Electron. Commun. Probab. **18** (2013), no. 30, 1–12.
DOI: 10.1214/ECP.v18-2751
ISSN: 1083-589X

**ELECTRONIC COMMUNICATIONS in PROBABILITY**

# Compound Poisson approximation with association or negative association via Stein's method

Fraser Daly*

**Abstract**

In this note we show how the properties of association and negative association can be combined with Stein's method for compound Poisson approximation. Applications include $k$–runs in iid Bernoulli trials, an urn model with urns of limited capacity and extremes of random variables.

**Keywords:** Compound Poisson approximation; Stein's method; (negative) association; runs; urn model with overflow; extremes.
**AMS MSC 2010:** 62E17.
Submitted to ECP on December 9, 2009, final version accepted on April 14, 2013.

## 1 Introduction and main results

In recent years Stein's method has proved to be an effective technique for probability approximation, often yielding explicit error bounds and working well in the presence of dependence. Stein's method may be applied in a wide variety of settings: in this note we consider compound Poisson approximation. See [1] and references therein for an introduction to Stein's method for compound Poisson approximation, and Stein's technique more generally.

Our purpose in this note is to show how assumptions of association or negative association may be combined with Stein's method in a compound Poisson approximation setting. This provides an analogue of the idea of a 'monotone coupling' in Stein's method for Poisson approximation. See, for example, [3, Section 2.1]. In a Poisson approximation setting, the existence of a monotone coupling means that error bounds obtained via Stein's method are often simpler to state and easier to evaluate in practice than they would otherwise be. The same is true if we make assumptions of association or negative association in a compound Poisson approximation setting, as will be demonstrated in the applications of Section 2.

This work is organised as follows. The remainder of Section 1 is devoted to introducing the notation and ideas we will need, and stating our main results. Applications of these results are discussed in Section 2, with the proof of our main theorems being given in Section 3.

Throughout this work, we assume that $X_1, \dots, X_n$ are (possibly dependent) non–negative integer valued random variables. We consider compound Poisson approximation for their sum $W = X_1 + \cdots + X_n$. Recall that $X_1, \dots, X_n$ are said to be associated if

$$E[f(X_i, 1 \leq i \leq n)g(X_i, 1 \leq i \leq n)] \geq E[f(X_i, 1 \leq i \leq n)]E[g(X_i, 1 \leq i \leq n)],$$

*Heilbronn Institute for Mathematical Research, School of Mathematics, University of Bristol, UK.
E-mail: fraser.daly@bristol.ac.uk

for all non–decreasing functions $f$ and $g$. On the other hand, $X_1, \dots, X_n$ are said to be negatively associated if

$$E[f(X_i, i \in \Gamma_1) g(X_i, i \in \Gamma_2)] \leq E[f(X_i, i \in \Gamma_1)] E[g(X_i, i \in \Gamma_2)],$$

for all non–decreasing functions $f$ and $g$, and all $\Gamma_1, \Gamma_2 \subseteq \{1, \dots, n\}$ with $\Gamma_1 \cap \Gamma_2 = \emptyset$. See [7] and references therein for further discussion of these properties.

We will say that $U \sim \mathrm{CP}(\lambda, \boldsymbol{\mu})$ has a compound Poisson distribution if $U \stackrel{d}{=} \sum_{i=1}^{N} Y_i$, where the $Y_i$ are positive integer valued random variables with $P(Y_i = j) = \mu_j$ for each $i$, $N \sim \mathrm{Po}(\lambda)$ has a Poisson distribution and each of these random variables are independent. We write $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots)$ and $\lambda_j = \lambda \mu_j$. Thus, $\lambda = \sum_{j \geq 1} \lambda_j$ and $\mu_j = \lambda^{-1} \lambda_j$.

We follow, for example, [13] and for each $i$ we consider a 'neighbourhood of dependence' consisting of those indices $j \in \{1, \dots, i-1, i+1, \dots, n\}$ for which $X_i$ and $X_j$ are strongly dependent, in some sense. These neighbourhoods of dependence are chosen to suit the problem at hand. We will denote by $\mathcal{J}(i)$ the neighbourhood of dependence of $X_i$. We then define

$$Z_i = \sum_{j \in \mathcal{J}(i)} X_j \,, \qquad \text{and} \qquad W_i = W - X_i - Z_i \,.$$

As in the work of Barbour et al. [1] or Roos [13], we then define our approximating compound Poisson random variable $U$ by setting

$$\lambda_j = \frac{1}{j} \sum_{i=1}^{n} E[X_i I(X_i + Z_i = j)] \,, \tag{1.1}$$

for each $j \geq 1$.

