Probability in the Engineering and Informational Sciences, **13**, 1999, 1–9. Printed in the U.S.A.

THE MEAN WAITING TIME FOR A PATTERN

SHELDON M. ROSS*
Department of Industrial Engineering and Operations Research
University of California at Berkeley
Berkeley, California 94720

Consider a sequence of independent and identically distributed random variables along with a specified set of k-vectors. We present an expression for $E[T]$, the mean time until the last k observed random variables fall within this set. Not only can this expression often be used to obtain bounds on $E[T]$, it also gives rise to an efficient way of approximating $E[T]$ by a simulation. Specific lower and upper bounds for $E[T]$ are also derived. These latter bounds are given in terms of a parameter, and a Markov chain Monte Carlo approach to approximate this parameter by a simulation is indicated. The results of this paper are illustrated by considering the problem of determining the mean time until a sequence of k-valued random variables has a run of size k that encompasses each value.

1. INTRODUCTION

Let $X_1, X_2, \ldots$ be a sequence of independent random variables having a common distribution function F. Let $\mathcal{A}$ be a fixed set of real valued vectors of size k, and let

$$N = \min\{n \geq k : (X_{n-k+1}, \ldots, X_n) \in \mathcal{A}\}.$$

Also, let

$$\alpha = P\{(X_1, \ldots, X_k) \in \mathcal{A}\}.$$

We are interested in obtaining bounds for $E[N]$ in situations where α is small. (When α is not small a straightforward simulation can be used to approximate $E[N]$. Such a simulation cannot be efficiently employed when α is small, as in this case N

*This research was supported by the National Science Foundation grant DMI-9610046 with the University of California.

© 1999 Cambridge University Press 0269-9648/99 $12.50

https://doi.org/10.1017/S0269964899131012 Published online by Cambridge University Press

will have an approximately geometric distribution with a standard deviation approximately equal to its large mean.)

In Section 2, we define a renewal process along with a reward structure that enables us to obtain an expression for $E[N]$. We also explain how this expression can be exploited when determining $E[N]$ by a simulation.

In Section 3, we suppose that the random variables X_i, $i \geq 1$, are discrete with k possible values with respective probabilities $p_1, \ldots, p_k$. With N equal to the first time that a run of k consecutive random variables includes all the k possible values, we use the expression derived in Section 2 to obtain a lower and an upper bound for $E[N]$. In addition, we show that $P\{N \geq n\}$ and $E[N]$ are Schur convex functions of the vector $(p_1, \ldots, p_k)$.

In Section 4, we return to the general problem and show how the renewal process can be used to derive lower and upper bounds on $E[N]$. These bounds are given in terms of a probability involving $2k-1$ data values, and, in Section 5, we indicate a general Markov chain Monte Carlo technique that can be used to evaluate it. We also show in this section how a much more efficient simulation procedure can be used to obtain this probability for the problem considered in Section 3.

2. USING A RENEWAL PROCESS WITH A REWARD STRUCTURE

Let $N_1 = N$ denote the first time that the vector of most recent k values is in $\mathcal{A}$. At this point, start over again and, without using any of the first N_1 data values, let N_2 denote the next time that the vector of most recent k values is in $\mathcal{A}$, and so on. For instance, if N is the first time that a value is equal to the sum of the previous two values and the successive X_i are 1,1,1,2,3,5,8, then $N_1 = 4$, $N_2 = 7$. In this way, we define the event times N_j, $j \geq 1$, of a renewal process.

Now, define a reward structure by supposing that a reward of 1 is earned at time n, $n \geq k$, if $X_{n-k+1}, \ldots, X_n$ is in $\mathcal{A}$. Thus, for instance, in the situation considered above, one unit rewards are earned at times 4,5,6, and 7. Let R_i denote the reward earned at time i, $i \geq k$. Since, for each j, $j = 0, \ldots, k-1$,

$$R_{k+j}, R_{2k+j}, R_{3k+j}, \ldots$$

are independent and identically distributed (i.i.d.) with mean α, it follows from the strong law of large numbers that, with probability 1,

$$\lim_n \frac{\sum_{i=1}^{n} R_i}{n} = \alpha. \tag{1}$$

Let R_i^* denote the reward earned between the ith and the $(i+1)$st renewal, $i \geq 1$. Then, with $\mathbf{V}_i$ equal to the ith renewal vector (for instance, $(X_{N-k+1}, \ldots, X_N)$ is the first renewal vector), it follows that R_i^* and $\mathbf{V}_{i+1}$ are dependent, as are $\mathbf{V}_{i+1}$ and R_{i+1}^*.

