It remains to be shown that k as determined by (8) equals l. This will be so if we can show that

$$(11) \qquad I_{\frac{1}{2}}(n - l + 1, l) \leq \frac{\alpha}{2} < I_{\frac{1}{2}}(n - l, l + 1).$$

Remembering that $I_x(p, q)$ is a monotonically increasing function of x we get with the help of (7) and (10)

$$\frac{\alpha}{2} = I_{1-\theta_{l-1}}(n - l + 1, l) \geq I_{\frac{1}{2}}(n - l + 1, l)$$

and

$$\frac{\alpha}{2} = I_{1-\theta_l}(n - l, l + 1) < I_{\frac{1}{2}}(n - l, l + 1)$$

which proves (11).

In conclusion it may be worth while pointing out that the formula

$$P\{Z_i < q_p < Z_j\} = I_p(i, n - i + 1) - I_p(j, n - j + 1)$$

given, e.g, in Wilks [1] for the continuous case can be obtained by a slight modification of (6).

REFERENCES

[1] S. S. Wilks, "Order statistics," *Am. Math. Soc. Bull.*, Vol. 54 (1948), pp. 6–50.
[2] H. Scheffé and J. W. Tukey, "Non-parametric estimation. I. Validation of order statistics," *Annals of Math. Stat.*, Vol. 16 (1945), pp. 187–192.
[3] C. J. Clopper and E. S. Pearson, "The use of confidence or fiducial limits illustrated in the case of the binomial," *Biometrika*, Vol. 26 (1934), pp. 404–413.
[4] H. Scheffé, "Note on the use of the tables of percentage points of the incomplete beta function to calculate small sample confidence intervals for binomial p," *Biometrika*, Vol. 33 (1944), p. 181.
[5] C. M. Thompson, E. S. Pearson, L. J. Comrie, and H. O. Hartley, "Tables of percentage points of the incomplete beta function," *Biometrika*, Vol. 32 (1941), pp. 151–181.
[6] W. R. Thompson, "On confidence ranges for the median and other expectation distributions for populations of unknown distribution form," *Annals of Math. Stat.*. Vol. 7 (1936), pp. 122–128.

A LOWER BOUND FOR THE EXPECTED TRAVEL AMONG m RANDOM POINTS

By Eli S. Marks

Bureau of the Census

In connection with cost determinations in sampling problems, it is frequently necessary to determine the amount of travel among m random sample points in an area. A lower bound for the expected value of this distance is found to be:

$$\sqrt{\frac{A}{2}} \frac{m - 1}{\sqrt{m}},$$

where A is the measure of the area from which the m random points are drawn.[1]

If in a finite area S we locate m points at random (see Figure 1), we can trace a continuous path among the m points by starting at some point and connecting the points by line segments. The points can be connected in any order so that the path touches each point only once (unless it intersects itself at one of the random points). We are interested in a lower bound for the expected value of the length of the shortest of the $m!$ possible paths.

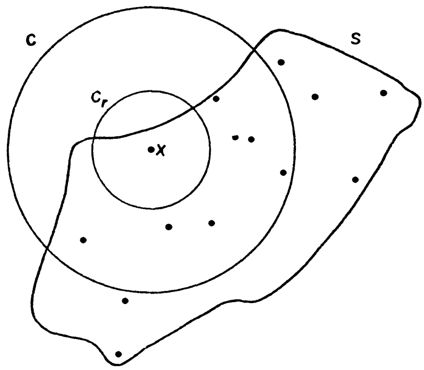

FIG. 1. m RANDOM POINTS IN S.

We have above an area S in which m random points have been selected (with $m = 14$).

The shortest path among the m points consists of $m - 1$ "links" (line segments) between two points. Each link can be assigned to one of its end points, leaving some pre-designated point (e.g., the m-th point selected) with no link assigned. The link assigned to the i-th random point ($x_{(i)}$) must be no less than $r_{(i)}$ the distance from $x_{(i)}$ to the nearest of the other ($m - 1$) points. If we denote the length of the shortest path by L:

$$L \geq \sum_{i=1}^{m-1} r_{(i)},$$

$$E(L) \geq \sum_{i=1}^{m-1} E(r_{(i)}).$$

Let $E_x(r_{(i)})$ be the expected value of $r_{(i)}$ conditional upon $x_{(i)}$ falling at the point x in S and let $F(r \mid x)$ be the conditional distribution function of $r_{(i)}$ for $x_{(i)} = x$. Thus $F(r \mid x)$ is the conditional probability of $r_{(i)} \leq r$ or the probability of