In this note we consider approximation in total variation distance, although our results may also be applied to approximation in other probability metrics. The total variation distance between random variables $W$ and $U$ supported on $\mathbb{Z}^+$ is defined by

$$d_{TV}(\mathcal{L}(W), \mathcal{L}(U)) = \sup_{A \subseteq \mathbb{Z}^+} |P(W \in A) - P(U \in A)| \,.$$

### 1.1 Stein's method for compound Poisson approximation

Before stating our main results, we give a brief outline of Stein's method for compound Poisson approximation. For further details, see [1]. Letting $U \sim \mathrm{CP}(\lambda, \boldsymbol{\mu})$, we begin by finding for each $A \subseteq \mathbb{Z}^+$ a function $f_A$ such that $f_A(0) = 0$ and

$$I(x \in A) - P(U \in A) = \sum_{j \geq 1} j \lambda_j f_A(x + j) - x f_A(x) \,,$$

for each $x \in \mathbb{Z}^+$. Note that the functions $f_A$ will depend on the choices of $\lambda$ and $\boldsymbol{\mu}$. Given such functions $f_A$, by replacing $x$ by $W$ and taking expectation in the above we may then write

$$d_{TV}(\mathcal{L}(W), \mathcal{L}(U)) = \sup_{A \subseteq \mathbb{Z}^+} \left| E\left[ \sum_{j \geq 1} j \lambda_j f_A(W + j) - W f_A(W) \right] \right| \,. \tag{1.2}$$

In proving our main results in Section 3, we proceed by bounding the right–hand side of (1.2). In doing so, it will be essential to have bounds on

$$H(\lambda, \boldsymbol{\mu}) = \sup_{A \subseteq \mathbb{Z}^+} \sup_{x \in \mathbb{Z}^+} |\Delta f_A(x)|,$$

where we use $\Delta$ to denote the forward difference operator, so that $\Delta f(x) = f(x+1) - f(x)$ for any function $f$.

Good bounds on $H(\lambda, \boldsymbol{\mu})$ can be hard to find. Barbour et al. [1, Theorem 4] show that

$$H(\lambda, \boldsymbol{\mu}) \le \min\left\{1, \frac{1}{\lambda_1}\right\} e^{\lambda},$$

and, furthermore, that we cannot do better than this in general. This bound is useful only for very small $\lambda$. However, under particular conditions we can find much better bounds. For example, if we define

$$\nu_i = i\lambda_i - (i+1)\lambda_{i+1}\,,$$

and assume that

$$\nu_i \ge 0 \qquad \forall i \ge 1\,, \tag{1.3}$$

then Barbour et al. [1, Theorem 5] show that

$$H(\lambda, \boldsymbol{\mu}) \le \min\left\{1, \frac{1}{\nu_1}\left(\frac{1}{4\nu_1} + \log^+(2\nu_1)\right)\right\}\,, \tag{1.4}$$

$\log^+$ denoting the positive part of the natural logarithm. Alternatively, under the assumption that

$$\theta = \frac{\sum_{i\ge 2} i(i-1)\lambda_i}{\sum_{i\ge 1} i\lambda_i} < \frac{1}{2}\,, \tag{1.5}$$

Barbour and Xia [5, Theorem 2.5] show that

$$H(\lambda, \boldsymbol{\mu}) \le \frac{1}{(1-2\theta)\sum_{i\ge 1} i\lambda_i}\,. \tag{1.6}$$

### 1.2 Main results

We are now in a position to state our main results. The proofs of Theorems 1.1 and 1.2 are deferred until Section 3.

**Theorem 1.1.** *Suppose that $X_1, \ldots, X_n$ are negatively associated, and define $\lambda_j$ as in (1.1) for each $j \ge 1$. Then*

$$d_{TV}(\mathcal{L}(W), CP(\lambda, \boldsymbol{\mu})) \le H(\lambda, \boldsymbol{\mu}) \left\{ \sum_{i=1}^{n} \sum_{j \in \mathcal{J}(i)\cup\{i\}} E[X_i X_j] - \mathrm{Var}(W) \right\}.$$

**Theorem 1.2.** *Suppose that $X_1, \ldots, X_n$ are associated, and define $\lambda_j$ as in (1.1) for each $j \ge 1$. Then*

$$d_{TV}(\mathcal{L}(W), CP(\lambda, \boldsymbol{\mu})) \le H(\lambda, \boldsymbol{\mu}) \left\{ \mathrm{Var}(W) - \sum_{i=1}^{n} \sum_{j \in \mathcal{J}(i)\cup\{i\}} E[X_i X_j] + 2\sum_{i=1}^{n} \sum_{j \in \mathcal{J}(i)} E[X_i]E[X_j] \right\}.$$

**Remark 1.3.** *We have stated our results for approximation in total variation distance, although they may easily be adapted for use with other probability metrics. For example, we may wish to use the Kolmogorov distance, defined for non–negative integer valued random variables by*

$$d_K(\mathcal{L}(W), \mathcal{L}(U)) = \sup_{j \ge 0} |P(W \ge j) - P(U \ge j)|\,.$$

*In this case, the bounds of Theorems 1.1 and 1.2 continue to hold, with $H(\lambda, \boldsymbol{\mu})$ replaced by*

$$K(\lambda, \boldsymbol{\mu}) = \sup_{A \in \mathcal{I}} \sup_{x \in \mathbb{Z}^+} |\Delta f_A(x)|\,,$$

*where $\mathcal{I} = \{[k, \infty) : k \in \mathbb{Z}^+\}$. Clearly $K(\lambda, \boldsymbol{\mu}) \leq H(\lambda, \boldsymbol{\mu})$. See [4, Section 3] and [6, Section 1] for further bounds on $K(\lambda, \boldsymbol{\mu})$.*

