ON A CLASS OF PROBLEMS RELATED TO THE RANDOM DIVISION OF AN INTERVAL

By D. A. Darling

Columbia University

Summary. Let $X_1, X_2, \cdots, X_n$ be n independent random variables each distributed uniformly over the interval (0, 1), and let $Y_0, Y_1, \cdots, Y_n$ be the respective lengths of the $n + 1$ segments into which the unit interval is divided by the $\{X_i\}$. A fairly wide class of statistical problems is related to finding the distribution of certain functions of the Y_j; these problems are reviewed in Section 1. The principal result of this paper is the development of a contour integral for the characteristic function (ch. fn.) of the random variable $W_n = \sum_{j=0}^{n} h_j(Y_j)$ for quite arbitrary functions $h_j(x)$, this result being essentially an extension of the classical integrals of Dirichlet. The cases of statistical interest correspond to $h_j(x) = h(x)$, independent of j. There is a fairly extensive literature devoted to studying the distributions for various functions $h(x)$. By applying our method these distributions and others are readily obtained, in a closed form in some instances, and generally in an asymptotic form by applying a steepest descent method to the contour integral.

1. Introduction. The statistical problems mentioned above are divided roughly into two classes: problems related to considerations of the Poisson stochastic process occurring in the study of infectious diseases, traffic flow, etc., and problems pertaining to certain nonparametric tests of the hypothesis that a given set of data came from a hypothetical cumulative distribution function (cdf) $F(x)$, which in turn are related to certain "goodness of fit" problems.

In 1946 Greenwood [6], in connection with a problem in epidemiology, posed the general problem of testing whether a given set of points on the unit interval could have arisen from the independent selection of points X_i described above, or whether the set of intervals Y_j they generate are too nearly equal for this hypothesis to be tenable. He suggested the statistic $W_n = \sum_{j=0}^{n} Y_j^2$ and gave a few properties of its distribution. Later Moran [11] proved that W_n had a limiting Gaussian distribution for $n \to \infty$.

If $U_0, U_1, \cdots, U_n$ are $n + 1$ independent random variables each having the density $\beta e^{-\beta x}$, $\beta > 0$, $x > 0$, and if $s_n = U_0 +, \cdots, + U_n$ it is a well known fact that the joint distribution of $\{U_j/s_n\}$, $j = 0, 1, \cdots, n$ is the same as the joint distribution of $\{Y_j\}$, $j = 0, 1, \cdots, n$, the successive lengths of the intervals into which the unit interval is divided by n random points. This correspondence has been used in studying the Poisson stochastic process (cf. [3] chap. 17) in which the interval between successive occurrences of the phenomenon are the U_j. In Greenwood's example these phenomena were the outbreaks of infectious disease.

Received 6/24/52.

In these examples the statistical problems can be reduced to evaluating the distribution of $W_n = \sum h(Y_j)$. In place of Greenwood's suggestion of $h(x) = x^2$, other suggestions were made (cf. the discussion of [6]). Kendall suggested that $h(x) = |x - 1/(n+1)|$ might be analytically more tractable and Irwin suggested $h(x) = (n+1)^{-1}(x - 1/(n+1))^2$. For an analysis of the distribution properties of the extreme Y_j (or U_j) it suffices to consider an $h(x)$ which is 1 for $\alpha < x < \beta$ and zero otherwise. A variety of problems can be reduced to determining the distribution of $W_n = \sum h(Y_j)$ for $h(x)$ of this form. Garwood [5] studied some extremal properties of the Y_j in connection with the occurrence of traffic vehicles on a highway. Fisher [4] had made a similar use in 1925 on the distribution of an extreme amplitude in a problem in harmonic analysis. Kendall made the suggestion of studying the difference (or quotient) of the largest and smallest Y_j as being a more sensitive test function for the equality of the Y_j than Greenwood's sum of squares.

Let $X_1, X_2, \cdots, X_n$ be independent identically distributed random variables with the common continuous cdf $F(x)$. Let them be relabeled so that $X_1' < X_2' <, \cdots, < X_n'$ and put $X_0' = -\infty$, $X_{n+1}' = +\infty$. Then, as is well known, the joint distribution of $\{F(X_{j+1}') - F(X_j')\}$, $j = 0, 1, \cdots, n$ is the same as the joint distribution of the $\{Y_j\}$, $j = 0, 1, \cdots, n$. Given a set of n data $x_1, x_2, \cdots, x_n$ arranged in increasing order (with $x_0 = -\infty$, $x_{n+1} = +\infty$) a possible test of the hypothesis H that they came from a population whose cdf is $F(x)$ consists in choosing a function $h(x)$ and rejecting H if $\sum h(F(x_{j+1}) - F(x_j))$ is sufficiently large or sufficiently small. Thus the basic problem is, as before, calculating the distribution of $W_n = \sum h(Y_j)$ for various functions h.

Kimball [7] suggested $h(x) = x^\alpha$, $\alpha > 0$, and gave some partial results for the case $\alpha = 2$. The asymptotic character of W_n for $\alpha = 2$ was later analyzed by Moran [11] who proved W_n has a limiting normal distribution for $n \to \infty$. Sherman [13] treated the case $h(x) = \frac{1}{2}|x - 1/(n+1)|$. It will be noted that these tests are somewhat related to the Kolmogoroff-Smirnov tests (cf. [1]) of the "goodness of fit" criteria. A discussion of the relative merits of these tests seems quite academic in view of the complete lack of information concerning their power.

In the present paper we give a unified treatment of these distributions. In Section 2 we develop a simple formula for the ch. fn. of the random variable $W_n = \sum h_j(Y_j)$ (Theorem 2.1) which is essentially an extension of the Dirichlet integral (Theorem 2.2). In Section 3 we study the joint distribution of the Y_j, finding the joint ch. fn. (Theorem 3.1) and the distribution of $Y_0 + Y_1 +, \cdots, + Y_\nu$. In Section 4 we put $W_n = \sum h(Y_j)$ and develop a few moments of W_n useful in the subsequent work, and in Section 5 we give the asymptotic distribution of W_n for $h(x) = x^\alpha$, the statistic of Greenwood, Moran and Kimball (Theorem 5.1). In Section 6 we analyze the distribution of Sherman and in Section 7 present two more possible test functions which yield readily to our methods.

