1095

行政院國家科學委員會專題研究計畫成果報告

圓周上隨間隔的機率問題之計算方法

A Computational Approach To Probabilities Which Involve Spacings On A Circle

計畫類別：☑個別型計畫　□整合型計畫

計畫編號：NSC 87—2118—M032—003

執行期間：86 年 8 月 1 日至 87 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∨）□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：淡江數學系

中華民國 87 年 10 月 1 日

A Computational Approach To Probabilities Which Invovle Spacings On A Circle

Chien-Tai Lin, Tamkang University, Department of Mathematics, Tamsui, Taiwan 251

Keywords: Circular Scan Statistic; Circular Spacings; Symbolic Computation.

Abstract

A specialized algorithm is developed, based on the recursions given by Huffer (1988) and the algorithm proposed by Lin (1993) and Huffer and Lin (1996). This approach can handle a variety of different multiple coverage problems and also give accurate answers. Application to the distribution of circular scan statistic is also discussed.

1 Introduction

Let $X_1, X_2, \ldots, X_{n+1}$ be $n+1$ points independently and uniformly drawn from a circle of unit circumference and let $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n+1)}$ denote these same points ordered in a clockwise fashion starting with $X_1 = X_{(1)}$. The circular spacings $S_1, S_2, \ldots, S_{n+1}$ are defined to be the successive arc-lengths between these points, that is, $S_i = X_{(i)} - X_{(i-1)}$ for $1 \leq i \leq n+1$ where we take $X_{(0)} = X_{(n+1)}$. Let S denote the vector of circular spacings; $S = (S_1, S_2, \ldots, S_{n+1})'$.

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

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

Probabilities of multiple coverage problems on a circle involve the joint distribution of overlapping sums of consecutive circular spacings, so that we are often required to calculate probabilities of the form

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

where the sets $\Delta_1, \Delta_2, \ldots, \Delta_r$ overlap and each of the sets Δ_i consists of a block of consecutive integers and is allowed to "wrap around" if necessary

In this report we present a specialized method for evaluating special probabilities of the form (1). Utilizing this specialized algorithm we can handle a variety of different multiple coverage problems much more quickly than the generalized program developed by Lin (1993).

The approach we use was sketched in Huffer (1988) and more fully developed in Lin (1993). It depends on repeated use of the recursion given in equations (2) below. This recursion is used to re-express a probability like that in (1) 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 Recursion and Definitions

Our approach is based on the following recursion.

Let Γ be an $r \times (n+1)$ real matrix. Let $S = (S_1, S_2, \ldots, S_{n+1})'$ be the vector of spacings between uniform random variables as defined in Section 1. For any $\xi \in \mathcal{R}^r$, 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*

$$P(\Gamma S \in B) = \sum_{i=1}^{n+1} c_i P(\Gamma_{i,\xi} S \in B). \quad (2)$$

for any measurable set $B \subset \mathcal{R}^r$.

This recursion was first obtained by Micchelli (1980) as a result about multivariate B-splines. It was rediscovered in the context of spacings by Huffer (1988). Examples showing how the basic recursion is applied in computations can be found in Huffer (1988) and Lin (1993).

Notation

We now introduce the notation we shall use for the problems.

Let A be any matrix having at most $n+1$ columns. Take Γ to be the matrix with $n+1$ columns obtained by padding A with columns of zeros; $\Gamma = (A|\ 0)$. Let r be the number of rows in A and define $Y = (Y_1, Y_2, \ldots, Y_r)'$ by $Y = \Gamma S$. For any value of d, we define

$$\{A\}_d^1 = P\{Y_i > d \text{ for all i}\} = P\{\min Y_i > d\}, \quad (3)$$

and

$$\{A\}_d^2 = P\{Y_i < d \text{ for all i}\} = P\{\max Y_i < d\}. \quad (4)$$

When the value of d is held fixed in an argument, we delete the subscript and just write $\{A\}^1$ or $\{A\}^2$. We also omit the superscript when convenient. The quantity $\{A\}$ is well defined so long as the number of spacings $n+1$ is greater than or equal to the number of columns in A.

