J. Appl. Prob. **43,** 282–288 (2006)
Printed in Israel

A COMPOUND POISSON APPROXIMATION INEQUALITY

EROL A. PEKÖZ,* *Boston University*

Abstract

We give conditions under which the number of events which occur in a sequence of m-dependent events is stochastically smaller than a suitably defined compound Poisson random variable. The results are applied to counts of sequence pattern appearances and to system reliability. We also provide a numerical example.

Keywords: Compound Poisson approximation; stochastic inequality; patterns in coin tossing; dependent events

2000 Mathematics Subject Classification: Primary 60E15
Secondary 62E17; 60J10

1. Introduction

Given a sequence of m-dependent indicator variables $X_1, X_2, \ldots, X_n$, we study the distribution of the sum $S = \sum_{i=1}^{n} X_i$ and give conditions under which S is stochastically smaller than a suitably defined compound Poisson random variable. A variable Y is stochastically smaller than a variable Z, written $Y \leq_{\mathrm{st}} Z$, if $\mathrm{P}(Y \geq k) \leq \mathrm{P}(Z \geq k)$ for all k. A variable C has a compound Poisson distribution if it can be written as

$$C = \sum_{i=1}^{N} A_i,$$

where N is a Poisson random variable with mean λ, and $A_1, A_2, \ldots$ are independent random variables each having the same distribution as the random variable A. We use the notation $\mathrm{CP}(\lambda, A)$ to denote this compound Poisson distribution. Random variables $X_1, X_2, \ldots, X_n$ are m-dependent if the vector $(X_{i_1}, X_{i_2}, \ldots, X_{i_a})$ is independent of the vector $(X_{j_1}, X_{j_2}, \ldots, X_{j_b})$ for all a and b and for all $i.$ and $j.$ such that

$$1 \leq i_1 < i_2 < \cdots < i_a \leq i_a + m < j_1 < j_2 < \cdots < j_b \leq n.$$

There is an extensive literature on Poisson and compound Poisson approximations for sums of variables. A Poisson approximation for the sum of independent indicator variables was studied by Le Cam (1960), and there have subsequently been many papers on assessing the accuracy of Poisson approximations for dependent indicator variables using the Stein–Chen method; see the survey by Barbour *et al.* (1992). Compound Poisson approximations, whose wide scope of applicability was surveyed by Aldous (1989), tend to be appropriate when positive values of the variables tend to occur in 'clumps'. There has also been much recent research directed towards assessing the accuracy of compound Poisson approximations using Stein's method;

Received 25 May 2005; revision received 27 September 2005.
* Postal address: School of Management, Boston University, 595 Commonwealth Avenue, Boston, MA 02215, USA. Email address: pekoz@bu.edu

see, for example, Barbour and Månsson (2002) and Barbour and Chryssaphinou (2001). These approaches generally give upper bounds on the total variation distance

$$d_{\mathrm{TV}}(S, C) = \sup_A |\,\mathrm{P}(S \in A) - \mathrm{P}(C \in A)|,$$

when C has either an appropriate Poisson or compound Poisson distribution.

In this paper, in contrast, we study conditions under which $S \leq_{\mathrm{st}} C$, for an approximating compound Poisson variable C. Upper bounds on tail probabilities of S can thus, in theory, be obtained by computing the tail probabilities of C using standard techniques; see, for example, Peköz and Ross (2004) for techniques for computing probabilities for the compound Poisson distribution. To get the stochastic inequality we must have $\mathrm{P}(S > 0) \leq \mathrm{P}(C > 0)$, and so computing the Poisson parameter for the approximating compound Poisson distribution may be complicated in practice. It is important to note that the approximating distributions we use are thus not necessarily the same as the ones studied above in the setting of total variation error bounds.

Some upper bounds for tail probabilities were given in terms of the tail probabilities of an approximating Poisson distribution in Barbour *et al.* (1992, p. 43) in the case of sums of negatively related indicators (variables are negatively related if increasing functions of disjoint sets of the variables are negatively correlated). Klass and Nowicki (2003) also gave a tail inequality for a Poisson approximation in the special case where the indicator variables are certain functions of independent or conditionally independent variables. Here, in contrast, we give results in the more general setting of the compound Poisson approximation.

