The Annals of Probability
1975, Vol. 3, No. 5, 894-896

A SIMPLER EXPRESSION FOR KTH NEAREST NEIGHBOR COINCIDENCE PROBABILITIES

BY RAYMOND J. HUNTINGTON[1] AND JOSEPH I. NAUS

A.T. & T., Long Lines Business Research and Rutgers University

Given N points distributed at random on [0, 1), let p_n be the size of the smallest interval that contains n points. Previous work finds $\Pr(p_n \leqq p)$ for all n, N, and rational p. The present note derives a new and considerably simplified formula for $\Pr(p_n \leqq p)$ for all n, N, and p.

1. Introduction. Given $N(\geqq 2)$ points distributed independently and uniformly on [0, 1), let p_n denote the size of the smallest interval containing $n(\leqq N)$ points. Wallenstein and Naus [5] gave an expression for $\Pr(p_n \leqq p)$ for all n, N, and rational p. The aim of this note is to give another expression for $\Pr(p_n \leqq p)$ that is computationally simpler. This enables us to give a simpler expression also for $\Pr(W_{n,t} \leqq T)$ where $W_{n,t}$ denotes the waiting time until the first occurrence within an interval of length t of n points of a Poisson process on $(0, \infty)$. The distribution of $W_{n,t}$ is of interest in certain models of neurone discharge; this and other applications are mentioned in [5] and [6]. Naus [4] shows how to adapt the proofs for $\Pr(p_n \leqq p)$ to solving a discrete generalized birthday problem. The proofs here similarly easily adapt; tables for both probabilities are given in Huntington [1].

2. A new formula for $\Pr(p_n \leqq p)$. For given integers n and N $(2 \leqq n \leqq N)$ and for given p in (0, 1) define $L = [p^{-1}]$, the largest integer in p^{-1}, and $b = 1 - pL =$ the fractional part of p^{-1}. The points ip and $b + ip$ $(i = 0, \cdots, L)$ partition [0, 1) into $2L + 1$ disjoint half-open intervals $I_1, \cdots, I_{2L+1}$ say, the $L + 1$ odd-numbered intervals being of length b and the other L intervals of length $p - b$. (Theorem 1 in Naus [3] gives an expression for the case $p = 1/L$; setting $b = 0$ here together with other simple modifications yield that case and result.) Define n_i to be the number of points in I_i.

THEOREM. *Given n, N integers, $2 \leqq n \leqq N$, and $0 < p < 1$,*

$$\Pr(p_n \leqq p) = 1 - \textstyle\sum_Q R \det |1/h_{ij}!| \det |1/l_{ij}!| \tag{1}$$

where the summation extends over the set Q of all partitions of N into $2L + 1$ integers n_i satisfying

Constraint C: $\qquad n_i + n_{i+1} < n \qquad i = 1, \cdots, 2L,$

Received October 21, 1974; revised February 5, 1975.

[1] This article is based in part on the author's Ph. D. dissertation.

AMS 1970 *subject classification.* Primary 60E05.

Key words and phrases. Coincidences, clusters, nearest neighbor distances, maximum clusters, smallest intervals, interarrival times.

$R = N!\, b^M(p-b)^{N-M}$ with $M = \sum_{k=0}^{L} n_{2k+1}$, and in the determinants of size $(L+1)\times(L+1)$ and $L \times L$ respectively, $1/\nu! = 0$ if $\nu < 0$ or $> N$, with

$$\begin{aligned} h_{ij} &= \textstyle\sum_{k=2j-1}^{2i-1} n_k - (i-j)n & L+1 \geqq i \geqq j \geqq 1\,, \\ &= -\textstyle\sum_{k=2i}^{2j-2} n_k + (j-i)n & 1 \leqq i < j \leqq L+1\,, \\ l_{ij} &= \textstyle\sum_{k=2j}^{2i} n_k - (i-j)n & L \geqq i \geqq j \geqq 1\,, \\ &= -\textstyle\sum_{k=2i+1}^{2j-1} n_k + (j-i)n & 1 \leqq i < j \leqq L\,. \end{aligned}$$

