J. Appl. Prob. **33**, 707–713 (1996)
Printed in Israel

STEIN'S METHOD FOR GEOMETRIC APPROXIMATION

EROL A. PEKÖZ,* *University of California, Los Angeles*

Abstract

The Stein–Chen method for Poisson approximation is adapted to the setting of the geometric distribution. This yields a convenient method for assessing the accuracy of the geometric approximation to the distribution of the number of failures preceding the first success in dependent trials. The results are applied to approximating waiting time distributions for patterns in coin tossing, and to approximating the distribution of the time when a stationary Markov chain first visits a rare set of states. The error bounds obtained are sharper than those obtainable using related Poisson approximations.

STEIN'S METHOD; GEOMETRIC APPROXIMATION; POISSON APPROXIMATION; HITTING TIMES; PATTERNS IN COIN TOSSING

AMS 1991 SUBJECT CLASSIFICATION: PRIMARY 60E15
SECONDARY 60C05; 60J10

1. Introduction

There have been many recent papers on the Stein–Chen method for bounding Poisson approximation errors (see [5] and references therein) and the application of the techniques to binomial, multinomial, compound Poisson and (recently and independently of the present work) geometric approximations: see [9], [14], [3], [4]. In this paper we further develop the Stein–Chen techniques to bound errors for a geometric approximation to the distribution of W, the number of failures before the first success in dependent trials. The results are applied to approximating the distribution of the number of biased coin flips needed until a given pattern first appears as a run, and the distribution of time until a stationary Markov chain first visits a rare set of states. This pattern distribution and its approximations have been studied in many places; see [5], [6], [7], [8], [10], [11], [16] and references therein. The application to Markov chain hitting times has also been studied in many places; see [5], [1], [2], [13].

The usual Poisson approximation approach to estimating $P(W>t)$ is to define indicators I_j, $j=1,2,\cdots$ each indicating the event of a success on trial j, and to use the Stein–Chen method to bound $|P(Y=0)-e^{-\lambda t}|$ where $Y=\Sigma_{j=1}^{t+1} I_j$, and $\lambda=E[Y]$. This approach is limited to bounding the error of tail probability estimates for W. To get bounds for estimating $P(W\in[i, j])$ the triangle inequality in the form $\sup_{i,j} |P(X\in[i, j])-P(Y\in[i, j])| \leqq 2 \sup_i |P(X>i)-P(Y>i)|$, can be used, but these bounds may be too large to be useful for estimating the probabilities of small intervals.

Received 18 July 1994; revision received 26 April 1995.
* Postal address: Department of Mathematics, University of California, Los Angeles, CA 90024, USA.

Our approach, in contrast, directly yields error bounds for estimating $\boldsymbol{P}(W \in B)$ for any set B; in the applications discussed these bounds specialized to tail probabilities are roughly half those obtainable using the usual Poisson approach, and specialized to point probabilities they are better than the square of bounds obtainable using the Poisson approach with the triangle inequality. Note that although the approximations we obtain are nearly indentical to Poisson approximations, our methods can yield better bounds. For cases where λ is small, see [15] for a method of improving the actual approximations.

In Section 2 we present a theorem bounding the errors for geometric approximations. In Section 3 we apply it to Markov chain hitting times, and in Section 4 we apply it to pattern waiting times. In Section 5 we give a method for improving the approximations in cases where neighboring trials are strongly positively correlated so that successes occur in 'clumps'.

2. Main result

Let W be the number of failures before the first success in a sequence of dependent Bernoulli trials with $p=1-q=\boldsymbol{P}(W=0)$. The main result below gives bounds on the error of the geometric approximation to the distribution of W. Let $1+V$ have the distribution of W, given $W>0$, and let G_p have a geometric distribution ('starting at zero') with parameter p so that, for $k \geqq 0$, $\boldsymbol{P}(G_p=k)=pq^k$.

Theorem 1. *With the above definitions,*
(a) $|\boldsymbol{P}(W \in B)-\boldsymbol{P}(G_p \in B)| \leqq qp^{-1}(1-q^{|B|})\boldsymbol{P}(W \neq V)$, *and*
(b) $d(W, G_p) \leqq qp^{-1}\boldsymbol{P}(W \neq V)$, *where* $d(X, Y) \equiv \sup_A |\boldsymbol{P}(X \in A)-\boldsymbol{P}(Y \in A)|$.

To prove the theorem we first need the following lemma; our argument closely follows the argument for the Stein–Chen method given in [5] for Poisson approximation.

