

A RANDOM VARIABLE RELATED TO THE SPACING OF SAMPLE VALUES

BY B. SHERMAN¹

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, \dots, x_n is an ordered sample of n values from the population $F(x)$ then y_1, y_2, \dots, 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 Smirnov [4]. This is

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

although it is slightly more generally defined in Smirnov'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 $y_i = F(x_i)$, $i = 1, 2, \dots, n$, from the midpoints of these intervals. Smirnov 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.

It is possible to construct a definition of deviation from uniform spacing which permits a broader investigation than these random variables. This is

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

where again $x_0 = -\infty$ and $x_{n+1} = +\infty$ and $F(x)$ is a continuous distribution function. (In Theorems 3 and 4 it is assumed additionally that $F'(x)$ exists and is continuous except for a finite number of points). It is to be noted that

$$0 \leq \omega_n \leq 1.$$

Generally speaking use of the absolute value in circumstances like this is an undesirable procedure, but it turns out that ω_n is relatively easy to handle, allowing a fairly simple calculation of its moments (which are independent of $F(x)$). These are ($\mu = \min(k, n)$)

$$\alpha_{nk} = E(\omega_n^k) = \binom{n+k}{k}^{-1} \sum_{s=0}^{\mu-1} \binom{n+1}{s+1} \binom{k-1}{s} \left(\frac{n-s}{n+1} \right)^{n+k}.$$

Thus in particular the mean of ω_n is

$$E(\omega_n) = \left(\frac{n}{n+1} \right)^{n+1} \rightarrow \frac{1}{e},$$

and the variance is

$$D^2(\omega_n) = E(\omega_n^2) - E^2(\omega_n) = \frac{2n^{n+2} + n(n-1)^{n+2}}{(n+2)(n+1)^{n+2}} - \left(\frac{n}{n+1} \right)^{2n+2} \sim \frac{2e-5}{e^2} \frac{1}{n}.$$

These results will be established in Theorem 1. From the moments the characteristic function of ω_n may be obtained, and indeed in finite terms. From the characteristic function the distribution function of ω_n may be readily calculated. The distribution function is written out explicitly at the end of Theorem 1.

To determine the asymptotic distribution of the standardized variable

$$\frac{\omega_n - E(\omega_n)}{D(\omega_n)},$$

it is sufficient to examine the behaviour as $n \rightarrow \infty$ of the moments of this variable or equivalently the moments of the variable

$$\left(\frac{ne^2}{2e-5} \right)^{1/2} \left(\omega_n - \frac{1}{e} \right).$$

For it is easy to show that if the moments of the standardized variable approach the moments of a unique distribution function $F(x)$ then the distribution function of the standardized variable approaches $F(x)$. In this manner it is proved

in Theorem 2 that the distribution function of the standardized variable approaches normality.

Since the asymptotic distribution of the standardized variable

$$\frac{\omega_n - E(\omega_n)}{D(\omega_n)}$$

is known it may be used as a test for goodness of fit if the number of sample values is large. Thus suppose x_1, x_2, \dots, x_n is an ordered sample of n values from some population and we wish to test the hypothesis that the population has the distribution function $F(x)$. Then we calculate the quantity

$$\left| \frac{1}{D(\omega_n)} \left[\frac{1}{2} \sum_{i=1}^{n+1} \left| F(x_i) - F(x_{i-1}) - \frac{1}{n+1} \right| - E(\omega_n) \right] \right| = X_n,$$

and if this quantity exceeds a certain value which depends on the level of significance at which we are working we reject the hypothesis. Let us say that $P(X_n > A) = B$. The probability of rejecting the hypothesis when it is indeed true is then precisely B and this is small if A is sufficiently large. But suppose that the hypothesis is false and the sample values come from a population whose distribution function $G(x) \neq F(x)$. Then we would desire the following property to hold for the random variable X_n , namely, for any fixed positive A the probability that X_n exceeds A approaches 1 as $n \rightarrow \infty$. For in this case (and when n is large) we are almost certain to reject the null hypothesis when it is false. A test for goodness of fit which satisfies this criterion, i.e. where the probability of rejection approaches 1 as $n \rightarrow \infty$ when the null hypothesis is false, is called consistent by Wald and Wolfowitz [5]. We wish to prove then that the test for goodness of fit which uses the random variable X_n is consistent. To express the matter formally we wish to prove that (the probability density element of x_1, x_2, \dots, x_n is $n! dG(x_1) dG(x_2) \dots dG(x_n)$ in the region

$$-\infty < x_1 < x_2 < \dots < x_n < +\infty$$

and zero outside that region).

$$\lim_{n \rightarrow \infty} \int \dots \int_{D_1} dG(x_1) \dots dG(x_n) = \begin{cases} \frac{2}{\sqrt{2\pi}} \int_A^\infty e^{-(x^2/2)} dx & \text{if } F(x) \equiv G(x), \\ 1 & \text{if } F(x) \neq G(x), \end{cases}$$

where D_1 is the domain

$$-\infty < x_1 < x_2 < \dots < x_n < +\infty,$$

$$\left| \frac{1}{D(\omega_n)} \left[\frac{1}{2} \sum_{i=1}^{n+1} \left| F(x_i) - F(x_{i-1}) - \frac{1}{n+1} \right| - E(\omega_n) \right] \right| > A.$$

The first assertion here is proved in Theorem 2. The second assertion is equivalent to proving that for any fixed positive A

$$(0.1) \quad \lim_{n \rightarrow \infty} \int \dots \int_{D_2} dG(x_1) dG(x_2) \dots dG(x_n) = 0,$$

where D_2 is the domain

$$-\infty < x_1 < x_2 < \cdots < x_n < +\infty,$$

$$E(\omega_n) - AD(\omega_n) < \frac{1}{2} \sum_{i=1}^{n+1} \left| F(x_i) - F(x_{i-1}) - \frac{1}{n+1} \right| < E(\omega_n) + AD(\omega_n),$$

when $F(x) \neq G(x)$. Now $D(\omega_n)$ is of order $n^{-1/2}$, $E(\omega_n) = e^{-1} +$ terms of order n^{-1} and A is fixed. Hence it is sufficient to show that, if x_1, x_2, \dots, x_n is an ordered sample of n values from a population with distribution function $G(x)$, then the random variable

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

(it is necessary to draw a distinction between ω_n and Ω_n since $F(x) \neq G(x)$) has a mean $L_n \rightarrow L \neq e^{-1}$ and a variance $D^2(\Omega_n) \rightarrow 0$. For then we have, when n is large enough so that the interval

$$[E(\omega_n) - AD(\omega_n), E(\omega_n) + AD(\omega_n)]$$

falls outside $[L - \frac{1}{2} |L - e^{-1}|, L + \frac{1}{2} |L - e^{-1}|]$ and $|L_n - L| < \frac{1}{4} |L - e^{-1}|$,

$$\begin{aligned} P(E(\omega_n) - AD(\omega_n) < \Omega_n < E(\omega_n) + AD(\omega_n)) \\ &\leq P(|\Omega_n - L| \geq \tfrac{1}{2} |L - e^{-1}|) \\ &\leq P(|\Omega_n - L_n| \geq \tfrac{1}{4} |L - e^{-1}|) \\ &\leq \frac{E(|\Omega_n - L_n|)}{\tfrac{1}{4} |L - e^{-1}|} \leq \frac{D(\Omega_n)}{\tfrac{1}{4} |L - e^{-1}|}, \end{aligned}$$

and this implies (0.1).

But now in Theorem 3 it is shown that the mean of the random variable Ω_n is (writing $k(x) = GF^{-1}(x)$, $k(x)$ a monotonic function such that $k(0) = 0$ and $k(1) = 1$)

$$\int_0^{n, n+1} \left[1 - k\left(x + \frac{1}{n+1}\right) + k(x) \right]^n dx.$$

This expression approaches

$$\int_0^1 e^{-k'(x)} dx$$

and this integral can assume the value e^{-1} , which is its minimum relative to the class of monotonic functions such that $k(0) = 0$ and $k(1) = 1$, only when $k(x) \equiv x$ i.e. $F(x) \equiv G(x)$. Finally in Theorem 4 we prove that $D^2(\Omega_n) \rightarrow 0$ and thus it is established that the test for goodness of fit based on X_n is consistent.

2. Moments and asymptotic distribution of ω_n .

THEOREM 1. *Let $F(x)$ be a continuous distribution function. If x_1, x_2, \dots, x_n is an ordered sample of n values from the population whose distribution function is $F(x)$ then the random variable*

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

where $x_0 = -\infty$ and $x_{n+1} = +\infty$, has the moments

$$\alpha_{nk} = E(\omega_n^k) = \binom{n+k}{k}^{-1} \sum_{s=0}^{n-1} \binom{n+1}{s+1} \binom{k-1}{s} \left(\frac{n-s}{n+1} \right)^{n+k},$$

where $\mu = \min(k, n)$.