In Section 8 we study the random variable $N_n(\alpha, \beta)$, the number of those Y_j satisfying $\alpha < Y_j < \beta$, $j = 0, 1, \cdots, n$. As special cases we obtain the limiting distributions of the number of intervals of "average" size, "small" size and "large" size (Theorems 8.1, 8.2 and 8.3, respectively) and the joint distribution of the largest and smallest Y_j for finite n (Theorem 8.4).

2. The fundamental formula. Let $Y_0, Y_1, \cdots, Y_n$ be the lengths of the $n + 1$ intervals into which the unit interval is divided by n random points. The following theorem is the basis for the subsequent analysis in this paper.

THEOREM 2.1. *Let $f_0(x), f_1(x), \cdots, f_n(x)$ be $n + 1$ real-valued functions for which the abscissas of convergence of the corresponding Laplace transforms are all less than c. Then*

$$(2.1)\quad E(f_0(Y_0)f_1(Y_1) \cdots f_n(Y_n)) = \frac{n!}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^z \prod_{j=0}^{n} \int_0^\infty e^{-r_j z} f_j(r_j)\, dr_j\, dz$$

the path of integration being the straight line Re $z = c$ (where Re z denotes the real part of z.)

PROOF. We have

$$(2.2)\quad E\left(\prod_{j=0}^{n} f_j(Y_j)\right) = n!\int_0^1\int_0^{x_n}\int_0^{x_{n-1}} \cdots \int_0^{x_3}\int_0^{x_2} f_0(x_1)f_1(x_2 - x_1) \cdots f_{n-1}(x_n - x_{n-1})f_n(1 - x_n)\, dx_1\, dx_2 \cdots dx_n$$

since the joint distribution of the n random points, when arranged in order, has a uniform density differential $n!\, dx_1\, dx_2 \cdots dx_n$ over the simplex $0 \leqq x_1 \leqq x_2 \leqq \cdots \leqq x_n \leqq 1$. The trick in "evaluating" this integral consists in considering the following function

$$F(r) = \int_0^r\int_0^{x_n}\int_0^{x_{n-1}} \cdots \int_0^{x_3}\int_0^{x_2} f_0(x_1)f_1(x_2 - x_1) \cdots f_n(x_n - x_{n-1})f_n(r - x_n)\, dx_1\, dx_2 \cdots dx_n$$

which we want to evaluate at $r = 1$. But it is clear that written this way $F(r)$ is merely the convolution $f_0 * f_1 * \cdots * f_n(r)$ where $g(x)*h(x) = \int_0^x g(x - t)h(t)\, dt$.

Since Laplace transforms multiply under convolution we obtain

$$\int_0^\infty F(r)e^{-zr}\, dr = \prod_{j=0}^{n} \int_0^\infty e^{-zr_j} f_j(r_j)\, dr_j$$

provided *Re* $z > c$. We now simply apply the complex inversion for the Laplace transform to obtain

$$F(x) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^{xz} \prod_{j=0}^{n} \int_0^\infty e^{-r_j z} f_j(r_j)\, dr_j\, dz,$$

and the theorem follows if we put $x = 1$ and supply the factor $n!$.

It is interesting to note that in (2.1) the value of the integral apparently depends on the value of the $f_j(r)$ for $r > 1$ while in (2.2) it does not. As a matter of fact the functions may be defined quite arbitrarily for $r > 1$ and not affect the value of (2.1).

THEOREM 2.2 *Let D be the domain in E_n defined by $t_i \geqq 0$, $\sum_{i=1}^{n} t_i \leqq 1$. Then for the $f_i(x)$ as in Theorem* 2.1

$$\iint_D \cdots \int f_0(t_1)f_1(t_2) \cdots f_{n-1}(t_n)f_n(1 - t_1 - t_2 - \cdots - t_n)\, dt_1\, dt_2 \cdots dt_n$$

$$= \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} e^z \prod_{j=0}^{n} \int_0^\infty e^{-r_j z} f_j(r_j)\, dr_j\, dz.$$

To prove the theorem we merely make the change of variables in (2.2) $t_1 = x_1$ $t_2 = x_2 - x_1, \cdots, t_n = x_n - x_{n-1}$ for which the Jacobian is 1.

Theorem 2.2 is, in a sense, a generalization of the integral of Dirichlet—that is, putting $f_0(x) = x^{\alpha_1 - 1}$, $f_1(x) = x^{\alpha_2 - 1}, \cdots, f_{n-1}(x) = x^{\alpha_n - 1}$, $\alpha_i > 0$, and $f_n(x) = f(x)$ we obtain

$$\iint_D \cdots \int t_1^{\alpha_1 - 1} t_2^{\alpha_2 - 1} \cdots t_n^{\alpha_n - 1} f(1 - \textstyle\sum t_j)\, dt_1\, dt_2 \cdots dt_n$$

$$\begin{aligned} &= \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} e^z \prod_{j=1}^{n} \int_0^\infty e^{-r_j z} r_j^{\alpha_j - 1}\, dr_j \int_0^\infty e^{-rz} f(r)\, dr\, dz \\ (2.3) \qquad &= \prod_{j=1}^{n} \Gamma(\alpha_j) \int_0^\infty f(r) \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} e^{z(1-r)} z^{-\Sigma\alpha_j}\, dz\, dr \\ &= \frac{\Pi\Gamma(\alpha_j)}{\Gamma(\Sigma\alpha_j)} \int_0^1 (1 - r)^{\Sigma\alpha_j - 1} f(r)\, dr \end{aligned}$$

since the inner complex integral in (2.3) is zero if $r > 1$ and is $(1 - r)^{\Sigma\alpha_j - 1}/\Gamma(\sum \alpha_j)$ if $0 \leqq r < 1$. This is the classical Dirichlet integral usually developed through the theory of the Beta functions, (cf. Whittaker and Watson [17]).