**Remark 1.4.** *Boutsikas and Koutras [7] also discuss compound Poisson approximation for a sum of associated or negatively associated random variables. The bounds we establish in this note offer a greater flexibility in the choice of approximating compound Poisson distribution than their results: we are able to choose the sets $\mathcal{J}(i)$ to suit the problem at hand. The approximating distribution chosen by [7] is the same as that obtained by us when setting $\mathcal{J}(i) = \emptyset$ for each $i$. Furthermore, our bounds have the advantage of including the so–called 'Stein factor' $H(\lambda, \boldsymbol{\mu})$, giving good bounds when this Stein factor is small.*

## 2 Applications

### 2.1 Independent summands

Suppose that $X_1, \ldots, X_n$ are independent, non–negative integer valued random variables. In line with the definitions of Section 1, we choose $\mathcal{J}(i) = \emptyset$ for each $i$, so that $\lambda_k = \sum_{i=1}^n P(X_i = k)$ for each $k \geq 1$.

Since $X_1, \ldots, X_n$ are independent, they are also negatively associated. We apply Theorem 1.1 to immediately obtain the bound

$$d_{TV}(\mathcal{L}(W), \text{CP}(\lambda, \boldsymbol{\mu})) \leq H(\lambda, \boldsymbol{\mu}) \sum_{i=1}^n (EX_i)^2\,.$$

This bound has also been obtained in the independent case by various other authors, for example [13, Corollary 1]. See also Lemma 8 of [7]. We do not concern ourselves with evaluating bounds on the Stein factor $H(\lambda, \boldsymbol{\mu})$ for this example, since better bounds in compound Poisson approximation for a sum of independent random variables are available by means other than Stein's method. See also Section 4.1 of [1].

### 2.2 $k$-runs

We now turn our attention to compound Poisson approximation for the number of runs in iid Bernoulli trials. This problem is discussed by Barbour et al. [2, Section 2.1]. We show that our Theorem 1.2 can be used to improve the bounds of their work.

We let $\xi_1, \ldots, \xi_n$ be iid Bernoulli random variables with $P(\xi_1 = 1) = p$. Fix some $k \geq 1$ and let $X_i = \xi_i \cdots \xi_{i+k-1}$, where all indices are written modulo $n$ to avoid edge effects. Thus, $W = X_1 + \cdots + X_n$ counts the number of $k$–runs in our Bernoulli trials. Our random variables $X_1, \ldots, X_n$ are associated, so Theorem 1.2 may be applied in this case.

Following [2], we choose $\mathcal{J}(i) = \{1 \leq j \leq n : 1 \leq |i - j| \leq k - 1\}$ and obtain

$$\lambda_j = \begin{cases} np^{k+j-1}(1-p)^2 & \text{if } j = 1, \ldots, k-1\,, \\ np^{k+j-1}j^{-1}(1-p)[2 + (2k - j - 2)(1-p)] & \text{if } j = k, \ldots, 2(k-1)\,, \\ np^{3k-2}(2k-1)^{-1} & \text{if } j = 2k-1\,. \end{cases}$$

We have that for each $i$, $EX_i = p^k$, $EZ_i = 2(k-1)p^k$ and that $EW = np^k$. We also have,

$$\text{Var}(W) = \sum_{i=1}^n \text{Var}(X_i) + 2\sum_{i<j} \text{Cov}(X_i, X_j) = \frac{np^k}{1-p}\left\{1 + p - p^k[2 + (2k-1)(1-p)]\right\}\,.$$

Furthermore, we also have that for each $i$

$$E[X_i(X_i+Z_i)] = p^k E[1+Z_i|X_i=1] = p^k\left[1+2\sum_{j=1}^{k-1}p^j\right] = p^k\left[1+\frac{2p(1-p^{k-1})}{1-p}\right],$$

so that

$$E\sum_{i=1}^{n}X_i(X_i+Z_i) = np^k\left[1+\frac{2p(1-p^{k-1})}{1-p}\right].$$

Combining these expressions, Theorem 1.2 easily yields the bound

$$d_{TV}(\mathcal{L}(W),\mathrm{CP}(\lambda,\boldsymbol{\mu})) \le H(\lambda,\boldsymbol{\mu})(2k-1)np^{2k}. \tag{2.1}$$

This improves upon the bound of [2], who show that the above total variation distance is bounded by $H(\lambda,\boldsymbol{\mu})(6k-5)np^{2k}$.

To conclude this example, we note that condition (1.3) is satisfied if $p \le 1/3$, and condition (1.5) holds if $p < 1/5$. See [2, Section 2.1]. Thus, when $p \le 1/3$, we may use the bound (1.4), with $\nu_1 = np^k(1-p)^2(1-2p)$ for $k > 2$. When $p < 1/5$, we may combine (1.6) with (2.1) to obtain the bound $d_{TV}(\mathcal{L}(W),\mathrm{CP}(\lambda,\boldsymbol{\mu})) \le (2k-1)p^k(1-p)(1-5p)^{-1}$.