The probability density element of the x_i is ([6], page 90)

$$n! dF(x_1) dF(x_2) \cdots dF(x_n)$$

in the domain $D_x: -\infty < x_1 < x_2 < \cdots < x_n < +\infty$ and zero outside of this domain. Then

$$\alpha_{nk} = n! \int \cdots \int_{D_x} \omega_n^k dF(x_1) dF(x_2) \cdots dF(x_n).$$

If we make the transformation $y_i = F(x_i)$, $i = 1, 2, \dots, n$, then

$$\alpha_{nk} = n! \int \cdots \int_{D_y} \left[\frac{1}{2} \sum_{i=1}^{n+1} \left| y_i - y_{i-1} - \frac{1}{n+1} \right| \right]^k dy_1 dy_2 \cdots dy_n,$$

where D_y is the domain $0 < y_1 < y_2 < \cdots < y_n < 1$, thus indicating that the moments of ω_n (and therefore also the distribution function of ω_n) are independent of $F(x)$. Here $y_0 = 0$ and $y_{n+1} = 1$. The transformation

$$\begin{aligned} u_1 &= y_1, & y_1 &= u_1, \\ u_2 &= y_2 - y_1, & y_2 &= u_1 + u_2, \\ \dots & & \dots & \\ u_n &= y_n - y_{n-1}, & y_n &= u_1 + u_2 + \cdots + u_n, \\ u_{n+1} &= y_{n+1} - y_n, & y_{n+1} &= u_1 + u_2 + \cdots + u_n + u_{n+1} = 1, \end{aligned}$$

whose Jacobian is 1, then yields

$$\begin{aligned} \alpha_{nk} &= n! \int \cdots \int_{D_u} \left[\frac{1}{2} \sum_{i=1}^{n+1} \left| u_i - \frac{1}{n+1} \right| \right]^k du_1 du_2 \cdots du_n \\ &= n! \int \cdots \int_{D_u} \left[\frac{1}{2} \sum_{i=1}^n \left| u_i - \frac{1}{n+1} \right| \right. \\ &\quad \left. + \frac{1}{2} \left| \frac{n}{n+1} - (u_1 + u_2 + \cdots + u_n) \right| \right]^k du_1 \cdots du_n, \end{aligned}$$

where D_u is the domain $\sum_{i=1}^n u_i < 1$, $u_i > 0$, $i = 1, 2, \dots, n$.

The domain D_u can be regarded as the union of $2^{n+1} - 2$ subdomains in the following way. First the hyperplane $u_1 + u_2 + \cdots + u_n = n/(n+1)$ divides the domain into two parts. In the part of the domain below the hyperplane, i.e. where $u_1 + u_2 + \cdots + u_n < n/(n+1)$, we have a subdomain defined by the statement: k of the variables u_i are greater than $(n+1)^{-1}$ and the

residual group of $n - k$ u_i are less than $(n + 1)^{-1}$. There are $\binom{n}{k}$ such subdomains and it is clear that, because of the symmetry in the u_i , the integral of $\left[\frac{1}{2} \sum_{i=1}^{n+1} \left| u_i - \frac{1}{n+1} \right| \right]^k$ over each such subdomain is the same. There are altogether $\sum_{k=0}^{n-1} \binom{n}{k} = 2^n - 1$ such subdomains, $k \neq n$ because of the inequality $u_1 + u_2 + \cdots + u_n < n/(n+1)$. In the part of the domain above the hyperplane

$$u_1 + u_2 + \cdots + u_n = n/(n+1),$$

i.e. where $u_1 + u_2 + \cdots + u_n > n/(n+1)$, the reasoning is exactly the same except that here $k \neq 0$. Thus we may write

$$\begin{aligned} \alpha_{nk} = n! \sum_{r=0}^{n-1} \binom{n}{r} \int \cdots \int_{D_{r1}} \left[\sum_{i=r+1}^n \left(\frac{1}{n+1} - u_i \right) \right]^k du_1 du_2 \cdots du_n \\ + n! \sum_{r=1}^n \binom{n}{r} \int \cdots \int_{D_{r2}} \left[\sum_{i=1}^r \left(u_i - \frac{1}{n+1} \right) \right]^k du_1 du_2 \cdots du_n, \end{aligned}$$

where D_{r1} is the domain

$$\begin{aligned} \sum_{i=1}^n u_i < \frac{n}{n+1}, \quad u_i > \frac{1}{n+1} \quad (i = 1, 2, \dots, r), \\ 0 < u_i < \frac{1}{n+1} \quad (i = r+1, \dots, n), \end{aligned}$$

and D_{r2} is the domain

$$\begin{aligned} \frac{n}{n+1} < \sum_{i=1}^n u_i < 1, \quad u_i > \frac{1}{n+1} \quad (i = 1, 2, \dots, r), \\ 0 < u_i < \frac{1}{n+1} \quad (i = r+1, \dots, n). \end{aligned}$$

If we introduce the variables

$$\begin{aligned} z_i = u_i - \frac{1}{n+1} \quad (i = 1, 2, \dots, r), \\ z_i = \frac{1}{n+1} - u_i \quad (i = r+1, \dots, n), \end{aligned}$$

we get

$$\begin{aligned} \alpha_{nk} = n! \sum_{r=0}^{n-1} \binom{n}{r} \int \cdots \int_{\Delta_{r1}} \left(\sum_{i=r+1}^n z_i \right)^k dz_1 \cdots dz_n \\ + n! \sum_{r=1}^n \binom{n}{r} \int \cdots \int_{\Delta_{r2}} \left(\sum_{i=1}^r z_i \right)^k dz_1 \cdots dz_n, \end{aligned}$$

where Δ_{r1} is the domain

$$\sum_{i=1}^r z_i < \sum_{i=r+1}^n z_i, \quad z_i > 0 \quad (i = 1, 2, \dots, r),$$

$$\frac{1}{n+1} > z_i > 0 \quad (i = r+1, \dots, n),$$

and Δ_{r2} is the domain

$$\sum_{i=r+1}^n z_i < \sum_{i=1}^r z_i < \frac{1}{n+1} + \sum_{i=r+1}^n z_i, \quad z_i > 0 \quad (i = 1, 2, \dots, r),$$

$$\frac{1}{n+1} > z_i > 0 \quad (i = r+1, \dots, n).$$

To effect the integrations with respect to the variables z_1, z_2, \dots, z_r we take as volume element in the r -space of z_1, z_2, \dots, z_r the volume between the hyperplanes $z_1 + z_2 + \dots + z_r = C, z_i > 0$ and $z_1 + z_2 + \dots + z_r = C + dC, z_i > 0$. This volume element is $d \frac{C^r}{r!} = \frac{C^{r-1}}{(r-1)!} dC$. Thus

$$\begin{aligned} \alpha_{nk} &= n! \sum_{r=0}^{n-1} \binom{n}{r} \int_0^{1/n+1} \dots \int_0^{1/n+1} \left[\int_0^{\sum_{i=r+1}^n z_i} \frac{C^{r-1}}{(r-1)!} dC \right] \left(\sum_{i=r+1}^n z_i \right)^k dz_{r+1} \dots dz_n \\ &\quad + n! \sum_{r=1}^n \binom{n}{r} \int_0^{1/n+1} \dots \int_0^{1/n+1} \left[\int_{\sum_{i=r+1}^n z_i}^{(1/n+1) + \sum_{i=r+1}^n z_i} \frac{C^{k+r-1}}{(r-1)!} dC \right] dz_{r+1} \dots dz_n \\ &= n! \sum_{r=0}^{n-1} \binom{n}{r} \int_0^{1/n+1} \dots \int_0^{1/n+1} \frac{1}{r!} (z_{r+1} + \dots + z_n)^{k+r} dz_{r+1} \dots dz_n \\ &\quad + n! \sum_{r=1}^n \binom{n}{r} \int_0^{1/n+1} \dots \int_0^{1/n+1} \frac{1}{(k+r)(r-1)!} \\ &\quad \quad \cdot \left(\frac{1}{n+1} + z_{r+1} + \dots + z_n \right)^{k+r} dz_{r+1} \dots dz_n \\ &\quad - n! \sum_{r=1}^n \binom{n}{r} \int_0^{1/n+1} \dots \int_0^{1/n+1} \frac{1}{(k+r)(r-1)!} \\ &\quad \quad \cdot (z_{r+1} + \dots + z_n)^{k+r} dz_{r+1} \dots dz_n. \end{aligned}$$

In order to perform these integrations we use the formula

$$\int_0^A \dots \int_0^A (B + x_1 + x_2 + \dots + x_n)^m dx_1 \dots dx_n$$

$$= \frac{m!}{(m+n)!} \sum_{q=0}^n (-1)^{n-q} \binom{n}{q} (B + qA)^{m+n},$$

which is established immediately by induction on n . Then

$$\begin{aligned}\alpha_{nk} = n! \sum_{r=0}^{n-1} \sum_{q=0}^{n-r} \frac{(-1)^{n-r-q}}{r!} \frac{(k+r)!}{(n+k)!} \binom{n}{r} \binom{n-r}{q} \left(\frac{q}{n+1}\right)^{n+k} \\ + n! \sum_{r=1}^n \sum_{q=0}^{n-r} \frac{(-1)^{n-r-q}}{(r-1)!} \frac{(k+r-1)!}{(n+k)!} \binom{n}{r} \binom{n-r}{q} \left(\frac{1+q}{n+1}\right)^{n+k} \\ - n! \sum_{r=1}^n \sum_{q=0}^{n-r} \frac{(-1)^{n-r-q}}{(r-1)!} \frac{(k+r-1)!}{(n+k)!} \binom{n}{r} \binom{n-r}{q} \left(\frac{q}{n+1}\right)^{n+k}.\end{aligned}$$

