CLUSTER PROBLEMS IN ONE DIMENSION

by

D. G. Hoel, Nat. Inst. Environ. Health Sciences

V. R. R. Uppuluri, Oak Ridge Nat. Laboratory*

M. Sobel**, University of Minnesota

Technical Report No. 216

August, 1973

*Operated by Union Carbide Corporation under contract with the U.S. Atomic Energy Commission.

**Supported by NSF Contract No. GP-28922X at the University of Minnesota, Minneapolis, Minnesota.

1. Introduction

Let us consider a one-dimensional lattice with n sites, which are indexed by $1, 2, \ldots, n$. The contiguous set $[i, i + 1, \ldots, i + s - 1]$, is called a cluster of size s. Let $B_i = B_i(s)$ denote the cluster $[i, i + 1, \ldots, i + s - 1]$, of size s. For a given one-dimensional lattice with n sites, we have a total of $M = n - s + 1$ clusters.

We keep n and s fixed in the following discussion. We may select a collection $(c > 0)$ of clusters, of common size, either with or without replacement. Let $X_c(Y_c)$ denote the number of points of the lattice which belong to the set-theoretic union (intersection) of c clusters. In each of these cases, we obtain expressions for the expected coverages $E(X_c)$ $(E(Y_c))$ using a method given in Robbins [2], and we also obtain expressions for the probability of a complete coverage with a collection of c clusters.

These problems were motivated in understanding the phenomenon of degradation of the DNA strand which is exposed to radiation. It is known that the DNA strand consists of nucleotide bases which are assumed by some to be arranged in a linear strand and by others to be in a circular strand. After irradiation a phenomenon described as degradation of the strand takes place and it is of interest to be able to predict the expected amount of degradation and the probability of total degradation. We can study these problems in the context of a linear model if the lattice points are arranged on a line or in the case of a circular model if the points are arranged on a circle. For further results on the distribution of $X_c(Y_c)$, the reader may refer to the report by Roth and Sobel [3].

2. Expected Coverage

In this section, in order to apply the known results from the theory of continuous variables, we will redefine the cluster as a half-open interval. We note that in doing this, we are still considering the n lattice sites, and the size of the cluster is the same as was defined before. Specifically, we consider the interval $[1, n + 1]$ which contains the n sites $i = 1,2,\ldots,n$. A cluster $B_i = B_i(s)$ of size s will be represented by an interval $[i, i + s]$ of length s. We will randomly choose c sites (cluster left endpoints) out of $n - s + 1$ (with or without replacement) and obtain their corresponding cluster intervals. Let $p(x)$, $x \in [1, n + 1)$, be the probability that the point x belongs to at least one of the c cluster intervals. By Robbins [2] we then have the expected number of sites in the set theoretic union X_c of c clusters given by

$$(2.1) \qquad E(X_c) = \int_1^{n+1} p(x)dx = \sum_{i=1}^{n} p_i$$

where p_i is the probability that site i belongs to at least one of the c cluster intervals. For the intersection coverage Y_c, we have as a corollary that

$$(2.2) \qquad E(Y_C) = \sum_{i=1}^{n} q_i$$

where q_i is the probability that the site i belongs to every one of the c cluster intervals.

2.1. Sampling with Replacement

We shall consider two geometric configurations of the n sites $i = 1,2,\ldots n$. The first configuration is the linear model which has been previously described and the second is the circular model. In this latter situation the sites lie on the circumference of a circle and a cluster of length s is again a set of s contiguous sites along the circle. We begin by giving the values of p_i with $i \le (n+1)/2$ for the linear model and when $i \ge (n+1)/2$ we use the relation $p_i = p_{n-i+1}$.

