ON DISTRIBUTIONS DETERMINED BY RANDOM VARIABLES DISTRIBUTED OVER THE $n$-CUBE

BY IAIN D. CURRIE

Heriot-Watt University

The distribution function of a random variable of the form $\sum_{i=1}^{n} a_i Y_i$ $\cdots$ $Y$, where $a_i > 0$ and $0 \leq Y_i \leq 1$, is considered. A geometric argument is used to obtain the distribution function as a repeated integral. The result is used first to obtain the distribution function of a linear combination of variables defined over the simplex $X_i \geq 0$, $\sum_{i=1}^{n} X_i \leq 1$. As a second application the distribution of certain quadratic forms over the simplex is obtained. This result yields a special case of the distribution of the internally studentized extreme deviate; the cases of normal and exponential samples are considered in detail and the required distributions obtained.

1. Introduction. Let $a_i > 0$, $i = 1, 2, \cdots, n$ be real numbers. We consider the distribution of the random variable

$$g(Y) = \sum_{i=1}^{n} a_i Y_i, \cdots, Y,$$

where $Y_1, Y_2, \cdots, Y_n$ are distributed over the unit $n$-cube, $0 \leq Y_i \leq 1$, $i = 1, 2, \cdots, n$. It is not assumed that $Y_1, Y_2, \cdots, Y_n$ are independent or identically distributed. We will obtain the distribution function of the random variable $g(Y)$ as a repeated integral; a statement of this theorem is given in Section 2 although we delay a proof until Section 4. Most of Section 2 is given over to an informal discussion of the geometry of random variables of the form (1.1) and an interpretation of the theorem and an outline of its proof are given in geometric terms.

The motivation for studying this random variable is as follows: we show in Section 3 that the solution to the distributional problem posed by (1.1) is equivalent to finding the distribution of a linear combination of variables defined over the simplex $X_i \geq 0$, $\sum_{i=1}^{n} X_i \leq 1$. This problem has as a special case the widely studied problem of obtaining the distribution of a linear combination of the order statistics of a sample from the uniform distribution, a problem approached from a geometric point of view by Hall (1927) and later by Dempster and Kleye (1968) and others. The Laplace transform provides a particularly effective way of tackling the problem of finding the distribution of a linear combination of order statistics from the uniform distribution, and recently Margolin (1977) used this method to obtain the distribution of a linear combination of Dirichlet variables. Closely related to this result is one first obtained by Fisher (1929), generalized first by Cochran (1941) and more recently by Lewis and Fieller (1979). Here we are concerned with the distribution of $X_{1n}/\sum_{i=1}^{n} X_i$, based on gamma samples. As well as linear forms, we may use the result on the distribution of a random variable of the form (1.1) to obtain the distribution of certain quadratic forms over the simplex. As a special case of this we consider the distribution of the extreme studentized deviate, a statistic commonly used in simple data monitoring; with an underlying normal population we have the problem studied by Pearson and Chandra Sekar (1936), Grubbs (1950) and Borenius (1959), while with an underlying exponential distribution we require the distribution of the statistic proposed by Shapiro and Wilk (1972) as a test of exponentiality.

Received August, 1979; revised August, 1980.


Key words and phrases. Geometric probability, order statistics, linear combinations, internally studentized deviate.
The paper concludes with a proof of the theorem and a few remarks on some of the geometric aspects of the results obtained in the main part of the paper.

2. The main theorem.

2.1 Geometrical Considerations. Random variables of the form (1.1) are in fact particularly convenient to deal with when the variables $Y_1$, $Y_2$, ..., $Y_n$ are defined over the $n$-cube. To illustrate the simplicity of the geometry involved consider an example with $n = 3$. Let $g(Y) = Y_1 + Y_2 + Y_3$, $0 \leq Y_i \leq 1$, $i = 1, 2, 3$ and suppose that $Y_1$, $Y_2$, $Y_3$ have a continuous joint probability density function $f(y)$. In order to compute $P(r(Y) \leq c)$ we must consider the surface $g(y_1, y_2, y_3) = c$. If $0 < c < 1$ this surface is shown in Figure 1 and the diagram tells us that

$$
(2.1) \quad P(r(Y_1, Y_2, Y_3) > c) = \int_0^1 \int_0^1 \int_{a_1(y_2, y_3)}^1 f(y_1, y_2, y_3) \, dy_1 \, dy_2 \, dy_3
$$

where

$$
(2.2) \quad a_1(y_2, y_3) = c/(1 + y_2 + y_3).
$$

The first critical value of $c$ is $c = 1$. Notice that $g(y) = 1$ along the edge $y_1 = 1$, $y_2 = 0$ and so the next case to consider is $1 < c < 2$ when the surface is shown in Figure 2.

This time we find

$$
(2.3) \quad P(r(Y_1, Y_2, Y_3) > c) = \int_0^1 \int_0^1 \int_{a_2(y_1, y_3)}^1 f(y_1, y_2, y_3) \, dy_1 \, dy_2 \, dy_3
$$

where

$$
(2.4) \quad a_2(y_3) = (c - 1)/(1 + y_3).
$$
The next critical value of $c$ is $c = 2$. The form of the function $g$ ensures that as $c$ increases through the value 2 the surface $g(y) = c$ changes form from that in Figure 2 straight to that in Figure 3.