The first of these double sums is equal to

$$\begin{aligned}\frac{n!k!}{(n+k)!} \sum_{q=1}^n \sum_{r=0}^{n-q} (-1)^{n-r-q} \binom{n}{q} \binom{n-q}{r} \binom{k+r}{k} \left(\frac{q}{n+1}\right)^{n+k} \\ = \binom{n+k}{k}^{-1} \sum_{q=1}^n \binom{n}{q} \left(\frac{q}{n+1}\right)^{n+k} \left[\sum_{r=0}^{n-q} (-1)^{n-r-q} \binom{n-q}{r} \binom{k+r}{k} \right].\end{aligned}$$

Let us assume first that $n \geq k$. The expression within the brackets is the coefficient of x^{n-q} in $(1-x)^{n-q}(1/(1-x)^{k+1}) = (1-x)^{n-q-k-1}$ and this is $\neq 0$ only when $q \geq n-k$ and then it has the value $\binom{k}{n-q}$. Thus the first double sum is equal to

$$\begin{aligned}\binom{n+k}{k}^{-1} \sum_{q=n-k}^k \binom{k}{n-q} \binom{n}{q} \left(\frac{q}{n+1}\right)^{n+k} \\ = \binom{n+k}{k}^{-1} \sum_{s=0}^k \binom{k}{s} \binom{n}{s} \left(\frac{n-s}{n+1}\right)^{n+k}.\end{aligned}$$

Similarly the second double sum is equal to

$$\binom{n+k}{k}^{-1} \sum_{s=0}^{k-1} \binom{k-1}{s} \binom{n}{s+1} \left(\frac{n-s}{n+1}\right)^{n+k},$$

and the third is equal to

$$\binom{n+k}{k}^{-1} \sum_{s=1}^k \binom{k-1}{s-1} \binom{n}{s} \left(\frac{n-s}{n+1}\right)^{n+k}.$$

Thus, using the identity

$$\binom{k}{s} \binom{n}{s} + \binom{k-1}{s} \binom{n}{s+1} - \binom{k-1}{s-1} \binom{n}{s} = \binom{n+1}{s+1} \binom{k-1}{s},$$

we get

$$\alpha_{nk} = \binom{n+k}{k}^{-1} \sum_{s=0}^{k-1} \binom{n+1}{s+1} \binom{k-1}{s} \left(\frac{n-s}{n+1}\right)^{n+k}.$$

If however $k > n$ then a similar argument shows that we get an expression for α_{nk} which differs from the above only in the upper limit of the summation, which is $n-1$ in this case. Thus the theorem is proved.

The distribution function of ω_n is

$$F(x) = 1 + \sum_{q=0}^{n-r-1} \sum_{p=0}^q (-1)^{q-p+1} \binom{n}{p} \binom{n+1}{q+1} \cdot \binom{n+q-p}{n} \left(\frac{n-q}{n+1}\right)^p \left(\frac{n-q}{n+1} - x\right)^{n-p},$$

where r is the non-negative integer determined by the inequality

$$\frac{r}{n+1} \leq x < \frac{r+1}{n+1}.$$

$F(x) = 0$ when $x \leq 0$, $F(x) = 1$ when $x \geq n/(n+1)$ and $F(x)$ is a polynomial of degree n in each of the intervals

$$\left(\frac{i-1}{n+1}, \frac{i}{n+1}\right), \quad i = 1, 2, \dots, n.$$

THEOREM 2. *The random variable ω_n is asymptotically normally distributed ($E(\omega_n)$, $D(\omega_n)$); i.e., the distribution function of the standardized variable*

$$\frac{\omega_n - E(\omega_n)}{D(\omega_n)}$$

approaches

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-(t^2/2)} dt.$$

It is sufficient to prove that the moments of the standardized variable approach the moments of the normal distribution. For in general it is known that if the moments α_{nk} of $F_n(x)$ approach the moments α_k of a uniquely determined distribution function $F(x)$, then $F_n(x)$ converges to $F(x)$ in every continuity point of the latter (M. G. Kendall, *Advanced Theory of Statistics*, Vol. 1, Third edition, Charles Griffin and Co., 1943, pp. 110-112).

Now $E(\omega_n) \rightarrow \frac{1}{e}$ and $D^2(\omega_n) \sim \frac{2e-5}{e^2} \frac{1}{n} = \frac{c}{n}$, so that the two variables $\frac{\omega_n - E(\omega_n)}{D(\omega_n)}$ and $\left(\frac{n}{c}\right)^{\frac{1}{2}} \left(\omega_n - \frac{1}{e}\right)$ have the same limiting distribution. Thus it is sufficient to prove that the moments of $\left(\frac{n}{c}\right)^{\frac{1}{2}} \left(\omega_n - \frac{1}{e}\right)$ tend to the moments of the normal distribution. In the following argument we take $\mu = k$ since $n \rightarrow \infty$.

$$\begin{aligned} E \left[\left(\frac{n}{c}\right)^{m/2} \left(\omega_n - \frac{1}{e}\right)^m \right] &= \left(\frac{n}{c}\right)^{m/2} \sum_{k=0}^m \binom{m}{k} \alpha_k \left(-\frac{1}{e}\right)^{m-k} \\ (2.1) \quad &= \frac{n^{m/2} m!}{(2e-5)^{m/2}} \left[\frac{(-1)^m}{m!} + \sum_{k=1}^m \sum_{s=0}^{k-1} \frac{(-1)^{m-k} n! e^k}{(n+k)! (m-k)!} \right. \\ &\quad \cdot \left. \binom{n+1}{s+1} \binom{k-1}{s} \binom{n-s}{n+1}^{n+k} \right]. \end{aligned}$$

Suppose now that it has been proved that $E\left[\left(\frac{n}{c}\right)^m\left(\omega_n - \frac{1}{e}\right)^{2m}\right]$ tends to a finite limit as $n \rightarrow \infty$, i.e., that the limiting moments of order $2m$ exist, $m = 1, 2, \dots$. If m is odd

$$\begin{aligned} & \left| E \left[\left(\frac{n}{c} \right)^{m/2} \left(\omega_n - \frac{1}{e} \right)^m \right] \right| \\ & \leq E \left[\left| \left(\frac{n}{c} \right)^{m/2} \left(\omega_n - \frac{1}{e} \right)^m \right| \right] \leq \left\{ E \left[\left(\frac{n}{c} \right)^m \left(\omega_n - \frac{1}{e} \right)^{2m} \right] \right\}^{1/2}. \end{aligned}$$

Hence, if m is odd, $E\left[\left(\frac{n}{c}\right)^{m/2}\left(\omega_n - \frac{1}{e}\right)^m\right]$ is bounded as $n \rightarrow \infty$. Now the expression in the bracket on the right of (2.1) can be expanded in a convergent power series in n^{-1} provided that $n > m$. Because of the factor $n^{m/2}$ and because the left hand side of (2.1) is bounded as $n \rightarrow \infty$ this power series must have $\frac{a_p}{n^p}$,

where $p \geq \frac{m+1}{2}$ (since m is odd), as its initial non-vanishing term. But then the left hand side of (2.1) must approach 0 as $n \rightarrow \infty$. Thus if the limiting moments of even order exist the limiting moments of odd order are zero. We may now restrict the discussion to even order moments.