3. The joint distribution of the $\{Y_j\}$. By means of Theorem 2.1 we can give certain properties of the joint distribution function of the Y_j, $j = 0, 1, \cdots, n$. For the ch. fn. of the Y_j we have the following theorem.

THEOREM 3.1. *If $t_i \neq t_j$ for $i \neq j$, and if $n \geqq 1$, then,*

$$E(e^{i(t_0 Y_0 + t_1 Y_1 + \cdots + t_n Y_n)}) = n! \sum_{j=0}^{n} \frac{e^{it_j}}{\prod_{k \neq j}(it_j - it_k)}$$

and is defined for other values of the t_i by continuity.

To prove the theorem we put $f_j(Y_j) = e^{it_j Y_j}$ in Theorem 2.1, giving

$$E(e^{i\Sigma t_j Y_j}) = \frac{n!}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{e^z\, dz}{\prod_{j=0}^{n}(z - it_j)}$$

and if all of the t_j are unequal the integral can be replaced by a contour integral surrounding the simple poles. A simple application of the theory of residues then establishes Theorem 3.1.

If some of the t_j are equal we proceed in the same manner. For instance if

$$t_i = \begin{cases} t, & i = 0, 1, \cdots, \nu - 1 \\ 0, & i = \nu, \nu + 1, \cdots, n \end{cases}$$

we obtain the ch. fn. of $x_\nu = \sum_{j=0}^{\nu-1} Y_j$, the νth smallest ordered observation from n observations taken from a rectangular population. Then

$$E(e^{itx_\nu}) = \frac{n!}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{e^z \, dz}{z^{n-\nu+1}(z - it)^\nu}$$

which again can be evaluated by residues, albeit somewhat awkwardly since the poles are no longer simple. But the density for x_ν is simple to calculate by considering

$$\frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{e^{xz} \, dz}{z^{n-\nu+1}(z - it)^\nu}$$

as the inversion of the product of two Laplace transforms

$$\frac{1}{\Gamma(n - \nu + 1)} \int_0^\infty e^{-sz} s^{n-\nu} \, ds = \frac{1}{z^{n-\nu+1}},$$

$$\frac{1}{\Gamma(\nu)} \int_0^\infty e^{-sz} e^{ist} s^{\nu-1} \, ds = \frac{1}{(z - it)^\nu}.$$

Consequently taking the convolution and putting $x = 1$

$$E(e^{itx_\nu}) = \frac{n!}{\Gamma(n - \nu + 1)\Gamma(\nu)} \int_0^1 e^{ist} s^{\nu-1}(1 - s)^{n-\nu} \, ds$$

and thus the density of x_ν is the Beta function $n!s^{\nu-1}(1 - s)^{n-\nu}/\Gamma(n - \nu + 1)\Gamma(\nu)$ as is well known. Other properties also related to the distribution of order statistics from a uniform distribution which have been proved recently by Malmquist [10] may be treated in a like manner.

An evaluation of the mixed moment $E(\prod_j Y_j^{\alpha_j})$ is, of course, easily given in terms of the Dirichlet integral of the preceding section.

4. The distribution of W_n. The statistical problems mentioned in Section 1 may all be reduced to finding the distribution of $W_n = \sum_{j=0}^{n} h(Y_j)$ for certain functions $h(x)$.

By putting $f_j(x) = e^{i\xi h(x)}$ in (2.1) we obtain

$$E(e^{i\xi W_n}) = \frac{n!}{2\pi i} \int_{c-i\infty}^{c+i\infty} e^W \left(\int_0^\infty e^{-rW + i\xi h(r)} \, dr \right)^{n+1} dW, \tag{4.1}$$

and from this expression we propose to study the distribution of $W_n = \sum h(Y_j)$.

As a preliminary we find the first two moments of W_n which will prove useful in the work to follow. If $\int_0^\infty h^k(r)\,dr$ is finite it is simple to see the (4.1) can be differentiated k times under the integral sign with respect to $i\xi$. Differentiating once and putting $\xi = 0$ we obtain

$$\mu_1 = E(W_n) = \frac{n!}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^W W^{-n}\,(n+1)\int_0^\infty e^{-rW}\,h(r)\,dr\,dW$$

$$(4.2) \qquad = (n+1)!\int_0^\infty h(r)\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^{W(1-r)}W^{-n}\,dW\,dr$$

$$= n(n+1)\int_0^1 (1-r)^{n-1}h(r)\,dr.$$

Similarly by differentiating twice and setting $\xi = 0$ we obtain the second moment

$$\mu_2 = E(W_n^2) = \frac{n!}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^W\left\{\frac{n+1}{W^n}\int_0^\infty e^{-rW}h^2(r)\,dr + \frac{n(n+1}{W^{n-1}}\left(\int_0^\infty e^{-rW}h(r)\,dr\right)^2\right\}dW$$

$$(4.3) \qquad = n(n+1)\int_0^1 (1-r)^{n-1}h^2(r)\,dr + n(n+1)!\int_0^\infty\int_0^\infty h(r_1)h(r_2)\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^{W(1-r_1-r_2)}\frac{dW}{W^{n-1}}\,dr_1\,dr_2$$

$$= n(n+1)\int_0^1 (1-r)^{n-1}h^2(r)\,dr + n^2(n^2-1)\iint_{\substack{0\leqq r_1+r_2\leqq 1\\ r_1\geqq 0,\, r_2\geqq 0.}} \cdot(1-r_1-r_2)^{n-2}h(r_1)h(r_2)\,dr_1\,dr_2.$$

From (4.2) and (4.3) we can calculate the variance $\sigma^2 = \mu_2 - \mu_1^2$, and proceeding in a similar fashion we can develop all moments if they exist.

5. The distributions of Greenwood, Moran and Kimball. Greenwood [6] suggested $h(x) = x^2$, and Irwin in the discussion of his paper suggested $h(x) = (n+1)^{-1}(x - 1/(n+1)^2$. Moran [11] later found the limiting distribution of Greenwood's statistic was normal. Kimball [7] proposed $h(x) = x^\alpha$ for $\alpha > 0$ and found some partial results for case $\alpha = 2$.