https://doi.org/10.1017/S0269964899131012 Published online by Cambridge University Press

As a result, R_i^* and R_{i+1}^* are dependent. However, as R_i^* is clearly independent of $\mathbf{V}_j$ for $j \geq i+2$, it follows that R_i^* is independent of R_j^* for $j \geq i+2$. Thus, we see that

$$R_1^*, R_3^*, R_5^*, \ldots$$

and

$$R_2^*, R_4^*, R_6^*, \ldots$$

are both i.i.d. sequences. As a result, with probability 1,

$$\lim_n \frac{R_1^* + \cdots + R_n^*}{N_{n+1}} = \lim_n \frac{(R_1^* + \cdots + R_n^*)/n}{N_{n+1}/n}$$

$$= \frac{E[R^*]}{E[N]}. \tag{2}$$

As the limits in Eqs. (1) and (2) are clearly equal, we obtain that

$$E[N] = \frac{E[R^*]}{\alpha}. \tag{3}$$

Remark: Equations (1) and (2) could also have been obtained from the ergodic theorem, as both R_i, $i \geq 1$ and R_i^*, $i \geq 1$ are stationary and ergodic.

Equation (3) can be used to determine $E[N]$ explicitly when $\mathcal{A}$ is a one-point set (Ross [4, p. 396]) and can often be used to obtain bounds on $E[N]$ (see Sect. 3). In addition, it can be very useful in approximating $E[N]$ by a simulation. Generating data until a renewal vector occurs at time N gives rise to the unbiased estimators N and

$$\frac{E[R^* | X_{N-k+2}, \ldots, X_N]}{\alpha} = \frac{1 + \sum_{j=1}^{k-1} P\{(X_{N-j+1}, \ldots, X_{N-j+k}) \in \mathcal{A} | X_{N-k+2}, \ldots, X_N\}}{\alpha}.$$

Thus, we can use an estimator of the form

$$cN + (1-c)E[R^* | X_{N-k+2}, \ldots, X_N]/\alpha.$$

The best value of c can be obtained from the simulation by standard methods (Ross [5, pp. 148–149]).

3. ON THE MEAN TIME FOR A MAXIMUM RUN OF DISTINCT VALUES

Suppose that the random variables X_i, $i \geq 1$, are finite valued random variables having probability mass function $p_j = P\{X_i = j\} > 0$, $j = 1, \ldots, k$, $\sum_{j=1}^k p_j = 1$. Let N denote the first time that a run of k consecutive values includes all the values $1, \ldots, k$; that is,

$$N = \min\{n \geq k : X_{n-k+1}, \ldots, X_n \text{ are all distinct}\}.$$

https://doi.org/10.1017/S0269964899131012 Published online by Cambridge University Press

As

$$\alpha = P\{X_1, \ldots, X_k \text{ are all distinct}\} = k! \prod_{j=1}^{k} p_j,$$

we have from Eq. (3) that

$$E[N] = \frac{E[R^*]}{k! \prod_{j=1}^{k} p_j}. \tag{4}$$

Now, define the set A_i by

$$A_i = \{X_{N-k+1}, \ldots, X_{N-k+i}\};$$

that is, A_i is the set of the first i values of a renewal vector. Also, let B_i equal the set of the first i values following this renewal vector. Then,

$$\begin{aligned} E[R^*] &= 1 + \sum_{i=1}^{k-1} E[\text{reward earned at time } i \text{ after a renewal}] \\ &= 1 + \sum_{i=1}^{k-1} P\{A_i = B_i\}. \end{aligned} \tag{5}$$

Now, suppose that the probabilities are numbered so that p_j is nondecreasing in j, $j = 1, \ldots, k$. Then,

$$P\{A_i = B_i | A_i = (j_1, \ldots, j_i)\} = i! \prod_{r=1}^{i} p_{j_r},$$

implying that

$$i! \prod_{j=1}^{i} p_j \le P\{A_i = B_i\} \le i! \prod_{j=k-i+1}^{k} p_j.$$

The preceding equation, along with Eqs. (4) and (5), gives the following result.