Replacing m by $2m$ in (2.1)

$$\begin{aligned} E \left[\left(\frac{n}{c} \right)^m \left(\omega_n - \frac{1}{e} \right)^{2m} \right] &= \frac{n^m (2m)!}{(2e-5)^m} \left[\frac{1}{(2m)!} \right. \\ &+ \sum_{k=1}^{2m} \sum_{s=0}^{k-1} \frac{(-1)^k n! e^k}{(n+k)!(2m-k)!} \binom{n+1}{s+1} \binom{k-1}{s} \left(\frac{n-s}{n+1} \right)^{n+k} \left. \right]. \end{aligned}$$

Let us introduce the index $q = k - s - 1$ which runs from 0 to $2m - 1$. Then

$$\begin{aligned} E \left[\left(\frac{n}{c} \right)^m \left(\omega_n - \frac{1}{e} \right)^{2m} \right] &= \frac{n^m (2m)!}{(2e-5)^m} \left[\frac{1}{(2m)!} \right. \\ &+ \sum_{q=0}^{2m-1} \sum_{k=q+1}^{2m} \frac{(-1)^k n! e^k}{(n+k)!(2m-k)!} \binom{n+1}{k-q} \binom{k-1}{q} \left(\frac{n-k+q+1}{n+1} \right)^{n+k} \left. \right] \\ &= \frac{n^m (2m)!}{(2e-5)^m} \left[a_0 + \frac{a_1}{n} + \frac{a_2}{n^2} + \dots + \frac{a_m}{n^m} + \frac{a_{m+1}}{n^{m+1}} + \dots \right]. \end{aligned}$$

In order for $\lim_{n \rightarrow \infty} E \left[\left(\frac{n}{c} \right)^m \left(\omega_n - \frac{1}{e} \right)^{2m} \right]$ to exist it is necessary to show that $a_i = 0$,

$i = 0, 1, 2, \dots, m-1$. Then $\lim_{n \rightarrow \infty} E \left[\left(\frac{n}{c} \right)^m \left(\omega_n - \frac{1}{e} \right)^{2m} \right] = \frac{a_m (2m)!}{(2e-5)^m}$. If we de-

termine the coefficient a_{iq} of n^{-i} in the expansion in powers of n^{-1} of

$$(2.2) \quad \sum_{k=q+1}^{2m} \frac{(-1)^k n! e^k}{(n+k)!(2m-k)!} \binom{n+1}{k-q} \binom{k-1}{q} \cdot \left(\frac{n-k+q+1}{n+1} \right)^{n+k} = \sum_{i=q}^{\infty} \frac{a_{iq}}{n^i},$$

we will then have

$$(2.3) \quad a_j = \sum_{q=0}^j a_{jq}, \quad j = 1, 2, \dots, m.$$

It can be established at once that $a_0 = 0$. For if we set $q = 0$ in (2.2) and let $n \rightarrow \infty$ then (2.2) has the limit $\sum_{k=1}^{2m} \frac{(-1)^k}{(2m-k)! k!} = \frac{1}{-(2m)!}$. To determine the expansion of (2.2) in powers of n^{-1} it is sufficient to focus attention on the expansion in powers of n^{-1} of

$$\begin{aligned} & \frac{n!}{(n+k)!} (n+1)(n) \cdots (n-k+q+2) \left(\frac{n-k+q+1}{n+1} \right)^{n+k} \\ &= \frac{(n+1)(n) \cdots (n-k+q+2)}{(n+k)(n+k-1) \cdots (n+1)} \left(\frac{n-k+q+1}{n+1} \right)^{n+k} \end{aligned}$$

or equivalently on the expansion in powers of x of the function

$$\begin{aligned} & \frac{\left(\frac{1}{x}+1\right)\left(\frac{1}{x}\right) \cdots \left(\frac{1}{x}-k+q+2\right)}{\left(\frac{1}{x}+k\right)\left(\frac{1}{x}+k-1\right) \cdots \left(\frac{1}{x}+1\right)} \left(\frac{\frac{1}{x}-k+q+1}{\frac{1}{x}+1} \right)^{(1/x)+k} \\ &= \frac{x^q(1-x)(1-2x) \cdots (1-(k-q-2)x)}{(1+2x)(1+3x) \cdots (1+kx)} \left(\frac{1-(k-q-1)x}{1+x} \right)^{(1/x)+k} \\ &= x^q(a_{kq0} + a_{kq1}x + a_{kq2}x^2 + \cdots) = x^q F(x). \end{aligned}$$

Here $a_{kq0} = e^{-k+q}$ and the other coefficients may be obtained by a recursion formula. Thus:

$$\begin{aligned} a_{kqp} &= \frac{1}{p!} D_{x=0}^{(p)} F(x) = \frac{1}{p!} D_{x=0}^{(p-1)} [F(x) D \log F(x)] \\ &= \frac{1}{p!} \sum_{s=0}^{p-1} \binom{p-1}{s} D_{x=0}^{(p-s-1)} F(x) D_{x=0}^{(s+1)} \log F(x). \end{aligned}$$

But

$$\begin{aligned} D_{x=0}^{(s+1)} \log F(x) &= D_{x=0}^{(s+1)} \left[\left(\frac{1}{x} + k \right) \log(1 - (k-q-1)x) \right. \\ &\quad \left. - \left(\frac{1}{x} + k \right) \log(1+x) + \sum_{i=1}^{k-q-2} \log(1-ix) - \sum_{i=2}^k \log(1+ix) \right] \\ &= s! \left[(k-q-1)^{s+1} \left(\frac{k-q-1}{s+2} - 2k+q+1 \right) \right. \\ &\quad \left. + (-1)^s \left(1 - k - \frac{1}{s+2} \right) - \sum_{i=1}^{k-q-2} i^{s+1} - \sum_{i=2}^k (-1)^s i^{s+1} \right] = s! b_{kqs}, \end{aligned}$$

so that

$$a_{kqp} = \frac{1}{p!} \sum_{s=0}^{p-1} \binom{p-1}{s} (p-s-1)! a_{kq(p-s-1)} s! b_{kqs} = \frac{1}{p} \sum_{s=0}^{p-1} a_{kq(p-s-1)} b_{kqs}.$$

Of b_{kqs} we need merely notice that it is a polynomial in k of degree $s+2$ and that $b_{kq0} = -\frac{5}{2}k^2 + Ak + B$, where A and B depend on q only. We wish to determine the value of $a_{kq(i-q)}$ and to this end we solve the system of linear equations

$$\begin{aligned} a_{kq0} &= e^{-k+q}, \\ \frac{1}{p} \sum_{s=0}^{p-1} a_{kq(p-s-1)} b_{kqs} - a_{kqp} &= 0, \quad p = 1, 2, \dots, i-q. \end{aligned}$$

$a_{kq(i-q)}$ is therefore a quotient of two determinants. The determinant in the denominator has the value $(-1)^{i-q}$ while the determinant in the numerator can be expanded by its last column and is therefore the product of $(-1)^{i-q} e^{-k+q}$ and a determinant B_{kqi} whose entries $d_{\alpha\beta}$, $\alpha, \beta = 1, 2, \dots, i-q$, can be described as follows. If $\beta > \alpha + 1$ then $d_{\alpha\beta} = 0$. $d_{\alpha(\alpha+1)} = -1$ and when $\beta \leq \alpha$, $d_{\alpha\beta} = \frac{1}{\alpha} b_{kq(\alpha-\beta)}$, a polynomial of degree $\alpha - \beta + 2$. Thus $a_{kq(i-q)} = e^{-k+q} B_{kqi}$.

The determinant B_{kqi} is a polynomial of degree $2(i-q)$ in k and the term of this degree comes only from the product of the diagonal elements. For $B_{kqi} = |d_{\alpha\beta}| = \sum \pm \prod_{\alpha=1}^{i-q} d_{\alpha\sigma(\alpha)}$ where $\sigma(\alpha) \leq \alpha + 1$ and $(\sigma(1), \sigma(2), \dots, \sigma(i-q))$

is a permutation of $(1, 2, \dots, i-q)$. The term $\prod_{\alpha=1}^{i-q} d_{\alpha\sigma(\alpha)}$ has degree $\sum_{\alpha=1}^{i-q} (\alpha - \sigma(\alpha) + \delta(\alpha)) = \sum_{\alpha=1}^{i-q} \delta(\alpha)$ where $\delta(\alpha) = 2$ if $\sigma(\alpha) \leq \alpha$ and $\delta(\alpha) = 1$

if $\sigma(\alpha) = \alpha + 1$. But $\sum_{\alpha=1}^{i-q} \delta(\alpha) = 2(i-q) \leftrightarrow \delta(\alpha) = 2 \leftrightarrow \sigma(\alpha) \leq \alpha \leftrightarrow \sigma(\alpha) = \alpha$,

so that it is the product of the diagonal terms and only that product which gives to the term of degree $2(i-q)$ in the expansion. Thus

$$\begin{aligned} B_{kqi} &= \frac{1}{(i-q)!} (b_{kq0})^{i-q} + \text{terms of lower degree in } k \\ &= \frac{1}{(i-q)!} \left(-\frac{5}{2}\right)^{i-q} k^{2(i-q)} + \sum_{j=0}^{2(i-q)-1} A_j k^j. \end{aligned}$$

We are now in position to evaluate a_{iq} .

$$\begin{aligned} a_{iq} &= \sum_{k=q+1}^{2m} \frac{(-1)^k e^k}{(2m-k)!(k-q)!} \binom{k-1}{q} a_{kq(i-q)} \\ &= \sum_{k=q+1}^{2m} \frac{(-1)^k e^q}{(2m-k)!(k-q)!} \binom{k-1}{q} B_{kqi} \\ (2.4) \quad &= \frac{e^q}{(i-q)!} \left(-\frac{5}{2}\right)^{i-q} \sum_{k=q+1}^{2m} \frac{(-1)^k k^{2(i-q)}}{(2m-k)!(k-q)!} \binom{k-1}{q} \\ &\quad + \sum_{k=q+1}^{2m} \frac{(-1)^k e^q}{(2m-k)!(k-q)!} \binom{k-1}{q} \left[\sum_{j=0}^{2(i-q)-1} A_j k^j \right]. \end{aligned}$$

