行政院國家科學委員會專題研究計畫成果報告

一般化留間隔机率問題計算的研究

Computations involve spacings with more general forms.

計畫類別：☑個別型計畫　□整合型計畫

計畫編號：NSC 88—2118—M032—002

執行期間：87年8月1日至88年7月31日

個別型計畫：計畫主持人：林千代
共同主持人：

整合型計畫：總計畫主持人：
子計畫主持人：

註：整合型計畫總報告與子計畫成果報告請分開編印各成一冊，彙整一起繳送國科會。

處理方式：□可立即對外提供參考
（請打ｖ）□一年後可對外提供參考
☑兩年後可對外提供參考
（必要時，本會得展延發表時限）

執行單位：淡江大學數學系

中華民國　88年10月25日

NSC Project Report: Computations Involve Spacings With More General Forms

Chien-Tai Lin
Department of Mathematics,Tamkang University
Number NSC 88-2118-M-032-002 Period 87/08/01 ~ 88/07/31

Abstract

We develop a new algorithm which can solve the problems involving linear combinations of spacings with the coefficients are rational.

Keywords: Order statistics; spacings.

1 Introduction

Let $X_1, X_2, \ldots, X_n$ be n points independently drawn from a uniform distribution on the interval [0, 1] and let $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ be the associated order statistics. The spacings $S_1, S_2, \ldots, S_{n+1}$ are defined to be the successive differences between the order statistics $S_i = X_{(i)} - X_{(i-1)}$, where we take $X_{(0)} = 0$ and $X_{(n+1)} = 1$. Let $\boldsymbol{S}$ denote the vector of spacings; $\boldsymbol{S} = (S_1, S_2, \ldots, S_{n+1})'$.

For a set $\Delta \subset \{1, 2, \ldots, n+1\}$ define

$$S(\Delta) = \sum_{i \in \Delta} S_i .$$

Many important situations involve the joint distribution of overlapping sums of spacings, so that we are often required to calculate probabilities of the form

$$P\left(\bigcap_{i=1}^{r}\{S(\Delta_i) > d\}\right) \text{ or } P\left(\bigcap_{i=1}^{r}\{S(\Delta_i) < d\}\right) \tag{1}$$

where the sets $\Delta_1, \Delta_2, \ldots, \Delta_r$ overlap. Probabilities of this type arise in many problems as described by Huffer and Lin (1997). Lin (1993) propose a general algorithm that attemps to calculate the probabilities involving spacings. Although this algorithm can deal with many cases of our interest, it will sometimes fail in others. Thus, a specialized algorithm is introduced. Huffer and Lin (1997) present a method for evaluating the distribution of the minimum or the maximum of linear combinations of spacings in (1). This algorithm is guaranteed to solve the problems much more quickly than is possible using the more general approach described in Lin (1993). However, there is still a need to develop a new methodology in order to handle more general configurations of the sets Δ_i, and can also solve some problems involving linear combinations of the spacings more complicated than simple sums like $S(\Delta)$.

In this report we present a method for computing probabilities which involve linear combinations of spacings. The approach we use depending on repeated use of the procedure of triangularization and the recursion given in equations (2) and (3) below. This recursion is used to re-express a probability like that in (2) by decomposing it into a sum of similar, but simpler components. The same recursion is then applied to each of these components and so on. The process is continued until we obtain components which are simple and easily expressed in closed form.

2 Basic Properties and Definitions

Let Γ be a $r \times (n+1)$ real matrix. Let $S = (S_1, S_2, \ldots, S_{n+1})'$ be the random vector of spacings. Let $\mathcal{P}(\Gamma)$ denote the probability measure of ΓS so that $P(\Gamma S \in B) = (\mathcal{P}(\Gamma))(B)$. For any $\xi \in \mathcal{R}^k$, define $\Gamma_{i,\xi}$ to be the $r \times (n+1)$ matrix obtained by replacing the i^{th} column of Γ by ξ. The basic recursion is the following.