For any set B and any $p=1-q$, we construct the function $f=f_{p,B}$ defined by $f(0)=0$ and for $k>0$, $qf(k+1)-f(k)=1_{(k \in B)}-\boldsymbol{P}(G_p \in B)$, and it can be easily verified that the solution to the above equations is given by $f(k)=\Sigma_{i \in B}\, q^i-\Sigma_{i \in B,\, i \geq k}\, q^{i-k}$.

Lemma 1. *For any* $j \geqq 0$, $k \geqq 0$, $|f(j)-f(k)| \leqq (1-q^{|B|})/p$.

Proof. Note that $f(j)-f(k)=\Sigma_{i \in B,\, i \geq k}\, q^{i-k}-\Sigma_{i \in B,\, i \geq j}\, q^{i-j}$, and since neither term on the right can be larger than $\Sigma_{i=0}^{|B|-1}\, q^i=(1-q^{|B|})/p$, the result follows.

Proof of Theorem 1. Substituting $k=W$ in the recurrence relation for f, taking expectations, and noting that $f(0)=0$, we obtain

$$\begin{aligned}
|\boldsymbol{P}(W \in B)-\boldsymbol{P}(G_p \in B)| &= |q\boldsymbol{E}[f(W+1)]-\boldsymbol{E}[f(W)]| \\
&= |q\boldsymbol{E}[f(W+1)]-\boldsymbol{E}[f(W) \mid W>0]\boldsymbol{P}(W>0)| \\
&= |q\boldsymbol{E}[f(W+1)-f(V+1)]| \\
&\leqq q\boldsymbol{E}\,|f(W+1)-f(V+1)|.
\end{aligned}$$

Since Lemma 1 gives $|f(W+1)-f(V+1)| \leqq (1-q^{|B|})/p \leqq 1/p$, the result follows.

3. Application to Markov chain hitting times

In this section we consider an ergodic Markov chain started according to its stationary distribution $\boldsymbol{\pi}$ and use a geometric distribution with parameter $p=1-q=\Sigma_{i\in A}\pi_i$ to approximate the distribution of the time W of the first visit to the set of states A. The result is as follows, where $P_{ij}^{(n)}$ denotes n-step transition probabilities for the Markov chain.

Theorem 2. *Let* $W=\min\{i: X_i\in A\}$ *be the time of the first visit to* A *for the stationary Markov chain* $\{X_i, i\geqq 0\}$, *and let* $p=1-q=\Sigma_{i\in A}\pi_i$. *Then*

$$|\boldsymbol{P}(W\in B)-\boldsymbol{P}(G_p\in B)|\leqq p^{-1}(1-q^{|B|})\sum_{i,\,j\in A}\pi_i\sum_{n\geqq 1}|P_{ij}^{(n)}-\pi_j|.$$

Proof. In light of Theorem 1, we proceed by creating a coupling to bound $\boldsymbol{P}(W\neq V)$. First, let $(Z, X_0, X_1,\cdots)$ be the desired Markov chain started according to its stationary distribution, and let $(Z', Y_0, Y_1,\cdots)$ be a coupled copy started according to the stationary distribution restricted to A^c. Letting $W=\min\{i: X_i\in A\}$ and $V=\min\{i: Y_i\in A\}$, note that W has the same distribution as is described in the theorem, and $1+V$ has the distribution of W given $W>0$. Next note that $\boldsymbol{P}(W\neq V)\leqq \Sigma_{j\in A}\Sigma_{n\geqq 0}[\boldsymbol{P}(X_n=j, Y_n\neq j)+\boldsymbol{P}(Y_n=j, X_n\neq j)]$.

We then couple (X_n, Y_n) as is done in [5], p. 166, using the maximal coupling of Griffeath [12] (see also [17]) so that $\boldsymbol{P}(X_n=Y_n=j)=\pi_j\wedge\boldsymbol{P}(Y_n=j)$ and therefore

$$\begin{aligned}\boldsymbol{P}(X_n=j, Y_n\neq j)&=\pi_j-\boldsymbol{P}(X_n=Y_n=j)\\&=\left[\pi_j-\sum_{i\in A^c}\frac{\pi_i}{q}P_{ij}^{(n+1)}\right]^+\\&=\left[\pi_j-\sum_{i}\frac{\pi_i}{q}P_{ij}^{(n+1)}+\sum_{i\in A}\frac{\pi_i}{q}P_{ij}^{(n+1)}\right]^+\\&=\left[\frac{1}{q}\sum_{i\in A}\pi_i(P_{ij}^{(n+1)}-\pi_j)\right]^+\\&\leqq\frac{1}{q}\sum_{i\in A}\pi_i[P_{ij}^{(n+1)}-\pi_j]^+\end{aligned}$$

