)\mathbb{P}(X_i \neq 0))$$
$$+ \sum_{i=1}^{n} \mathbb{P}(X_i \neq 0)^2 + c_{\mathbf{X}}(B), \tag{9}$$

where $\mathbf{X} = (X_1, X_2, \ldots, X_n)$, $\lambda_i = \mathbb{P}(X_i \neq 0)$, $F_i(x) = \mathbb{P}(X_i \leqslant x | X_i \neq 0)$ *and* $c_{\mathbf{X}}(B)$ *is given by* (7).

Remark 1. By employing arguments similar to those used in Arratia *et al.* (1989: 22), the above inequality could be extended from the finite carrier space Γ_n to an infinite countable carrier space Γ for the process $(X_i)_{i\in\Gamma}$.

Remark 2. It is worth stressing that $c_{\mathbf{X}}(B) = 0$ when each X_i is independent of X_b, $b \in B_i^c$ (B_i is the *neighbourhood of dependence* of X_i). However, if each X_i is 'weakly' dependent on X_b, $b \in B_i^c$ (B_i is the *neighbourhood of strong dependence* of X_i) then $c_{\mathbf{X}}(B)$ could be bounded from above. For example, if $X_1, X_2, \ldots$ is a ϕ-*mixing* sequence of integer-valued random variables, then on choosing $B_i \cap \Gamma_i = \{i - s + 1, \ldots, i - 1\}$ for some $s > 1$ and denoting $\mathbf{Y} = (X_b)_{b \in B_i^c \cap \Gamma_i} = (X_1, \ldots, X_{i-s})$, we deduce that

$$d_{\mathrm{TV}}(\mathcal{L}(\mathbf{Y}, X_i), \mathcal{L}(\mathbf{Y}, X_i^{\perp})) = \sup_A \left| \sum_j \{\mathbb{P}((\mathbf{Y}, j) \in A | X_i = j) - \mathbb{P}((\mathbf{Y}, j) \in A)\} \mathbb{P}(X_i = j) \right| \leqslant \phi(s),$$

where

$$\phi(s) = \sup_k \{\sup\{|\mathbb{P}(B|C) - \mathbb{P}(B)|, C \in \sigma(X_i; i \leqslant k), B \in \sigma(X_i; i \geqslant k + s)\}\} \underset{s\to\infty}{\longrightarrow} 0$$

(e.g. for a Doeblin irreducible Markov chain, $\phi(s) \leqslant ab^s$ for some $a > 0$, $0 \leqslant b < 1$). In a similar way we can treat α-*mixing* (strongly mixing) or other types of weakly dependent sequences.

Remark 3. Theorem 6 or Corollary 7 can be used to prove weak convergence for any function of the process $\mathbf{X}$. More specifically, from (1) and (9) we conclude that

$$\sup_{A \in \mathcal{B}(\mathbb{R}^k)} |\mathbb{P}(f(\mathbf{X}) \in A) - \mathbb{P}(f(\mathbf{Y}) \in A)| = d_{\mathrm{TV}}(\mathcal{L}(f(\mathbf{X})), \mathcal{L}(f(\mathbf{Y}))) \leqslant UB,$$

for any measurable function $f : \mathbb{R}^{|\Gamma|} \to \mathbb{R}^k$ where $\mathbf{Y} \sim \prod_{i\in\Gamma} CP(\lambda_i, F_i)$. If, for example, we choose $f(\mathbf{x}) = \sum_i x_i$, then it readily follows that

$$d_{\mathrm{TV}}\left(\mathcal{L}\left(\sum X_i\right), CP\left(\lambda, \sum \frac{\lambda_i}{\lambda} F_i\right)\right) \leqslant UB,$$

a generalization of the *Khinchine–Doeblin* inequality (here $\lambda = \sum_i \lambda_i$, and UB is given by (9)). Other choices of f could, for example, be $f(\mathbf{x}) = (\sum_{i\in C_1} x_i, \sum_{i \in C_2} x_i)$ with $C_1, C_2 \subset \Gamma$, or $f(\mathbf{x}) = (\max_i x_i, \sum_i x_i)$.