PROOF. Observe first that if $n < 1 + N/([p^{-1} - 0] + 1)$, $\Pr(p_n \leqq p) = 1$, and that (as it ought) Q is vacuous. To establish (1) otherwise, let y_p be the number of points in $[y, y+p)$. Let $n_p = \sup_{0 \leqq y < 1-p}\{y_p\}$. Observe that $\Pr(p_n \leqq p) = \Pr(n_p \geqq n)$. Let $S_1 = \bigcup_{i=1}^{L} I_{2i-1}$, $S_2 = \bigcup_{i=1}^{L-1} I_{2i}$, and write $m_k = \sup_{y \in S_k}\{y_p\}$ $(k = 1, 2)$, so that $n_p = \max(m_1, m_2)$. Given $\{n_i\}$, m_1 and m_2 are independent, and to derive the conditional distribution of m_1 (that of m_2 is found analogously), write $y_{2i+1}(t)$ to denote the number of points in $[ip, ip+t)$. Then $m_1 < n$ provided constraint C is satisfied and further that for all $0 < t < b$ and each $i = 1, \cdots, L$,

$$n_{2i-1} - y_{2i-1}(t) + y_{2i+1}(t) < n - n_{2i}\,.$$

Apply Barton and Mallows' [1, page 243] corollary as in Naus [3, proof of Theorem 1], to find $\Pr(m_1 < n \mid \{n_i\}) = \det|1/h_{ij}!| \prod_{i=1}^{L+1} n_{2i-1}!$. (In Barton and Mallows' corollary substitute $y_{2i-1}(t)$ for $A_i(m)$, n_{2i-1} for a_i, and $n_{2r-1} - (n - n_{2r})$ for $\alpha_{r+1} - \alpha_r$, or summing, h_{ij} for $a_i + \alpha_i - \alpha_j$, in their equations (18) and (19).) To complete the proof average $\prod_{j=1}^{2} \Pr(m_j < n \mid \{n_i\})$ over the multinomial distribution of $\{n_i\}$: $\Pr(\{n_i\}) = R/\prod_{i=1}^{2L+1} n_i!$.

By using the displayed expression in Section 1 of [6] with equation (1) we have:

COROLLARY. *Given a Poisson process with rate $\lambda(> 0)$,*

$$\Pr(W_{n,t} \leqq T) = 1 - \textstyle\sum_{Q^*} R^* \det|1/h_{ij}!| \det|1/l_{ij}!| \tag{2}$$

where $p = t/T$, Q^* is the set of all $2L+1$ nonnegative integers n_i satisfying Constraint C *and $R^* = \mathrm{R}e^{-\lambda p}\lambda^N/N!$.*

REMARK. Comparison of the derivation of equation (1) with that of the result in Wallenstein and Naus [5] shows that the present result is both logically and computationally simpler. Both results involve the summation of many terms and require computer calculation. The new formula sums over a simpler set of partitions (dramatically simpler for cases such as $p = 0.333$, n and N large).

REFERENCES

[1] BARTON, D. E. and MALLOWS, C. L. (1965). Some aspects of the random sequence. *Ann. Math. Statist.* **36** 236–260.

[2] HUNTINGTON, R. J. (1974). Distributions and expectations for clusters in continuous and discrete cases, with applications. Ph. D. dissertation, Rutgers University.

[3] NAUS, J. I. (1966). Some probabilities, expectations, and variances for the size of largest clusters and smallest intervals. *J. Amer. Statist. Assoc.* **61** 1191–1199.

[4] NAUS, J. I. (1974). Probabilities for a generalized birthday problem. *J. Amer. Statist. Assoc.* **69** 810–815.

[5] WALLENSTEIN, S. R. and NAUS, J. I. (1973). Probabilities for a kth nearest neighbor problem on the line. *Ann. Probability* **1** 188-190.

[6] WALLENSTEIN, S. R. and NAUS, J. I. (1974). Probabilities for the size of largest clusters and smallest intervals. *J. Amer. Statist. Assoc.* **69** 690-697.

258 PARK AVENUE
BOUND BROOK, N. J. 08805

STATISTICS CENTER
RUTGERS UNIVERSITY
NEW BRUNSWICK, N. J. 08903