To complete the evaluation of a_{iq} we observe that

$$(2.5) \quad \sum_{k=q+1}^{2m} \frac{(-1)^k k^l}{(2m-k)!(k-q)!} \binom{k-1}{q} = \begin{cases} \frac{1}{q!} & \text{if } l = 2(m-q), \\ 0 & \text{if } l < 2(m-q). \end{cases}$$

(2.5) implies that $a_{iq} = 0$ if $i < m$ and therefore $a_j = 0$ if $j < m$. The proof of (2.5) is brief. We note that $k^{l-1} = \sum_{j=0}^{l-1} c_j \binom{k+j}{j}$, where c_j is independent of k and $c_{l-1} = (l-1)!$. Then

$$\begin{aligned} \sum_{k=q+1}^{2m} \frac{(-1)^k k^l}{(2m-k)!(k-q)!} \binom{k-1}{q} &= \sum_{k=q+1}^{2m} \frac{(-1)^k k^{l-1} k!}{q!(k-q-1)!(2m-q)!} \binom{2m-q}{k-q} \\ &= \sum_{j=0}^{l-1} \sum_{k=q+1}^{2m} (-1)^k \frac{c_j k!}{(2m-q)! q!(k-q-1)!} \binom{2m-q}{k-q} \binom{k+j}{j} \\ &= \sum_{j=0}^{l-1} \frac{c_j (j+q+1)!}{(2m-q)! j! q!} \left[\sum_{k=q+1}^{2m} (-1)^k \binom{2m-q}{k-q} \binom{k+j}{q+j+1} \right]. \end{aligned}$$

The expression within the brackets is the coefficient of x^{2m-q-1} in $(1-x)^{2m-q} \frac{1}{(1-x)^{q+j+2}} = (1-x)^{2m-2q-j-2}$ and this is zero if $j < 2(m-q) - 1$ and 1 if $j = 2(m-q) - 1$. Accordingly

$$\begin{aligned} &\sum_{k=q+1}^{2m} \frac{(-1)^k k^l}{(2m-q)!(k-q)!} \binom{k-1}{q} \\ &= \begin{cases} 0 & \text{if } l = 2(m-q), \\ \frac{[2(m-q)-1]![2(m-q)-1+q+1]!}{(2m-q)![2(m-q)-1]!q!} = \frac{1}{q!} & \text{if } l = 2(m-q), \end{cases} \end{aligned}$$

and (2.5) is established. Returning to (2.4), $a_{iq} = 0$ when $i < m$, while

$$a_{mq} = \frac{e^q}{(m-q)!q!} \left(-\frac{5}{2}\right)^{m-q} = \frac{1}{m!} \binom{m}{q} \left(-\frac{5}{2}\right)^{m-q} e^q;$$

and now applying this expression to (2.3)

$$a_m = \sum_{q=0}^m \frac{1}{m!} \binom{m}{q} \left(-\frac{5}{2}\right)^{m-q} e^q = \frac{1}{m! 2^m} (2e - 5)^m.$$

Thus

$$\lim_{n \rightarrow \infty} E \left[\left(\frac{n}{c} \right)^m \left(\omega_n - \frac{1}{e} \right)^{2m} \right] = \frac{a_m (2m)!}{(2e-5)!} = \frac{(2m)!}{m! 2^m},$$

and these are precisely the even order moments of the normal distribution.

Thus $\left(\frac{n}{c} \right)^{1/2} \left(\omega_n - \frac{1}{e} \right)$ is asymptotically normal and so is $\frac{\omega_n - E(\omega_n)}{D(\omega_n)}$.

The skewness $\beta_1 = \left(\frac{\mu_3}{\sigma^3}\right)^2$ and kurtosis $\beta_2 = \frac{\mu_4}{\sigma^4}$ of the standardized variable $\frac{\omega_n - E(\omega_n)}{D(\omega_n)}$ are

$$\beta_1 = \frac{1}{n} \frac{(6e^2 - 42e + 70)^2}{(2e - 5)^3} + O(n^{-2}) = \frac{.356 \dots}{n} + O(n^{-2}),$$

$$\beta_2 = 3 + \frac{1}{n} \frac{24e^3 - 336e^2 + 1368e - 1718}{(2e - 5)^2} + O(n^{-2}) = 3 - \frac{1.05 \dots}{n} + O(n^{-2}).$$

3. Consistency. According to previous discussion in order to prove the consistency of the test for goodness of fit based on the asymptotically normal variable $\frac{\omega_n - E(\omega_n)}{D(\omega_n)}$ it is sufficient to show that, if x_1, x_2, \dots, x_n is an ordered sample from a population whose distribution function is $G(x)$, then the limiting mean of the random variable $\frac{1}{2} \sum_{i=1}^{n+1} |F(x_i) - F(x_{i-1}) - \frac{1}{n+1}|$ is not equal to e^{-1} if $F(x) \neq G(x)$ and the limiting variance of this variable is zero. This is the content of the next two theorems. In connection with these theorems it is to be observed that, when $y = F(x)$ is continuous, $F^{-1}(y)$, $0 \leq y \leq 1$, can be defined unambiguously by writing $F^{-1}(y) = [\text{Sup } x: y = F(x)]$ except for $y = 0$, and $F^{-1}(0) = -\infty$. The function $k(x) = GF^{-1}(x)$ is then a non-decreasing function mapping $[0, 1]$ into $[0, 1]$ and such that $k(0) = 0$ and $k(1) = 1$. Now if $F'(x)$ exists for all but a finite number of points and is never zero then $F^{-1}(x)$ is continuous and so is $k(x)$. If further $G'(x)$ and $F'(x)$ exist and are continuous except for a finite number of points then $(F'(x) \neq 0)k'(x)$ enjoys the same property. These remarks justify the substitutions and partial integrations that are effected in the course of the next two theorems.

THEOREM 3. *Let $F(x)$ and $G(x)$ be continuous distribution functions whose derivatives exist and are continuous except for a finite number of points. If x_1, x_2, \dots, x_n is an ordered sample of n values from the population whose distribution function is $G(x)$ then $(k(x) = GF^{-1}(x))$*

$$\begin{aligned} E(\Omega_n) &= E \left(\frac{1}{2} \sum_{i=1}^{n+1} \left| F(x_i) - F(x_{i-1}) - \frac{1}{n+1} \right| \right) \\ &= \int_0^{n/(n+1)} \left[1 - k \left(x + \frac{1}{n+1} \right) + k(x) \right]^n dx \rightarrow \int_0^1 e^{-k'(x)} dx. \end{aligned}$$

The integral $\int_0^1 e^{-k'(x)} dx$ has, relative to the class of monotonic functions such that $k(0) = 0$ and $k(1) = 1$, the minimum value e^{-1} and assumes that value only when $k(x) \equiv x$ i.e. $F(x) \equiv G(x)$.

Let us suppose first that $F'(x) \neq 0$. Then $F^{-1}(x)$ is continuous and it is differentiable at all but a finite number of points as is also the function $GF^{-1}(x) = k(x)$.

$$\begin{aligned}
 E(\Omega_n) &= \frac{1}{2} \sum_{i=1}^{n+1} E \left(\left| F(x_i) - F(x_{i-1}) - \frac{1}{n+1} \right| \right) \\
 (3.1) \quad &= \frac{1}{2} E \left(\left| F(x_i) - \frac{1}{n+1} \right| \right) + \frac{1}{2} E \left(\left| 1 - F(x_n) - \frac{1}{n+1} \right| \right) \\
 &\quad + \frac{1}{2} \sum_{i=2}^n E \left(\left| F(x_i) - F(x_{i-1}) - \frac{1}{n+1} \right| \right).
 \end{aligned}$$

The joint probability density element of x_{i-1} and x_i is

$$\frac{n!}{(i-2)!(n-i)!} G(x_{i-1})^{i-2} (1 - G(x_i))^{n-i} dG(x_{i-1}) dG(x_i)$$

in the domain $-\infty < x_{i-1} < x_i < +\infty$ and zero outside that domain. Hence

$$\begin{aligned}
 &\frac{1}{2} \sum_{i=2}^n E \left(\left| F(x_i) - F(x_{i-1}) - \frac{1}{n+1} \right| \right) \\
 &= \frac{1}{2} \sum_{i=2}^n \int_{-\infty}^{\infty} \int_{-\infty}^{x_i} \left| F(x_i) - F(x_{i-1}) - \frac{1}{n+1} \right| \\
 &\quad \cdot \frac{n!}{(i-2)!(n-i)!} G(x_{i-1})^{i-2} (1 - G(x_i))^{n-i} dG(x_{i-1}) dG(x_i) \\
 &= \frac{1}{2} n(n-1) \int_{-\infty}^{\infty} \int_{-\infty}^Y \left| F(Y) - F(X) - \frac{1}{n+1} \right| \\
 &\quad \cdot [1 - G(Y) + G(X)]^{n-2} dG(X) dG(Y),
 \end{aligned}$$

and making the transformation $y = F(Y)$ and $x = F(X)$ the integral on the right can be written

$$\begin{aligned}
 &\frac{1}{2} n(n-1) \int_0^1 \int_0^y \left| y - x - \frac{1}{n+1} \right| [1 - k(y) + k(x)]^{n-2} dk(x) dk(y) \\
 &= \frac{1}{2} n(n-1) \int_0^1 \int_0^y \left(x - y + \frac{1}{n+1} \right) [1 - k(y) + k(x)]^{n-2} dk(x) dk(y) \\
 &+ n(n-1) \int_{1/n+1}^1 \int_0^{y-(1/n+1)} \left(y - x - \frac{1}{n+1} \right) [1 - k(y) + k(x)]^{n-2} dk(x) dk(y).
 \end{aligned}$$