5. Applications

As already mentioned in the Introduction, the bound of Corollary 7 has almost the same form as the bounds developed with the aid of the Stein–Chen method (see Arratia *et al.* 1989, 1990, Barbour *et al.* 1992) for Poisson approximation. Consequently, (9) can almost directly be applied to many of the problems where the Stein–Chen method has been applied in the past. These models include problems from graph theory, extreme value theory, run and scan statistics, biomolecular sequence analysis, risk theory, and reliability theory. Moreover, bound (9) is almost identical to the bounds offered by Boutsikas and Koutras (2001) for compound Poisson approximation (via the Kolmogorov distance). The same authors applied these bounds to problems related to scan statistics in risk theory (Boutsikas and Koutras 2002). Next, we present two simple applications elucidating the techniques employed when applying Corollary 7.

Example 1 *Compound Poisson process approximation for overlapping success runs in i.i.d. trials*. Let $\{Z_i\}_{i\in\mathbb{Z}}$ be a sequence of i.i.d. binary trials with $\mathbb{P}(Z_i=1)=p$, $\mathbb{P}(Z_i=0)=q$, $p+q=1$. We are interested in the appearances of overlapping success runs (runs of 1s) of length k in trials $1, 2, \ldots, n$. This model has been studied by many authors in the past; see, for example, Barbour *et al.* (1992) and the relevant references therein.

Define $X_i=\prod_{j=i}^{i+k-1} Z_j$, $i=1, 2, \ldots, n-k+1$. The random vector $\mathbf{X}=(X_1, X_2, \ldots, X_{n-k+1})\in\{0, 1\}^{n-k+1}$ indicates the starting points of the observed overlapping success runs. In this case X_i is dependent only on $X_{i-k+1}, \ldots, X_{i+k-1}$, and therefore we can conveniently choose

$$B_i=\{\max\{1, i-k+1\}, \ldots, i-1, i+1, \ldots, \min\{n-k+1, i+k-1\}\}.$$

With the above choice, $c_{\mathbf{X}}(B)=0$ and since $F_i(x)=I(x\geqslant 1)$ we see that $CP(\lambda_i, F_i)\equiv Po(\lambda_i)=Po(p^k)$ (Poisson distribution with parameter p^k). From Corollary 7 we readily obtain that

$$d_{\mathrm{TV}}\left(\mathcal{L}(\mathbf{X}), \prod_{i=1}^{n-k+1} Po(p^k)\right)\leqslant 2\sum_{i=2}^{n-k+1}\sum_{b=\max\{1,i-k+1\}}^{i-1}\left(\mathbb{P}(X_b=1, X_i=1)+p^{2k}\right)+\sum_{i=1}^{n-k+1} p^{2k}$$

$$\leqslant (n-k+1)p^k\left(2p\frac{1-p^{k-1}}{1-p}+(2k-1)p^k\right),$$

and therefore $\mathcal{L}(\mathbf{X})$ can be approximated by a Poisson process with intensity p^k on the carrier space Γ_n when n is large, p is small (k is fixed) and $np^k\to\lambda$ (the upper bound is of order $O(p)$).

The above bound cannot be used when we assume that $n\to\infty$, $k\to\infty$ and p is fixed. Under these conditions, the success runs tend to occur in 'clumps' (clusters of adjacent success runs). The occurrences of these clumps are rare and asymptotically independent while each clump consists of a random number of overlapping success runs. This situation readily calls for a compound Poisson approximation result. To achieve this, let $Y_1, Y_2, \ldots, Y_{n-k+1}$ represent the sizes of the clumps started at trials $1, 2, \ldots, n-k+1$ respectively, that is,

$$Y_i=(1-Z_{i-1})\sum_{r=0}^{n-i-k+1}\prod_{j=i}^{i+k+r-1} Z_j,\quad i=2, 3, \ldots, n-k+1,\qquad Y_1=\sum_{r=0}^{n-k}\prod_{j=1}^{k+r} Z_j.$$