THEOREM 1: *If* $p_1 \le p_2 \le \cdots \le p_k$ *then*

$$\frac{1 + \sum_{i=1}^{k-1} i! \prod_{j=1}^{i} p_j}{k! \prod_{j=1}^{k} p_j} \le E[N] \le \frac{1 + \sum_{i=1}^{k-1} i! \prod_{j=k-i+1}^{k} p_j}{k! \prod_{j=1}^{k} p_j}.$$

When $p_j = 1/k, j = 1, \ldots, k,$

$$E[N] = \frac{k^k}{k!} \sum_{i=0}^{k-1} i!/k^i.$$

https://doi.org/10.1017/S0269964899131012 Published online by Cambridge University Press

THEOREM 2:

(a) *$P\{N > n\}$ is a Schur convex function of $p_1, \ldots, p_k$.*

(b) *$E[N]$ is a Schur convex function of $p_1, \ldots, p_k$.*

PROOF: To prove (a), fix j and r, $1 \le j < r \le k$, and compute $P\{N > n\}$ by conditioning on the values of Y_i, $i = 1, \ldots, n$, where

$$Y_i = X_i \quad \text{if } X_i \neq j \text{ and } X_i \neq r$$

$$Y_i = 0 \quad \text{otherwise.}$$

Each conditional probability will be of the form

$$P\{N > n | Y_1, \ldots, Y_n\} = \prod_{i=1}^{s} (p^{m_i} + (1-p)^{m_i}),$$

where $p = p_j/(p_j + p_r)$ and the values of s and $m_1, \ldots, m_s$ depend on $Y_1, \ldots, Y_n$. The quantity m_1 refers, for instance, to a set of m_1 indices whose Y value is 0 and which are such that $N \le n$ if the X values in any two of those indices are unequal. Because the distribution of the vector $Y_1, \ldots, Y_n$ depends on p_j, p_r only through $p_j + p_r$, this shows that $P\{N > n\}$ is a Schur convex function of the vector p_j, p_r. As j and r are arbitrary, it follows by the Ostrowski theorem (Marshall and Olkin [3, p. 57]) that $P\{N > n\}$ is a Schur convex function of $p_1, \ldots, p_k$.

Part (b) is an immediate consequence of part (a). ■

In a simulation, if we let $J_i = X_{N-k+i}$, then unbiased estimators from a simulation run are N and

$$\frac{1 + \sum_{i=1}^{k-1} i! \prod_{r=1}^{i} p_{J_r}}{k! \prod_{j=1}^{k} p_j}.$$

4. USING THE RENEWAL PROCESS TO OBTAIN BOUNDS

If we let M_n denote the event that a renewal occurs at time n, then, from Blackwell's theorem, we obtain

$$E[N] = \frac{1}{P(M_\infty)} \tag{6}$$

where

$$P(M_\infty) = \lim_{n \to \infty} P(M_n).$$

https://doi.org/10.1017/S0269964899131012 Published online by Cambridge University Press

Let A_n be the event that $(X_{n-k+1},\ldots,X_n) \in \mathcal{A}$. As a renewal at time n implies that A_n occurs, it follows that

$$P(M_n) \le \alpha, \tag{7}$$

which shows that $E[N] \ge 1/\alpha$.

To obtain more precise bounds for $E[N]$, note that the event $A_n A_{n-1}^c \cdots A_{n-k+1}^c$ implies that there is a renewal at time n. Thus, if we define

$$\omega \equiv P(A_n A_{n-1}^c \cdots A_{n-k+1}^c), \tag{8}$$

then we obtain the inequality

$$\omega \le P(M_n). \tag{9}$$

To obtain a bound in the other direction, let B be the event that the pattern does not appear at any of the times $n-k+1$ through $n-1$. That is, $B = A_{n-1}^c \cdots A_{n-k+1}^c$. Then, for $n \ge 2k$,

$$\begin{aligned} P(M_n) &= P(M_n B) + P(M_n B^c) \\ &= \omega + P(M_n B^c). \end{aligned}$$