(a) Linear Model:

The values of the p_i's are easily seen to be as follows:

Case 1: $n \ge 2s-1$

$$p_i = \begin{cases} 1 - \left(\frac{n-s+1-i}{n-s+1}\right)^c & \text{for } 1 \le i \le s \\ 1 - \left(\frac{n-2s+1}{n-s+1}\right)^c & \text{for } s \le i \le \frac{n+1}{2} \end{cases}$$

Case 2: $n \le 2s-1$

$$p_i = \begin{cases} 1 - \frac{n-s+1-i}{n-s+1} & \text{for } 1 \le i \le n-s+1 \\ 1 & \text{for } n-s+1 \le i \le \frac{n+1}{2} \end{cases} .$$

It then follows from (2.1) that

Case 1: $n \ge 2s-1$

$$E(X_c) = n - (n-s+1)^{-c} \left\{ 2\sum_{i=1}^{s-1} (n-s-i+1)^c + (n-2s+2)(n-2s+1)^c \right\} . \tag{2.3}$$

Case 2: $n \le 2s-1$

$$E(X_c) = n - 2(n-s+1)^{-c} \sum_{i=1}^{n-s} (n-s-i+1)^c . \tag{2.4}$$

To obtain $E(Y_c)$ it is easy to see that for $i \le (n+1)/2$

Case 1: $n \ge 2s-1$

$$q_i = \begin{cases} \{i/(n-s+1)\}^c & \text{for } i \le s \\ \{s/(n-s+1)\}^c & \text{for } i \ge s \end{cases}$$

then by (2.2) and the relation $q_i = q_{n-i+1} (i \ge (n+1)/2)$

(2.5) $$E(Y_c) = (n-s+1)^{-c} \left\{ 2\sum_{i=1}^{s-1} i^c + (n - 2s + 2)s^c \right\} .$$

Similarly, for

Case 2: $n \leq 2s - 1$ we have

$$q_i = \begin{cases} [i/(n-s+1)]^c & i \leq n-s+1 \\ 1 & i \geq n-s+1 \end{cases}$$

and thus

(2.6) $$E(Y_c) = 2(n-s+1)^{-c} \sum_{i=1}^{n-s+1} i^c + 2(s-1)-n.$$

(b) Circular Model

The circular model is simpler since the adjustments for the internal endpoints are not needed. For

Case 1: $n \geq s$, we have $p_i = 1 - (\frac{n-s}{n})^c$ and thus

(2.7) $$E(X_c) = n \left\{ 1-(1- \frac{s}{n})^c \right\} .$$

For $E(Y_c)$ we use $q_i = (s/n)^c$ so that

(2.8) $$E(Y_c) = s(\frac{s}{n})^{c-1} .$$

For

Case 2: $n \leq s$, we have $p_i = 1$ and $q_i = 0$ so that

(2.9) $$E(X_c) = n \text{ and } E(Y_c) = 0.$$

2.2. Sampling without Replacement

The solution to this problem is analogous to sampling with replacement except that the probabilities p_i and q_i change. Here $c \leq n-s+1$ and we define $\binom{j}{c}$ to be zero if $0 \leq j < c$.

(a) Linear Model:

We again give the p_i, q_i values for $i \leq (n+1)/2$ and use the relations

$p_i = p_{n-i+1}$ and $q_i = q_{n-i+1}$.

Case 1: $n \geq 2s-1$

$$p_i = \begin{cases} 1 - \binom{n-s-i+1}{c} / \binom{n-s+1}{c} & 1 \leq i \leq s \\ 1 - \binom{n-2s+1}{c} / \binom{n-s+1}{c} & s \leq i \leq \frac{n+1}{2} \end{cases}$$

so that

$$E(X_c) = n - \binom{n-s+1}{c}^{-1} \left\{ 2 \sum_{i=1}^{s-1} \binom{n-s-i+1}{c} + (n-2s+2)\binom{n-2s+1}{c} \right\} \quad (2.10)$$

Case 2: $n \leq 2s-1$

$$p_i = \begin{cases} 1 - \binom{n-s-i+1}{c} / \binom{n-s+1}{c} & 1 \leq i \leq n-s-c+1 \\ 1 & n-s-c+1 \leq i \leq \frac{n+1}{2} \end{cases}$$

and thus

$$E(X_c) = n - 2 \sum_{j=c}^{n-s} \binom{j}{c} / \binom{n-s+1}{c} . \quad (2.11)$$

To obtain $E(Y_c)$ we again consider the two cases.