Theorem 1 *Suppose $c = (c_1, c_2, \ldots, c_{n+1})'$ satisfies $\sum_{i=1}^{n+1} c_i = 1$. Let $\xi = \Gamma c$. Then*

$$\mathcal{P}(\Gamma) = \sum_{i=1}^{n+1} c_i \mathcal{P}(\Gamma_{i,\xi}). \tag{2}$$

Notation

For integer $n > 0$ and real values $\lambda \geq 0$ and $d \geq 0$, define

$$P(A, b, \lambda, n, d) = P(N_{(0,\lambda d)}^{(n)} = 0, AT^{(n)} > bd),$$

where A be any matrix having at most $n+1$ columns with rational entires, b be any $r \times 1$ vector with rational entires. Here n is the number of random points placed in $(0,1)$, $N_{(0,\lambda d)}^{(n)}$ is the number of points (from n random points) in the interval $(0, \lambda d)$, and $T^{(n)}$ is a ramdom vector of spacings (from n random points) in the interval $(\lambda d, 1)$. By rescaling rows of A by positive values, we may assume the entries in b are ± 1 or 0. Also, we define

$$Q(A, b, \lambda, i) = \frac{n!}{(n-i)!} d^i P(A, b, \lambda, n-i, d)$$

for integers $i \geq 0$.

It can be shown that if a matrix $A = (a_{ij})$ satisfies (for some $k \geq 1$) the following four conditions:

(C1) $a_{1j} = 0$ for $j > k$,

(C2) $a_{ij} = a_{i1} \equiv a_i$ for $j \leq k$, i.e., the first k columns of A are identical (note that (C2) is vacuous when $k = 1$),

(C3) $a_1 > 0$, and **(C4)** $b_1 > 0$,

then $Q(A, b, \lambda, p)$

$$= \sum_{\ell=0}^{k-1} \frac{1}{\ell!} (\frac{b_1}{a_1})^\ell Q(A^*{}_{(-\ell)}, b^* - \frac{b_1}{a_1} a^*, \lambda + \frac{b_1}{a_1}, p + \ell), \tag{3}$$

where A^* is a matrix obtained by deleting the first row of A, $A^*_{(-\ell)}$ is a matrix obtained by deleting the first ℓ columns of A^*, b^* is a vector obtained by deleting the first entry of b, and a^* is a vector obtained by taking the first column of A and deleting the first entry.

When specialized to the problem using the operator Q, the basic recursion (2) becomes the following: If $c = (c_1, c_2, \ldots, c_{n+1})'$ satisfies $\sum_{i=1}^{n+1} c_i = 1$ and $\xi = Ac$. Then

$$Q(A, b, \lambda, p) = \sum_{i=1}^{n+1} c_i Q(A_{i,\xi}, b, \lambda, p). \tag{4}$$

Useful Properties

For any permutation matrix $\boldsymbol{G}$,

(E1) $Q(\boldsymbol{A}, \boldsymbol{b}, \lambda, p) = Q(\boldsymbol{AG}, \boldsymbol{b}, \lambda, p)$,

(E2) $Q(\boldsymbol{A}, \boldsymbol{b}, \lambda, p) = Q(\boldsymbol{GA}, \boldsymbol{Gb}, \lambda, p)$.

For any diagonal matrix $\boldsymbol{D}$ with strictly positive entries on the diagonal

(E3) $Q(\boldsymbol{A}, \boldsymbol{b}, \lambda, p) = Q(\boldsymbol{DA}, \boldsymbol{Db}, \lambda, p)$.

These properties ensure us that we can always make the entries in $\boldsymbol{b}$ to be ± 1 or 0.

The value of Q remains the same when we delete redundant inequalities from $(\boldsymbol{A}, \boldsymbol{b})$. If any inequalities in $(\boldsymbol{A}, \boldsymbol{b})$ are contradictory, then $Q(\boldsymbol{A}, \boldsymbol{b}, \lambda, p) = 0$. In particular, the value of Q remains the same when

(S1) the i^{th} row of $\boldsymbol{A}$ dominates (is componentwise greater than or equal to) the j^{th} row, and the j^{th} entry of $\boldsymbol{b}$ is greater than or equal to the i^{th} entry;

(S2) the minimum entry of the i^{th} row of $\boldsymbol{A}$ is greater than or equal to 0 and the i^{th} entry of $\boldsymbol{b}$ is less than 0;

(S3) if the i^{th} entry of $\boldsymbol{b}$ is equal to 0 and the minimum entry of the i^{th} row of $\boldsymbol{A}$ is greater than 0;

