The Annals of Mathematical Statistics
1971, Vol. 42, No. 2, 704–709

THE DISTRIBUTION OF LINEAR COMBINATIONS OF ORDER STATISTICS FROM THE UNIFORM DISTRIBUTION

BY HERBERT WEISBERG

New York University

1. Introduction. In this paper we derive an algorithm for computing the distribution function of an arbitrary linear combination of order statistics from a uniform distribution. Suppose $U_{(i)}$ is the ith smallest observation from a sample of size n from the uniform distribution on [0, 1], with the convention $U_0 \equiv 0$, $U_{n+1} \equiv 1$. Consider a set of S integers $\{k_i\}$ such that

(1.1) $$k_0 = 0 < k_1 < k_2 < \cdots < k_S \leqq n.$$

For any set of constants $d_i > 0$ and any x, we seek

$$P\{\sum_{s=1}^{S} d_s U_{(k_s)} \leqq x\}.$$

Our approach is to generalize a formula derived by Dempster and Kleyle (1968).

2. Derivation of the algorithm. Let $X_i = U_{(i)} - U_{(i-1)}$, $i = 1, 2, \cdots n$. Let $c_{n+1} = 0$.

Define $c_1, c_2, \cdots c_n$ by

$$c_{k_i} = c_{k_i+1} + d_i \qquad \text{for} \quad i = 1, \cdots S$$

$$c_j = c_{j+1} \qquad \text{for} \quad j \notin (k_1, \cdots k_S).$$

Then we have

(2.1) $$\sum_{s=1}^{S} d_s U_{(k_s)} = \sum_{i=1}^{n} c_i X_i.$$

For the special case $S = n$, Dempster and Kleyle (1968) have shown that

(2.2) $$P\left\{\sum_{i=1}^{n} c_i X_i \leqq x\right\} = 1 - \sum_{j=1}^{r} \frac{(c_j - x)^n}{c_j \prod_{i \neq j}(c_j - c_i)}$$

for $0 \leqq x \leqq c_1$, where r is the largest positive integer such that $x \leqq c_r$. In the general case $S \leqq n$, we wish to allow

$$c_{k_{s-1}+1} = c_{k_{s-1}+2} = \cdots = c_{k_s} = c_{(s)},$$

for $s = 1, 2, \cdots S$.

Let $k_s - k_{s-1} = r_s$, $s = 1, \cdots S$; and $n - k_S = r_{S+1}$. Then we wish to let the first r_1 c_1's take the value $c_{(1)}$, the next r_2 take the value $c_{(2)}$, etc. Let $c_{(s+1)} = 0$. In this situation (2.2) is not applicable unless $r_s = 1$ for all s.

Received March 6, 1970; revised October 19, 1970.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Mathematical Statistics.*
www.jstor.org

Suppose, however, that we define

$$b_{(k_{s-1}+i)}(h) = c_{(s)}+(r_s-i)h, \qquad \text{for} \quad h > 0,\ i = 1, \cdots r_s;\ s = 1 \cdots S. \tag{2.3}$$

$$b_{(k_S+i)}(h) = (r_{S+1}+1-i)h \qquad i = 1, 2, \cdots r_{S+1}.$$

Then we have

LEMMA 1. $\lim_{h\to 0} P\{\sum_{i=1}^n b_i(h)X_i \leqq x\} = P\{\sum_{s=1}^S d_s U_{(k_s)} \leqq x\}$.

PROOF. Let $A_r = \{\sum_{i=1}^n b_i(1/r)X_i \leqq x\}$

$$A = \{\sum_{s=1}^n d_s U_{(k_s)} \leqq x\} = \{\sum_{i=1}^{k_1} c_{(1)}X_i + \sum_{i=k_1+1}^{k_2} c_{(2)}X_i + \cdots \leqq x\}.$$

Now $\left\{\sum_{i=1}^n b_i\left(\frac{1}{r}\right) X_i \leqq x\right\} \Rightarrow \left\{\sum_{i=1}^n b_i\left(\frac{1}{r+1}\right) X_i \leqq x\right\}$, so that

$A_r \subset A_{r+1}$ and $A = \bigcup_{r=1}^{\infty} A_r$. Therefore

$$P(A) = P(\bigcap_{r=1}^{\infty} A_r) = \lim_{r\to\infty} P\{A_r\}.$$

Suppressing for convenience the dependence of b_i on h we have from (2.2) that

$$P\left\{\sum_{i=1}^n b_i X_i \leqq x\right\} = 1 - \sum_{j=1}^{k_1} \frac{(b_j-x)^n}{b_j \prod_{i\neq j}(b_j-b_i)} - \sum_{j=k_1+1}^{k_2} \frac{(b_j-x)^n}{b_j \prod_{i\neq j}(b_j-b_i)}$$

$$- \cdots - \sum_{j=k_{m-1}+1}^{k_m} \frac{(b_j-x)^n}{b_j \prod_{i\neq j}(b_j-b_i)}$$

where m is the largest integer such that $x \leqq c_{(m)}$. Let

$$T_s = \sum_{j=k_{s-1}+1}^{k_s} \frac{(b_j-x)^n}{b_j \prod_{i\neq j}(b_j-b_i)}$$

so that

$$P\{\sum_{i=1}^n b_i X_i \leqq x\} = 1 - \sum_{s=1}^m T_s. \tag{2.4}$$

LEMMA 2.