The organization of this paper is as follows. In Section 2, we present the main result and its proof and, in Section 3, we present applications of the main result to the distribution of the number of overlapping or non-overlapping patterns occurring in a sequence of coin flips, and the reliability of the m-consecutive-k-of-n system. In Section 4, we give a numerical example illustrating the approximations.

2. Main result

Here we present our main result.

Theorem 1. *Given an m-dependent sequence of nonnegative random variables*

$$X_1, X_2, \ldots, X_n$$

and a random variable Y, called the 'clump size' variable, independent of all else and satisfying

$$\left(\sum_{i=k}^{n} X_i \;\middle|\; X_k > 0,\ X_1, \ldots, X_{k-1}\right) \leq_{\mathrm{st}} Y + \sum_{i=k+m+1}^{n} X_i, \quad \textit{for all } k > 0, \tag{1}$$

we have $S \leq_{\mathrm{st}} C$, where $S = \sum_{i=1}^{n} X_i$ and C has the compound Poisson distribution $\mathrm{CP}(-\ln \mathrm{P}(S = 0), Y)$ *as defined above.*

Remark 1. Notice that this choice of the parameter $\lambda = -\ln \mathrm{P}(S = 0)$ for the compound Poisson distribution gives

$$\mathrm{P}(C = 0) = \mathrm{e}^{\ln \mathrm{P}(S=0)} = \mathrm{P}(S = 0);$$

thus, it is the smallest value of λ that could possibly be used and still have the stochastic inequality $S \leq_{\mathrm{st}} C$.

https://doi.org/10.1239/jap/1143936260 Published online by Cambridge University Press

Remark 2. We can think of a 'clump' as beginning whenever a positive value appears in the sequence, and continuing until a situation arises where no additional positive value can occur within m subsequent positions. Condition (1) says that the sum of the values occurring during a 'clump' is stochastically smaller than Y. Note that the left-hand side of (1) represents the conditional distribution of $\sum_{i=k}^{n} X_i$, given the values of $X_1, \ldots, X_{k-1}$ and the fact that $X_k > 0$.

Before we prove the main result we need the following lemma about the compound Poisson distribution.

Lemma 1. *Let $C(\lambda)$ have the distribution* $\mathrm{CP}(\lambda, Y)$*, and let U be a* uniform$(0, 1)$ *random variable independent of all else. Then, for $k > 0$, we obtain*

$$\mathrm{P}(C(\lambda) \geq k) = \mathrm{P}(U \geq \mathrm{e}^{-\lambda},\ C(\lambda + \ln U) \geq k - Y).$$

Proof. Consider a unit-rate Poisson process on the interval $(0, \lambda)$, where each Poisson event is marked with an independent draw from the distribution Y. Construct $C(\lambda)$ by equating it to the sum of all the marks. The result follows by conditioning on the time of the first event, and by noting that $-\ln U$ has an exponential distribution with mean 1.

Next we prove the main result.

Proof of Theorem 1. Let $C(\lambda)$ be a random variable with the $\mathrm{CP}(\lambda, Y)$ distribution, and let

$$S_j = \sum_{i=j}^{n} X_i.$$

We show, by backwards induction on j, that $S_j \leq_{\mathrm{st}} C(-\ln \mathrm{P}(S_j = 0))$, for all j, and thus prove our main result, $S \leq_{\mathrm{st}} C$.

The cases where $j > n - m - 1$ are immediate from (1). Then, given some $j < n - m - 1$, we assume as our induction hypothesis that, for all $i > j$, $S_i \leq_{\mathrm{st}} C(-\ln \mathrm{P}(S_i = 0))$. Picking any i such that $j \leq i + 1 < n - m - 1$ and letting $T = \min\{k \geq j : X_k > 0\}$, we obtain

$$\begin{aligned}
\mathrm{P}(S_{i+1} = 0) &= \mathrm{P}(S_j = 0,\ S_{i+1} = 0) + \mathrm{P}(S_j > 0,\ S_{i+1} = 0) \\
&= \mathrm{P}(S_j = 0) + \mathrm{P}(S_{i+1} = 0,\ T \leq i) \\
&\geq \mathrm{P}(S_j = 0) + \mathrm{P}(S_{i+1} = 0,\ T \leq i - m) \\
&= \mathrm{P}(S_j = 0) + \mathrm{P}(S_{i+1} = 0)\,\mathrm{P}(T \leq i - m),
\end{aligned}$$

where the fourth line follows from the fact that the variables are m-dependent. By rearranging this equation, we obtain

$$\mathrm{P}(S_{i+1} = 0) \geq \frac{\mathrm{P}(S_j = 0)}{1 - \mathrm{P}(T \leq i - m)} = \frac{\mathrm{P}(S_j = 0)}{1 - F(i - m)}, \tag{2}$$

where we construct the function $F(x) = \mathrm{P}(T - U \leq x)$ using U, a uniform$(0, 1)$ random variable which is independent of all else.