Case 1: $n \geq 2s-1$

$$q_i = \begin{cases} \binom{i}{c} / \binom{n-s+1}{c} & 1 \leq i \leq s \\ \binom{s}{c} / \binom{n-s+1}{c} & s \leq i \leq (n+1)/2 \end{cases}$$

and thus

$$E(Y_c) = \binom{n-s+1}{c}^{-1} \left\{ 2 \sum_{i=c}^{s-1} \binom{i}{c} + (n-2s+2)\binom{s}{c} \right\} . \quad (2.12)$$

Case 2: $n \le 2s - 1$

$$q_i = \begin{cases} \binom{i}{c}/\binom{n-s+1}{c} & 1 \le i \le n-s+1 \\ 1 & n-s+1 \le i \le (n+1)/2 \end{cases}$$

and hence

$$E(Y_c) = 2\binom{n-s+1}{c}^{-1} \sum_{i=c}^{n-s} \binom{i}{c} + 2s-n. \tag{2.13}$$

(b) Circular Model:

For this case we have by similar reasoning that for

Case 1: $n \ge s$, it follows that $p_i = 1 - \binom{n-s}{c}/\binom{n}{c}$ and hence

$$E(X_c) = n \left\{1 - \binom{n-s}{c}/\binom{n}{c}\right\}. \tag{2.14}$$

For $E(Y_c)$ we use $q_i = \binom{s}{c}/\binom{n}{c}$ and thus

$$E(Y_c) = c\binom{s}{c}/\binom{n-1}{c-1} \quad . \tag{2.15}$$

Case 2: $n > s$

We have that $p_i = 1$ and $q_i = 0$ which yields

$$E(X_c) = n \text{ and } E(Y_c) = 0. \tag{2.16}$$

3. Probability of Complete Coverage

To develop a formula for the probability of a complete coverage in the linear case with replacement, we start with the case of odd s and point out later that the same result holds for even s. For odd s, we can identify each cluster with its center point. For $n=s$, we always get complete coverage. To obtain complete coverage for $n > s$, we need first of all to to select the two end clusters, whose center points are distinct; we denote them by C_1 and C_2. Our analysis focuses attention on the n-s spaces between C_1 and C_2. We obtain complete coverage if and only if the remaining c-2 center points to be selected partition these n-s spaces into c-1 parts so that each part is of size at most s. We separate these partitions first according to the number of zeros in the partition and then according to the sizes of the subsets of contiguous zeros in the partition. Thus $A_{3,2,1}(n-s,c-1,s)$ denotes the number of ordered partitions with a total of 6 zeros of which 3 are contiguous in one set, 2 are contiguous in another and one zero is separated from both sets. This term has to be multiplied by $c!/(4!3!2!)$ since we are selecting c clusters in all and each set of j contiguous zeros corresponds to j selections of the same cluster. Thus the probability $P\{CC\}$ of complete coverage is given by

$$P\{CC\} = \frac{c!}{M^c}\left(A_o(n-s,c-1,s) + \frac{A_1(n-s,c-1,s)}{2!} + \left\{\frac{A_2(n-s,c-1,s)}{3!} + \frac{A_{1,1}(n-s,c-1,s)}{(2!)^2}\right\} \left\{ + \frac{A_3(n-s,c-1,s)}{4!} + \frac{A_{2,1}(n-s,c-1,s)}{3!2!} + \frac{A_{1,1,1}(n-s,c-1,s)}{(2!)^3}\right\} + \cdots \right) \tag{3.1}$$

where $M = (n-s+1)^c$ is the total possible number of ordered partitions. The terms indicated by braces in (3.1) terminate as soon as the total number of zeros J is such that $(c-1-J)s < n-s$ or $J > c - \frac{n}{s}$, i.e., we sum on J up to $[c - \frac{n}{s}]$, the integer part of $c - \frac{n}{s}$. Let $\underset{\sim}{j} = (j_1, j_2, \ldots, j_r)$ denote r subsets of zeros, the <u>ith</u> one containing j_i contiguous zeros and let $J = \sum_{i=1}^{r} j_i$. The problem of finding ordered partitions $A_{\underset{\sim}{j}}(t,p,m)$ with J zeros is equivalent to finding partitions of t into p-J parts, with each part at most m and no zeros, except for the positions of the sets of zeros. Hence we obtain for $p \geq J$