The final form for the distribution function of $g(Y)$ is given for $2 < c < 3$ by
(2.5) \[ \Pr(g(Y_1, Y_2, Y_3) > c) = \int_{y_n}^1 \int_{y_{n-1}}^1 \int_{y_{n-2}}^1 f(y_1, y_2, y_3) \, dy_1 \, dy_2 \, dy_3 \]

where

(2.6) \[ \alpha_3 = c - 2. \]

This example illustrates a common feature of the distribution function of a random variable of the form \( g(Y) \), namely, that it has a different functional form in each of \( n \) intervals. Distributions of this type are not uncommon in statistics but often they are awkward to deal with. There are two features of the variable \( g(Y) \) that make it easy to study. The functional form of the distribution function is determined by the surface \( g(y) = c \); as \( c \) increases this form changes whenever the surface passes through a vertex of the cube. What is not so obvious is that the functional form should change only \( n \) times; this is a consequence of the 'nested' form of \( g(y) \) which ensures that the surface \( g(y) = c \) coincides with whole faces of the \( n \)-cube for certain values of \( c \). This dramatically reduces the number of cases to be considered. The other feature of the geometry of this problem worth noting is that the condition \( \alpha_i > 0 \) ensures that the surface \( g(y) = c \) consists of a single piece inside the cube; this simplifies the regions of integration involved and means that the upper limits of integration are always unity.

2.2 Statement of the Main Theorem. We will obtain the distribution function of the random variable \( g(Y) \) defined in (1.1) in terms of functions \( \alpha_i = \alpha_i(n; x_1, y_1, \ldots, y_i) \); these functions are natural generalizations of those defined in (2.2), (2.4) and (2.6) and enable the distribution function of \( g(Y) \) to be obtained as integrals of the form (2.1), (2.3) and (2.5).

Let \( \alpha_0 = 0 \) and define

(2.7) \[ A(i) = \alpha_i + \sum_{j=1}^{i} \alpha_j y_j y_{j+1} \cdots y_i, \quad i = 0, 1, \ldots, n \]

where we have adopted the convention that \( \sum_{j=1}^{0} \) with \( i > n \) is identically zero. Clearly we have \( A(0) = g(y) \) from (1.1), \( A(n) = a_n \) and

(2.8) \[ A(i) = \alpha_i + \sum_{j=1}^{i} \alpha_j A(i + 1), \quad i = 0, 1, \ldots, n - 1. \]

Thus the functions \( A(i) \) enable us to deal conveniently with the nested form of \( g(y) \).

Further, define

(2.9) \[ b_i = \sum_{j=0}^{i} \alpha_j, \quad i = 0, 1, 2, \ldots, n. \]

Then, if \( x \geq b_{i-1} \), define

(2.10) \[ \alpha_i = (x - b_{i-1})/A(i), \quad i = 1, 2, \ldots, n. \]

Notice that the functional form of \( \alpha_i \) factorises into a function of \( x \) and a function of \( y_1, y_2, \ldots, y_i \); this has important consequences when the evaluation of the integrals is performed in particular applications. Observe also that we have simplified notation by dropping the variables on which \( A(i) \) and \( \alpha_i \) depend. Finally, let

(2.11) \[ I_i(i) = [b_{i-1}, b_i], \quad i = 1, 2, \ldots, n. \]

The set of intervals \( \{I_i(i)\}_{i=1}^{n} \) covers the range of \( g(Y) \) and these are nonoverlapping (apart from the endpoints). We can now state our main result.

Theorem. If \( Y_1, Y_2, \ldots, Y_n \) have a continuous joint probability density function \( f(y) = f(y_1, y_2, \ldots, y_n) \) and if \( x \in I_i(k) \) then

(2.12) \[ \Pr(g(Y) \geq x) = \int_{y_n}^1 \int_{y_{n-1}}^1 \cdots \int_{y_1}^1 \int_{0}^1 f(y) \, dy \]
The proof of this theorem is in two parts: first, we show how to define the set required in the integral (2.12); secondly, we show that this set coincides with the required set \( \{ y; g(y) \geq x \} \). The proof will be given in Section 4.

We now give a number of simple generalizations of the main theorem.

**Remark 1.** The condition that \( Y_1, Y_2, \ldots, Y_n \) have a density function is not necessary. The result holds for an arbitrary distribution for \( Y_1, Y_2, \ldots, Y_n \). We have that \( P(g(Y) \geq x) \) is given simply by adding up all the probability inside and on the boundary of the region of integration in the theorem. However, the applications given in this paper all have \( Y \) with a continuous probability density function.

**Remark 2.** The condition \( a_i > 0, i = 1, \ldots, n \) may be relaxed to \( a_i \geq 0, i = 1, \ldots, n \) with no difficulty. If only \( r \) of the \( n \) coefficients \( a_i \) are nonzero then the problem is easily expressed in the standard form for \( r \) variables with all \( a_i > 0 \).