### 2.3 An urn model with overflow

We consider the following model of [8]. Suppose that $n$ balls are distributed into $m$ urns, with each ball equally likely to be assigned to any urn. We fix some $k \ge 2$, and assume that each of our $m$ urns can hold at most $k-1$ balls. If a ball is assigned to an urn which is already full, that ball is placed in an additional 'overflow urn' of unlimited capacity. We consider a compound Poisson approximation for $W$, the number of balls allocated to the overflow urn. Boutsikas and Koutras [8, pg 278] give a bound for such an approximation in Kolmogorov distance. Here, we consider approximation in the stronger total variation distance.

We write $W = X_1+\cdots+X_m$, where, for each $1 \le j \le m$, $X_j = (S_j-k+1)I(S_j \ge k)$ and $S_j$ is the number of balls allocated to urn $j$. Our random variables $X_1,\ldots,X_m$ are negatively associated, so we may apply Theorem 1.1. See [8, pg 278].

For notational convenience, we write $p = m^{-1}$. For each $i = 1,\ldots,n-k+1$ and $j = 1,\ldots,m$ we have that

$$P(X_j=i) = P(S_j=i+k-1) = \binom{n}{i+k-1}p^{i+k-1}(1-p)^{n-i-k+1}.$$

We choose $\mathcal{J}(j) = \emptyset$ for each $j = 1,\ldots,m$, so that

$$\lambda_i = \sum_{j=1}^{m}P(X_j=i) = \binom{n}{i+k-1}p^{i+k-2}(1-p)^{n-i-k+1}, \tag{2.2}$$

for each $i = 1,\ldots,n-k+1$.

Now, for each $j = 1,\ldots,m$ we have that

$$EX_j = \sum_{i=1}^{n-k+1}iP(X_j=i) = \sum_{i=k}^{n}(i-k+1)\binom{n}{i}p^i(1-p)^{n-i}.$$

Furthermore, Boutsikas and Koutras [8, (4.16)] show that

$$-2\sum_{1\le i<j\le m}\mathrm{Cov}(X_i,X_j)=m(m-1)\left\{\left(\sum_{i=k}^{n}(i-k+1)\binom{n}{i}p^i(1-p)^{n-i}\right)^2 - \sum_{i=k}^{n-k}\sum_{j=k}^{n-i}(i-k+1)(j-k+1)\binom{n}{i,j}p^{i+j}(1-2p)^{n-i-j}\right\}.$$

Combining the above expressions, Theorem 1.1 gives us that

$$d_{TV}(\mathcal{L}(W),\mathrm{CP}(\lambda,\boldsymbol{\mu}))\le H(\lambda,\boldsymbol{\mu})\left\{\sum_{j=1}^{m}(EX_j)^2-2\sum_{1\le i<j\le m}\mathrm{Cov}(X_i,X_j)\right\}=H(\lambda,\boldsymbol{\mu})\Theta\,, \tag{2.3}$$

where

$$\Theta=m^2\left(\sum_{i=k}^{n}(i-k+1)\binom{n}{i}p^i(1-p)^{n-i}\right)^2 - m(m-1)\sum_{i=k}^{n-k}\sum_{j=k}^{n-i}(i-k+1)(j-k+1)\binom{n}{i,j}p^{i+j}(1-2p)^{n-i-j}\,.$$

To conclude this example, we note that it can easily be shown using (2.2) that condition (1.3) is satisfied if and only if $i(i+k-np)\ge p(n-k+1)$ for $1\le i\le n-k$. Under this condition, we may combine (1.4) with (2.3), where (2.2) gives us that $\nu_1=\binom{n}{k}p^{k-1}(1-p)^{n-k-1}\left\{1-p-2(n-k)p(k+1)^{-1}\right\}$.

### 2.4 Extremes

Suppose that $\xi_1,\dots,\xi_n$ is a stationary sequence of negatively associated random variables. For simplicity, we will assume that each of the $\xi_i$ have the same distribution function, with $F(x)=P(\xi_i\le x)$ for each $i=1,\dots,n$. Note that throughout this section we will treat all indices modulo $n$.

We fix some $a_1,\dots,a_n\in\mathbb{R}$ and let $X_i=I(\xi_i>a_i)$. Since $\xi_1,\dots,\xi_n$ are negatively associated, $X_1,\dots,X_n$ also have this property. See [11, page 288]. We define $W=\sum_{i=1}^n X_i$, so that $W$ counts the number of the $\xi_i$ exceeding the threshold $a_i$.

In line with our earlier work, we consider the approximation of $W$ by a compound Poisson distribution $\mathrm{CP}(\lambda,\boldsymbol{\mu})$, with $\lambda_j$ given by (1.1) for $j\ge1$. For further discussion on the choice and calculation of the $\lambda_j$ see [10] and references therein.