$$(3.2) \qquad A_{\underset{\sim}{j}}(t,p,m) = \binom{p-J+1}{r}\left(\!\!\binom{r}{\alpha_1,\alpha_2,\ldots,\alpha_a}\!\!\right)A_o(t,p-J,m)$$

$$=\left(\!\!\binom{p-J+1}{\alpha_1,\alpha_2,\ldots,\alpha_a,p-J+1-r}\!\!\right)A_o(t,p-J,m),$$

where α_1 is the frequency of the number 1 in the vector $\underset{\sim}{j},\ldots,\alpha_a$ is the frequency of the number a, in $\underset{\sim}{j}$, so that $\alpha_1+\alpha_2+\ldots+\alpha_a = r$, and double parentheses indicate the usual multinomial coefficient. Thus for $J = 3$ we have

$$(3.3) \qquad \begin{aligned} A_3(t,p,m) &= \binom{p-2}{1}A_o(t,p-3,m) \\ A_{2,1}(t,p,m) &= 2\binom{p-2}{2}A_o(t,p-3,m) \\ A_{1,1,1}(t,p,m) &= \binom{p-2}{3}A_o(t,p-3,m) \end{aligned}$$

Substituting these in (3.1) gives the result

$$(3.4) \qquad P\{CC\} = \frac{c!}{M^c}\sum_{J=o}^{[c-\frac{n}{s}]}\Sigma\left(\!\!\binom{c-J}{\alpha_1,\alpha_2,\ldots,\alpha_a,c-J-r}\!\!\right)\frac{A_o(n-s,c-1-J,s)}{\prod_{i=1}^{r}(j_i+1)!}$$

where the inside sum in (3.4) is over all the unordered partitions of J into positive parts. In more detail this can be written as

$$(3.5)\quad P\{CC\} = \frac{c!}{M^c}\Bigg(A_o(n-s,c-1,s) + A_o(n-s,c-2,s)\left\{\frac{\binom{c-1}{1}}{2!}\right\}$$

$$+ A_o(n-s,c-3,s)\left\{\frac{\binom{c-2}{1}}{3!} + \frac{\binom{c-2}{2}}{(2!)^2}\right\} + A_o(n-s,c-4,s)\left\{\frac{\binom{c-3}{1}}{4!} + \frac{2\binom{c-3}{2}}{2!3!} + \frac{\binom{c-3}{3}}{(2!)^3}\right\}$$

$$+ A_o(n-s,c-5,s)\left\{\frac{\binom{c-4}{1}}{5!} + \frac{2\binom{c-4}{2}}{2!4!} + \frac{\binom{c-4}{2}}{(3!)^2} + \frac{3\binom{c-4}{3}}{3!(2!)^2} + \frac{\binom{c-4}{4}}{(2!)^4}\right\} + \cdots\Bigg);$$

this shows the terms for $J \le 4$ and J continues up to $[c - \frac{n}{s}]$.

The quantity $A_o(t,p,m)$ used throughout (3.5) is the number of ordered partitions of t into p positive parts each at most m; it is well known (see for example [1], pp. 227-228) that for $p \ge 1$

$$(3.6)\qquad A_o(t,p,m) = \sum_{i=0}^{\min([\frac{t-p}{m}],p)} (-1)^i \binom{p}{i}\binom{t-1-mi}{p-1}.$$

For $t \ge p$ and $m \ge 1$ the formula (3.6) also gives the number of ordered partitions of t-p into p nonnegative parts, each part being of size at most m-1.

For even values of the integer s we select the spaces (associating the middle space with each cluster) and partition the n-s points from the $(\frac{s}{2} + 1)^{st}$ to the $(n - \frac{s}{2} - 1)^{st}$. We again select c-2 spaces and partition the n-s points into c-1 parts, and for complete coverage each part must be at most s. Hence, the derivation is exactly the same for both s odd and s even.