https://doi.org/10.1239/jap/1143936260 Published online by Cambridge University Press

Letting $\lambda = -\ln \mathrm{P}(S_j = 0)$ and letting Y be independent of all else, we have, for $k > 0$,

$$\begin{aligned}
\mathrm{P}(S_j \geq k) &\leq \mathrm{P}(Y + S_{T+m+1} \geq k,\ T \leq n) \\
&\leq \mathrm{P}(C(-\ln \mathrm{P}(S_{T+m+1} = 0)) \geq k - Y,\ T \leq n) \\
&\leq \mathrm{P}(C(\lambda + \ln(1 - F(T))) \geq k - Y,\ F(T) \leq F(n)) \\
&\leq \mathrm{P}(C(\lambda + \ln(1 - F(T - U)) \geq k - Y,\ F(T - U) \leq F(n)) \\
&= \mathrm{P}(C(\lambda + \ln U \geq k - Y,\ 1 - U \leq 1 - \mathrm{e}^{-\lambda}) \\
&= \mathrm{P}(C(\lambda + \ln U) \geq k - Y,\ U \geq \mathrm{e}^{-\lambda}), \qquad (3)
\end{aligned}$$

where the first line follows from (1), the second line from the induction hypothesis after conditioning on Y and T, the third line from (2), and the fourth line from the fact that $C(\lambda)$ is stochastically increasing in λ. The last two lines follow from the fact that $T - U$ is a continuous random variable with continuous cumulative distribution function $F(x)$, and that, for any x, $0 \leq x \leq 1$, we have $\mathrm{P}(F(T - U) \leq x) = x$ and, thus, the pair of variables $F(T - U)$ and $1 - F(T - U)$ have the same joint distribution as the pair $1 - U$ and U. We also use the fact that, by the definition of λ, $1 - F(n) = \mathrm{e}^{-\lambda}$. By applying Lemma 1 to (3), we obtain $S_j \leq_{\mathrm{st}} C(-\ln \mathrm{P}(S_j = 0))$, and so the induction is complete and the theorem is proved.

3. Applications

In this section, we apply Theorem 1 to counts of sequence patterns and the reliability of the m-consecutive-k-of-n system.

Suppose that a coin for which the probability of heads is equal to p is flipped n times. Let S denote the number of times a given pattern appears as a run, including overlapping runs. For example, the pattern HHHH (i.e. four heads) appears twice in the sequence HHHHH. The study of approximations for this classical problem goes back at least to von Mises (1921), and is often used to test the effectiveness of Poisson or compound Poisson approximations; see, for example, Barbour *et al.* (1992), Arratia *et al.* (1989), Erhardsson (2000), Chryssaphinou *et al.* (2001), Chryssaphinou and Papastavridis (1988), and Geske *et al.* (1995). There is also an extensive literature on this type of problem in the context of reliability; see the survey by Chang *et al.* (2000). In the reliability setting, this is called the m-consecutive-k-of-n system, a system of n independent components which fails if there are at least m runs of at least k failed components. Some approximations for this problem in the case $m = 1$ were studied in Peköz (1996), and an exact formula was given in Peköz and Ross (1995).

Corollary 1. *If a coin for which the probability of heads is equal to p is flipped n times, and S denotes the number of times k heads appear in a row (including overlapping runs), then $S \leq_{\mathrm{st}} C$, where C has a* $\mathrm{CP}(-\ln \mathrm{P}(S = 0), Y)$ *distribution, and Y has a geometric distribution with parameter $1 - p$.*

Proof. Let $X_i = 1$ if a run of length k ends with flip number i, and let $S = \sum_{i=1}^{n} X_i$. It is clear that the indicator variables are $(k - 1)$-dependent. Given that a run appears at some position, the number of subsequent overlapping runs plus the initial run follows a geometric distribution with parameter $1 - p$. After the first tails appears, no subsequent run can appear for at least k additional flips. Thus, it is clear that the choice of the geometric$(1 - p)$ distribution for the clump size Y will satisfy (1). The result then follows from Theorem 1.