Since $\sum_{i=1}^n(EX_i)^2=\sum_{i=1}^n(1-F(a_i))^2$, we use the definition of $Z_i$ and immediately obtain from Theorem 1.1 that

$$d_{TV}(\mathcal{L}(W),\mathrm{CP}(\lambda,\boldsymbol{\mu}))\le H(\lambda,\boldsymbol{\mu})\Lambda\,, \tag{2.4}$$

where

$$\Lambda=\sum_{i=1}^{n}(1-F(a_i))^2+2\sum_{i=1}^{n}\sum_{\substack{j\in\mathcal{J}(i)\\ i<j}}E[X_i]E[X_j] + \sum_{i=1}^{n}\sum_{j\in\mathcal{J}(i)}\left\{E[X_iX_j]I(j<i)-E[X_iX_j]I(i<j)\right\}\,.$$

**Remark 2.1.** *Suppose that $a_i = a \in \mathbb{R}$ for each $i$. We note that $W = 0$ if and only if $\xi_1, \dots, \xi_n$ are all at most $a$. Hence, an immediate corollary of (2.4) is the bound*

$$\left| P(\max_i \xi_i \le a) - e^{-\lambda} \right| \le H(\lambda, \boldsymbol{\mu})\Lambda \,,$$

*where $\lambda$, $\boldsymbol{\mu}$ and $\Lambda$ are as above. Similar bounds apply if we wish to consider the probability that exactly $k$ of our random variables $\xi_1, \dots, \xi_n$ exceed $a$.*

We consider now an example in which we may compare (2.4) with a result of [10].

**Example 2.2.** *Assume, in the setting of this section, that the random variables $\xi_1, \dots, \xi_n$ are also $m$–dependent. That is, we may choose $\mathcal{J}(i) = \{i-m, \dots, i-1, i+1, \dots, i+m\}$ for $i = 1, \dots, n$ and we have that for $j \neq i$, $\xi_i$ is independent of $\xi_j$ for all $j \notin \mathcal{J}(i)$. We will assume, for simplicity, that $a_i = a \in \mathbb{R}$ for each $i = 1, \dots, n$. In this case, we have that*

$$\sum_{i=1}^n \sum_{\substack{j \in \mathcal{J}(i) \\ i<j}} E[X_i]E[X_j] \le mn(1 - F(a))^2 \,,$$

*and*

$$\sum_{i=1}^n \sum_{j \in \mathcal{J}(i)} \{E[X_i X_j] I(j < i) - E[X_i X_j] I(i < j)\}$$
$$\le \sum_{i=1}^n \sum_{j \in \mathcal{J}(i)} E[X_i X_j] \le 2mn(1 - F(a))^2 \,,$$

*where this final inequality uses the negative association property. We thus obtain the bound*

$$d_{TV}(\mathcal{L}(W), CP(\lambda, \boldsymbol{\mu})) \le H(\lambda, \boldsymbol{\mu})(1 + 4m)n(1 - F(a))^2 \,. \tag{2.5}$$

*We compare this to the bound of Proposition 2.1 of [10], which states that if $\xi_1, \dots, \xi_n$ is a stationary $m$–dependent (but not necessarily negatively associated) sequence of random variables then*

$$d_{TV}(\mathcal{L}(W), CP(\lambda, \boldsymbol{\mu})) \le H(\lambda, \boldsymbol{\mu})(1 + 6m)n(1 - F(a))^2 \,,$$

*where the same approximating compound Poisson distribution is used as in (2.5). When $\xi_1, \dots, \xi_n$ are negatively associated in addition to being $m$–dependent, our bound (2.5) slightly improves upon this result. Note also that we do not need a condition of $m$–dependence in order to apply the more general bound (2.4).*

*For further details in this example, including comments on the Stein factor $H(\lambda, \boldsymbol{\mu})$, we refer the reader to [10].*

*Finally, we note that it is straightforward to construct examples with $m$–dependence where (2.4) does better than (2.5). For example, if $m = 1$ take pairs of random variables $(\xi_i, \eta_i)$ for $i = 1, \dots, n/2$ such that for each $i$, $\xi_i$ and $\eta_i$ are independent of all random variables except each other. Now take $W = \sum_{i=1}^{n/2} \{I(\xi_i > a) + I(\eta_i > a)\}$. If our negative association condition holds, an argument analogous to that used to derive (2.5) gives a bound smaller than that result.*

In the setting of this section, if we suppose that our random variables $\xi_1, \dots, \xi_n$ are associated rather than negatively associated, we may employ Theorem 1.2 rather than than Theorem 1.1 to obtain a bound analogous to (2.4). One application of such a bound would be to extremes of moving average processes. We illustrate this by considering a special case in the following example.

**Example 2.3.** *We suppose that* $\eta_1, \dots, \eta_n$ *are iid uniform random variables on* $(0,1)$*, and define* $\xi_i = \eta_{i-1} + \eta_i$*. Following the notation of this section, we define* $X_i = I(\xi_i > a)$*. We have that* $\lambda = EW = nP(\eta_1 + \eta_2 > a)$*. We choose* $a = 2 - \sqrt{2\lambda/n}$ *(where* $n \geq 2\lambda$*, so* $1 \leq a \leq 2$*). As usual,* $W = \sum_{i=1}^n X_i$*. Note that* $EX_i = \lambda/n$ *for each* $i$*.*

*Since each* $X_i$ *is an increasing function of independent random variables* $\eta_1, \dots, \eta_n$*, we have that* $X_1, \dots, X_n$ *are associated, so Theorem 1.2 may be applied, in line with which we choose* $\mathcal{J}(i) = \{i-1, i+1\}$*.*

*Barbour et al. [3, Section 8.3] consider Poisson approximation for* $W$ *in this case, and use the fact that*

$$\text{Var}(W) = \lambda - \frac{\lambda^2}{n} + \frac{2\lambda(n-1)}{n}\left(\frac{\sqrt{8}}{3}\sqrt{\frac{\lambda}{n}} - \frac{\lambda}{n}\right).$$