Now, in order to have a renewal at time n and, in addition, to have the pattern appear in at least one of the preceding $k-1$ vectors, it is necessary that there be a renewal at one of the $k-1$ earlier times $n-k,\ldots,n-2k+2$, otherwise there would be a renewal at one of the $k-1$ times immediately preceding n, and thus no renewal at time n. Therefore, we see that

$$\begin{aligned} P(M_n) &\le \omega + P\left(\bigcup_{i=n-2k+2}^{n-k} M_i M_n\right) \\ &= \omega + \sum_{i=n-2k+2}^{n-k} P(M_i M_n) \\ &= \omega + \sum_{i=n-2k+2}^{n-k} P(M_i) P(M_n | M_i) \\ &\le \omega + \sum_{i=n-2k+2}^{n-k} \alpha P(M_i). \end{aligned}$$

Letting $n \to \infty$ shows that

$$P(M_\infty) \le \frac{\omega}{1-(k-1)\alpha}. \tag{10}$$

Hence, using Eqs. (6), (9), and (10), we see that

$$\frac{1-(k-1)\alpha}{\omega} \le E[N] \le \frac{1}{\omega}. \tag{11}$$

https://doi.org/10.1017/S0269964899131012 Published online by Cambridge University Press

5. DETERMINING ω

A simulation will normally be needed to approximate ω. Now,

$$\begin{aligned}\omega &= P(A_{2k}A_{2k-1}^c \cdots A_{k+1}^c)\\ &= P(A_k A_{k+1}^c \cdots A_{2k-1}^c)\\ &= P(A_k)P(A_{k+1}^c \cdots A_{2k-1}^c | A_k)\\ &= \alpha\beta,\end{aligned}$$

where the second equality follows from the fact that the X_i are exchangeable, and where

$$\beta \equiv P(A_{k+1}^c \cdots A_{2k-1}^c | A_k).$$

If α is computable and we can easily simulate $(X_1,\ldots,X_k)$ conditional on it being in $\mathcal{A}$, then we can directly obtain a simulation estimate of β.

Example: For the model of Section 3, A_k is the event that $X_1,\ldots,X_k$ are all distinct. It easily follows that conditional on A_k all $k!$ possible orderings are equally likely. Hence, to obtain a simulation estimator of β, first generate a random permutation $P_1,\ldots,P_k$ of $1,\ldots,k$, then generate random variables $X_{k+1},\ldots,X_{2k-1}$. The estimator of β from this run is 1 if, for each $j = 2,\ldots,k$, none of the vectors $P_j,\ldots,P_k,X_{k+1},\ldots,X_{k+j-1}$ have all distinct values, and the estimator is 0 otherwise. The average value of the estimators over many simulation runs is the final estimator of β.

If α is computable but we cannot easily simulate $(X_1,\ldots,X_k)$ conditional on it being in $\mathcal{A}$, then we can use the Gibbs sampler, a Markov chain Monte Carlo technique, to obtain an estimate of β. For instance, with $m = 2k-1$, consider the following procedure:

1. Let $J = 0$.
2. Choose a vector $(x_1,\ldots,x_k) \in \mathcal{A}$.
3. Generate $X_{k+1},\ldots,X_m$ according to the distribution F. Let their values be $x_{k+1},\ldots,x_m$.
4. Generate a random number U and set $I = \text{Int}(mU+1)$.
5. Generate X having distribution F.
6. If $I \le k$ and $(x_1,\ldots,x_{I-1},X,x_{I+1},\ldots,x_k) \in \mathcal{A}$, let $x_I = X$.
7. If $I > k$ let $x_I = X$.
8. If $(x_i,\ldots,x_{i+k-1}) \notin \mathcal{A}$ for all $i = 2,\ldots,k$, let $J = J+1$.
9. Return to 4.

Provided that the Markov chain of successive states $(x_1,\ldots,x_m)$ is ergodic, the final value of J divided by the number of iterations of the algorithm that are performed can then be used to estimate β.

https://doi.org/10.1017/S0269964899131012 Published online by Cambridge University Press

In performing the preceding Monte Carlo, note that we only change the value of one of the first k values if this change still leaves the vector in $\mathcal{A}$. When possible, it is more efficient to generate the new value conditional on the event that the first k values remain in $\mathcal{A}$. We illustrate by an example.