Explicit Formulas

For explicit calculations, we rely on the following formula. Let $\Delta_1, \Delta_2, \ldots, \Delta_r$ be nonempty subsets of $\{1, 2, \ldots, n+1\}$ with cardinalities $|\Delta_i| = 1 + \ell_i$. Define $\boldsymbol{\ell} = (\ell_1, \ldots, \ell_r)$. Let $0 < d < 1$. If the sets Δ_i, $i = 1, \ldots, r$, are disjoint, then

$$P\left(\bigcap_{i=1}^{r}\{S(\Delta_i) > d\}\right) \quad (5)$$
$$= \sum_{0 \le \boldsymbol{k} \le \boldsymbol{\ell}} \binom{n}{k_1, \ldots, k_r} d^{\Sigma k_i} (1 - rd)_+^{n - \Sigma k_i}.$$

where $\boldsymbol{k} = (k_1, \ldots, k_r)$ is an r-tuple of integers and $0 \le \boldsymbol{k} \le \boldsymbol{\ell}$ means that $0 \le k_i \le \ell_i$ for all i. Here we use $(x)_+$ to denote the positive part, that is, $(x)_+ = \max(x, 0)$. For our purposes, a very convenient way to rewrite this formula is as follows. For integers $j \ge 0$ and real values $\lambda \ge 0$, define

$$R(j, \lambda) = \begin{cases} \binom{n}{j} d^j (1 - \lambda d)^{n-j} & \text{for } \lambda d < 1, \\ 0 & \text{for } \lambda d \ge 1. \end{cases}$$

The dependence of R on n and d can be left implicit because these values are fixed in any given application of our methods. In terms of R, formula (5) becomes: If the sets Δ_i, $i = 1, \ldots, r$ are disjoint, then

$$P\left(\bigcap_{i=1}^{r}\{S(\Delta_i) > d\}\right)$$
$$= \sum_{0 \le \boldsymbol{k} \le \boldsymbol{\ell}} \binom{\sum_i k_i}{k_1, \ldots, k_r} R(\textstyle\sum_i k_i, r). \quad (6)$$

To evaluate $\{A\}$, we continue decomposing matrices using (2) until we reach "simple" terms which can be evaluated using (6). Writing our answers in terms of R allows us to obtain very compact expressions for $\{A\}$ by suppressing the dependence on n and d. (See the examples in Section 4.)

3 Proposed Algorithm

In this section we describe a specialized algorithm for computing $\{A\}$ when A is a binary matrix where the 1's in each row of the top part form a contiguous block, and the 0's in each row of the bottom part form a contiguous block. Our description will be somewhat sketchy for the algorithm remains under development. The current version of the algorithm can evaluate $\{A\}$ for many cases of interest.

Our approach for evaluating $\{A\}$ is roughly as follows. The main idea behind this algorithm is to delete the bottom part of the matrix and form a matrix with consecutive nonzero integers in each row.

Before describing our algorithm we must introduce some notation. Let $\mathcal{B}$ be the class of matrices in which the 1's in each row form a contiguous block, $\mathcal{C}$ be the class of matrices in which nonzero integers in each row form a contiguous block, and $\mathcal{D}$ be the class of matrices in which nonzero integers in each row form a contiguous block and is allowed to "wrap around". Denote $A \in \mathcal{D}$ be an $(r+s) \times p$ matrix with $r \ge 2$ and $s \ge 1$, where r and s represent the number of rows in the top and bottom part of the matrix respectively.

In order to maintain a canonical "descending" form of matrix, we have the following rules:

(R1) We always move the replaced vector $(0, \boldsymbol{\xi})'$ to the last column or the replaced vector $(e_r, \boldsymbol{\xi})'$ to the same group of the columns in the matrix.

(R2) If there occurs the first row which is not overlapping to the second row in the top part of matrix, we will push the second row to the last row of matrix to be the new first $(r+s)-1$ rows of the matrix and pack the original first row to be the new last row of matrix. We then count the number of 0's before the starting 1 in the first row, say k, push the last $p-k$ columns of the matrix to be the new first $p-k$ columns of the matrix, and pack the original k columns to be the new last k columns of the matrix.

We can now describe our algorithm. We compute $\{A\}$ as follows:

1. If $A \in \mathcal{B}$, we evaluate $\{A\}$ using the program of Huffer and Lin (1996).

2. If $A \in \mathcal{C}$, we evaluate $\{A\}$ using the program of Lin (1993).

3. Otherwise, locate the first block in the top part of $\boldsymbol{A}_i$ having at least two rows and decompose $\boldsymbol{A}_i$ using the procedure in Huffer and Lin (1996) for matrices in $\mathcal{D}$.