In this section we find the limiting distribution for the case $h(x) = x^\alpha$. We have the following theorem.

THEOREM 5.1. *The random variable* $W_n = \sum Y_j^\alpha$, $\alpha > 0$, $\alpha \neq 1$, *has a limiting normal distribution with the limiting mean and variance*

$$\mu_n \sim \frac{\Gamma(\alpha + 1)}{n^{\alpha-1}},$$

$$\sigma_n^2 \sim \frac{1}{n^{2\alpha-1}} \{\Gamma(2\alpha + 1) - (\alpha^2 + 1)\Gamma^2(\alpha + 1)\},$$

respectively, that is,

$$\lim_{n\to\infty} Pr\left\{\frac{W_n - \mu_n}{\sigma_n} < x\right\} = \frac{1}{\sqrt{2\pi}} \int_{\infty}^{x} e^{-t^2/2}\, dt$$

for μ_n *and* σ_n^2 *as above.*

Of course if $\alpha = 0$ or $\alpha = 1$ we have $\sigma_n = 0$ and it might be proper to speak of W_n as having a *degenerate* normal distribution.

This theorem will follow by applying a slight variation of the method of steepest descent to the integral (4.1). The proof is given in a fair amount of detail and will serve as a model for the later distributions whose treatment follows essentially the same pattern and for which we give considerably less detailed proofs.

Substituting $h(x) = x^\alpha$ in (4.2) and (4.3) we obtain for the first two moments

$$\mu_1 = \Gamma(\alpha + 1) \frac{\Gamma(n + 2)}{\Gamma(n + \alpha + 1)},$$

$$\mu_2 = \frac{\Gamma(n + 2)}{\Gamma(n + 2\alpha + 1)} (\Gamma(2\alpha + 1) - n\Gamma^2(\alpha + 1)).$$

The asymptotic character of these moments is easily obtained through the formula

$$\frac{\Gamma(n)}{\Gamma(n + \beta)} = \frac{1}{n^\beta} - \frac{\beta(\beta - 1)}{2n^{\beta+1}} + o\left(\frac{1}{n^{\beta+1}}\right) \quad \beta \geqq 0, \qquad n \to \infty$$

giving

$$\mu_1 = \frac{\Gamma(\alpha + 1)}{(n + 2)^{\alpha-1}} - \Gamma(\alpha + 1) \frac{(\alpha - 1)(\alpha - 2)}{2(n + 2)^\alpha} + o\left(\frac{1}{n^\alpha}\right)$$

$$\mu_2 = \frac{\Gamma^2(\alpha + 1)}{(n + 2)^{2\alpha-2}} + \frac{1}{(n + 2)^{2\alpha-1}} (\Gamma(2\alpha + 1) - \Gamma^2(\alpha + 1)(2\alpha^2 - 3\alpha + 3)) + o\left(\frac{1}{n^{2\alpha-1}}\right)$$

from which we deduce

$$\mu_n \sim \frac{\Gamma(\alpha + 1)}{n^{\alpha-1}} \tag{5.1}$$

$$\sigma_n^2 \sim \frac{1}{n^{2\alpha-1}} (\Gamma(2\alpha + 1) - (\alpha^2 + 1)\Gamma^2(\alpha + 1)).$$

Thus for $\alpha > \frac{1}{2}$, $\sigma_n^2 \to 0$ and for $\alpha < \frac{1}{2}$, $\sigma_n^2 \to \infty$ while in the transitional case $\alpha = \frac{1}{2}$ we have $\sigma_n^2 \to 1 - 5\pi/16$.

Using (4.1) we obtain for the ch. fn. of $W_n = \sum Y_j^\alpha$

$$\varphi_n(\xi) = E(\exp(i\xi W_n)) = \frac{n!}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^W \left(\int_0^\infty e^{-rW+i\xi r^\alpha}\,dr\right)^{n+1} dW$$

for $c > 0$. Letting $\xi = (n + 1)^{\alpha-\frac{1}{2}}t$, $W = (n + 1)z$ and shifting the contour parallel with itself we find

$$\varphi_n((n + 1)^{\alpha-\frac{1}{2}}t) = \frac{(n + 1)!}{(n + 1)^{n+1}}\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^{(n+1)z}z^{-n-1}(B_n(z, t))^{n+1}\,dz \tag{5.2}$$

where

$$B_n(z, t) = (n + 1)z\int_0^\infty e^{-r(n+1)z+it(n+1)^{\alpha-\frac{1}{2}}r^\alpha}\,dr. \tag{5.3}$$

Now it will turn out that $(B_n(z, t))^{n+1}$ is, aside from a multiplied factor depending on n and t but independent of z, actually a bounded function approaching a limit as $n \to \infty$ for $|t|$ bounded and z arbitrary. This suggests that relative to the dominant term $e^{(n+1)z}z^{-n-1}$ this factor will cause negligible interference when $n \to \infty$, (cf. Szegö [14], p. 220 who treats an example very similar to this.)

If we write $e^{(n+1)z}z^{-n-1} = e^{(n+1)f(z)}$ then $f(z) = z - \log z$ where $\log z$ is real when z is real and positive. Then since $f'(1) = 0$, $f''(1) = 1$, the saddle point is $z = 1$ with the critical direction parallel to the imaginary axis. Hence in (5.2) we merely take $c = 1$ to get the contour of steepest descent.

Thus we put

$$z = 1 + \frac{iy}{\sqrt{n + 1}}, \qquad dz = \frac{idy}{\sqrt{n + 1}} \tag{5.4}$$

for y in the domain

$$-(n + 1)^\delta < y < (n + 1)^\delta, \qquad 0 < \delta < \tfrac{1}{2} \tag{5.5}$$

and the entire integral has its essential contribution in this range—the value of the integral extended over the range complementary to (5.5) becoming negligible as $n \to \infty$ after we have modified $B_n(z, t)$ by a factor independent of z. With the substitution (5.4) we find

$$\begin{aligned}\varphi_n((n + 1)^{\alpha-\frac{1}{2}}t) &= \frac{(n + 1)!e^{n+1}}{(n + 1)^{n+3/2}2\pi}\int_{-(n+1)^\delta}^{(n+1)^\delta} e^{-y^2/2}(B_n(z, t))^{n+1}\,dy(1 + o(1)) \\ &= \frac{1}{\sqrt{2\pi}}\int_{-(n+1)^\delta}^{(n+1)^\delta} e^{-y^2/2}(B_n(z, t))^{n+1}\,dy(1 + o(1))\end{aligned} \tag{5.6}$$

by using Stirling's formula for $(n + 1)!$.