**Remark 3.** Although we have stated the main result for a random variable of the form (1.1) where \( 0 \leq Y_i \leq 1, i = 1, 2, \ldots, n \) the essential assumption is that \( Y_i \geq 0 \). For example, if we have \( 0 \leq Y_i \leq \delta \), then the substitution \( Y_i' = Y_i / \delta \), recovers the problem in the standard form. If however we have only that \( Y_i \geq 0 \) then the geometry simplifies and it is easily seen that \( P(g(Y) \geq x) \) is given by the form
\[
\int_0^\infty \cdots \int_0^\infty f(y) \, dy.
\]
whatever the value of \( x \).

3. Applications.

**Example 1.** We consider the distribution of \( X = \sum_{i=1}^{n} c_i X_i \), where \( X_i = (X_1, X_2, \ldots, X_n) \) is distributed over the simplex \( X_i \geq 0, \sum_{i=1}^{n} X_i = 1 \). We assume without loss of generality that \( c_1 > c_2 > \ldots > c_n > 0 \) and transform \( X \rightarrow Y \) as follows:
\[
X_i = 1 - Y_i, \quad \quad i = 2, \ldots, n.
\]

The Jacobian of this transformation is \( y_{1}^{2-1}y_{2}^{2-2} \cdots y_{n-1} \) and the joint probability density function of \( Y_1, Y_2, \ldots, Y_n \) is easily found. The region \( X_i \geq 0, \sum_{i=1}^{n} X_i \leq 1 \) transforms into \( 0 \leq Y_i \leq 1, i = 1, 2, \ldots, n \) and the random variable \( X \) transforms to
\[
X \rightarrow d_0 - \sum_{i=1}^{n} d_i Y_i, \quad Y \leq Y \quad \quad \quad \quad \quad Y \leq Y.
\]
where
\[
d_0 = c_1; \quad \quad d_i = c_i - c_{i+1}, \quad i = 1, 2, \ldots, n - 1; \quad d_n = c_n.
\]
Thus
\[
P(X \leq x) = P(\sum_{i=1}^{n} d_i Y_i, Y_n \geq d_0 - x)
\]
and since \( d_i > 0 \) (2.12) may be applied immediately. In this case we have \( \alpha_i = (c_i - x) / A(i) \) where
\[
A(i) = d_i + \sum_{i=1}^{n} d_i Y_i + \sum_{i=2}^{n} Y_i.
\]

In any particular application we still have the problem of evaluating the \( n \)-dimensional repeated integral in (2.12). This involves firstly the computation of the joint probability density function of \( Y_1, Y_2, \ldots, Y_n \) and secondly the evaluation of the resulting integral. This evaluation may well have to be attempted numerically but there do exist a number
of problems of interest where these integrals may be obtained explicitly. We illustrate the kind of calculation involved by considering the distribution of the sum in a sample from the uniform distribution on [0, 1]. This is a well-known problem whose solution goes back to Lagrange. To apply the present method we note that if \( X_{11}, X_{12}, \ldots, X_{1n} \) are the order statistics from a sample of size \( n \) from the uniform distribution on [0, 1] with \( X_{11} \leq \cdots \leq X_{1n} \) and if we let \( U_i = X_{1i} - X_{(i-1)}, i = 2, 3, \ldots, n \) then we have \( f(u_1, u_2, \ldots, u_n) = n! \) where \( u_1, u_2, \ldots, u_n \) are distributed over the simplex \( u_i \geq 0, \sum_{i=1}^{n} u_i \leq 1 \) as required. In terms of \( u_1, u_2, \ldots, u_n \) we have \( \sum_{i=1}^{n} X_i = \sum_{i=1}^{n} X_{1i} = \sum_{i=1}^{n} (n - i + 1) U_i \) and we may apply the preceding general result with \( c_i = n - i + 1 \). We find from (3.2) and (3.3) that

\[
\sum_{i=1}^{n} X_i \rightarrow n - Y_1 - Y_1 Y_2 - \cdots - Y_1 Y_2 \cdots Y_n
\]

where \( f(y_1, y_2, \ldots, y_n) = n! y_1^{y_1-1} y_2^{y_2-2} \cdots y_{n-1} \) over \( 0 \leq y_i \leq 1, i = 1, \ldots, n \). Applying the main result (2.12) as in (3.4) we find when \( x \in [n - i, n - n + 1] \), \( i = 1, 2, \ldots, n \) that

\[
P(\sum_{i=1}^{n} X_i \leq x) \text{ is given by}
\]

\[
\int_{0}^{1} \int_{0}^{1} \cdots \int_{0}^{1} \int_{0}^{1} \cdots \int_{0}^{1} n! y_1^{y_1-1} y_2^{y_2-2} \cdots y_{n-1} \ dy_1 \ dy_2 \cdots \ dy_n,
\]

where \( a_n = (n - k - 1 - x)/(1 + \sum_{j=1}^{n} y_{k+1} y_{k+2} \cdots y_j), j = 1, \ldots, i \). We omit the evaluation of this integral and remark only that the required result may be obtained by using a technique similar to that used in Currie (1978, page 40 ff). We also remark that the more general results of Dempster and Kleyle (1968), Weissberg (1971), and Margolin's (1977) result for Dirichlet variables, may all be obtained in a similar fashion, though the calculations involved are rather formidable.