*With our choice of* $\mathcal{J}(i)$ *we have that* $EZ_i = 2\lambda/n$ *for each* $i$*. Further straightforward calculations show that*

$$\sum_{i=1}^n E[X_i(X_i + Z_i)] = \lambda\left\{1 + 2P(\xi_2 > a|\xi_1 > a)\right\} = \lambda\left\{1 + \frac{4}{3}\sqrt{\frac{2\lambda}{n}}\right\}.$$

*Combining the above expressions, Theorem 1.2 gives*

$$d_{TV}(\mathcal{L}(W), CP(\lambda, \boldsymbol{\mu})) \leq H(\lambda, \boldsymbol{\mu})\left\{\frac{3\lambda^2}{n} - \frac{4\lambda}{3n}\sqrt{\frac{2\lambda}{n}} + \frac{2\lambda^2}{n^2}\right\}. \tag{2.6}$$

*Similar calculations to the above also give us the parameters of our approximating compound Poisson distribution. Using (1.1) we have that*

$$\lambda_1 = \frac{n(2-a)^2(12a - a^2 - 8)}{24}, \qquad \lambda_2 = \frac{n(2-a)^2(20 - 8a - a^2)}{36}, \qquad \lambda_3 = \frac{5n(2-a)^4}{72},$$

*and* $\lambda_j = 0$ *for* $j \geq 4$*. Recalling that we must have* $1 \leq a \leq 2$*, it is now straightforward to check that (1.3) holds whenever* $6\sqrt{35} - 34 \leq a \leq 2$*, while (1.5) does not hold for any valid choice of* $a$*. When* $6\sqrt{35} - 34 \leq a \leq 2$ *we may combine (2.6) with the bound (1.4) on* $H(\lambda, \boldsymbol{\mu})$*, where* $\nu_1 = n(2-a)^2(a^2 + 68a - 104)/72$*.*

## 3 Proofs of Theorems 1.1 and 1.2

Our proofs are based on techniques developed by [12] and [9]. We begin by defining the size–biased distribution. For any non–negative integer valued random variable $X$ with $EX > 0$, we let $X^\star$ have the $X$–size–biased distribution. That is, $X^\star$ satisfies

$$E[Xg(X)] = (EX)Eg(X^\star), \tag{3.1}$$

for all functions $g$ for which the above expectations exist. We have that

$$P(X^\star = j) = (EX)^{-1} jP(X = j),$$

for all $j \geq 1$. Throughout this section we let $\lambda_j$ be defined by (1.1) for each $j \geq 1$ and $Y$ be a random variable, independent of all else, with $P(Y = j) = \mu_j$ for $j \geq 1$, where $\mu_j = \lambda^{-1}\lambda_j$ and $\lambda = \sum_{j\geq 1}\lambda_j$. We note that with these choices of parameters, we have that

$$\sum_{i\geq 1} i\lambda_i = EW, \quad \text{and} \quad \sum_{i\geq 1} i^2\lambda_i = E\sum_{j=1}^n X_j(X_j + Z_j). \tag{3.2}$$

Consider now the relation (1.2). Using the above definitions and relations, we obtain

$$E\left[\sum_{j\geq 1} j\lambda_j f_A(W+j) - Wf_A(W)\right] = E\left[\lambda Y f_A(Y+W) - Wf_A(W)\right]$$
$$= E\left[\lambda(EY)f_A(Y^\star+W) - (EW)f_A(W^\star)\right] = (EW)E\left[f_A(Y^\star+W) - f_A(W^\star)\right].$$

Recalling that $f_A(0)=0$, we write $f_A(j) = \Delta f_A(0) + \cdots + \Delta f_A(j-1)$. Substituting this in the above and interchanging the order of summation, we have that

$$E\left[\sum_{j\geq 1} j\lambda_j f_A(W+j) - Wf_A(W)\right] = (EW)\sum_{k=0}^{\infty} \Delta f_A(k)\left\{P(Y^\star+W>k) - P(W^\star>k)\right\}.$$

Then, using (1.2), we obtain

$$d_{TV}(\mathcal{L}(W), \mathrm{CP}(\lambda, \boldsymbol{\mu})) \leq H(\lambda, \boldsymbol{\mu})(EW)\sum_{k=0}^{\infty} \left|P(Y^\star+W>k) - P(W^\star>k)\right|. \tag{3.3}$$

To proceed further with our proofs, we need the following lemmas. Lemma 3.1 treats the case of negative association, while Lemma 3.2 considers the case of association. The proofs of these lemmas are given in Sections 3.1 and 3.2, respectively, before which we show how they are used to prove Theorems 1.1 and 1.2. Note that, since we are assuming $Y$ is independent of all other random variables, the size–biased version $Y^\star$ used in Lemmas 3.1 and 3.2 is, in particular, independent of $W$.