Next we turn to $B_n(z, t)$ as given by (5.3). By standard methods we obtain an asymptotic expansion (cf. Watson [16])

$$B_n(z, t) = (n+1)z \int_0^\infty e^{-r(n+1)z}$$

$$\{1 + it(n+1)^{\alpha-\frac{1}{2}} r^\alpha + \tfrac{1}{2}(it)^2 (n+t)^{2\alpha-1} r^{2\alpha} + \cdots\}\, dr$$

$$= 1 + \frac{it\Gamma(\alpha+1)}{(n+1)^{\frac{1}{2}} z^\alpha} + \tfrac{1}{2}(it)^2 \frac{\Gamma(2\alpha+1)}{(n+1)z^{2\alpha}} + o(1/n)$$

and for z as in (5.4) and y in the range (5.5)

$$\frac{1}{z^\alpha} = \left(1 + \frac{iy}{(n+1)^{\frac{1}{2}}}\right)^{-\alpha} = 1 - \frac{i\alpha y}{(n+1)^{\frac{1}{2}}} + o(1/n^{\frac{1}{2}})$$

$$\frac{1}{z^{2\alpha}} = \left(1 + \frac{iy}{(n+1)^{\frac{1}{2}}}\right)^{-2\alpha} = 1 + o(1)$$

so that

$$(n+1) \log B_n(z, t)$$

$$= (n+1) \log\left\{1 + \frac{it\Gamma(\alpha+1)}{(n+1)^{\frac{1}{2}} z^\alpha} + \tfrac{1}{2}(it)^2 \frac{\Gamma(2\alpha+1)}{(n+1)z^{2\alpha}} + o(1/n)\right\}$$

$$= it\Gamma(\alpha+1) \frac{(n+1)^{\frac{1}{2}}}{z^\alpha} + \tfrac{1}{2}(it)^2\{\Gamma(2\alpha+1) - \Gamma^2(\alpha+1)\} \frac{1}{z^{2\alpha}} + o(1)$$

$$= (n+1)^{\frac{1}{2}} it\Gamma(\alpha+1) - ty\alpha\Gamma(\alpha+1) + \tfrac{1}{2}(it)^2 (\Gamma(2\alpha+1) - \Gamma^2(\alpha+1)) + o(1).$$

Using this estimate in (5.6) we obtain

$$\varphi_n((n+1)^{\alpha-\frac{1}{2}} t) \exp(-(n+1)^{\frac{1}{2}} it\Gamma(\alpha+1)) = \exp\{-\tfrac{1}{2}t^2(\Gamma(2\alpha+1) - \Gamma(\alpha+1))\}$$

$$\cdot \frac{1}{\sqrt{2\pi}} \int_{-(n+1)^\delta}^{(n+1)^\delta} e^{-y^2/2 - ty\alpha\Gamma(\alpha+1)}\, dy(1 + o(1))$$

and hence

$$\lim_{n\to\infty} E[\exp(it((n+1)^{\alpha-\frac{1}{2}} W_n - (n+1)^{\frac{1}{2}}\Gamma(\alpha+1))]$$

$$= \lim_{n\to\infty} E\left(\exp\left(it \frac{W_n - (n+1)^{-\alpha+1}\Gamma(\alpha+1)}{(n+1)^{-\alpha+\frac{1}{2}}}\right)\right)$$

$$= \exp\{-\tfrac{1}{2}t^2(\Gamma(2\alpha+1) - \Gamma^2(\alpha+1))\} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2 - ty\alpha\Gamma(\alpha+1)}\, dy$$

$$= \exp\{-\tfrac{1}{2}t^2(\Gamma(2\alpha+1) - (\alpha^2+1)\Gamma^2(\alpha+1))\}$$

which establishes the theorem, and gives an independent derivation for the asymptotic moments.

6. The distribution of Sherman. To avoid some of the difficulties pertaining to the case $h(x) = x^\alpha$ Sherman [13] considered the case $h(x) = \frac{1}{2} \mid x - 1/(n+1) \mid$.

Kendall ([6], discussion) had suggested that such a function might be easier to treat because of the simplification of the geometry of the integration. Sherman gave the distribution of $W_n = \frac{1}{2}\sum | Y_j - 1/(n+1) |$ and proved it had a limiting Gaussian distribution.

In this section we develop the distribution of W_n using (4.1). Here the inner integration can be performed explicitly and the analysis is much simpler. We have in fact

$$\int_0^\infty e^{-rz+\frac{1}{2}i\xi|r-1/(n+1)|}\,dr = \frac{e^{i\xi/2(n+1)} - e^{-z/(n+1)}}{z+\frac{1}{2}i\xi} + \frac{e^{-z/(n+1)}}{z-\frac{1}{2}i\xi} \tag{6.1}$$

so that using (4.1)

$$\begin{aligned}\varphi_n(\xi) &= \frac{n!}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^z\left\{\frac{e^{i\xi/2(n+1)} - e^{-z/(n+1)}}{z+\frac{1}{2}i\xi} + \frac{e^{-z/(n+1)}}{z-\frac{1}{2}i\xi}\right\}^{n+1} dz \\ &= \frac{n!}{2\pi i}\int_{c-i\infty}^{c+i\infty}\left\{\frac{1}{z-\frac{1}{2}i\xi} + \frac{e^{(n+1)^{-1}(z+\frac{1}{2}i\xi)}-1}{z+\frac{1}{2}i\xi}\right\}^{n+1} dz \\ &= \frac{n!}{2\pi i}\int_{c-i\infty}^{c+i\infty}\sum_{j=0}^{n+1}\binom{n+1}{j}\left(\frac{e^{(n+1)^{-1}(z+\frac{1}{2}i\xi)}-1}{z+\frac{1}{2}i\xi}\right)^j \frac{dz}{(z-\frac{1}{2}i\xi)^{n+1-j}} \\ &= n!\sum_{j=1}^{n}\binom{n+1}{j}\frac{1}{(n-j)!}\frac{d^{n-j}}{d(i\xi)^{n-j}}\left(\frac{e^{i\xi/(n+1)}-1}{i\xi}\right)^j\end{aligned}$$

by a simple application of the theory of residues. From this ch. fn. we can easily deduce the density for W_n.