Integrating partially with respect to x , the expression on the right becomes

$$\begin{aligned}
 &\frac{n}{2} \int_0^1 \frac{1}{n+1} dk(y) - \frac{n}{2} \int_0^1 \left(-y + \frac{1}{n+1} \right) [1 - k(y)]^{n-1} dk(y) \\
 &\quad - \frac{n}{2} \int_0^1 \int_0^y [1 - k(y) + k(x)]^{n-1} dx dk(y) \\
 &\quad - n \int_{1/n+1}^1 \left(y - \frac{1}{n+1} \right) [1 - k(y)]^{n-1} dk(y) \\
 &\quad + n \int_{1/n+1}^1 \int_0^{y-(1/n+1)} [1 - k(y) + k(x)]^{n-1} dx dk(y),
 \end{aligned}$$

and now integrating with respect to y

$$\begin{aligned} \frac{1}{2} \sum_{i=2}^n E \left(\left| F(x_i) - F(x_{i-1}) - \frac{1}{n+1} \right| \right) &= -\frac{1}{n+1} + \frac{1}{2} \int_0^1 [1 - k(x)]^n dx \\ &\quad + \frac{1}{2} \int_0^1 k(x)^n dx - \int_{1/n+1}^1 [1 - k(x)]^n dx - \int_0^{n/n+1} k(x)^n dx \\ &\quad + \int_0^{n/n+1} \left[1 - k \left(x + \frac{1}{n+1} \right) + k(x) \right]^n dx. \end{aligned}$$

The other two terms in (3.1) are treated similarly. The probability density element of x_1 is $n(1 - G(x_1))^{n-1} dG(x_1)$ so that

$$\begin{aligned} \frac{1}{2} E \left(\left| F(x_1) - \frac{1}{n+1} \right| \right) &= \frac{n}{2} \int_{-\infty}^{\infty} \left| F(x) - \frac{1}{n+1} \right| (1 - G(x))^{n-1} dG(x) \\ &= \frac{n}{2} \int_0^1 \left| x - \frac{1}{n+1} \right| (1 - k(x))^{n-1} dk(x) \\ &= \frac{1}{2(n+1)} - \frac{1}{2} \int_0^{1/n+1} (1 - k(x))^n dx \\ &\quad + \frac{1}{2} \int_{1/n+1}^1 (1 - k(x))^n dx. \end{aligned}$$

Similarly we find that

$$\begin{aligned} \frac{1}{2} E \left(\left| 1 - F(x_n) - \frac{1}{n+1} \right| \right) &= \frac{1}{2(n+1)} \\ &\quad + \frac{1}{2} \int_0^{n/n+1} k(x)^n dx - \frac{1}{2} \int_{1/n+1}^1 k(x)^n dx. \end{aligned}$$

Thus

$$E(\Omega_n) = \int_0^{n/n+1} \left[1 - k \left(x + \frac{1}{n+1} \right) + k(x) \right]^n dx.$$

This result is, however, independent of the hypothesis $F'(x) \neq 0$. For if $F'(x)$ is sometimes zero we may select a sequence of distribution functions $F_m(x)$, $m = 1, 2, \dots$, which converges everywhere to $F(x)$ and which is such that $F'_m(x) \neq 0$. The $F_m(x)$ otherwise satisfy the conditions of the theorem. If Ω_{mn} is that function of x_1, x_2, \dots, x_n obtained by replacing $F(x)$ by $F_m(x)$ in Ω_n then Ω_{mn} converges to Ω_n for every fixed set of x_1, x_2, \dots, x_n and $E(\Omega_{mn})$ converges to $E(\Omega_n)$ since both Ω_{mn} and Ω_n are bounded by 1. Furthermore if x_0 is any value such that $F'(x_0) \neq 0$ and $y_0 = F(x_0)$ then $F_m^{-1}(y_0)$ converges to $F^{-1}(y_0) = x_0$. For if x_1 is a cluster point of the set $F_m^{-1}(y_0)$, then there exists, for a given ϵ , a sufficiently large m such that $|F(x_1) - F_m(x_1)| < \epsilon$ (because $F_m(x) \rightarrow F(x)$) while, for the same m , $|F_m(x_1) - y_0| < \epsilon$ because of the continuity of $F_m(x)$. Thus $|F(x_1) - y_0| < 2\epsilon$ and, since ϵ is arbitrary, $y_0 = F(x_1) = F(x_0)$. So $x_1 = x_0$ since $F'(x_0) \neq 0$. Thus $F_m^{-1}(y) \rightarrow F^{-1}(y)$ for any

value of y such that if x is mapped into y by $F(x)$ then $F'(x) \neq 0$. This set on the y axis however includes all y except for a set of measure zero and so $F_m^{-1}(y) \rightarrow F^{-1}(y)$ almost everywhere. So $k_m(y) = GF_m^{-1}(y) \rightarrow GF^{-1}(y) = k(y)$ almost everywhere and

$$\left[1 - k_m\left(y + \frac{1}{n+1}\right) + k_m(y)\right]^n \rightarrow \left[1 - k\left(y + \frac{1}{n+1}\right) + k(y)\right]^n$$

almost everywhere. Then

$$\begin{aligned} \int_0^{n/n+1} \left[1 - k_m\left(x + \frac{1}{n+1}\right) + k_m(x)\right]^n dx \\ \rightarrow \int_0^{n/n+1} \left[1 - k\left(x + \frac{1}{n+1}\right) + k(x)\right]^n dx \end{aligned}$$

since both integrands are bounded by 1. Therefore the equality

$$E(\Omega_{mn}) = \int_0^{n/n+1} \left[1 - k_m\left(x + \frac{1}{n+1}\right) + k_m(x)\right]^n dx$$

is preserved as $m \rightarrow \infty$.

Now $k(x)$ is a monotonic function and hence has a derivative almost everywhere. Then

$$\begin{aligned} \left[1 - k\left(x + \frac{1}{n+1}\right) + k(x)\right]^n \\ = \left[1 - \frac{1}{n+1} \left(k\left(x + \frac{1}{n+1}\right) - k(x)\right) / \frac{1}{n+1}\right]^n \end{aligned}$$

converges to $e^{-k'(x)}$ almost everywhere. If we write

$$H_n(x) = \left[1 - k\left(x + \frac{1}{n+1}\right) + k(x)\right]^n$$

when $0 \leq x \leq \frac{n}{n+1}$ and $H_n(x) = 0$ when $\frac{n}{n+1} < x \leq 1$, then

$$\int_0^1 H_n(x) dx = \int_0^{n/n+1} \left[1 - k\left(x + \frac{1}{n+1}\right) + k(x)\right]^n dx \rightarrow \int_0^1 e^{-k'(x)} dx$$

as $n \rightarrow \infty$. The curve $y = e^{-x}$ lies always above its tangents and the tangent at $x = 1$ is $y = -\frac{1}{e}x + \frac{2}{e}$. Thus $e^{-x} \geq -\frac{1}{e}x + \frac{2}{e}$ for all x , equality holding only when $x = 1$, and therefore $e^{-k'(x)} \geq -\frac{1}{e}k'(x) + \frac{2}{e}$, equality holding only when $k'(x) = 1$.

So

$$\int_0^1 e^{-k'(x)} dx \geq -\frac{1}{e} \int_0^1 k'(x) dx + \frac{2}{e},$$

equality holding if and only if $k'(x) = 1$ almost everywhere. But for any monotonic non-decreasing function

$$\int_0^1 k'(x) dx \leq k(1) - k(0),$$

equality holding if and only if $k(x)$ is absolutely continuous. Hence

$$\int_0^1 e^{-k'(x)} dx \geq -\frac{1}{e} \int_0^1 k'(x) dx + \frac{2}{e} \geq \frac{1}{e},$$

and the equality runs through if and only if $k(x)$ is an absolutely continuous function such that $k'(x) = 1$ almost everywhere. But this is true of $k(x)$ if and only if $k(x) \equiv x$ and this in turn is true if and only if $F(x) \equiv G(x)$.