To illustrate the computation, suppose we want the $P\{CC\}$ for $n = 11$, $s = 3$ and $c = 5$. The $[c - \frac{n}{s}] = 1$ and needs only 2 terms. Considering the 3 partitions (3,3,1,1),(3,2,2,1) and (2,2,2,2) with their multiplicities (or by using (3.6)), we obtain $A_o(8,4,3) = 19$ and from the partition (3,3,2) with its multiplicity we obtain $A_o(8,3,3) = 3$. Hence, for this case

$$(3.7)\qquad P\{CC\} = \frac{5!}{9^5}\{19 + 3(\tfrac{4}{2})\} = \frac{3000}{59049} = .0508\ldots\ .$$

3.1 Circular With Replacement

In the case of the circular chain with replacement (with n points, c clusters, each of size s), we again consider s odd and later show that the same result holds for s even. The analysis is reduced to that of a linear chain with replacement by breaking the circle at the center point of the first cluster selected and linearizing it. We arbitrarily assign this midpoint to the right end of the linear chain and, since we are partitioning the n spaces by selecting points, the partition cannot start with a zero. A set of Z contiguous zeros in general indicates that a particular cluster was selected Z+1 times, but a set of Z terminal zeros indicates that the initial cluster was selected a total of Z+1 times or Z times after the first choice.

As before, A(n,c,s) denotes the number of ordered partitions of n into c nonnegative parts with each part of size at most s. We need the number of ordered partitions of the n spaces (there is no space after the last point) by c-1 points into c nonnegative parts, each part at most s, and with no zero at the outset. This is given by

$$B(n,c,s) = A(n,c,s) - A(n,c-1,s) \quad (3.8)$$

the total number of ordered partitions of n into c parts, each at most s, and with the first part positive. To separate these according to the structure of the non-terminal zeros, we first note that the number of ordered partitions with no zeros in B(n,c,s) is

$$B_o(n,c,s) = A_o(n,c,s) = \sum_{i=o}^{[\frac{n-c}{s}]} (-1)^i \binom{c}{i} \binom{n-1-si}{c-1} . \quad (3.9)$$

By substracting 1 from each part of any partition we note that $A_o(n,c,s) = A(n-c,c,s-1)$. In order to include with the non-zero cases in (3.9) those cases with only 1 terminal zero, we define

$$(3.10) \qquad B_o^*(n,c,s) = B_o(n,c,s) + B_o(n,c-1,s) = A_o(n,c,s) + A_o(n,c-1,s).$$

To obtain $B_1^*(n,c,s)$, the number of ordered partitions with exactly 1 non-extreme zero and at most 1 terminal zero <u>or</u> with 2 terminal zeros, we note that we can insert a new zero in exactly c-1 places in the partitions associated with the two B_o functions in (3.10). This is because each extreme is either ruled out or leads to no new partition. Hence, the number of unordered partitions corresponding to $B_1^*(n,c,s)$ is $(c-1)!B_1^*(n,c,s)/2!$. Similarly, let $B_2^*(n,c,s)$ denote the number of partitions either with two zeros which are contiguous and both non-extreme <u>or</u> with 3 terminal zeros. This is multiplied by $(c-1)!/3!$ to get the unordered partitions. Hence, the probability $P\{CC\}$ of a complete coverage is given by

$$(3.11) \qquad P\{CC\} = \frac{(c-1)!}{n^{c-1}}\left(B_o^*(n,c,s) + \frac{B_1^*(n,c,s)}{2!} + \left\{ \frac{B_2^*(n,c,s)}{3!} + \frac{B_{1,1}^*(n,c,s)}{(2!)^2} \right\} + \left\{ \frac{B_3^*(n,c,s)}{4!} + \frac{B_{2,1}^*(n,c,s)}{3!2!} + \frac{B_{1,1,1}^*(n,c,s)}{(2!)^3} \right\} + \cdots \right).$$