We rewrite the preceding expression

$$\varphi_n(\xi) = n!\sum_{j=1}^{n}\binom{n+1}{j}\frac{1}{(n-j)!(n+1)^j}\frac{d^{n-j}}{d(i\xi)^{n-j}}\left(\frac{e^{i\xi/(n+1)}-1}{i\xi/(n+1)}\right)^j$$

and invert termwise. Let $X_1, X_2, \cdots$ be independent and uniformly distributed over (0, 1). The density of $X_1 + X_2 + \cdots + X_j$ is then (cf. Cramér [2], p. 245)

$$f_j(x) = \frac{1}{(j-1)!}\sum_{0\le k<x}(-1)^k\binom{j}{k}(x-k)^{j-1} \qquad 0 < x < j. \tag{6.2}$$

Then the density for $1/(n+1)(X_1 + X_2 + \cdots + X_j)$ is $(n+1)f_j((n+1)x)$ and the ch. fn. for it is

$$\left(\frac{e^{i\xi/(n+1)}-1}{i\xi/(n+1)}\right)^j.$$

Hence

$$\frac{d^{n-j}}{d(i\xi)^{n-j}}\left(\frac{e^{i\xi/(n+1)}-1}{i\xi/(n+1)}\right)^j = \int_{-\infty}^{\infty} e^{i\xi x}x^{n-j}(n+1)f_j((n+1)x)\,dx.$$

Having inverted the typical term in $\varphi_n(\xi)$ we obtain for the density of W_n

$$n!\sum_{j=1}^{n}\binom{n+1}{j}\frac{x^{n-j}f_j((n+1)x)}{(n-j)!(n+1)^{j-1}}$$

with $f_j(x)$ as in (6.2).

It is also simple to get an asymptotic distribution for W_n following the pattern of Section 5 exactly. If we put $z = (n + 1) + (n + 1)^{\frac{1}{2}}iy$, $\xi \doteq (n + 1)^{\frac{1}{2}}t$ in (6.1) we obtain after some easy estimates

$$\int_0^\infty e^{-rz+(i\xi/2)|r-1/(n+1)|}\,dr$$

$$= \frac{1}{z}\left\{1 + \frac{ite^{-1}}{(n+1)^{\frac{1}{2}}} + \frac{ty}{n+1}(2e^{-1} - \tfrac{1}{2}) - \frac{t^2}{8(n+1)} + o(1/n)\right\},$$

and we choose the same contour as before with $c = 1$. These same substitutions yield

$$n!e^z\,dz/z^{n+1} = i\sqrt{2\pi}e^{-y^2/2}\,dy(1 + o(1))$$

as in the preceding example so that

$$\varphi_n((n + 1)^{\frac{1}{2}}t)e^{-it(n+1)^{\frac{1}{2}}e^{-1}} \to \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-y^2/2+ty(2e^{-1}-\frac{1}{2})-t^2/8}\,dy$$

$$E(e^{it(n+1)^{\frac{1}{2}}(W_n-e^{-1})}) \to e^{-t^2/2(2e^{-1}-4e^{-2})}$$

which exhibits the approach of W_n to the normal distribution.

7. Other possibilities. If we put $h(x) = \log x$ we can evaluate $\varphi_n(\xi)$ explicitly, obtaining,

$$\varphi_n(\xi) = \frac{n!}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^z\left(\int_0^\infty e^{-rz+i\xi\log r}\,dr\right)^{n+1} dz = \frac{\Gamma(n + 1)\Gamma^{n+1}(i\xi + 1)}{\Gamma((n + 1)(i\xi + 1))}.$$

Setting $\xi = (n + 1)^{-\frac{1}{2}}t$ and using Stirling's formula we get

$$\log \varphi_n((n + 1)^{-\frac{1}{2}}t) = -it(n + 1)^{\frac{1}{2}}(\log n + \gamma) - \tfrac{1}{2}t^2(\pi^2/6 - 1) + o(1)$$

and it follows that $\sum \log Y_j$ is asymptotically normally distributed with asymptotic mean and variance $-(n + 1)(\log n + \gamma)$ and $(n + 1)(\pi^2/6 - 1)$ respectively, γ being Euler's constant, $\gamma = .577 \cdots$.

In the preceding examples we have always obtained a limiting normal distribution and it seems a reasonable conjecture in analogy with the central limit theorem that we will generally obtain the asymptotic Gaussian distribution when the two moments (4.2) and (4.3) exist. But it appears very difficult to prove a theorem of this generality. We next give an example for which we do not obtain the normal distribution.

Let $h(x) = 1/x$; then since

$$\int_0^\infty e^{-rz-\xi/r}\,dr = \frac{1}{z}2\sqrt{\xi z}\,K_1(2\sqrt{\xi z}), \qquad \xi > 0, \quad Re\, z > 0$$

where $K_1(x)$ is the Bessel function, (cf. Watson [16]), we have for the Laplace transform of the density of $W_n = \sum Y_j^{-1}$

$$E(e^{-\xi W_n}) = \frac{n!}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^z z^{-n-1}\,(2\sqrt{\xi z}\,K_1(2\sqrt{\xi z}))^{n+1}\,dz.$$

Again letting $z = (n + 1) + iy(n + 1)^{\frac{1}{2}}$ and $\xi = t(n + 1)^{-1}$ we have $2\sqrt{\xi z}K_1(2\sqrt{\xi z}) \to 2\sqrt{t}K_1(2\sqrt{t})$ and this expression is the Laplace transform for the density whose cdf is $e^{-1/x}$, $0 < x < \infty$. It will follow then that $W_n/(n + 1)$ has the same limiting distribution as the sum of $(n + 1)$ independent random variables each having a cdf $e^{-1/x}$, and thus that this limiting distribution is a quasi-stable law of exponent 1 (cf. Lévy [8], p. 208).