4. Now apply this same process to all the daughter matrices obtained in the previous step.

For any matrix $\boldsymbol{A} \in \mathcal{D}$, this algorithm in general leads to expressions of the form

$$\{\boldsymbol{A}\} = \sum_i w_i \{\boldsymbol{B}_i\}\,.$$

where the values w_i are integers and the matrices $\boldsymbol{B}_i$ belong to $\mathcal{E}$ (If $\boldsymbol{B} \in \mathcal{E}$, then $\{\boldsymbol{B}\}$ is given in a simple closed form by equation (6)). Evaluating each of the terms $\{\boldsymbol{B}_i\}$, we obtain our final answer in the form

$$\{\boldsymbol{A}\} = \sum_i c_i R(j_i, \lambda_i)\,.$$

where the values c_i and λ_i are all integers.

Special Case: Binary Matrices with One Zero in Each Row

Let $\boldsymbol{A}$ be a $(m+1) \times (m+1)$ matrix as

$$\begin{pmatrix} \multicolumn{9}{c}{\overbrace{\hphantom{1\ 1\ 1\ 1\ \dots\ 1\ 1\ 1}}^{m}} \\ 1 & 1 & 1 & 1 & \dots & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & \dots & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & \dots & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & \dots & 1 & 1 & 1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 1 & 1 & \dots & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & \dots & 1 & 1 & 0 & 1 \end{pmatrix}$$

and $n \geq m$.

The probabilities of

$$\{\boldsymbol{A}\}_d^1 = P(X_{(i+m)} - X_{(i)} > d \text{ for } i = 1, \dots, m+1)$$

and

$$\{\boldsymbol{A}\}_d^2 = P(X_{(i+m)} - X_{(i)} < d \text{ for } i = 1, \dots, m+1)$$

are discussed respectively in the followings.

Evaluating $\{\boldsymbol{A}\}^1$

Upon repeated applications of the recursion we have

$$\{\boldsymbol{A}\}^1 = \left\{ \begin{array}{cccccc} \frac{m}{m+1} & \frac{m-1}{m} & \dots & \frac{3}{4} & \frac{2}{3} & 1 \quad 0 \\ \frac{m}{m+1} & \frac{m-1}{m} & \dots & \frac{3}{4} & \frac{2}{3} & 0 \quad 1 \end{array} \right\}.$$

Carrying on the application of the recursion with $\boldsymbol{\xi} = (0,0)'$ gives

$$\{\boldsymbol{A}\}_d^1 = (-1)^{m-1}\binom{m+1}{2} P\{S_1 > d, S_2 > d\}$$
$$+ \sum_{j=0}^{m-2} (-1)^j (j+2)^2 P\{\frac{m}{m+1} S_1 + \frac{m-1}{m} S_2$$
$$+ \cdots + \frac{j+3}{j+4} S_{m-j-2} + \frac{j+2}{j+3} S_{m-j-1} > d\}\,. \quad (7)$$

We can now evaluate (7) using the program of Lin (1993).

Evaluating $\{\boldsymbol{A}\}^2$

It is clear that

$$\{\boldsymbol{A}\}_d^2 = 1 - P\{\frac{m}{m+1} S_1 + S_2 + \cdots + S_{m+1} \geq d\}\,.$$

Continually applying the recursion with $\xi = 0$ we obtain

$$P\{\frac{m}{m+1} S_1 + S_2 + \cdots + S_{m+1} \geq d\}$$
$$= \sum_{j=0}^{m-1} (-1)^j m^j (m+1) P\{S_1 + \ldots + S_{m-j} \geq d\}$$
$$+ (-1)^m m^m P\{\frac{m}{m+1} S_1 \geq d\}\,.$$

Writing out the expression in terms of R in (6) leads to

$$\{\boldsymbol{A}\}^2 = 1 - \sum_{j=0}^{m-1} (-1)^j m^j (m+1) \sum_{k=0}^{m-j-1} R(k, 1)$$
$$-(-1)^m m^m R(0, \frac{m+1}{m})\,.$$

4 Examples

This section contains two examples of computations carried out using the algorithm in Section 3. The problems in the examples have been chosen to be small enough so that the answers fit conveniently on the page; our programs can handle larger problems.