Closely related to the distribution of the sum in uniform samples is the distribution of \( T_n = X_{1n}/(X_{11} + \cdots + X_{1n}) \). The distribution of \( T_n \) for an exponential population was obtained first by Fisher (1929) using an ingenious geometrical argument. Cochran (1941) extended this result to a gamma population. Recently, Lewis and Fieller (1979) gave a recursive algorithm that obtains these results in a very neat way. The link between the distribution of \( T_n \) and the distribution of the sum in uniform samples was given by Darling (1952). Using the Laplace transform, he found that the distribution of \( T_n^{-1} \), based on a sample from the uniform distribution on [0, \( a_n \)], was the same as \( 1 + S_{n-1} \), where \( S_{n-1} \) has the distribution of the sum of the observations in a sample of size \( n - 1 \) from the uniform distribution on [0, 1]. This curious result has an important consequence in the present context: since we know that \( S_{n-1} \) has a distribution of the form (1.1) we can conclude that \( T_n^{-1} \) also has a distribution of this form. The result (2.12) can now be applied to give the distribution of \( T_n^{-1} \) as a repeated integral. The integrand depends on the sampled population. If we supply the integrands appropriate for uniform, exponential or gamma sampling we obtain the results of Darling, Fisher and Cochran respectively. Again it is only fair to remark that the integrations required are rather daunting.

**Example 2.** In the previous example we considered linear functions of variables distributed over the simplex; we turn now to the distribution of certain quadratic forms in such variables. To be precise, we consider the distribution of the random variable \( X \) where

\[
X = a_1 + \sum_{i=1}^{n} a_i X_{1i}^2 - 2 \sum_{i=1}^{n} a_i X_{1i} + 2 \sum_{i=1}^{n} \sum_{j=i+1}^{n} a_i a_j X_{1i} X_{1j}
\]

where \( a_1 > a_2 > \cdots > a_n > 0 \) and \( X_{1i} \geq 0, \sum_{i=1}^{n} X_{1i} \leq 1 \). This quadratic form has particular statistical interest since it turns out that the studentized extreme deviate, \( (X - X_{11})/s \) can be written in this way. We first show that, by using suitable transformations the quadratic form can be put into the form (1.1) and so the main result can be applied. Secondly, we show that \( (X - X_{11})/s \) does indeed have a distribution of the type (1.1) by transforming it into an example of the quadratic form being discussed here. Lastly, we specialize to samples from (a) normal and (b) exponential populations and give explicit results for the distribution of \( (X - X_{11})/s \) in both cases.
We have the following identity for $X$ as given by (3.6):

$$X = \sum_{i=1}^{n} b_i (1 - x_1 - x_2 - \cdots - x_i)^2,$$

where $a_i = \sum_{j=1}^{i} b_j$. We now apply the transformation (3.1) and find

$$X = \sum_{i=1}^{n} b_i Y_i^2 Y_i^3 \cdots Y_i^l$$

which has the required form (1.1) and the distribution function is obtained.

As an illustration of this result we consider the distribution of

$$W = \frac{n(\bar{X} - X_{(n)})^2}{(n - 1)S^2}$$

where $W$ is based on a random sample $X_1, X_2, \ldots, X_n$ with $\bar{X} = \frac{1}{n} \sum X_i$, $S^2 = \frac{1}{n} \sum (X_i - \bar{X})^2$ and $X_{(1)} = \min(X_1, X_2, \ldots, X_n)$. This statistic, in the equivalent form $R = (\bar{X} - X_{(1)})/S$, is widely used in detecting various forms of abnormality in the sample; for example, it can be used in the detection of outliers; in this case the sampled population is assumed to be normal. This distribution was studied first by Pearson and Chandra Sekar (1936) who obtained upper tail probabilities. Grubbs (1950) obtained an integral form for the distribution function of the statistic $\theta$, where $\sin^2 \theta = 1 - W$; Grubbs evaluated his integrals numerically. Borenius (1959) effectively obtained the density function of $W$. When the underlying distribution is assumed to be exponential, we have the distributional problem for $W$, the test statistic proposed by Shapiro and Wilk (1972) as a test for the exponential distribution.

To obtain $W$ in the required form (3.6) or (3.7) we proceed as follows. If $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ are the order statistics let $G_1 = X_{(1)}; G_2 = X_{(2)} - X_{(1)}, i = 2, \ldots, n$. Then $W$ depends only on $G_2, G_3, \ldots, G_n$. In fact $W^{-1} = (n - 1) Y/Z^2$ where

$$Z = n(\bar{X} - X_{(n)}) = \sum_{i=1}^{n} q_i G_i$$

and

$$Y = nS^2 = \sum_{i=2}^{n} p_i q_i G_i^2 + 2 \sum_{i=2}^{n} \sum_{j=i+1}^{n} p_i q_i G_i G_j$$

where $p_i = i - 1$ and $q_i = n - i + 1$. Now transform as follows:

$$U_i = \sum_{j=1}^{i} q_j G_j; U_i U_i = q_i G_i, \quad i = 2, \ldots, n - 1.$$

Then $W$ depends only on $U_2, \ldots, U_{n-1}$, and we find

$$W^{-1} = (n - 1)(n - 1 - 2n \sum_{i=2}^{n} a_i U_i + n \sum_{i=2}^{n} a_i U_i^2 + 2n \sum_{i=2}^{n} \sum_{j=i+1}^{n} a_i U_i U_j)$$

where $a_i = (n - i)/(n - i + 1), i = 2, \ldots, n - 1$ and $N = n - 1$. It is a routine calculation to show that $U_2, \ldots, U_{n-1}$ are distributed over the simplex $U_i \geq 0, \sum_{i=1}^{n-1} U_i \leq 1$ and so the problem is essentially of the form (3.6). It is worth remarking that the transformation $X \rightarrow G$ which gives $W$ in terms of $G_2, \ldots, G_n$ uses the location invariance of $W$; the transformation $G \rightarrow U$ which gives $W$ in terms of $U_2, \ldots, U_{n-1}$ depends on the scale invariance of $W$.

We now have the following integral form for the distribution function of $W$. For $k = 1, 2, \ldots, n - 2$ let

$$I_n(k) = \left[ \begin{array}{c} n - k - 1 \\ (k + 1)(n - 1) \end{array} \right] \left[ \begin{array}{c} n - k \\ k(n - 1) \end{array} \right].$$

Then for $w \in I_n(k)$ we have

$$P(W \leq w) = \int_0^1 \cdots \int_0^1 \int_{a_{n-1}}^{a_{n-2}} \cdots \int_{a_1}^{a_1} f(y_2, \ldots, y_{n-1}) dy_2 \cdots dy_{n-1}$$

where $a_t = \psi_t(w)/A(k)$ with
\[ \psi(w)^2 = (n-2)/(n)w^{-1} - 1; \]
\[ \psi_k(w)^2 = \psi(w)^2 - (n-2)(k-1)/(n-k), \quad k = 2, \ldots, n-1 \]

and

\[ A(k)^2 = \phi(k)^2 + \sum_{i=k+2}^{n} \phi(i-1)^2 y_i^2 \cdot \ldots \]

with

\[ \phi(k)^2 = (n-2)(n-1)/(n-k-1)(n-k), \quad k = 1, \ldots, n-2. \]

We remark that (3.10) gives the distribution function of \( W \) whatever the underlying sampled population is. We now specialise to the exponential and normal cases. If \( X_1, \ldots, X_n \) is a random sample from the exponential distribution, then we may show that the joint distribution of \( Y_2, \ldots, Y_{n-1} \) is

\[ f(y_2, \ldots, y_{n-1}) = (n-2)! \cdot y_2^{r-2} \cdot y_{n-2}^{-1} \cdot \ldots \cdot 0 \leq y_i \leq 1. \]

In this case the variables \( Y_2, \ldots, Y_{n-1} \) are independently distributed over the \( n-2 \) cube. The integral (3.10) may now be evaluated in the same way as the integral (3.5). Details are contained in Currie (1978, page 40 ff). We report only the final result. In (3.10) we first replace \( J_0 \) by \( J_0^2 \) and hence obtain (3.10) in terms of \( 2^l \) integrals of the form

\[ J_{\gamma}(k) = J_{\gamma}(k_1, k_2, \ldots, k_l) = J_{\gamma}(k_l). \]

Now define

\[ T_\gamma(l) = \{ k_l : k_l = (k_1, k_2, \ldots, k_l) ; k_l \geq 1, \sum_{l=1}^{l-1} k_l \leq k \}. \]

Then the integral (3.10) is given by

\[ 1 + \sum_{l=1}^{\gamma} (-1)^l \sum_{\gamma'_{l+1}} J_\gamma(k_l) \]

with the density function \( f \) given by (3.11). The integral (3.12) can be evaluated as a product of \( l \) functions and it is worth noting that this factorisation depends critically on the factorisation of \( \alpha_l \), remarked on earlier.

For \( n \geq 4, k = 1, 2, \ldots, n-3 \) and \( x \geq 0 \) we define

\[ G_{n,1}(x) = 1 \]

\[ G_{n,x}(x) = \int_{-\infty}^{\tan^{-1}(x)} G_{n,k-1} \left( \frac{n-k+1}{n-k-1} \sec \theta \right) \sin^{n-k-1}(\theta) \, d\theta. \]

It is interesting to note that the functions \( G_{n,x}(x) \) are very similar to the functions \( \beta_p(r) \) used by Borelius (1959) in deriving the density function of \( W \) in the normal case. We note that \( G_{n,x}(x) \) is essentially a \( k \) dimensional integral. For any \( k_l = (k_1, \ldots, k_l) \) let \( \kappa = \sum_{l=1}^{l-1} k_l \) and

\[ \gamma_k(k_l) = \frac{2(n-k_l)^{1/2}}{\Gamma((n-k_l-1)/2)} \cdot \frac{(n-k_l-1)!(n-k_l-2)!}{(n-k!)^2} \cdot \frac{w^{n-k-1/2}}{((n-1)(n-2))^{(n-k-1)/2}}, \]

then
\[ J_n(k_i) = \gamma_n(k_i) \psi_{\frac{1}{2}}(w) \sum_{i=1}^{n-k-1} \frac{\psi_{\frac{i}{2}}(w)}{\phi(w(i+1) - 1)} \]

Note that we have adopted the convention that a product \( \prod_{i} \) with \( k > l \) is identically one.