8. The number of intervals satisfying certain inequalities. Let $N_n(\alpha, \beta)$ be the number of those Y_j which satisfy $\alpha < Y_j < \beta$ for $j = 0, 1, \cdots, n$. A number of statistical problems relate to the distribution of $N_n(\alpha, \beta)$ as we have outlined in Section 1.

If we put

$$h(r) = \begin{cases} 1 & \alpha < r < \beta \\ 0 & \text{otherwise} \end{cases}$$

then $N_n(\alpha, \beta) = \sum h(Y_j)$ and our preceding discussion is applicable in studying the distribution of this random variable.

Using (4.1) we have

$$\begin{aligned} E(e^{i\xi N_n(\alpha,\beta)}) &= \frac{n!}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^z\left\{\int_0^\infty e^{-rz+i\xi h(r)}\,dr\right\}^{n+1} \\ &= \frac{n!}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^z z^{-n-1}\{1 + (e^{i\xi} - 1)(e^{-z\alpha} - e^{-z\beta})\}^{n+1}\,dz \end{aligned} \tag{8.1}$$

and this expression will be a basis for the analysis of N_n.

"Most" of the Y_j are presumably of the order of magnitude $(n + 1)^{-1}$, and we first find the asymptotic distribution of the number of those Y_j which lie between $a/(n + 1)$ and $b/(n + 1)$.

THEOREM 8.1. *The random variable* $N_n(a/(n + 1), b/(n + 1))$ *is asymptotically normally distributed with an asymptotic mean and variance*

$$\mu_n \sim (n + 1)(e^{-a} - e^{-b})$$
$$\sigma_n^2 \sim (n + 1)(e^{-a} - e^{-b} - (ae^{-a} - be^{-b})^2).$$

The proof parallels the analysis of Section 5. Putting $z = (n + 1) + (n + 1)^{\frac{1}{2}}iy$, $\xi = (n + 1)^{-\frac{1}{2}}t$ in (8.1) we deduce easily

$$\begin{aligned} &1 + (e^{i\xi} - 1)(e^{-za/(n+1)} - e^{-zb/(n+1)}) \\ &= 1 + \frac{it}{(n + 1)^{\frac{1}{2}}}(e^{-a} - e^{-b}) - \frac{t^2(e^{-a} - e^{-b})}{2(n + 1)} + \frac{ty}{n + 1}(ae^{-a} - be^{-b}) + o(1/n) \end{aligned}$$

and thus

$$E\left(\exp\left\{i\frac{t}{(n + 1)^{\frac{1}{2}}}N_n(a/(n + 1), b/(n + 1)) - it(e^{-a} - e^{-b})(n + 1)^{\frac{1}{2}}\right\}\right)$$
$$\to e^{-t^2/2(e^{-a}-e^{-b})}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-y^2/2+ty(ae^{-a}-be^{-b})}\,dy$$

$$E\left(\exp\left\{it\,\frac{N_n(a/(n+1), b/(n+1)) - (n+1)(e^{-a} - e^{-b})}{(n+1)^{\frac{1}{2}}}\right\}\right)$$

$$\to e^{-t^2/2(e^{-a}-e^{-b}-(ae^{-a}-be^{-b})^2)},$$

which proves the theorem.

We next analyze the distribution of the number of "small" Y_j. It turns out that with probability 1 only finitely many are of the order of magnitude $(n+1)^{-2}$ as $n \to \infty$.

THEOREM 8.2. *$N_n(a/(n+1)^2, b/(n+1)^2)$ has an asymptotic Poisson distribution with parameter $(b - a)$. That is*

$$\lim_{n\to\infty} Pr\{N_n(a/(n+1)^2, b/(n+1)^2) = k\} = e^{-(b-a)}\frac{(b-a)^k}{k!}, \quad k = 0, 1, \cdots.$$

To prove the theorem we put $\alpha = a/(n+1)^2$, $\beta = b/(n+1)^2$ and $z = (n+1) + (n+1)^{\frac{1}{2}}iy$ in (8.1), giving

$$(1 + (e^{i\xi} - 1)(e^{-z\alpha} - e^{-z\gamma}))^{n+1} = e^{(b-a)(e^{i\xi}-1)}(1 + o(1)),$$

and we have, arguing as before,

$$E(e^{i\xi N_n(a/(n+1)^2, b/(n+1)^2)}) \to e^{(b-a)(e^{i\xi}-1)}$$

which establishes the theorem.

The distribution of the number of "large" Y_j proceeds in a similar way.

THEOREM 8.3. *$N_n((\log(n+1) + a)/(n+1), (\log(n+1) + b)/(n+1))$ has an asymptotic Poisson distribution with parameter $(e^{-a} - e^{-b})$;*

$$\lim_{n\to\infty} Pr\left\{N_n\left(\frac{\log(n+1) + a}{n+1}, \frac{\log(n+1) + b}{n+1}\right) = k\right\}$$

$$= e^{-(e^{-a}-e^{-b})}\frac{(e^{-a} - e^{-b})^k}{k!}, \qquad \kappa = 0, 1, \cdots.$$

Thus only finitely many intervals are as large as $\log n/n$ asymptotically with probability 1. To prove the theorem we put

$$\alpha = \frac{\log\dfrac{n+1}{b}}{n+1}, \qquad \beta = \frac{\log\dfrac{n+1}{a}}{n+1} \qquad a < b$$

in (8.1) and take $z = (n+1) + (n+1)^{\frac{1}{2}}iy$ giving

$$(1 + (e^{i\xi} - 1)(e^{-z\alpha} - e^{-zb}))^{n+1} \to e^{(b-a)(e^{i\xi}-1)}$$

and the rest of the proof proceeds as before.