[1] The lower bound obtained is similar in form to the expression for distance traveled among a set of random points used by Mahalanobis [2] and Jessen [1].

one or more of the $(m - 1)$ random points other than $x_{(i)}$ falling inside a circle, C_r, with radius r and center at x (see Fig. 1). Then, we have:

$$E_x(r_{(i)}) = \int_0^{+\infty} r\,dF(r|x),$$

$$F(r \mid x) = 1 - \left\{\frac{M(S) - M(SC_r)}{M(S)}\right\}^{m-1},$$

where $M(S)$ and $M(SC_r)$ are the measures of S and SC_r, so that $\frac{M(SC_r)}{M(S)}$ is the probability of a random point in S falling into C_r.

Let $A = M(S)$ and construct a circle C with center at x and radius $\rho = \sqrt{\frac{A}{\pi}}$. Then $M(C) = A = M(S)$. Let d be the distance from x to the nearest of $(m - 1)$ points selected at random from C and let $G(r)$ be the distribution function of d. Then we have:

$$E(d) = \int_0^{+\infty} r\,dG(r),$$

$$G(r) = 1 - \left\{\frac{M(C) - M(CC_r)}{M(C)}\right\}^{m-1}.$$

For $r \leq \rho$,

$$M(C_rC) = M(C_r) \geq M(SC_r).$$

For $r > \rho$,

$$M(C_rC) = M(C) = M(S) \geq M(SC_r).$$

Thus, since $M(C_rC) \geq M(SC_r)$, we have for all x in S:

$$G(r) \geq F(r \mid x),$$

and thus,

$$E(d) \leq E_x(r_{(i)}).$$

Since $E(d) \leq E_x(r_{(i)})$ for all x in S:

$$E(d) \leq E(r_{(i)}),$$

$$(m - 1)E(d) \leq \sum_{i=1}^{m-1} E(r_{(i)}) \leq E(L).$$

It only remains to evaluate $E(d)$, the expected distance from the center of a circle to the nearest of $(m - 1)$ random points. This can be done very easily by substituting in the expression for $G(r)$:

$$A = M(C),$$

$$\pi r^2 = M(C_r C), \qquad \text{when } r \leq \rho = \sqrt{\frac{A}{\pi}},$$

to give:

$$G(r) = 1 - \left\{\frac{A - \pi r^2}{A}\right\}^{m-1},$$

$$G'(r) = \frac{2\pi r}{A}(m-1)\left\{\frac{A - \pi r^2}{A}\right\}^{m-2},$$

$$E(d) = \int_0^\rho rG'(r)\,dr = \frac{1}{2}\sqrt{\frac{A}{\pi}}\,[B(m, \tfrac{1}{2})],$$

where $B(m, \frac{1}{2})$ is the complete Beta function.
Since $\sqrt{m}\,[B(m, \frac{1}{2})] \geq \sqrt{\pi}$:

$$E(d) \geq \frac{1}{2}\sqrt{\frac{A}{m}}.$$

Thus, we have:

$$E(L) \geq \tfrac{1}{2}\sqrt{A}\,\frac{m-1}{\sqrt{m}}.$$

It is obvious that the development is general and applies to m random points in any bounded two-dimensional Borel set. However, the lower bound obtained will, in general, be useful only when S is a connected region.

REFERENCES

[1] RAYMOND J. JESSEN, "Statistical investigation of a sample survey for obtaining farm facts," *Iowa State College Research Bulletin* 304 (1942).
[2] P. C. MAHALANOBIS, "A sample survey of the acreage under jute in Bengal," *Sankhyā*, Vol. 4 (1940), pp. 511–530.

A MATRIX ARISING IN CORRELATION THEORY[1]

BY H. M. BACON

Stanford University

1. Introduction. In the study of time series, it is frequently desirable to consider correlations between observations made in different years. Let x_{i1}, x_{i2}, $\cdots$, x_{im} be m values of the variable x_i, expressed as deviations from their arithmetic mean, where x_i is a variable observed in the ith year $(i = 1, 2, \cdots, n)$.

[1] A linear correlogram is considered by Cochran in his paper, "Relative accuracy of systematic and stratified random samples for a certain class of populations," (*Annals of Math. Stat.*, Vol. 17 (1946), pp. 164–177) in which $\rho_\mu = 1 - \frac{\mu}{L}$. Setting $\mu = |i - j|$ and $L = 1/p$, we have the case considered above.