**Lemma 3.1.** *Suppose that $X_1, \ldots, X_n$ are negatively associated. Then, for all non–decreasing functions $g$,*

$$Eg(W^\star) \leq Eg(Y^\star + W).$$

**Lemma 3.2.** *Suppose that $X_1, \ldots, X_n$ are associated and let $V$ be a random index, independent of all else, chosen according to the distribution $P(V=i) = (EW)^{-1}EX_i$ for $i = 1, \ldots, n$. Then, for all non–decreasing functions $g$,*

$$Eg(W^\star) \geq Eg(Y^\star + W - X_V - Z_V).$$

To complete the proof of Theorem 1.1, we combine (3.3) with Lemma 3.1. Since $I(j>k)$ is non–decreasing in $j$ we get that, in the negatively associated case,

$$d_{TV}(\mathcal{L}(W), \mathrm{CP}(\lambda, \boldsymbol{\mu})) \leq H(\lambda, \boldsymbol{\mu})(EW)E[Y^\star + W - W^\star]. \tag{3.4}$$

Using (3.1) and (3.2), we have that

$$E[W^\star] = \frac{E[W^2]}{EW}, \quad \text{and} \quad E[Y^\star] = \frac{E[Y^2]}{EY} = \frac{\sum_{i\geq 1} i^2\lambda_i}{\sum_{i\geq 1} i\lambda_i} = \frac{E\sum_{j=1}^n X_j(X_j+Z_j)}{EW}. \tag{3.5}$$

Combining these expressions with (3.4), we easily obtain that

$$d_{TV}(\mathcal{L}(W), \mathrm{CP}(\lambda, \boldsymbol{\mu})) \leq H(\lambda, \boldsymbol{\mu})\left\{E\sum_{j=1}^n X_j(X_j+Z_j) - \mathrm{Var}(W)\right\},$$

from which the bound of Theorem 1.1 follows. To complete the proof of Theorem 1.2, we use the triangle inequality to obtain from (3.3) that

$$\begin{aligned} &d_{TV}(\mathcal{L}(W), \mathrm{CP}(\lambda, \boldsymbol{\mu})) \\ &\quad \leq H(\lambda, \boldsymbol{\mu})(EW) \left\{ \sum_{k=0}^{\infty} |P(Y^\star + W > k) - P(Y^\star + W - X_V - Z_V > k)| \right. \\ &\qquad \left. + \sum_{k=0}^{\infty} |P(Y^\star + W - X_V - Z_V > k) - P(W^\star > k)| \right\}, \end{aligned}$$

where $V$ is as in Lemma 3.2. Since $I(j > k)$ is non–decreasing in $j$, we may apply Lemma 3.2 to get that in the associated case

$$d_{TV}(\mathcal{L}(W), \mathrm{CP}(\lambda, \boldsymbol{\mu})) \leq H(\lambda, \boldsymbol{\mu})(EW)E[2(X_V + Z_V) + W^\star - W - Y^\star].$$

Using (3.5), this is easily shown to yield

$$d_{TV}(\mathcal{L}(W), \mathrm{CP}(\lambda, \boldsymbol{\mu})) \leq H(\lambda, \boldsymbol{\mu}) \left\{ 2(EW)E[X_V + Z_V] + \mathrm{Var}(W) - E\sum_{j=1}^{n} X_j(X_j + Z_j) \right\}.$$

Employing the definition of $V$ then gives us the bound of Theorem 1.2. To establish our Theorems 1.1 and 1.2, it therefore remains only to prove Lemmas 3.1 and 3.2.

### 3.1 Proof of Lemma 3.1

Our proof is based upon that of Lemma 3.1 of [12]. We begin by recalling the definition of $W_i$ from Section 1 and observing that, under the conditions of Theorem 1.1,

$$E[X_i I(X_i + Z_i = j) g(j + W_i)] \leq E[X_i I(X_i + Z_i = j)] Eg(j + W_i), \tag{3.6}$$

for any $j \in \mathbb{Z}^+$, $i = 1, \ldots, n$ and $g$ non–decreasing. This can be seen by conditioning on $X_i + Z_i$ and using the negative association property.

To prove Lemma 3.1, we note that it is enough to show that under our negative association assumption

$$E[Wg(W)] \leq \left( \sum_{k \geq 1} k\lambda_k \right) Eg(Y^\star + W),$$

for all non–decreasing functions $g$. Using (3.1) and (3.2), this is equivalent to the statement of our lemma. We have that

$$\begin{aligned} E[Wg(W)] &= E\sum_{i=1}^{n} X_i g(W) = E\sum_{i=1}^{n}\sum_{j\geq 1} X_i I(X_i + Z_i = j) g(W) \\ &= E\sum_{i=1}^{n}\sum_{j\geq 1} X_i I(X_i + Z_i = j) g(j + W_i) \leq \sum_{i=1}^{n}\sum_{j\geq 1} E[X_i I(X_i + Z_i = j)] Eg(j + W_i), \end{aligned}$$

where we use (3.6) for this inequality. Now, since $g$ is non–decreasing and $W_i \leq W$ almost surely,

$$\begin{aligned} E[Wg(W)] &\leq \sum_{i=1}^{n}\sum_{j\geq 1} E[X_i I(X_i + Z_i = j)] Eg(j + W) = \sum_{j \geq 1} j\lambda_j Eg(j + W) \\ &= \lambda(EY)\sum_{j\geq 1} P(Y^\star = j) Eg(j + W) = \left( \sum_{k\geq 1} k\lambda_k \right) Eg(Y^\star + W). \end{aligned}$$

This completes the proof of Lemma 3.1.