where the next to last equality follows from $\Sigma_i\pi_iP_{ij}^{(n+1)}=\pi_j$. A similar calculation gives $\boldsymbol{P}(Y_n=j, X_n\neq j)\leqq(1/q)\Sigma_{i\in A}\pi_i[\pi_j-P_{ij}^{(n+1)}]^+$. We can then deduce $\boldsymbol{P}(W\neq V)\leqq(1/q)\Sigma_{i,\,j\in A}\pi_i\Sigma_{n\geqq 1}|P_{ij}^{(n)}-\pi_j|$, and the result follows upon application of Theorem 1.

Remark. The bound given by Theorem 2 can be compared with the related bound of Theorem 8.H presented in [5] (for a Poisson approximation to the distribution of the number of visits to A by time n) in the case where tail probabilities of W are of interest.

Example. This example is taken from [5], p. 168. Let $X_1, X_2, \cdots, X_m$ be a collection of m independent two-state Markov chains on $\{0, 1\}$ each with $\boldsymbol{P}(X_i(n+1)=0 \mid X_i(n)=0)=p+\varepsilon(1-p)$ and $\boldsymbol{P}(X_i(n+1)=1 \mid X_i(n)=1)=\varepsilon p+(1-p)$ for some $0 \leqq \varepsilon < 1$. Note that when $\varepsilon=0$ the sequence $\{X_i(j)\}_{j \geqq 1}$ is just a sequence of i.i.d. Bernoulli random variables, but for $\varepsilon > 0$ there is a greater tendency for runs to occur. Starting from stationarity, we are interested in the distribution W, the time when all of the X_i are simultaneously zero. Next, define a Markov chain with state space $\{0, 1\}^m$ and n-step transition probability $P_{ij}^{(n)}$ so that coordinate i at time n stores the value of X_i at time n. Calculations from [5] show $\boldsymbol{P}(W=0)=p^m$ and, letting state j be the state where all the X_i are zero, that $\Sigma_{n \geqq 1} |P_{jj}^{(n)}-p^m| \leqq 2\varepsilon p^m/(1-\varepsilon)$ when $\varepsilon \leqq p/((m-1)(1-p))$ and $m \geqq 2$, so Theorem 2 gives $d(W, G_{p^m}) \leqq 2\varepsilon p^m/(1-\varepsilon)$ under the above conditions on ε and m.

4. Application to sequence patterns

A biased coin is flipped in succession. What is the distribution of the number of flips preceding the first appearance of a given pattern of heads and tails? Here we approximate this distribution using a geometric distribution. Brown and Ge [7], using reliability theory methods, obtain the result of Corollary (1b) below for the restricted class of non-overlapping patterns. The results of [4, Theorem 8.F] and [10] apply to general patterns, but only directly bound the errors when estimating tail probabilities of the waiting time distribution. Though these can then be used with the triangle inequality to bound other probabilities, Theorem 3 can give much better bounds.

Let $X_0, X_1, \cdots$ be a sequence of i.i.d. Bernoulli trials and let some desired k-digit binary pattern be specified. Let I_i be the indicator of the event that the desired pattern appears in $(X_i, X_{i+1}, \cdots, X_{i+k-1})$ and let $W=\min\{i : I_i=1\}$ be the number of trials preceding the first appearance of the pattern. Also, for $1 \leqq i \leqq k-1$, let $c_i=\boldsymbol{P}(I_i=1 \mid I_0=1)$ be the probability that, given an occurrence of the pattern, there is also an overlapping occurrence i trials later. For example, with Bernoulli(.5) trials, the pattern 11011 gives $c_1=c_2=0$, $c_3=1/8$, $c_4=1/16$.

Theorem 3. *For k-digit patterns with* $p=1-q=\boldsymbol{P}(W=0)$,
(a) $|\boldsymbol{P}(W \in B)-\boldsymbol{P}(G_p \in B)| \leqq (1-q^{|B|})\Sigma_{i=1}^{k-1} |c_i-p|$, *and*
(b) $d(W, G_p) \leqq \Sigma_{i=1}^{k-1} |c_i-p|$.

Proof. Define a Markov chain with 2^k states so that the state at time n encodes the outcomes of the k consecutive Bernoulli trials starting with trial n. Let $A=\{i\}$ consist of the state corresponding to the desired pattern. The result then follows from Theorem 2, since

$$P_{ii}^{(n)}=c_n \quad \text{when } 1 \leqq n \leqq k-1$$
$$=p \quad \text{when } k \leqq n.$$

We next give an immediate corollary for non-overlapping patterns, patterns which have $c_i=0$ for all i.