Theorem 1 also applies to non-overlapping patterns.

https://doi.org/10.1239/jap/1143936260 Published online by Cambridge University Press

Corollary 2. *If a coin for which the probability of heads is equal to p is flipped n times, and S denotes the number of times a given pattern, which cannot overlap with itself, appears (e.g. TTTH), then $S \leq_{\mathrm{st}} C$, where C has a Poisson distribution with parameter $\lambda = -\ln \mathrm{P}(S = 0)$.*

Proof. Again let $X_i = 1$ if the pattern ends with flip number i, and let $S = \sum_{i=1}^{n} X_i$. Given a pattern appearance, no other pattern can appear for at least k additional flips. Thus, it is clear that $Y = 1$ will satisfy (1) and, since $\mathrm{CP}(\lambda, 1)$ is a Poisson(λ) distribution, the result follows from Theorem 1.

A similar result applies to any pattern which can overlap with itself. The clump size variable is still geometric, but with a different parameter. For concreteness, we consider the pattern HTHT.

Corollary 3. *If a coin for which the probability of heads is equal to p is flipped n times, and S denotes the number of times the pattern HTHT appears, then $S \leq_{\mathrm{st}} C$, where C has a $\mathrm{CP}(-\ln \mathrm{P}(S = 0), Y)$ distribution and Y has a geometric distribution with parameter $1/(1+p)$.*

Proof. Let $X_i = 1$ if the pattern ends with flip number i, and let $S = \sum_{i=1}^{n} X_i$. Suppose that the pattern HTHT has just appeared; let Y be equal to the total number of times HTHT appears (including this initial one) before two tails in a row (TT) appears. Once the pattern TT appears, the next appearance of HTHT cannot overlap with it and cannot appear for at least four additional flips. It can be seen that Y satisfies

$$Y \stackrel{\mathrm{D}}{=} \begin{cases} 1 & \text{if the next flip is T,} \\ 1 + Y & \text{if the next two flips are HT,} \\ Y & \text{if the next two flips are HH,} \end{cases}$$

where '$\stackrel{\mathrm{D}}{=}$' denotes equality in distribution. This gives the moment generating function

$$\begin{aligned} \Phi(t) &= \mathrm{E}[\mathrm{e}^{tY}] \\ &= (1-p)\mathrm{e}^{t} + p(1-p)\mathrm{e}^{t}\Phi(t) + p^{2}\Phi(t) \\ &= \frac{(1/(1+p))\mathrm{e}^{t}}{1-(1-1/(1+p))\mathrm{e}^{t}}, \end{aligned}$$

which is the moment generating function of a geometric random variable with parameter $1/(1+p)$.

4. A numerical example

Here we give a numerical example and compare the approximations obtained to the exact values. Let S be equal to the number of times four heads appear in a row in ten flips of a fair coin. Let C_1 have the compound Poisson distribution $\mathrm{CP}(-\ln \mathrm{P}(S = 0), \text{geometric}(\frac{1}{2}))$, which is our approximation from Theorem 1. For the parameter we calculated that $-\ln \mathrm{P}(S = 0) = 0.281$.

For the purposes of comparison, we also compute the usual compound Poisson approximation used in the literature in the context of Stein's method. For this type of problem, Erhardsson (2000, Theorem 3.1) used the $\mathrm{CP}(\lambda, \text{geometric}(\frac{1}{2}))$ distribution with

$$\lambda = (n - r + 1)p^{r}(1-p) = 0.219,$$

where $n = 10$, $r = 4$, and $p = \frac{1}{2}$. This approximation used the 'declumping' idea, i.e. that the appearances of the pattern THHHH should approximately follow a Poisson process; thus, the

TABLE 1: Distribution of S, C_1, and C_2.

k	$P(S = k)$	$P(C_1 = k)$	$P(C_2 = k)$
0	0.7549	0.7549	0.8035
1	0.1328	0.1061	0.0879
2	0.0635	0.0605	0.0487
3	0.0293	0.0343	0.0270
4	0.0117	0.0194	0.0149
5	0.0049	0.0109	0.0082
6	0.0020	0.0061	0.0045
7	0.0010	0.0034	0.0025

