The Annals of Mathematical Statistics
1971, Vol. 42, No. 1, 1–11

RECTANGLE PROBABILITIES FOR UNIFORM ORDER STATISTICS AND THE PROBABILITY THAT THE EMPIRICAL DISTRIBUTION FUNCTION LIES BETWEEN TWO DISTRIBUTION FUNCTIONS[1]

BY G. P. STECK

Sandia Laboratories

1. Introduction. The principal result of this paper is a simple determinant for the probability that the order statistics from a sample of uniform random variables all lie in a multi-dimensional rectangle. Immediate applications of this result give: (i) the probability that the empirical distribution function lies between two other distribution functions; (ii) very general confidence regions for an unknown continuous distribution function; (iii) the power of tests based on the empirical distribution function. These applications, and others, are discussed in Section 4.

Let $X_1, X_2, \cdots, X_m$ be independent random variables with a continuous distribution function F and let F_m denote the empirical distribution function. Let

$$P_m(gF, hF \mid F) = P(g\{F(x)\} \leqq F_m(x) \leqq h\{F(x)\}, \text{ for all } x \mid F), \tag{1.1}$$

where g and h are distribution functions on [0, 1] with g continuous to the left and h continuous to the right.

Since the random variables $F(X_i)$ are uniform random variables with empirical distribution function $F_m F^{-1}$, it follows that

$$P_m(gF, hF \mid F) = P_m(g, h \mid F(x) = x) = P_m(g, h), \quad \text{say.}$$

Also, since $F_m F^{-1}$ passes through the points $(0, 0)$, $(U^{(1)}, 1/m)$, $(U^{(2)}, 2/m), \cdots$, $(U^{(m-1)}, (m-1)/m)$, $(1, 1)$, it follows that

$$P_m(u, v) \equiv P_m(g, h) = P(u_i \leqq U^{(i)} \leqq v_i, \ i = 1, 2, \cdots, m)$$

where $U^{(1)}, \cdots, U^{(m)}$ are the order statistics from a sample of m independent uniform random variables, $u_i = h^{-1}(i/m)$ and $v_i = g^{-1}((i-1)/m)$, $i = 1, 2, \cdots, m$.

In this paper we show that $P_m(u, v)$ is a determinant whose ijth element is $\binom{j}{j-i+1}(v_i - u_j)_+^{j-i+1}$ or 0 according as $j-i+1$ is nonnegative or negative and $(x)_+ = \max(0, x)$. Thus the determinant is of Hessenberg form with ones on the first subdiagonal and zeros below the first subdiagonal.

After this paper had been accepted for publication, I found that this result had been anticipated by Epanechnikov (1968), who proved an equivalent recurrence as a tool for studying the power of the Kolmogorov one-sample test.

With few exceptions, most notably Epanechnikov (1968), all the previous results concerning $P_m(g, h)$ have been for the special case where g and h are linear. If $g(x) = \max(0, ax-b)$ and $h(x) = \min(1, cx+d)$, with $a, b, c, d \geqq 0$, let $P_m(g, h)$ be denoted by $P_m(a, b; c, d)$. In particular, if $a = c = 1$ then $P_m(1, b; 1, d) = P(D_m^+ \leqq b, D_m^- \leqq d)$, where D_m^+ and D_m^- are the two one-sided Kolmogorov statistics.

Received November 21, 1969.

[1] This work was supported by the United States Atomic Energy Commission.