### 3.2 Proof of Lemma 3.2

Our proof of Lemma 3.2 is built upon the work of Daly et al. [9, Section 4.2]. Analogously to (3.6), we observe that with our association assumption,

$$E[X_i I(X_i + Z_i = j) g(j + W_i)] \geq E[X_i I(X_i + Z_i = j)] Eg(j + W_i)\,, \tag{3.7}$$

for any $j \in \mathbb{Z}^+$, $i = 1, \dots, n$ and $g$ non–decreasing. Similarly to the proof of Lemma 3.1, we establish our result by showing that

$$E[Wg(W)] \geq \left(\sum_{k\geq 1} k\lambda_k\right) Eg(Y^\star + W - X_V - Z_V)\,,$$

for all $g$ non–decreasing. We begin by defining for each $j \geq 1$ a random index $V(j)$ defined by

$$P(V(j) = i) = \frac{E[X_i I(X_i + Z_i = j)]}{j\lambda_j}\,, \qquad 1 \leq i \leq n\,. \tag{3.8}$$

It is straightforward to check that the random variable $V(Y^\star)$ has the same distribution as $V$, the random index defined in the statement of Lemma 3.2. Following the proof of Lemma 1.1, but employing (3.7) in place of (3.6), we have that

$$E[Wg(W)] = E\sum_{i=1}^{n}\sum_{j\geq 1} X_i I(X_i + Z_i = j) g(j + W_i) \geq \sum_{i=1}^{n}\sum_{j\geq 1} E[X_i I(X_i + Z_i = j)] Eg(j + W_i)\,.$$

Using the definition (3.8), we thus obtain

$$E[Wg(W)] \geq \sum_{j\geq 1} j\lambda_j \sum_{i=1}^{n} P(V(j) = i) Eg(j + W_i) = \sum_{j\geq 1} j\lambda_j Eg(j + W - X_{V(j)} - Z_{V(j)})$$

$$= \lambda(EY) \sum_{j\geq 1} P(Y^\star = j) Eg(j + W - X_{V(j)} - Z_{V(j)}) = \left(\sum_{k\geq 1} k\lambda_k\right) Eg(Y^\star + W - X_V - Z_V)\,.$$

This completes the proof of Lemma 3.2.

## References


[1] Barbour, A. D., Chen, L. H. Y. and Loh, W.–L.: Compound Poisson approximation for nonnegative random variables via Stein's method. *Ann. Probab.* **20**, (1992), 1843–1866. MR-1188044

[2] Barbour, A. D., Chryssaphinou, O. and Vaggelatou, E.: Applications of compound Poisson approximation. In Charalambides, C. A., Koutras, M. V. and Balakrishnan, N. (Eds.): *Probability and Statistical Models with Applications*, 41–62. Chapman and Hall, Boca Raton, 2001.

[3] Barbour, A. D., Holst, L. and Janson, S.: *Poisson Approximation*. Oxford Univ. Press, Oxford, 1992. MR-1163825

[4] Barbour, A. D. and Utev, S.: Solving the Stein equation in compound Poisson approximation. *Adv. Appl. Probab.* **30**, (1998), 449–475. MR-1642848

[5] Barbour, A. D. and Xia, A.: Poisson perturbations. *ESAIM Probab. Stat.* **3**, (1999), 131–159. MR-1716120

[6] Barbour, A. D. and Xia, A.: Estimating Stein's constants for compound Poisson approximation. *Bernoulli* **6**, (2000), 581–590. MR-1777683

[7] Boutsikas, M. V. and Koutras, M. V.: A bound for the distribution of the sum of discrete associated or negatively associated random variables. *Ann. Appl. Probab.* **10**, (2000), 1137–1150. MR-1810868

[8] Boutsikas, M. V. and Koutras, M. V.: On the number of overflown urns and excess balls in an allocation model with limited urn capacity. *J. Statist. Plann. Inference* **104**, (2002), 259–286. MR-1906011

[9] Daly, F., Lefèvre, C. and Utev, S.: Stein's method and stochastic orderings. *Adv. Appl. Probab.* **44**, (2012), 343–372. MR-2977399

[10] Hashorva, E. and Hüsler, J.: Remarks on compound Poisson approximation of Gaussian random sequences. *Stat. Prob. Lett.* **57**, (2002), 1–8. MR-1911806

[11] Joag–Dev, K. and Proschan, F.: Negative association of random variables, with applications. *Ann. Stat.* **11**, (1983), 286–295. MR-0684886

[12] Papadatos, N. and Papathanasiou, V.: Approximation for a sum of dependent indicators: an alternative approach. *Adv. Appl. Probab.* **34**, (2002), 609–625. MR-1929600

[13] Roos, M.: Stein's method for compound Poisson approximation: the local approach. *Ann. Appl. Probab.* **4**, (1994), 1177–1187. MR-1304780

**Acknowledgments.** The author gratefully acknowledges the financial support of the Swiss SNF and Universität Zürich, and comments made by a referee and associate editor which greatly improved the presentation of the paper.