For each $B^*(n,c,s)$ in (3.11) we count the number of ways we can put in the zeros and note that

$$(3.12) \qquad \begin{aligned} B_1^*(n,c,s) &= \binom{c-2}{1} B_o^*(n,c-1,s) \\ B_2^*(n,c,s) &= \binom{c-3}{1} B_o^*(n,c-2,s) \\ B_{1,1}^*(n,c,s) &= \binom{c-3}{2} B_o^*(n,c-2,s) \\ B_3^*(n,c,s) &= \binom{c-4}{1} B_o^*(n,c-3,s) \end{aligned}$$

$$B^*_{2,1}(n,c,s) = 2\binom{c-4}{2}B^*_o(n,c-3,s)$$

$$B^*_{1,1,1}(n,c,s) = \binom{c-4}{3}B^*_o(n,c-3,s), \text{ etc.}$$

In general, we can write

$$B^*_{\underset{\sim}{j}}(n,c,s) = \binom{c-J-1}{r}\left(\!\!\binom{r}{\alpha_1,\alpha_2,\ldots,\alpha_r}\!\!\right) B^*_o(n,c-J,s) \tag{3.13}$$

$$=\left(\!\!\binom{c-J-1}{\alpha_1,\alpha_2,\ldots,\alpha_r,c-J-1-\sum_{i=1}^{a}\alpha_i}\!\!\right) B^*_o(n,c-J,s)$$

where $\underset{\sim}{j} = (j_1,j_2,\ldots,j_r)$, $J = \sum_{i=1}^{r} j_i$ and α_i is the frequency of i contiguous zeros (non-extreme) or $i+1$ terminal zeros $(i = 1,2,\ldots,a)$ as in (3.9) and $\sum_{i=1}^{a} \alpha_i = r$. Hence the probability of complete coverage is

$$P\{CC\} = \frac{(c-1)!}{n^{c-1}} \sum_{J=o}^{[c-\frac{n}{s}]} \sum\left(\!\!\binom{c-J-1}{\alpha_1,\alpha_2,\ldots,\alpha_a,c-J-1-r}\!\!\right) \frac{B^*_o(n,c-J,s)}{\prod_{i=1}^{r}(j_i+1)!} \tag{3.14}$$

where the inside sum is over all the unordered partitions of the J zeros into positive parts. In more detail this can be written as

$$P\{CC\} = \frac{(c-1)!}{n^{c-1}}\left(B^*_o(n,c,s) + \left\{\frac{\binom{c-2}{1}}{2!}\right\} B^*_o(n,c-1,s) \right. \tag{3.15}$$

$$+\left\{\frac{\binom{c-3}{1}}{3!} + \frac{\binom{c-3}{2}}{(2!)^2}\right\} B^*_o(n,c-2,s)$$

$$\left. +\left\{\frac{\binom{c-4}{1}}{4!} + \frac{2\binom{c-4}{2}}{2!3!} + \frac{\binom{c-4}{3}}{(2!)^3}\right\} B^*_o(n,c-3,s) + \ldots \right) .$$

For example, for n=10, c=4, and s=3, we note that

$$
(3.16)\quad \begin{cases} A_o(10,4,3) = \sum_{i=o}^{2} (-1)^i \binom{4}{i} \binom{9-3i}{3} = 84-80+6 = 10, \\ B_o^*(10,4,3) = o \quad \text{for } x < 4; \quad B_o^*(10,4,3) = 10, \\ P\{CC\} = \frac{3!}{10^3}(10) = .06 . \end{cases}
$$

The ten partitions in this example are (3,3,3,1) with multiplicity 4 and (3,3,3,2) with multiplicity 6.

For even s we select spaces and partition the points. Except for this, the duality is complete and the resulting formula is the same for all s. As another example we take n=5, c=4, s=3 and obtain $B_o^*(5,4,3) = 10$, $B_o^*(5,3,3) = 8$, $B_o^*(5,2,3) = 2$ and hence

$$
(3.17)\qquad P\{CC\} = \frac{6}{5^3}\left(10 + 8 + \frac{2}{6}\right) = \frac{110}{125} = \frac{22}{25} = .88.
$$

This can also be (and has been) checked by independent calculations.

In addition to the formula for $A_o(n,c,s)$ in (3.9) we also note that $A_o(n,c,s)$ is the coefficient of x^{n-c} in

$$
(3.18)\qquad \left(\frac{1-x^s}{1-x}\right)^c = \left(1 + x + \ldots + x^{s-1}\right)^c .
$$

A similar formula is also obtained by David and Barton [1, p. 227, 228] for the probability that the maximum white ball run $\leq m$ when r_1 white balls and $r_2 = r - r_1$ red balls are randomly aligned. They obtained

$$(3.19) \qquad f(m,r,r_2) = \binom{r}{r_2}^{-1} \sum_{i=o}^{a} (-1)^i \binom{r_2+1}{i} \binom{r-(m+1)i}{r_2}$$

where $a = \min(r_2+1, (r-r_2)/(m+1))$ and hence we find that

$$(3.20) \qquad f(m,r,r_2) = A_o(r+1,\ r_2+1,\ m+1)/\binom{r}{r_2},$$ where A_o is given by (3.6).