From 8.2 and 8.3 we can find the asymptotic distribution of the largest Y_j and the smallest Y_j. Let, in fact, $U_n = \min(Y_0, Y_1, \cdots, Y_n)$ and $V_n =$

max $(Y_0, Y_1, \cdots, Y_n)$. Then putting $a = 0$, $k = 0$ in Theorem 8.2 and $b = \infty$, $k = 0$ in Theorem 8.3 we obtain

$$\lim_{n\to\infty} Pr\{U_n > b/(n+1)^2\} = e^{-b}, \qquad 0 < b < \infty$$

$$\lim_{n\to\infty} Pr\left\{V_n < \frac{\log (n+1) + a}{n+1}\right\} = e^{-e^{-a}}, \qquad -\infty < a < \infty.$$

These two expressions were given by Lévy [9] using geometrical arguments. It is possible to show that U_n and V_n are, besides, asymptotically independent. If we put $\alpha = a/(n+1)^2$ and $\beta = \log\frac{n+1}{b} \Big/ n+1$ in (8.1) and duplicate the above reasoning we get

$$\lim_{n\to\infty} Pr\left\{U_n > a/(n+1)^2, V_n < \frac{\log (n+1) - \log b}{n+1}\right\} = e^{-(a+b)}$$

However by taking a different attack we can get more precise information about the joint distribution of U_n and V_n.

THEOREM 8.4.

$$\begin{aligned} Pr\{U_n > \alpha, V_n < \beta\} &= Pr\{\alpha < Y_j < \beta, j = 0, 1, \cdots, n\} \\ &= \sum_{(j)}^{*} \binom{n+1}{j} (-1)^j (1 - \alpha(n+1-j) - \beta j)^n \end{aligned} \tag{8.2}$$

where $\sum^*$ *means to include only those terms for which* $1 - \alpha(n+1-j) - \beta j$ *is positive,* $j = 0, 1, \cdots$.

The required probability is clearly the probability that $N_n(\alpha, \beta)$ is equal to $(n+1)$. Hence in (8.1) if we expand the factor in braces and select the coefficient of $e^{i\xi(n+1)}$ we get

$$\begin{aligned} Pr\{N_n(\alpha, \beta) = n+1\} &= \frac{n!}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^z z^{-n-1} (e^{-z\alpha} - e^{-z\beta})^{n+1}\, dz \\ &= \sum_{j=0}^{n} \binom{n+1}{j} (-1)^j \frac{n!}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^{z(1-\alpha(n+1-j)-\beta j)} z^{-n-1}\, dz \end{aligned}$$

and this is equal to (8.2) by a direct application of the residue theorem.

Putting $\alpha = 0$ we obtain the probability that all intervals Y_j are less than β

$$Pr\{V_n < \beta\} = \sum_{0 \le j < 1/\beta} \binom{n+1}{j} (-1)^j (1 - \beta j)^n$$

a result going back to Whitworth [18] and used by Fisher [4] in studying the significance of the largest amplitude in harmonic analysis, and by Garwood [5] in traffic studies. Setting $\beta = 1$ in (8.2) we have only the term corresponding to $j = 0$ in the series, and the distribution of the minimum of the Y_j becomes

$$Pr\{U_n > \alpha\} = (1 - (n+1)\alpha)^n \qquad \alpha < 1/(n+1)$$

which is also a result of considerable age.

There are also interesting relationships between the distributions of U_n and V_n with the work of Robbins [12] and Votow [15] on the measure of a random set.

By using (8.2) it would be easy to find the distribution of $V_n - U_n$ or V_n/U_n and, as suggested by Kendall ([6], discussion), these might be better statistics to test for the equality of the Y_j than the statistics W_n discussed in Sections 5, 6, and 7 above.

REFERENCES

[1] T. W. Anderson and D. A. Darling, "Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 193–212.

[2] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 1946.

[3] W. Feller, *An Introduction to Probability Theory*, John Wiley and Sons, 1950.

[4] R. A. Fisher, "Tests of significance in harmonic analysis," *Proc. Roy. Soc. Edinburgh, Sect. A.*, Vol. 125 (1929), pp. 54–59.

[5] F. Garwood, "An application of the theory of probability to the operation of vehicular-controlled traffic signals," *J. Roy. Stat. Soc. Suppl.*, Vol. 7 (1940), pp. 65–77.

[6] M. Greenwood, "The statistical study of infectious diseases," *J. Roy. Stat. Soc.*, Vol. 109 (1946), pp. 85–110.

[7] B. F. Kimball, "Some basic theorems for developing tests of fit for the case of non-parametric probability distribution functions," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 540–548.

[8] P. Lévy, *Theorie de l'Addition des Variables Aleatoires*, Gauthier-Villars, Paris, 1937.

[9] Paul Lévy, "Sur la division d'un segment par des points choisis au hasard," *C. R. Acad. Sci. Paris*, Vol. 208 (1939), pp. 147–149.

[10] S. Malmquist, "On a property of order statistics from a rectangular distribution," *Skand. Aktuarietids.*, Vol. 33 (1950), pp. 214–222.

[11] P. Moran, "The random division of an interval," *J. Roy. Stat. Soc. Suppl.*, Vol. 9 (1947), pp. 92–98.

[12] H. E. Robbins, "On the expected value of two statistics," *Ann. Math. Stat.*, Vol. 15 (1944), pp. 321–323.

[13] B. Sherman, "A random variable related to the spacing of sample values," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 339–361.

[14] G. Szegö, "Orthogonal polynomials," *Amer. Math. Soc. Colloq. Publ.*, Vol. 23 (1939).

[15] D. F. Votaw, "The probability distribution of the measure of a random set," *Ann. Math. Stat.*, Vol. 17 (1946), pp. 240–244.

[16] G. N. Watson, *Theory of Bessel Functions*, Cambridge University Press, 1944.

[17] E. T. Whittaker and G. N. Watson, *Modern Analysis*, Cambridge University Press, 1927.

[18] W. A. Whitworth, *Choice and Chance*, Cambridge University Press, 1897.