Corollary 1. *For a non-overlapping k-digit pattern,*
(a) $|\boldsymbol{P}(W \in B) - \boldsymbol{P}(G_p \in B)| \leqq (k-1)(1-q^{|B|})p$, *and*
(b) $d(W, G_p) \leqq (k-1)p$.

Remark. The result of Corollary (1b) was obtained by Brown and Ge [7]. Note that for point probabilities, Corollary (1a) gives $|\boldsymbol{P}(W=k) - pq^k| \leqq (k-1)p^2$.

For patterns where many overlaps are possible, the bound from Theorem 3 can be improved using a geometric with larger mean. The following corollary illustrates this, as does Corollary 3 below.

Corollary 2. *Let W count the number of trials preceding the first appearance of a run of k consecutive* 1*'s in a sequence of i.i.d. Bernoulli(a) trials. Letting* $p = (1-a)a^k$, *we have* $d(W, G_p) \leqq (k+1)p$.

Proof. Consider instead the non-overlapping pattern consisting of a 0 followed by k 1's. We then apply Corollary 1 and add a^k because this new pattern's appearance time differs from $W-1$ only if the first k trials are 1's. Finally note that a result similar to Theorem 3 and Corollary 1 can be established using $1+W$ instead of W.

Remark. The bound in Corollary 2 is half the bound in [5], p. 164. For a simple derivation of the exact distribution of W defined as in Corollary 2, see [16].

5. Improving approximations when 'clumping' occurs

For a Markov chain X_n when visits to A generally occur in 'clumps', that is when $P_{ij}^{(n)}$ is much larger than π_j for $i, j \in A$ and n small, the bound of Theorem 3 can be poor. In this case the hitting time generally has mean much larger than $1/p$ where $p = \Sigma_{i \in A} \pi_i$. In the case where the Markov chain is m-dependent (i.e. when X_i is independent of X_j for $|i-j| > m$), much better bounds can be obtained as follows in the same spirit as Corollary 2 above using a geometric with a larger but less easily calculable mean.

Theorem 4. *Let* $W = \min\{i : X_i \in A\}$ *be the time of the first visit to A for the stationary m-dependent Markov chain* $\{X_i, i \geqq 0\}$, *let* $p = 1 - q = \Sigma_{i \in A} \pi_i$, *and let* $a = \boldsymbol{P}(W=m)$. *Then* $d(W, G_a) \leqq 2mp + ma$.

Proof. Consider instead starting the chain at time $-m$ and let $Z = \min\{i \geqq 0 : X_{i-m} \notin A, X_{i-m+1} \notin A, \cdots, X_{i-1} \notin A, X_i \in A\}$; note that it can be viewed as the hitting time on a set of states $\hat{A}$ for a related stationary $2m$-dependent Markov chain where the state at time n now encodes the values $(X_{n-m}, X_{n-m+1}, \cdots, X_n)$.

We next apply Theorem 2 to this related new chain (letting $\hat{P}_{ij}^{(n)}$ and $\hat{\pi}_j$ refer to this new chain) while noting $\boldsymbol{P}(Z=0) = \boldsymbol{P}(W=m) = a$. Since this chain is $2m$-dependent we have $\hat{P}_{ij}^{(n)} = \hat{\pi}_j$ for $n > 2m$, and also, for $i, j \in \hat{A}$, we have $\hat{P}_{ij}^{(n)} = 0$ when $n \leqq m$.

Note that every state in $\hat{A}$ is of the form $(a_0, \cdots, a_m)$ where $a_0 \notin A, \cdots, a_{m-1} \notin A$, $a_m = k \in A$ and for a given state $j = (a_0, \cdots, a_m) \in \hat{A}$ there is a corresponding state $k = (a_m) \in A$.

Since the original chain was m-dependent we have, for $m < n \leqq 2m$ and $i, j \in \hat{A}$, that $\hat{P}_{ij}^{(n)} \leqq P(X_n = k) = \pi_k \geqq \hat{\pi}_j$, where k is the state in A corresponding to $j \in \hat{A}$.

These together imply

$$\sum_{j \in \hat{A}} |\hat{P}_{ij}^{(n)} - \hat{\pi}_j| \leqq p \quad \text{when } 1 \leqq n \leqq m$$
$$\leqq a \quad \text{when } m < n \leqq 2m$$
$$= 0 \quad \text{when } n > 2m.$$

Theorem 2 then gives $d(Z, G_a) \leqq mp + ma$ and the result follows noting that Z differs from W only if a visit to A happens before time 0 in the original chain, an event of probability at most mp.

