The Annals of Mathematical Statistics
1968, Vol. 39, No. 2, 398–411

THE PROBABILITY THAT THE SAMPLE DISTRIBUTION FUNCTION LIES BETWEEN TWO PARALLEL STRAIGHT LINES[1]

BY J. DURBIN

London School of Economics and Political Science

1. Introduction. Suppose that $0 \leqq x_1 \leqq \cdots \leqq x_n \leqq 1$ is an ordered sample of n independent observations from the uniform (0, 1) distribution. The sample distribution function is

$$\begin{aligned} F_n(x) &= r/n, \qquad x_r \leqq x < x_{r+1}, \\ &= 0, \qquad x < x_1, \\ &= 1, \qquad x \geqq x_n. \end{aligned}$$

Let S denote the sample path of $F_n(x)$ as x moves from 0 to 1. In this paper we consider the probability $p_n(a, b, c)$ that S lies entirely in the region R between the lines $ny = a + (n + c)x$ and $ny = -b + (n + c)x$, $(a, b, a + c, b - c > 0)$. A knowledge of this probability is important for problems arising in tests of goodness of fit, tests for the Poisson process and tests of serial independence.

For the case $a = b =$ an integer and $c = 0$, a set of simultaneous recurrence relations were obtained by Kolmogorov (1933) as a preliminary step in the development of the asymptotic form of $p_n(a, a, 0)$. These were solved by Massey (1950) to give a linear difference equation of order $2a - 1$ in the quantity $p_n(a, a, 0)n^n/n!$. In Section 2 we obtain the general form of Kolmogorov's relations and from them deduce a generalization of Massey's difference equation expressed in terms of the quantities $q_n(a, b, c) = p_n(a, b, c)(n + c)^n/n!$. Surprisingly, the coefficients of this difference equation do not depend on c or n and in fact depend only on $a + b$. Initial conditions are given whence values of $p_n(a, b, c)$ can be obtained by repeated applications of the difference equation.

For the case where c is a positive integer an explicit generating function for $q_n(a, b, c)$ is given in Section 3, generalizing results of Kemperman (1961) for the case $c = 0$. This is applied to the study of the distribution of the two-sided Kolmogorov statistic $C_n = \max_j |x_j - j/(n + 1)|$ derived from Pyke's (1959) modified sample distribution function. The methods used are different from Kemperman's and are more elementary.

The asymptotic form of $p_n(a, b, 0)$ was obtained by Doob (1949) by methods based on the reflection principle. In Section 5 we consider the application of a variant of this principle to the finite-sample case. It turns out that while the technique does not give exact results in a simple form, some sharp inequalities

Received September 7, 1967.

[1] This report was supported by the Office of Naval Research. Reproduction in whole or in part is permitted for any purpose of the United States Government.