Obviously, $\sum_{i=1}^{n-k+1} Y_i$ is equal to $\sum_{i=1}^{n-k+1} X_i$, the total number of overlapping success runs within trials 1, 2, ..., n. Instead of Y_i, it is more convenient to use the random variables

$$Y_i{}' = (1 - Z_{i-1}) \sum_{r=0}^{k-1} \prod_{j=i}^{i+k+r-1} Z_j, \qquad i = 1, 2, \ldots, n-k+1,$$

which represent the *truncated* sizes of clumps ($Y_i{}' \leqslant k$) starting at positions $1, 2, \ldots, n-k+1$ (to obtain stationarity, we have also allowed the last clumps to extend further than trial n). When k, n increase while the expected number of runs $(n-k+1)p^k$ remains bounded, the processes $\mathbf{Y} = (Y_1, \ldots, Y_{n-k+1})$, $\mathbf{Y}' = (Y_1', \ldots, Y_{n-k+1}')$ rarely differ. This is expressed by the inequality

$$\begin{aligned} d_{\mathrm{TV}}(\mathcal{L}(\mathbf{Y}), \mathcal{L}(\mathbf{Y}')) \leqslant \mathbb{P}(\mathbf{Y} \neq \mathbf{Y}') &\leqslant \mathbb{P}\left(\bigcup_{i=1}^{n-2k+1} [Z_{i-1} = 0, Z_i = \ldots = Z_{i+2k-1} = 1]\right) \\ &\quad + \mathbb{P}(Z_0 = \ldots = Z_k = 1) \\ &\quad + \mathbb{P}\left(\bigcup_{i=n-2k+3}^{n-k+1} [Z_{i-1} = 0, Z_i = \ldots = Z_{n+1} = 1]\right) \\ &\leqslant (n-2k+1)qp^{2k} + p^{k+1} + p^{k+1}\left(1 - p^{k-1}\right) \leqslant (n-2k+1)qp^{2k} + 2p^{k+1}. \end{aligned}$$

Now, consider $B_i \cap \Gamma_i = \{\max\{1, i-2k+1\}, \ldots, i-1\}$ and apply Corollary 7 to obtain the inequality

$$\begin{aligned} &d_{\mathrm{TV}}\left(\mathcal{L}(\mathbf{Y}'), \prod_{i=1}^{n-k+1} CP(\lambda_i, F)\right) \\ &\quad \leqslant 2 \sum_{i=2}^{n-k+1} \sum_{b=\max\{1, i-2k+1\}}^{i-1} (\mathbb{P}(Y_i{}' \neq 0, Y_b{}' \neq 0) + \mathbb{P}(Y_i{}' \neq 0)\mathbb{P}(Y_b{}' \neq 0)) \\ &\qquad + \sum_{i=1}^{n-k+1} \mathbb{P}(Y_i{}' \neq 0)^2, \end{aligned}$$

where $\lambda_i = \mathbb{P}(Y_i{}' \neq 0)$ and $F(x) = \mathbb{P}(Y_i{}' \leqslant x | Y_i{}' \neq 0)$, $x \in \mathbb{R}$ ($c_{\mathbf{X}}(B) = 0$). It is easy to check that

$$\begin{aligned} &\mathbb{P}(Y_i{}' \neq 0) = qp^k, && i = 1, 2, \ldots, n-k+1, \\ &\mathbb{P}(Y_i{}' \neq 0, Y_b{}' \neq 0) = 0, && b = i-k, \ldots, i-1, \\ &\mathbb{P}(Y_i{}' \neq 0, Y_b{}' \neq 0) \leqslant q^2 p^{2k}, && b = i-2k+1, \ldots, i-k-1, \end{aligned}$$

and $\lambda_i = qp^k$, $F(x) = 1 - p^x$, $x = 1, 2, \ldots, k$. Using the above and the triangle inequality we obtain

$$d_{\mathrm{TV}}\left(\mathcal{L}(\mathbf{Y}), \prod_{i=1}^{n-k+1} CP(\lambda_i, F)\right) \leqslant 2(n-k)(k-1)q^2 p^{2k} + 2(n-k)(2k-1)q^2 p^{2k}$$
$$+ (n-k+1)q^2 p^{2k} + (n-2k+1)qp^{2k} + 2p^{k+1}$$
$$\leqslant \lambda p^k(1+(6k-3)q) + 2p^{k+1},$$