Example 1

For our first example we evaluate $\{\boldsymbol{A}\}_d^2$ for the 7×7 matrix

$$\boldsymbol{A} = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

This matrix has a pattern that the entries in the last few rows are "wrap around". Thus, the matrix does belong to $\mathcal{D}$ so that an expression for $\{A\}_d^2$ can be rapidly obtained by the algorithm in Section 3. Our final answer is

$$\begin{aligned}\{A\}_d^2 &= 1 + 728R(0,2) - 729R(0,7/3) + 7R(1,1) \\ &\quad - 252R(1,2) - 7R(1,2) + 70R(2,2) \\ &\quad - 35R(3,2) + 7R(4,2).\end{aligned}$$

We may now use this expression to compute $\{A\}_d^2$ for arbitrary n and d. When $n = 6$, the value of $\{A\}_d^2$ can be interpreted as a "multiple coverage" probability for random points on a circle, and $S_1, S_2, \ldots, S_7$ can be viewed as the circular spacings between 7 random points on a circle with circumference equal to 1. For 7 random points on a circle, $\{A\}_d^2$ is the probability that every arc of length d contains at least 3 of these points. An equivalent interpretation is the following. Suppose 7 arcs of length d are placed at random on a circle. Then $\{A\}_d^2$ is the probability that every point on the circle is covered by at least 3 of these arcs. These multiple coverage probabilities have been studied by Holst (1980) and Deken (1981).

Example 2

The other type of multiple coverage problems which have also received some attention in the literature is the problems of covering the unit circle by random arcs of fix length. The circular coverage probabilities are closely related to the clustering probabilities (See Glaz and Naus (1979) eq. 1). The clustering probabilities have been extensively studied particularly in connection with the distribution of the circular scan statistic, a test for the presence of nonrandom clustering in a circle of unit circumference. Much work has been done on the exact calculation of these clustering probabilities; see Ajne (1968), Rothman (1969), and Wallenstein (1971). Other work relates to the asymptotic results; see Ajne (1968) or Flatto (1973). Some of the clustering probabilities can be evaluated by the current version of the algorithm. We conclude by giving an example of one such probability.

In this example we suppose that there are 10 random points on a circle of unit circumference. We want to compute

$$\{A\}_d^1 = P(X_{(i+6)} - X_{(i)} > d, \text{ for } i = 1, \ldots, 10).$$

This is simply the probability that *no* arc of length d contains more than 6 of the 10 random points $X_1, X_2, \ldots, X_{10}$. That is, $1 - \{A\}_d^1$ is the probability that there exists a cluster of 7 or more points in an arc of length d. The final answer is

$$\begin{aligned}&-4290894R(0,5/3) + 4587520R(0,7/4) \\ &-296625R(0,2) + 153090R(1,5/3) + 163840R(1,7/4) \\ &-33510R(1,2) + 6070R(2,2) - 5500R(3,2) \\ &+2150R(4,2) + 560R(5,2) + 80R(6,2),\end{aligned}$$

which is valid for all d and all $n \geq 6$.

References

Ajne, B. (1968). A simple test for uniformity of a circular distribution. *Biometrika* **55**, 343–354.

Deken, J. G. (1981). Exact distributions for gaps and stretches. Technical Report **295**, Dept. of Statistics, Stanford University.

Flatto, L. 91973). A limit theorem for random coverings of a circle. *Israel J. Math.* **15**, 167–184.

Glaz, J., and Naus, J. (1979). Multiple coverage of the line. *Ann. Prob.* **7**, 900–906.

Holst, L. (1980). On multiple covering of a circle with random arcs. *J. Appl. Prob.* **17**, 284–290.

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

Huffer, F., and Lin, C. T. (1996). Computing the exact distribution of the extremes of linear combinations of spacings. Technical Report, Dept. of statistics, Florida State University, Tallahassee, FL.

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

Micchelli, C. A. (1980). A constructive approach to Kergin Interpolation in $\mathcal{R}^k$: Multivariate B-splines and Lagrange interpolation. *Rocky Mountain J. Math.* **10**, 485–497.

Roos, M. (1993). Compound Poisson approximations for the number of extreme spacings. *Adv. Appl. Prob.* **25**, 847–874.

Rothman, E. D. (1972). Tests for uniformity of a circular distribution. *Sankhyā A* **34**, 23–32.

Wallenstein, S. R. (1971). Coincidence probabilities used in nearest neighbor problems on the line and circle. Ph. D. Thesis, Rutgers University.