(S4) if the i^{th} entry of $\boldsymbol{b}$ and the minimum entry of the i^{th} row of $\boldsymbol{A}$ are both equal to 0, but the maximum entry of the same row is not equal to 0;

and the value of Q is 0 when

(S5) if the i^{th} row of $\boldsymbol{A}$ is dominated by (is componentwise less than or equal to) the j^{th} row with all negative-signed entries, and the i^{th} entry of $\boldsymbol{b}$ is greater than or equal to the negative i^{th} entry;

(S6) if the maximum entry of the i^{th} row of $\boldsymbol{A}$ is less than or equal to 0 and the i^{th} entry of $\boldsymbol{b}$ is greater than or equal to 0;

(S7) if the i^{th} entry of $\boldsymbol{b}$ is greater than 0 and the maximum entry of the i^{th} row of $\boldsymbol{A}$ is less than 0;

(S8) if the i^{th} entry of $\boldsymbol{b}$ and the maximum entry of the i^{th} row of $\boldsymbol{A}$ are both equal to 0, but the minimum entry of the same row is not equal to 0.

For explicit calculations, we rely on the following formula. Let ϕ be an empty set. Then it can verify that for integers $j \geq 0$ and real values $\lambda \geq 0$,

$$Q(\phi, \phi, \lambda, p)$$
$$= (1 - \lambda d)_+^{n-p} d^p \frac{n!}{(n-p)!} I(n \geq p) = p! R(p, \lambda), \quad (5)$$

where

$$R(j, \lambda) = \binom{n}{j} d^j (1 - \lambda d)^{n-j} \text{for} \lambda d < 1$$

To evaluate Q, we continue decomposing matrices using the procedure of triangularization (presented in Section 3) and the recursion (3) until we reach "simple" terms which can be evaluated using (5). Writing our answers in terms of R allows us to obtain very compact expressions for Q by suppressing the dependence on n and d.

3 Algorithm for Triangularization

Let $\boldsymbol{A}$ be an arbitrary matrix with rational entries. We shall now describe a general approach for reducing $\boldsymbol{A}$ to a triangular form with zeros entries in the right upper region of the matrix. Our approach makes us always to increase the region of zeros in $\boldsymbol{A}$. We start with an matrix in the following form:

$$\begin{pmatrix} 6 & 6 & 0 & 0 & 0 \\ 0 & 0 & 6 & 6 & 0 \\ 12/5 & 12/5 & 0 & 0 & 12/5 \\ 0 & 0 & 12/5 & 12/5 & 12/5 \\ 3/2 & 3/2 & 0 & 3/2 & 3/2 \\ 0 & 3/2 & 3/2 & 3/2 & 3/2 \end{pmatrix} \qquad (6)$$

This example illustrates an matrix with three zones, including three frontier blocks and three interior blocks. The organization is the followings. (i) Boundary marks last nonzero entry in each row. (ii) The rows are grouped into zones, i.e., rows with the identical positions of the last nonzero entry form a zone. (iii) Matrix is broken into blocks. Each zone has two kinds of blocks, interior blocks and frontier block (lie along the boundary). (iv) Blocks are broken into chunks, with each chunk lying entirely within one block.

We shall now describe the procedure of triangularization. We search the blocks in order. The ordering follows the picture below.

1			
2	5		
3	6	8	
4	7	9	10

In such a way that a block is never searched until all the blocks lying directly above it have already been searched. For each block we find the first chunk which contains two distinct values. Move the row containing this chunk to the top of its zone. If block containing chunk is a frontier block, use $\mathbf{c}$ with 2 nonzero coefficients at the positions of the two distinct values in chunk. If block containing chunk is an interior block, use $\mathbf{c}$ with 3 nonzero coefficients at the position of the two distinct values and the last nonzero entry in that row.

References

Huffer, F., Divided differences and the joint distribution of linear combinations of spacings, *J. Appl. Prob.* 25 (1988) 346–354.

Huffer, F., and C.T. Lin, Approximating the distribution of the scan statistic using moments of the number of clumps, *J. Amer. Stat. Assoc.* 92 (1997) 1466–1475.

Huffer, F., and C.T. Lin, Computing the exact distribution of the extremes of sums of consecutive spacings. *Comput. Stat. Data Analysis* 26 (1997) 117–132.

Lin, C.T., The computation of probabilities which involve spacings, with applications to the scan statistic, Ph.D. Dissertation (Dept. of Statistics, Florida State University, Tallahassee, FL, 1993).