Example: Suppose that F is the exponential distribution with rate λ, and suppose that N is the first time that a data value exceeds the sum of the preceding $k-1$ values. That is,

$$N = \min\left\{ n \geq k : X_N > \sum_{i=n-k+1}^{n-1} X_i \right\}.$$

In this case

$$\alpha = P\left\{ X_k > \sum_{i=1}^{k-1} X_i \right\} = (1/2)^{k-1}.$$

The preceding Monte Carlo can be altered by generating exponentials conditional on the event that the sum of the first $k-1$ values is less than the kth value. For instance, if $I < k$, then we need to generate an exponential X conditioned on the event that

$$\sum_{i=1}^{k-1} x_i - x_I + X < x_k;$$

that is, an exponential X is generated conditional on the event that it is less than $c \equiv x_k - \sum_{i=1}^{k-1} x_i + x_I$, which can be accomplished by generating a random number U and setting

$$X = -\frac{1}{\lambda} \log(1 - U(1 - e^{-\lambda c})).$$

If $I = k$, then we need to generate X conditional on the event that it exceeds the sum of the first $k-1$ values. But this can be accomplished by generating a random number U and then setting

$$X = -\frac{1}{\lambda} \log(U) + \sum_{i=1}^{k-1} x_i.$$

In both cases, the present value of x_I is reset to X.

When the Markov chain of successive states $(x_1, \ldots, x_m)$ is not ergodic, a different Markov chain Monte Carlo variant is needed. For instance, suppose that the data are discrete and N is the first time that the sum of the most recent k values is equal to c. In this case, we could run a Markov chain Monte Carlo by initially choosing $x_1, \ldots, x_k$ whose sum is c, and then generating $X_{k+1}, \ldots, X_m$ according to F. We would then generate two random indices, I and J between 1 and m; when they are both less than or equal to k, we would then generate random variables X and Y distributed according to F but conditional on their sum being $x_I + x_J$. These generated values would then become the new values of x_I and x_J. If I is greater than k, then

https://doi.org/10.1017/S0269964899131012 Published online by Cambridge University Press

we generate a random variable according to F and let that be the new value of x_J, and similarly if $J > k$.

If α is not easily computed, then a simulation can still be used to estimate ω. However, as precisely estimating a small probability by simulation can be very time consuming, we would hope to be able to employ some variance reduction techniques. Also, for a particular pattern set $\mathcal{A}$ and distribution function F, a more efficient simulation procedure than Markov chain Monte Carlo may be possible.

6. FINAL REMARKS

When $\mathcal{A}$ is a one-point set, then $E[N]$ can be explicitly computed by a martingale argument (Li [2]) or by a renewal theory argument [4]. If $\mathcal{A}$ is a finite set of s elements then we can use either of these two methods to set up a set of s linear equations whose solution gives $E[N]$ [4, Sect. 7.9]. However, there are no general techniques available for explicitly solving for $E[N]$ when $\mathcal{A}$ is arbitrary. A natural approach is to regard $E[N]$ as a mean first passage time of the Markov chain that continually keeps track of the most recent $k-1$ values of the $X's$ until the pattern appears. However, as this results in a Markov chain with a $k-1$ dimensional state space, the analysis is only computationally feasible for very small values of k. The renewal theoretic approach used in Section 4 to obtain bounds on $E[N]$ was initiated in Hwang and Trietsch [1], where $1/\omega$ was suggested as an approximation for $E[N]$.

References

1. Hwang, F. & Trietsch, D. (1996). A simple relationship between the pattern probability and the rate of false signals in control charts. *Probability in the Engineering and Informational Sciences* 11(3): 315–324.
2. Li, S.Y.R. (1980). A Martingale approach to the study of the occurrence of patterns in repeated experiments. *Annals of Probability* 8(6): 1171–1175.
3. Marshall, A.W. & Olkin, I. (1979). *Inequalities: Theory of majorization and its applications*. San Diego: Academic Press.
4. Ross, S.M. (1997). *Introduction to probability models*, 6th ed. Chestnut Hill, MA: Academic Press.
5. Ross, S.M. (1997). *Simulation*, 2nd ed. San Diego: Academic Press.

https://doi.org/10.1017/S0269964899131012 Published online by Cambridge University Press