We recall that \( G_{n,k}(x) \) is effectively a \( k \)-dimensional integral. This limits the practical usefulness of the above result but exact percentage points for \( W \) have been found for small values of \( n \) and for part of the range of \( W \) for larger values of \( n \). These results are reported in Table 1. We remark that the original values of Shapiro and Wilk (1972) obtained by simulation agree with these exact results within the limits of their experimental error.

We now specialise in (3.10) to the case when the underlying sampled population is normal. We may then show that the joint distribution of \( Y_1, Y_2, \ldots, Y_{n-1} \) is

\[ f(y_1, \ldots, y_{n-1}) = c_n \frac{y_2^{2}\cdot y_3^{2}\cdots y_{n-2}^{2}}{(1 + n(n-1) \sum_{i=2}^{n} y_i^2)^{N/2}} \]

with \( b_i = ((n-i)(n-i+1))^{-1}, N = n - 1 \) and

\[ c_n = \frac{1}{n^{N/2}} \Gamma(N/2) \]

over the region \( 0 \leq y_i \leq 1, i = 2, \ldots, n - 1 \).

In this integral form, the result is equivalent to that obtained by Grubbs (1950). We can however evaluate the integral in terms of the function \( G_{n,k}(x) \) defined in (3.15); for details

<table>
<thead>
<tr>
<th>Table 1</th>
<th>The Percentage Points of W exponential</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n )</td>
<td>0.5</td>
</tr>
<tr>
<td>3</td>
<td>0.2519</td>
</tr>
<tr>
<td>4</td>
<td>0.1242</td>
</tr>
<tr>
<td>5</td>
<td>0.0831</td>
</tr>
<tr>
<td>6</td>
<td>0.0638</td>
</tr>
<tr>
<td>7</td>
<td>0.0528</td>
</tr>
<tr>
<td>8</td>
<td>0.0455</td>
</tr>
<tr>
<td>9</td>
<td>0.0544</td>
</tr>
<tr>
<td>10</td>
<td>0.0574</td>
</tr>
<tr>
<td>11</td>
<td>0.0624</td>
</tr>
<tr>
<td>12</td>
<td>0.1010</td>
</tr>
<tr>
<td>13</td>
<td>0.0922</td>
</tr>
<tr>
<td>14</td>
<td>0.0848</td>
</tr>
<tr>
<td>15</td>
<td>0.0784</td>
</tr>
<tr>
<td>16</td>
<td>0.1185</td>
</tr>
<tr>
<td>17</td>
<td>0.1096</td>
</tr>
<tr>
<td>18</td>
<td>0.1018</td>
</tr>
<tr>
<td>19</td>
<td>0.0950</td>
</tr>
<tr>
<td>20</td>
<td>0.0890</td>
</tr>
<tr>
<td>21</td>
<td>0.0837</td>
</tr>
<tr>
<td>22</td>
<td>0.0890</td>
</tr>
<tr>
<td>23</td>
<td>0.0930</td>
</tr>
<tr>
<td>24</td>
<td>0.0967</td>
</tr>
<tr>
<td>25</td>
<td>0.0933</td>
</tr>
<tr>
<td>26</td>
<td>0.0884</td>
</tr>
<tr>
<td>27</td>
<td>0.0906</td>
</tr>
<tr>
<td>28</td>
<td>0.0880</td>
</tr>
<tr>
<td>29</td>
<td>0.0819</td>
</tr>
<tr>
<td>30</td>
<td>0.0781</td>
</tr>
</tbody>
</table>
of this evaluation see Currie (1978, page 64 ff). We denote the integral (3.12) by $J^*(k_i)$ when the density function (3.17) is used. Let

$$
\delta_\nu(k_i) = \frac{n!}{(n - k)!} \frac{\Gamma((n - 1)/2)}{\Gamma((n - k - 1)/2)} e^{-x^2}
$$

Then

$$
J^*(k_i) = \delta_\nu(k_i) \prod_{i=1}^{n} G_{a_{\nu+1}, \ldots, a_{\nu+k-1}} \left( \frac{\psi_{\nu+1}(i)}{\phi(i + 1)} \right)
$$

where $x(0) = 0$.

The distribution function of $W$ in the normal case is now given by (3.14) with $J^*(k_i)$ replaced by $J^*(k_i)$. It is interesting that the formulae (3.16) and (3.18) for $J^*(k_i)$ and $J^*(k_i)$ each consist of a product of $l$ functions of $u, l - 1$ of which are identical. This result for the distribution function of $W$ is equivalent to the result of Borenius (1959) for the density function of $W$.