As mentioned before, the bound from Theorem 3 can be poor for long patterns where many overlaps are possible (e.g. the pattern 101010101). To improve the bound in this case, Theorem 4 can be immediately applied (a situation with $(k-1)$-dependence).

Corollary 3. *Let W count the number of i.i.d. Bernoulli trials preceding the first appearance of a given k-digit binary pattern with $p = \boldsymbol{P}(W=0)$ and $a = \boldsymbol{P}(W=k-1)$; $d(W, G_a) \leqq 2(k-1)p + (k-1)a$.*

Remark. A result related to Corollary 3 for Poisson approximation appears as Theorem 2.4 in [11].

To compute the value of $a = \boldsymbol{P}(W=k-1)$ for k-digit patterns, the recursion in the following lemma can be used.

Lemma 2. *For k-digit patterns with c_i defined as in Theorem* 3, *the following recursion holds for $0 \leqq i \leqq k-1$: $\boldsymbol{P}(W=i) = p - \Sigma_{j=0}^{i-1} c_{i-j} \boldsymbol{P}(W=j)$.*

Proof. Writing $\{X_i \in A\}$ for the event that the pattern occurs (not necessarily for the first time) starting with trial i,

$$\boldsymbol{P}(X_i \in A) = p = \boldsymbol{P}(X_i \in A,\, W \geqq i) + \sum_{j=0}^{i-1} \boldsymbol{P}(X_i \in A,\, W=j)$$
$$= \boldsymbol{P}(W=i) + \sum_{j=0}^{i-1} c_{i-j} \boldsymbol{P}(W=j)$$

and the result follows upon re-arranging.

Acknowledgements

I would like to thank the referee for many comments which greatly improved the presentation of this paper. I would also like to thank Sheldon Ross and Michael Klass for very many helpful comments and inspirational discussions, and thanks are also due to Jim Pitman, David Aldous and David Freedman for their encouragement and comments.

References

[1] Aldous, D. (1989) *Probability Approximations via the Poisson Clumping Heuristic*. Springer, Berlin.

[2] Aldous, D. (1982) Markov chains with almost exponential hitting times. *Stoch. Proc. Appl.* **13**, 305–310.

[3] Barbour, A. D., Chen L. H. Y. and Loh, W.-L. (1992) Compond Poisson approximation for non-negative random variables via Stein's method. *Ann. Prob.* **20**, 1843–1866.

[4] Barbour, A. D. and Grübel, R. (1995) The first divisible sum. *J. Theoret. Prob.* **8**, 39–47.

[5] Barbour, A. D., Holst, L. and Janson, S. (1992) *Poisson Approximation.* Oxford University Press, Oxford.

[6] Blom, G. and Thorburn, D. (1982) How many random digits are required until given sequences are obtained? *J. Appl. Prob.* **19**, 518–531.

[7] Brown, M. and Ge, G. (1984) On the waiting time for the first occurrence of a pattern. In *Reliability Theory and Models.* ed. M. Abdel-Hameed, E. Çinlar and J. Quinn.

[8] Chryssaphinou, O. and Papastavridis, S. (1988) A limit theorem for the number of non-overlapping occurrences of a pattern in a sequence of independent trials. *J. Appl. Prob.* **25**, 428–431.

[9] Ehm, W. (1991) Binomial approximation to the Poisson binomial distribution. *Statist. Prob. Lett.* **11**, 7–16.

[10] Godbole, A. P. (1991) Poisson approximations for runs and patterns of rare events. *Adv. Appl. Prob.* **23**, 851–865.

[11] Godbole, A. P. and Schaffner, A. A. (1993) Improved Poisson approximations for word patterns. *Adv. Appl. Prob.* **25**, 334–347.

[12] Griffeath, D. (1976) Partial coupling and loss of memory for Markov chains. *Ann. Prob.* **4**, 550–558.

[13] Keilson, J. (1979) *Markov Chain Models — Rarity and Exponentiality*. Springer, Berlin.

[14] Loh, W.-L. (1992) Stein's method and multinomial approximation. *Ann. Appl. Prob.* **2**, 536–554.

[15] Peköz , E. and Ross, S. M. (1994) Improving Poisson approximations. *Prob. Eng. Inf. Sci.* **8**, 449–462.

[16] Peköz, E. 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.

[17] Pitman, J. W. (1976) On coupling of Markov chains. *Z. Wahrscheinlichkeitsth.* **35**, 315–322.