where $\lambda = (n-k+1)qp^k$. Obviously, if $n, k \to \infty$ so that $(n-k+1)qp^k \to \lambda_0 \in (0, \infty)$ then the upper bound vanishes and the law of the clump process $\mathbf{Y}$ can be approximated by a compound Poisson process (the convergence rate being of order $O(kp^k)$). Therefore, according to Remark 3 above, we can obtain weak convergence results (as $n, k \to \infty$) for any functional $f(\mathbf{Y})$ of the process $\mathbf{Y}$. For example, choosing $f(\mathbf{y}) = \sum y_i$, the distance $d_{\mathrm{TV}}(\mathcal{L}(\sum Y_i), CP(\lambda, F))$ is bounded above by the same quantity $\lambda p^k(1+(6k-3)q) + 2p^{k+1}$ and, therefore, $\sum Y_i$ (the total number of overlapping success runs) follows asymptotically (as $n, k \to \infty$) a compound Poisson with geometric compounding distribution (Pólya–Aeppli distribution). Note, though, that for this special case better bounds can be obtained via the Stein–Chen method that include the so-called 'magic factor' – see, for example, Barbour *et al.* (2001).

Example 2 *Compound Poisson process approximation for the total excess amount above a high threshold for moving sums of i.i.d. random variables*. In this example we are interested in the exceedances of the moving sum (r-scan process)

$$S_i = \sum_{j=i}^{i+r-1} X_j, \qquad i = 1, 2, \ldots, n-r+1,$$

above a threshold b, where $X_1, X_2, \ldots, X_n$ is a sequence of i.i.d. non-negative unbounded random variables with a common distribution function F. More specifically we are interested in the process of excess values ('peaks over threshold b')

$$Y_i = \max\{S_i - b, 0\}, \qquad i = 1, 2, \ldots, n-r+1.$$

Obviously, the Y_i are locally dependent and if we choose b to be 'high' (so that the Y_i are rarely non-zero) then it is clear that the process of excess values $\mathbf{Y}$ can be approximated by an appropriate compound Poisson process. Dembo and Karlin (1992) studied the number of exceedances $\sum_{i=1}^{n-r+1} I(S_i > b)$ and proved (using the Stein–Chen method) that, under appropriate conditions, the number of exceedances converges to a Poisson distribution.

In a more general set-up, Rootzén *et al.* (1998) considered strongly mixing stationary sequences $\{X_i\}$ and offered results pertaining to the asymptotic distribution of tail array sums of the general form $\sum \psi(X_i - b)$ for a class of real functions ψ (which includes the case $\psi(x) = \max\{0, x\}$ considered above). They proved that, under appropriate conditions, tail array sums converge to a compound Poisson distribution (for very high levels of b). Note, though, that their approach does not provide any bounds or convergence rates, and the parameters of the limiting $CP(\lambda, G)$ were not explicitly described.

Boutsikas and Koutras (2001) proved that the sum of excess values converges to a compound Poisson distribution. Here, following essentially the same steps, we employ Corollary 7 to obtain a compound Poisson process approximation for the law of the process of excess values $\mathbf{Y} = (Y_i)_{i\in\Gamma_{n-r+1}}$. As in Boutsikas and Koutras (2001), we first choose the left neighbourhoods of dependence $B_i \cap \Gamma_i = \{\max\{i - r + 1, 1\}, \ldots, i - 1\}$ and then apply Corollary 7 to obtain

$$d_{\mathrm{TV}}\left(\mathcal{L}(\mathbf{Y}), \prod_{i=1}^{n-r+1} CP(\lambda_i, G_b)\right) \leqslant 2\lambda\sum_{m=2}^{r}\mathbb{P}(S_m > b|S_1 > b) + 2\lambda r(1 - F^{(r)}(b)) := \lambda\varepsilon(r, b), \tag{10}$$

where $\lambda_i = \mathbb{P}(S_i > b) = (1 - F^{(r)}(b))$, $\lambda = (n - r + 1)(1 - F^{(r)}(b))$ and $F^{(m)}$ denotes the m-fold convolution of F. The compounding distribution G_b is given by the expression

$$G_b(x) = \mathbb{P}(S_i \leqslant x + b|S_i > b) = 1 - \frac{1 - F^{(r)}(b + x)}{1 - F^{(r)}(b)}, \qquad x \geqslant 0.$$