4. Proof of the theorem. First, we define

$$
(4.1)\quad S(x) = \{y' = (y_1, y_2, \ldots, y_n) : g(y) \geq x, 0 \leq y_1 \leq 1, 1 \leq i \leq n\}
$$

and observe that $S(x)$ is the set over which we require to integrate the density $f(y)$ in order to evaluate the required probability. The set $S(x)$ is given implicitly by the condition $g(y) \geq x$; we show first how to construct a set $S_\nu(x)$ for $x \in I_\nu(k)$ that will turn out to be $S(x)$. The construction of $S_\nu(x)$ is aimed at producing the set of points in the region of integration in the integral (2.12). We choose the coordinates of the points $y$ to be included in $S_\nu(x)$ in the order $y_n, y_{n-1}, \ldots, y_1$ and finally $y_1$. We need to check that if $y_n, y_{n-1}, y_{n-2}, \ldots, y_1$ have been chosen using the construction where $f \leq k$, then $0 \leq a_{\nu+1} \leq 1$.

Suppose that $x \in I_\nu(k)$ for some $k = 1, 2, \ldots, n - 1$ where $n \geq 2$. Let $y \in [0, 1]$, $i = k + 1, \ldots, n$. Let $a_k = a_k(n; x; y_{n-1}, \ldots, y_1)$ be defined as in (2.10). Then

$$
a_k = \frac{x - b_{k-1}}{A(k)} \leq \frac{b_k - b_{k-1}}{A(k)} = \frac{a_k}{a_k + y_{k-1}A(k + 1)} \leq 1.
$$

Clearly, $a_k \geq 0$ and so $0 \leq a_k \leq 1$ as required. Similarly, if $x \in I_\nu(n)$ then $0 \leq a_n \leq 1$.

Assume now that $y_n, y_{n-1}, \ldots, y_1$ have been chosen and are such that $a_{n-1} = a_{n-1}(n; x; y_{n-1}, \ldots, y_1)$ is defined and $0 \leq a_{n-1} \leq 1$ for some $l$ where $0 \leq l \leq n - 3, n \geq 3$. Let $y_{n-1} \in [a_{n-1}, 1]$. Then we have

$$
a_{n-2} \geq 0 \Rightarrow x - b_{n-2} \geq 0 \quad \text{by (2.10)}
$$

$$
\Rightarrow x - b_{n-2} \geq a_{n-2} > 0 \quad \text{by (2.9)}
$$

$$
\Rightarrow a_{n-2} > 0 \quad \text{by (2.10)}.
$$

We also have, using (2.10), (2.9) and (2.8) that

$$
a_{n-1} = \frac{x - b_{n-2} + a_{n-2}}{a_{n-2} + y_{n-2}A(n - l - 1)}
$$

$$
= \frac{a + a_{n-1}}{a + y_{n-1}} \quad \text{where} \quad a = \frac{a_{n-2}}{A(n - l - 1)} > 0
$$

$$
\leq 1 \quad \text{since} \quad y_{n-1} \geq a_{n-1} \geq 0.
$$

And so $0 \leq a_{n-2} \leq 1$ as required. In a similar way if $a_n$ is defined and $0 \leq a_n \leq 1$ where $n \geq 2$, and if $y_n \in [a_n, 1]$ then $0 < a_{n-1} \leq 1$.

These results enable us to define the sets $S_\nu(x)$ that will turn out to give the required set $S(x)$. 


DEFINITION. (a) Let \( x \in L_n(k) \), for some \( k = 1, 2, \ldots, n - 1 \). The set \( S_k(x) \) consists of all the points \( y' = (y_1, y_2, \ldots, y_n) \) constructed as follows: for \( i = k + 1, \ldots, n \) choose any \( y_i \in [0, 1] \). Now form \( \alpha_k \) using (2.10). We have just proved that \([\alpha_k, 1] \subseteq [0, 1]\). Choose any \( y_k \in [\alpha_k, 1] \) and form \( \alpha_{k+1} \). Again we know \([\alpha_{k+1}, 1] \subseteq [0, 1]\). We may continue in this fashion choosing \( y_j \in [\alpha_j, 1] \) and forming \( \alpha_{j+1} \) in the order \( j = k - 1, k - 2, \ldots, 2 \). At each stage we are ensured that \([\alpha_{k-1}, 1] \subseteq [0, 1]\) and the choice \( y_{k-1} \in [\alpha_{k-1}, 1] \) can be made. Lastly, choose any \( y_1 \in [\alpha_1, 1] \).

(b) If \( x \in L_n(n) \) then \( S_n(x) \) consists of all the points \( y' = (y_1, y_2, \ldots, y_n) \) constructed as follows: we know that \( 0 \leq \alpha_n \leq 1 \) and so we may choose any \( y_n \in [\alpha_n, 1] \) and form \( \alpha_{n-1} \). Then \([\alpha_{n-1}, 1] \subseteq [0, 1]\) and the construction of \( S_n(x) \) is completed as in (a).

We can now prove that the set of interest \( S(x) \) is given by \( S_k(x) \) whenever \( x \in L_n(k) \). We begin by proving the following: if \( x \geq b_m \) then \( S(x) \subseteq S_m(x) \). Suppose \( y \in S(x) \). We must show that \( y_i \in [\alpha_i, 1] \) for \( i = 1, 2, \ldots, k \). We prove this by induction on \( k \).