THEOREM 4. *The random variable Ω_n has limiting variance zero; i.e., $\lim_{n \rightarrow \infty} E(\Omega_n^2) = \left[\int_0^1 e^{-k'(x)} dx \right]^2$.*

As before we assume first that $F'(x) \neq 0$. Then

$$\begin{aligned} E(\Omega_n^2) &= E \left[\left(\frac{1}{2} \sum_{i=2}^n \left| F(x_i) - F(x_{i-1}) - \frac{1}{n+1} \right| \right)^2 \right] \\ (4.1) \quad &+ E \left[\left| F(x_1) - \frac{1}{n+1} \right| \Omega_n \right] + E \left[\left| 1 - F(x_n) - \frac{1}{n+1} \right| \Omega_n \right] \\ &- E \left[\frac{1}{4} \left(\left| F(x_1) - \frac{1}{n+1} \right| + \left| 1 - F(x_n) - \frac{1}{n+1} \right| \right)^2 \right]. \end{aligned}$$

Suppose $[\text{Sup } x: k(x) = 0] = a$ and $[\text{Inf } x: k(x) = 1] = b$. We may then obtain $\lim_{n \rightarrow \infty} E \left[\left| F(x_1) - \frac{1}{n+1} \right| \Omega_n \right]$ in the following manner:

$$\begin{aligned} \left| E \left[\left| F(x_1) - \frac{1}{n+1} \right| \Omega_n \right] - E[a\Omega_n] \right| &\leq E \left[\left| \left| F(x_1) - \frac{1}{n+1} \right| - a \right| \Omega_n \right] \\ (4.2) \quad &\leq E \left[\left| F(x_1) - \frac{1}{n+1} - a \right| \Omega_n \right] \\ &\leq \left[E \left(F(x_1) - \frac{1}{n+1} - a \right)^2 \right]^{1/2} [E(\Omega_n^2)]^{1/2}. \end{aligned}$$

But $\Omega_n \leq 1$ so that $E(\Omega_n^2)$ is bounded as $n \rightarrow \infty$. On the other hand

$$\begin{aligned} E \left(F(x_1) - \frac{1}{n+1} - a \right)^2 &= n \int_{-\infty}^{\infty} \left(F(x_1) - \frac{1}{n+1} - a \right)^2 (1 - G(x_1))^{n-1} dG(x_1) \\ &= n \int_0^1 \left(x - a - \frac{1}{n+1} \right)^2 (1 - k(x))^{n-1} dk(x) \\ &= \left(a + \frac{1}{n+1} \right)^2 + \int_0^1 2 \left(x - a - \frac{1}{n+1} \right) (1 - k(x))^n dx. \end{aligned}$$

As $n \rightarrow \infty$ the expression on the right tends to $a^2 + \int_0^a 2(x-a) dx = 0$. Thus the expression on the right of (4.2) goes to zero as $n \rightarrow \infty$ and therefore

$$(4.3) \quad \lim_{n \rightarrow \infty} E \left[\left| F(x_1) - \frac{1}{n+1} \right| \Omega_n \right] = \lim_{n \rightarrow \infty} E [a\Omega_n] = a \int_0^1 e^{-k'(x)} dx.$$

In a similar manner we obtain

$$(4.4) \quad \lim_{n \rightarrow \infty} E \left[\left| 1 - F(x_n) - \frac{1}{n+1} \right| \Omega_n \right] = (1-b) \int_0^1 e^{-k'(x)} dx$$

and

$$(4.5) \quad \lim_{n \rightarrow \infty} -E \left[\frac{1}{4} \left(\left| F(x_1) - \frac{1}{n+1} \right| + \left| 1 - F(x_n) - \frac{1}{n+1} \right| \right)^2 \right] \\ = -\frac{1}{4}(a+1-b)^2$$

The first term on the right of (4.1) remains to be investigated. We have

$$(4.6) \quad E \left[\left(\frac{1}{2} \sum_{i=2}^n \left| F(x_i) - F(x_{i-1}) - \frac{1}{n+1} \right| \right)^2 \right] \\ = \frac{1}{4} E \left[\sum_{i=2}^n \left(F(x_i) - F(x_{i-1}) - \frac{1}{n+1} \right)^2 \right] \\ + \frac{1}{2} E \left[\sum_{i=2}^{n-2} \sum_{j=i+2}^n \left| F(x_i) - F(x_{i-1}) - \frac{1}{n+1} \right| \left| F(x_j) - F(x_{j-1}) - \frac{1}{n+1} \right| \right] \\ + \frac{1}{2} E \left[\sum_{i=2}^{n-1} \left| F(x_i) - F(x_{i-1}) - \frac{1}{n+1} \right| \left| F(x_{i+1}) - F(x_i) - \frac{1}{n+1} \right| \right].$$

The joint probability density element of x_{i-1} and x_i is

$$\frac{n!}{(i-2)!(n-i)!} (1 - G(x_{i-1}))^{i-2} G(x_i)^{n-i} dG(x_{i-1}) dG(x_i)$$

so that

$$\frac{1}{4} E \left[\sum_{i=2}^n \left(F(x_i) - F(x_{i-1}) - \frac{1}{n+1} \right)^2 \right] \\ = \frac{1}{4} n(n-1) \iint_{-\infty < X < Y < \infty} \left(F(Y) - F(X) - \frac{1}{n+1} \right)^2 \\ [1 - G(Y) + G(X)]^{n-2} dG(X) dG(Y) \\ = \frac{1}{4} n(n-1) \int_0^1 \int_0^y \left(y - x - \frac{1}{n+1} \right)^2 [1 - k(y) + k(x)]^{n-2} dk(x) dk(y).$$

In this latter double integral we integrate first with respect to x and then with respect to y obtaining

$$\begin{aligned} \frac{-n-3}{4(n+1)^2} - \frac{1}{2} \int_0^1 \left(y - \frac{1}{n+1} \right) [1 - k(y)]^n dy - \frac{1}{2} \int_0^1 \left(\frac{n}{n+1} - x \right) k(x)^n dx \\ + \frac{1}{2} \iint_{0 < x < y < 1} [1 - k(y) + k(x)]^n dx dy, \end{aligned}$$

and proceeding to the limit

$$\begin{aligned} (4.7) \quad \lim_{n \rightarrow \infty} \frac{1}{4} E \left[\sum_{i=2}^n \left(F(x_i) - F(x_{i-1}) - \frac{1}{n+1} \right)^2 \right] \\ = -\frac{1}{2} \int_0^a y dy - \frac{1}{2} \int_b^1 (1-x) dx + \frac{1}{2} \iint_{\substack{0 < x < y < 1 \\ k(x)=k(y)}} dx dy \\ = -\frac{1}{4} a^2 - \frac{1}{4} (1-b)^2 + \frac{1}{2} \iint_{\substack{0 < x < y < 1 \\ k(x)=k(y)}} dx dy. \end{aligned}$$

The joint probability density element of $x_{i-1}, x_i, x_{j-1}, x_j$ when $j > i+1$ is

$$\frac{n!}{(i-2)!(j-i-2)!(n-j)!} G(x_{i-1})^{i-2} [G(x_{j-1}) - G(x_i)]^{j-i-2} \\ [1 - G(x_j)]^{n-j} dG(x_{i-1}) dG(x_i) dG(x_{j-1}) dG(x_j),$$

so

$$\begin{aligned} (4.8) \quad \frac{1}{2} E \left[\sum_{i=2}^{n-2} \sum_{j=i+2}^n \left| F(x_i) - F(x_{i-1}) - \frac{1}{n+1} \right| \left| F(x_j) - F(x_{j-1}) - \frac{1}{n+1} \right| \right] \\ = \frac{1}{2} n(n-1)(n-2)(n-3) \iiint_{0 < x < y < u < v < 1} \left| F(Y) - F(X) - \frac{1}{n+1} \right| \\ \cdot \left| F(V) - F(U) - \frac{1}{n+1} \right| [1 - G(V) + G(U) \\ - G(Y) + G(X)]^{n-4} dG(X) dG(Y) dG(U) dG(V) \\ = \frac{1}{2} n(n-1)(n-2)(n-3) \iiint_{0 < x < y < u < v < 1} \left| y - x - \frac{1}{n+1} \right| \left| v - u - \frac{1}{n+1} \right| \\ \cdot [1 - k(v) + k(u) - k(y) + k(x)]^{n-4} dk(x) dk(y) dk(u) dk(v). \end{aligned}$$

The joint probability density element of x_{i-1}, x_i, x_{i+1} is

$$\frac{n!}{(i-2)!(n-i-1)!} G(x_{i-1})^{i-2} [1 - G(x_{i+1})]^{n-i-1} dG(x_{i-1}) dG(x_i) dG(x_{i+1})$$

and so

$$\begin{aligned}
 & \frac{1}{2} E \left[\sum_{i=2}^{n-1} \left| F(x_i) - F(x_{i-1}) - \frac{1}{n+1} \right| \left| F(x_{i+1}) - F(x_i) - \frac{1}{n+1} \right| \right] \\
 &= \frac{1}{2} n(n-1)(n-2) \iiint_{0 < x < y < v < 1} \left| F(Y) - F(X) - \frac{1}{n+1} \right| \\
 (4.9) \quad & \cdot \left| F(V) - F(Y) - \frac{1}{n+1} \right| [1 - G(V) + G(X)]^{n-3} dG(X) dG(Y) dG(V) \\
 &= \frac{1}{2} n(n-1)(n-2) \iiint_{0 < x < y < v < 1} \left| y - x - \frac{1}{n+1} \right| \\
 & \cdot \left| v - y - \frac{1}{n+1} \right| [1 - k(v) + k(x)]^{n-3} dk(x) dk(y) dk(v).
 \end{aligned}$$