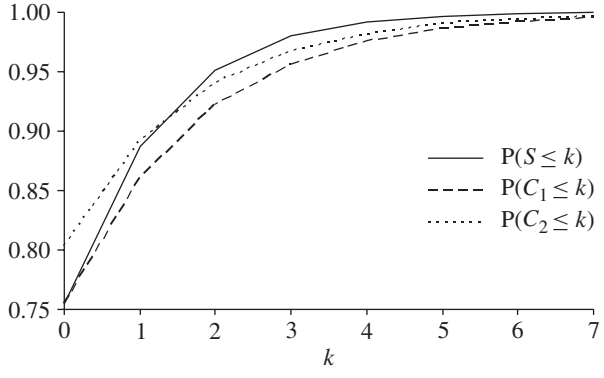

FIGURE 1: Cumulative distribution function of S, C_1, and C_2.

subsequent appearances of the pattern (the clump size) should have a geometric distribution with parameter $\frac{1}{2}$. We let C_2 have this compound Poisson distribution.

We use EXCEL® to calculate the exact distribution of S, C_1, and C_2; the results are displayed in Table 1. There S has the exact distribution, C_1 is our (stochastically larger) approximation, and C_2 is the usual approximation given in the literature. It can be seen that our approximation is better at the low end of the distribution, while the usual approximation is better further out in the tail of the distribution. In Figure 1, we plot the cumulative distribution functions of these three distributions, and see that the cumulative distribution function for our approximation C_1, as expected, goes below the ones for C_2 and S. This indicates, as expected from Theorem 1, that C_1 is stochastically larger than C_2 and S.

Acknowledgement

I would like to thank an anonymous referee, who found several errors in an earlier version of this paper.

References

ALDOUS, D. J. (1989). *Probability Approximations via the Poisson Clumping Heuristic.* Springer, New York.

ARRATIA, R., GOLDSTEIN, L. AND GORDON, L. (1989). Two moments suffice for Poisson approximations: the Chen–Stein method. *Ann. Prob.* **17,** 9–25.

Barbour, A. D. and Månsson, M. (2002). Compound Poisson process approximation. *Ann. Prob.* **30,** 1492–1537.
Barbour, A. D. and Chryssaphinou, O. (2001). Compound Poisson approximation: a user's guide. *Ann. Appl. Prob.* **11,** 964–1002.
Barbour, A. D., Holst, L. and Janson, S. (1992). *Poisson Approximation.* Oxford University Press.
Chang, G. J., Cui, L. and Hwang, F. K. (2000). *Reliabilities of Consecutive-k Systems* (Network Theory Appl. **4**). Kluwer, Dordrecht.
Chryssaphinou, O. and Papastavridis, S. (1988). A limit theorem on the number of overlapping appearances of a pattern in a sequence of independent trials. *Prob. Theory Relat. Fields* **79,** 129–143.
Chryssaphinou, O., Papastavridis, S. and Vaggelatou, E. (2001). Poisson approximation for the non-overlapping appearances of several words in Markov chains. *Combin. Prob. Comput.* **10,** 293–308.
Erhardsson, T. (2000). Compound Poisson approximation for counts of rare patterns in Markov chains and extreme sojourns in birth–death chains. *Ann. Appl. Prob.* **10,** 573–591.
Geske, M. X. *et al.* (1995). Compound Poisson approximation for word patterns under Markovian hypotheses. *J. Appl. Prob.* **32,** 877–892.
Klass, M. J. and Nowicki, K. (2003). An optimal bound on the tail distribution of the number of recurrences of an event in product spaces. *Prob. Theory Relat. Fields* **126,** 51–60.
Le Cam, L. (1960). An approximation theorem for the Poisson binomial distribution. *Pacific J. Math.* **10,** 1181–1197.
Peköz, E. A. (1996). Stein's method for geometric approximation. *J. Appl. Prob.* **33,** 707–713.
Peköz, E. A. and Ross, S. M. (1995). A simple derivation of exact reliability formulas for linear and circular consecutive-k-of-n:F systems. *J. Appl. Prob.* **32,** 554–557.
Peköz, E. A. and Ross, S. M. (2004). Compound random variables. *Prob. Eng. Inf. Sci.* **18,** 473–484.
Von Mises, R. (1921). Das Problem der Iterationen. *Z. Angew. Math. Mech.* **1,** 298–307.