A RANDOM VARIABLE RELATED TO THE SPACING OF SAMPLE VALUES

BY B. SHERMAN[1]

University of Southern California

1. Introduction and summary. Let x be a random variable with continuous distribution function $F(x)$. Then $y = F(x)$ is a random variable uniformly distributed over [0, 1]. If $x_1, x_2, \cdots, x_n$ is an ordered sample of n values from the population $F(x)$ then $y_1, y_2, \cdots, y_n$ $(y_i = F(x_i))$ is an ordered sample of n values from a uniform distribution over [0, 1]. For n large it is reasonable to expect that the y_i should be fairly uniformly spaced. Measures of the deviation from uniform spacing can be devised in various ways. Thus Kimball [2] has studied the random variable

$$\alpha = \sum_{i=1}^{n+1} \left(F(x_i) - F(x_{i-1}) - \frac{1}{n+1}\right)^2,$$

where $x_0 = -\infty$ and $x_{n+1} = +\infty$, conjecturing that $\alpha^{\frac{1}{2}}$ is asymptotically normally distributed. Moran [3] has studied the random variable

$$\beta = \sum_{i=1}^{n+1} (F(x_i) - F(x_{i-1}))^2,$$

which differs from α only by the quantity $-2/(n + 1) + (n + 1)^{-2}$, and has proved that β is asymptotically normally distributed. Somewhat related to these two random variables is the quantity ω^2 introduced by Smirnoff [4]. This is

$$\omega^2 = n \int_{-\infty}^{\infty} (F(x) - F^*(x))^2 \, dF(x),$$

although it is slightly more generally defined in Smirnoff's paper. Here $F^*(x)$ is the sample distribution function ([1], page 325) of a sample of n values from the population with continuous distribution function $F(x)$. The variable ω^2 may be written ([1], page 451)

$$\omega^2 = \frac{1}{12n} + \sum_{i=1}^{n} \left(F(x_i) - \frac{2i-1}{2n}\right)^2.$$

$(2i - 1)/2n$ is the midpoint of the interval $((i - 1)/n, i/n)$. Thus, if [0, 1] is partitioned into n equal subintervals then ω^2 measures the deviation of the sample values $y_i = F(x_i)$, $i = 1, 2, \cdots, n$, from the midpoints of these intervals. Smirnoff has investigated the asymptotic behavior of ω^2 obtaining a rather complicated non-normal asymptotic distribution.

[1] I wish to thank Professors J. W. Tukey and S. S. Wilks for their helpful suggestion and criticism.