We introduce the symbol $S(p, q)$ as follows

$$S(p, q) = \begin{cases} 1 & \text{if } q \leq p + \frac{1}{n+1}, \\ -1 & \text{if } q > p + \frac{1}{n+1}. \end{cases}$$

Then in the integral on the right of (4.8) we perform a partial integration with respect to u and add to the integral on the right of (4.9). We get

$$\begin{aligned}
 & \frac{1}{2} n(n-1)(n-2) \iiint_{0 < x < y < v < 1} \frac{1}{n+1} \left| y - x - \frac{1}{n+1} \right| \\
 & \cdot [1 - k(y) + k(x)]^{n-3} dk(x) dk(y) dk(v) \\
 & - \frac{1}{2} n(n-1)(n-2) \iiint_{0 < x < y < u < v < 1} S(u, v) \left| y - x - \frac{1}{n+1} \right| \\
 & \cdot [1 - k(v) + k(u) - k(y) + k(x)]^{n-3} dk(x) dk(y) dk(v) du,
 \end{aligned}$$

and now integrating with respect to v in the triple integral and performing partial integrations with respect to x and collecting terms the sum of (4.8) and (4.9) becomes

$$\begin{aligned}
 & \frac{n(n-1)}{4(n+1)^2} - \frac{n(n-1)}{2(n+1)} \int_0^1 \left| y - \frac{1}{n+1} \right| [1 - k(y)]^{n-1} dk(y) - \frac{2n(n-1)}{n+1} \\
 & \cdot \iint_{0 < x < y < 1} S(x, y) [1 - k(y) + k(x)]^{n-1} dx dk(y) + \frac{1}{2} n(n-1) \\
 & \cdot \iiint_{0 < y < u < v < 1} S(u, v) \left| y - \frac{1}{n+1} \right| [1 - k(v) + k(u) - k(y)]^{n-2} \\
 & \quad \cdot dk(y) dk(v) du + \frac{1}{2} n(n-1) \\
 & \cdot \iiint_{0 < x < y < u < v < 1} S(u, v) S(x, y) [1 - k(v) + k(u) - k(y) + k(x)]^{n-2} dk(y) dk(v) dx du.
 \end{aligned}$$

Now some tedious, although in principle straightforward, calculations show that the first three terms of this expression approach

$$(4.10) \quad -\frac{1}{4} - \frac{1}{2}a - \frac{1}{2}(1-b) + \int_0^1 e^{-k'(x)} dx,$$

that the triple integral approaches

$$(4.11) \quad \frac{1}{2}a + \frac{1}{2}a(1-b) + \frac{1}{2}a^2 - a \int_0^1 e^{-k'(u)} du,$$

and that the quadruple integral approaches

$$(4.12) \quad 2 \iint_{0 < x < u < 1} e^{-k'(x)-k'(u)} dx du - \int_0^1 e^{-k'(x)} dx - (1-b) \int_0^1 e^{-k'(x)} dx \\ - \frac{1}{2} \iint_{\substack{0 < x < u < 1 \\ k(x)=k(u)}} dx du + (1-b)^2 + \frac{1}{2}b(1-b) + \frac{1}{4}.$$

Thus collecting the results of (4.3), (4.4), (4.5), (4.7), (4.10), (4.11), and (4.12) we have

$$\lim_{n \rightarrow \infty} E(\Omega_n^2) = 2 \iint_{0 < x < u < 1} e^{-k'(x)-k'(u)} dx du.$$

Since the integrand is symmetrical in the variables u and x we may write

$$(4.13) \quad \lim_{n \rightarrow \infty} E(\Omega_n^2) = \iint_{\substack{0 < x < 1 \\ 0 < u < 1}} e^{-k'(x)-k'(u)} dx du = \left[\int_0^1 e^{-k'(x)} dx \right]^2,$$

and this proves the theorem in the case $F'(x) \neq 0$.

Using the procedure of theorem 3 we may however extend the theorem to include the possibility that $F'(x)$ is sometimes zero. But it must be shown additionally that the sequence $F_m(x)$ can be so chosen that Ω_{mn} converges to Ω_n uniformly in n , i.e. that, for a given ϵ , $|\Omega_{mn} - \Omega_n| < \epsilon$ for m sufficiently large and for any value of n . If this is true then, observing that $0 \leq \Omega_{mn} + \Omega_n \leq 2$, $|\Omega_{mn}^2 - \Omega_n^2| < 2\epsilon$ and

$$|E(\Omega_{mn}^2) - E(\Omega_n^2)| \leq E(|\Omega_{mn}^2 - \Omega_n^2|) \leq 2\epsilon$$

independently of n . Letting $n \rightarrow \infty$

$$\left| \left[\int_0^1 e^{-k'_m(x)} dx \right]^2 - \lim_{n \rightarrow \infty} E(\Omega_n^2) \right| \leq 2\epsilon,$$

and now letting $m \rightarrow \infty$ (the $F_m(x)$ constructed below are such that $k'_m(x) \rightarrow k'(x)$)

$$\left| \left[\int_0^1 e^{-k'(x)} dx \right]^2 - \lim_{n \rightarrow \infty} E(\Omega_n^2) \right| \leq 2\epsilon.$$

Since ϵ is arbitrary this implies (4.13), so that the theorem is extended to include the possibility that $F'(x)$ is sometimes zero. That the sequence $F_m(x)$ can be chosen so that Ω_{mn} converges to Ω_n uniformly in n can be shown as follows. The set of points on the x axis for which $F'(x) = 0$ maps into a set of points on the y axis of measure zero. For any m we may enclose this set on the y axis in an open set S of measure less than $\frac{1}{m}$. S is the union of disjoint open intervals S_i , $i = 1, 2, \dots$. The sets $T_i = F^{-1}(S_i)$ on the x axis are disjoint open intervals. Now we may construct a distribution function $F_m(x)$ which coincides with $F(x)$ outside ΣT_i , is such that $F'_m(x) \neq 0$, and otherwise satisfies the conditions of the theorem (stated explicitly in Theorem 3). The sequence $F_m(x)$ converges to $F(x)$. Furthermore

$$\begin{aligned}
 & |\Omega_{mn} - \Omega_n| \\
 &= \left| \frac{1}{2} \sum_{i=1}^{n+1} \left| F(x_i) - F(x_{i-1}) - \frac{1}{n+1} \right| - \frac{1}{2} \sum_{i=1}^{n+1} \left| F_m(x_i) - F_m(x_{i-1}) - \frac{1}{n+1} \right| \right| \\
 (4.14) \quad &\leq \frac{1}{2} \sum_{i=1}^{n+1} \left| \left| F(x_i) - F(x_{i-1}) - \frac{1}{n+1} \right| - \left| F_m(x_i) - F_m(x_{i-1}) - \frac{1}{n+1} \right| \right| \\
 &\leq \frac{1}{2} \sum_{i=1}^{n+1} | [F(x_i) - F(x_{i-1})] - [F_m(x_i) - F_m(x_{i-1})] |.
 \end{aligned}$$

For any particular set of values of x_1, x_2, \dots, x_n some (possibly none or possibly all) of the x_i will fall into intervals of the ΣT_i . If this finite set of intervals, each containing at least one x_i , is say T_1, T_2, \dots, T_k , then a simple analysis of the sum on the right of (4.14) shows that it is less than twice the total length of the intervals $F(T_1), F(T_2), \dots, F(T_k)$ and this total length is less than $\frac{1}{m}$.

Thus $|\Omega_{mn} - \Omega_n| < \frac{1}{m}$ and this result is independent of n .

REFERENCES

- [1] H. CRAMÉR, *Mathematical Methods of Statistics*, Princeton University Press, 1946.
- [2] B. F. KIMBALL, "Some basic theorems for developing tests of fit for the case of the non-parametric probability distribution function. I," *Annals of Math. Stat.*, Vol. 18 (1947), pp. 540-548.
- [3] P. A. P. MORAN, "The random division of an interval," *Jour. Royal Stat. Soc., Supp.*, Vol. 9 (1947), pp. 92-98.
- [4] N. SMIRNOFF, "Sur la distribution de ω^2 ," *Compte Rendus de l'Academie des Sciences*, Paris, 202 (1932), p. 449.
- [5] A. WALD AND J. WOLFOWITZ, "On a test of whether two samples are from the same population," *Annals of Math. Stat.*, Vol. 11 (1940), pp. 147-162.
- [6] S. S. WILKS, *Mathematical Statistics*, Princeton University Press, 1943.