From the left side of (3.18) we easily obtain (3.9) and from the right side we note that

$$(3.21) \qquad A_o(n,c,s) = \begin{cases} 0 & \text{for } n < c \text{ and for } n > cs \\ 1 & \text{for } n = c \text{ and any } s \geq 1 \\ >1 & \text{for } c < n < cs \end{cases}$$

and by (3.9) the same result holds for $B_o(n,c,s)$. It follows from (3.10) that for $s \geq 1$

$$B_o^*(n,c,s) = \begin{cases} 0 & \text{for } n < c-1 \text{ and for } n > cs \\ 1 & \text{for } n = c-1, \text{ for } n = c \text{ and for } n = cs \\ >1 & \text{for } c < n < cs \end{cases} \quad . \tag{3.22}$$

Hence, we have the two special results

$$P(CC) = \begin{cases} 0 & \text{for } n > cs. \\ (c-1)!/n^{c-1} & \text{for } n = cs \end{cases} \quad . \tag{3.23}$$

3.2. Linear Case Without Replacement

Assuming s is odd again, in order to have complete coverage we must first of all select the two clusters with centers $(s+1)/2$ and $n-(s-1)/2$. Between these two centers there are $n-s$ spaces which we have to partition (with $c-2$ additional selections) into $c-1$ parts in such a way that each part is at most s. Each partition (or selection) has probability $\binom{n-s+1}{c}^{-1}$ since there are $n-s+1$ clusters in all and we are selecting c of them. Hence, using (3.9)

$$\begin{aligned} P(CC) &= A_o(n-s,c-1,s)/\binom{n-s+1}{c} \\ &= \binom{n-s+1}{c}^{-1} \sum_{i=o}^{\left[\frac{n-s-c+1}{s}\right]} (-1)^i \binom{c-1}{i} \binom{n-s(i+1)-1}{c-2} \quad , \end{aligned} \tag{3.24}$$

where it is understood that $\binom{a}{b}=o$ for $o \le a < b$.

For even s we again select spaces and partition points, and by this duality it is easily seen that the same result holds for odd and even values of s.

In particular, we note from (3.21) and (3.24) that

$$P(CC) = \begin{cases} 0 & \text{for } n > cs \\ 1 & \text{for } c = n-s+1 \end{cases} \tag{3.25}$$

3.3 Circular Case Without Replacement

In this case we use the first cluster to unfold or linearlize the circle. Hence, (assuming s odd), we wish to partition the n spaces using the c-1 additional selections into c parts in such a manner that each part is at most s. The total number of possible selections is $\binom{n-1}{c-1}$ and the number of partitions is $A_o(n,c,s)$. Hence,

$$P\{CC\} = A_o(n,c,s)/\binom{n-1}{c-1}$$

(3.26)
$$= \binom{n-1}{c-1}^{-i} \sum_{i=o}^{\left[\frac{n-c}{s}\right]} (-1)^i \binom{c}{i}\binom{n-si-1}{c-1} \quad .$$

Again, the same result holds for s even. In particular, we note as above that

(3.27)
$$P(CC) = \begin{cases} 0 & \text{for } n > cs \\ 1 & \text{for } c \geq n-s+1 \end{cases}$$

where c can only go up to n if we sample clusters without replacement.

References

[1] David, F. N. and Barton, D. E. (1962). Combinatorial Chance. Charles Griffin and Company, London.

[2] Robbins, H. E. (1944). On the measure of a random set I. Ann. Math. Statist. 15 70-74.

[3] Roth, A. and Sobel, M. (1973). Distribution of the Size of Unions and Intersections of Clusters in one Dimension. Univ. of Minnesota Technical Report No. 217.