It can be shown (see Theorem 3 of Dembo and Karlin 1992) that $\epsilon(r, b) \to 0$ for any fixed r, provided that for each constant $K > 0$,

$$\frac{1 - F(b - K)}{1 - F^{(2)}(b)} \to 0 \qquad \text{as } b \to \infty. \tag{11}$$

According to Dembo and Karlin (1992), condition (11) holds for any distribution F which is a finite or infinite convolution of exponentials of any scale parameters or has a log-concave density. Therefore, if the common distribution F of the X_i satisfies (11) then the law of the process $\mathbf{Y}$ of excess values can be approximated by a compound Poisson process and, as in the previous cases, we may establish weak convergence for any functional of the process $\mathbf{Y}$. For example, we can obtain that the sum of excess values $\sum_{i=1}^{n-r+1} Y_i$ converges to a compound Poisson distribution $CP(\lambda, G)$ with a convergence rate given by (10) provided that $n, b \to \infty$ (r is fixed) so that $n(1 - F^{(r)}(b)) \to \lambda \in (0, \infty)$ and $G_b(x) \to_{b\to\infty} G(x)$.

References

Arratia, R., Goldstein, L. and Gordon, L. (1989) Two moments suffice for Poisson approximations: the Chen–Stein method. *Ann. Probab.*, **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 Chryssaphinou O. (2001) Compound Poisson approximation: a user's guide. *Ann. Appl. Probab.*, **11**, 964–1002.

Barbour, A.D. and Mansson, M. (2002) Compound Poisson process approximation. *Ann. Probab.*, **30**, 1492–1537.

Barbour, A.D., Holst, L. and Janson, S. (1992) *Poisson Approximation*. Oxford: Clarendon Press.

Barbour, A.D., Chryssaphinou, O. and Vaggelatou, E. (2001) Applications of compound Poisson approximation. In Ch. A. Charalambides, M.V. Koutras and N. Balakrishnan (eds), *Probability and Statistical Models with Applications*, pp. 41–62. Boca Raton, FL: Chapman & Hall.

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. Probab.*, **10**, 1137–1150.

Boutsikas, M.V. and Koutras, M.V. (2001) Compound Poisson approximation for sums of dependent random variables. In Ch.A. Charalambides, M.V. Koutras, N. Balakrishnan (eds), *Probability and Statistical Models with Applications*, pp. 63–86. Boca Raton, FL: Chapman & Hall.

Boutsikas, M.V. and Koutras, M.V. (2002) Modelling claim exceedances over thresholds. *Insurance Math. Econom.*, **30**, 67–83.

Boutsikas, M.V. and Vaggelatou, E. (2002) On the distance between convex-ordered random variables. *Adv. Appl. Probab.*, **34**, 349–374.

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

Chen, L.H.Y. and Xia, A. (2004) Stein's method, Palm theory and Poisson process approximation. *Ann. Probab.*, **32**, 2545–2569.

Dembo, A. and Karlin, S. (1992) Poisson approximations for *r*-scan processes. *Ann. Appl. Probab.*, **2**, 329–357.

Freedmen, D. (1974) The Poisson approximation for dependent events. *Ann. Probab.*, **2**, 256–269.

Rootzén, H., Leadbetter, M.R. and De Haan, L. (1998) On the distribution of tail array sums for strongly mixing stationary sequences. *Ann. Appl. Probab.*, **8**, 868–885.

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

Serfling, R.J. (1978) Some elementary results on Poisson approximations in a sequence of Bernoulli trials. *SIAM Rev.*, **20**, 567–579.

Serfozo, R. (1986) Compound Poisson approximation for sums of random variables *Ann. Probab.*, **14**, 1391–1398.

Vellaisamy, P. and Chaudhuri, B. (1999) On compound Poisson approximation for sums of random variables. *Statist. Probab. Lett.*, **41**, 179–189.

Wang, Y.H. (1986) Coupling methods in approximations. *Canad. J. Statist.*, **14**, 69–74.

Wang, Y.H. (1989) From Poisson to compound Poisson approximations. *Math. Sci.*, **14**, 38–49.

Received November 2004 and revised September 2005