Case 1. \( k = 1 \). If \( y \in S(x) \) then by definition \( x \leq g(y) = y_1 A(1) \) and so \( y_i \leq x/\alpha_i \). Since we also have \( y_i \leq 1 \) we have \( y_i \in [\alpha_i, 1] \) as required.

Case 2. We assume the result is true for \( k = m < n \). We further assume that \( x \geq b_m \); hence \( x - b_m \geq \alpha_m > 0 \) and so by induction \( y_i \in [\alpha_i, 1] \), \( i = 1, \ldots, m \). Further

\[
\alpha_m \leq 1
\]

yields

\[
\frac{x - b_m}{\alpha_m + y_m A(m + 1)} \leq 1
\]

\[
\Rightarrow x - b_m - \alpha_m \leq y_m A(m + 1)
\]

\[
\Rightarrow y_m \geq \frac{x - b_m}{A(m + 1)} = \alpha_{m+1}
\]

since we have assumed \( x - b_m \geq 0 \). Since \( y_{m+1} \leq 1 \) we have \( y_{m+1} \in [\alpha_{m+1}, 1] \) as required.

Now suppose that \( x \in L_n(k) \). Then \( x \geq b_{k-1} \) and so we have \( S(x) \subseteq S_{k-1}(x) \). Conversely, suppose that \( y \in S_{k-1}(x) \). Then by the construction of \( S_{k-1}(x) \) we have \( y_i \in [\alpha_i, 1] \). Thus by (2.10) \( y_i \geq (x - b_i)/A(1) \Rightarrow y_i A(1) \geq x \) since \( b_0 = 0 \); and so by (1.1) and (2.7) \( g(y) \geq x \) and this is the condition (4.1) that \( y \in S(x) \). Thus \( S_k(x) \subseteq S(x) \) and so \( x \in L_n(k) \) implies that \( S(x) = S_k(x) \).

We can now prove the main result. We assume \( x \in L_n(k) \) and so \( P(g(Y) \geq x) \) is given by

\[
\int_{S_n(x)} \cdots \int_{S_{n-1}} f(y) \, dy = \int_{S_{n+1}} \cdots \int_{S_1} f(y) \, dy
\]

By the construction of \( S_k(x) \) we may replace the multiple integral over \( S_k(x) \) with the repeated integral in (2.12) and this proves the result.

5. Discussion. In Section 2.1 we noted the connection between the changing form of the density function of \( g(Y) \) and the geometry of a surface passing through a cube. We now give a statistical interpretation of this for one of our examples. Pearson and Chandra Sekar (1936) obtained the upper tail of the probability density function of the extreme studentized deviate from a normal population as follows: let \( x_1, x_2, \ldots, x_n \) be a series of observed values of a variable \( x \). Let \( \bar{x} = \frac{\sum x_i}{n}, s^2 = \frac{\sum (x_i - \bar{x})^2}{n} \) and \( \tau_i = (x_i - \bar{x})/s \). Let \( \tau_1 \leq \tau_2 \leq \cdots \leq \tau_n \) be the ordered values of \( \tau_1, \tau_2, \ldots, \tau_n \). Pearson and Chandra Sekar show that \( \max \tau_{n-1} < \max \tau_0 \) and they deduce from the general relation \( n f(t) = \sum f(r) \) that, for \( \tau > \max \tau_{n-1} \), \( f_\tau(r = n f(t) \) where \( f_\tau(t) \) represents the (unknown) probability.
density function of \( \tau_i \), and \( f(\tau) \) the (known) probability density function of the unordered \( \tau_i \). It is simple to verify that the value of \( \max \tau_{\alpha-1} \) coincides with the left end of the interval \( I_0(1) \) defined in (3.9). This correspondence extends as follows: we may show as in Currie (1978, page 77 ff) that \( \max \tau i \) = \((i - 1)/(n - i + 1)\) \( i \); it is now a simple matter to verify that this value coincides with the left end of the interval \( I_0(n - i) \). We do not have to look far for an explanation of this apparent set of coincidences. From the relation \( nf(\tau) = \sum f_i(\tau) \) we have \( nf(\tau) = f_\tau(\tau) \) for \( \tau \geq \max \tau_{\alpha-1} \), and \( nf(\tau) = f_\tau(\tau) + f_{\beta-1}(\tau) \) for \( \max \tau_{\alpha-1} \geq \tau \geq \max \tau_{\beta-2} \). Now \( f(\tau) \) has a single functional form over its entire range, at least in sampling from a normal population. Hence, for \( f(\tau) \) to maintain its single form, \( f_\tau(\tau) \) must change form at exactly the point that \( f_{\beta-1}(\tau) \) becomes nonzero, and an obvious extension of the argument provides the general explanation.

REFERENCES


HALL, P. (1927). The distribution of means for samples of size \( N \) drawn from a population in which the variate takes values between 0 and 1, all such values being equally probable. Biometrika 19 240-244.


DEPARTMENT OF ACTUARIAL MATHEMATICS AND STATISTICS
HERIOT-WATT UNIVERSITY
EDINBURGH EH14 4AS, SCOTLAND