$$T_s = \frac{\Delta^{r_s-1} f_s(c_{(s)})}{h^{r_s-1}(r_s-1)!}$$

where

$$f_s(c) = \frac{(c-x)^n}{c \prod_{i\leqq k_{s-1},\, i>k_s}(c-b_i)},$$

and Δ *is the forward difference operator defined by*

$$\Delta^k f(x) = \Delta^{k-1} f(x+h) - \Delta^{k-1} f(x), \qquad k = 1, 2, \cdots.$$

PROOF.

$$T_s = \sum_{j=k_{s-1}+1}^{k_s} \frac{(b_j - x)^n}{b_j \prod_{i \neq j}(b_j - b_i)}$$

$$= \sum_{j=k_{s-1}+1}^{k_s} \frac{(b_j - x)^n}{b_j \prod_{i \leqq k_{s-1},\, i > k_s}(b_j - b_i) \prod_{k_{s-1} < i < j \leqq k_s,\, k_{s-1} < j < i \leqq k_s}(b_j - b_i)}$$

$$= \sum_{j=k_{s-1}+1}^{k_s} \frac{f_s(b_j)}{\prod_{k_{s-1} < i < j \leqq k_s,\, k_{s-1} < j < i \leqq k_s}(b_j - b_i)}$$

$$= \sum_{\rho=1}^{r_s} \frac{f_s(c_{(s)} + (r_s - \rho)h)}{h^{r_s - 1}(r_s - \rho)!(\rho - 1)!}(-1)^{\rho - 1}.$$

Making the transformation $j' = r_s - \rho$, this can be written

$$\sum_{j'=0}^{r_s - 1} \frac{f_s(c_{(s)} + j'h)}{h^{r_s - 1}(r_s - 1)!} \binom{r_s - 1}{j'} (-1)^{r_s - 1 - j'},$$

which is equivalent to (see, for example [3] page 46)

$$\frac{\Delta^{r_s - 1} f_s(c_{(s)})}{h^{r_s - 1}(r_s - 1)!}.$$

We are now ready to prove the main result.

THEOREM.

$$P\left\{\sum_{s=1}^{S} d_s U_{(k_s)} \leqq x\right\} = 1 - \sum_{s=1}^{m} \frac{g_s^{(r_s - 1)}(c_{(s)})}{(r_s - 1)!}$$

where m is the largest integer such that $x \leqq c_{(m)}$ and

$$g_s(c) = \frac{(c - x)^n}{c \prod_{\mu \neq s}(c - c_{(\mu)})^r}.$$

PROOF. It is clear that for any function f whose kth derivative exists at x,

$$\lim_{h \to 0} \frac{\Delta^k f(x)}{h^k} = f^{(k)}(x).$$

Note also that

$$\lim_{h \to 0} b_{k_\mu + i} = c_{(\mu)} \quad \text{for} \quad i = 1, 2, \cdots r_\mu.$$

It follows from Lemma 2 that

$$\lim_{h \to 0} T_s = \frac{g_s^{(r_s - 1)}(c)}{(r_s - 1)!}$$

evaluated at $c_{(s)}$. The theorem then follows from Lemma 1 and (2.4).

To make use of this formula in practice we must be able to evaluate the high order derivatives of the functions g_s. We can write

$$\log g_s(c) = n \log(c-x) - \log c - \textstyle\sum_{\mu \neq s} r_\mu \log(c - c_{(\mu)}).$$

Differentiating both sides we obtain

$$g_s'(c) = g_s(c)h(c) \tag{2.5}$$

where

$$h(c) = \frac{n}{c-x} - \frac{1}{c} - \sum_{\mu \neq s} \frac{r_\mu}{c - c_{(\mu)}}. \tag{2.6}$$

Using Leibniz's rule for the kth derivative of a product, we obtain the recurrence relation:

$$\begin{aligned} g_s^{(k)}(c) &= \frac{d^{k-1}}{dc^{k-1}} g_s'(c) = \frac{d^{k-1}}{dc^{k-1}} (g_s(c)h(c)) \\ &= \textstyle\sum_{i=0}^{k-1} \binom{k-1}{i} g_s^{(i)}(c) h^{(k-1-i)}(c). \end{aligned} \tag{2.7}$$

We also have from (2.6)

$$h^{(i)}(c) = (-1)^i i! \left[\frac{n}{(c-x)^{i+1}} - \frac{1}{c^{i+1}} - \sum_{\mu \neq s} \frac{r_\mu}{(c - c_{(\mu)})^{i+1}} \right].$$

Thus (2.7) can be used recursively to obtain $g_s^{(k)}(c)$ for any k.

Note that although we have been assuming $d_i > 0$ for all i, the general problem can be handled by reordering and shifting variables, making use of the symmetry in the situation. For example $2U_{(3)} - U_{(1)} = X_1 + 2X_2 + 2X_3$ has the same distribution as $2X_1 + 2X_2 + X_3 = U_{(2)} + U_{(3)}$.

3. Application. Following the notation of Wilks (1962) we define the $(k-1)$—variate Dirichlet distribution $D(v_1, v_2, \cdots v_{k-1}; v_k)$ by the density

$$\begin{aligned} f(x_1, \cdots x_{k-1}) &= \frac{\Gamma(v_1 + v_2 + \cdots + v_k)}{\prod_{i=1}^{k} \Gamma(v_i)} \prod_{i=1}^{k-1} x_i^{v_i - 1} \left(1 - \sum_1^{k-1} x_i\right)^{v_k - 1} \\ &\qquad \text{for} \quad x_i \geqq 0,\ i = 1, \cdots k \quad \textstyle\sum_1^{k-1} x_i \leqq 1 \\ &= 0 \quad \text{otherwise.} \end{aligned}$$

It is easily shown that the joint distribution of $U_{(k_1)}$, $U_{(k_1+k_2)} - U_{(k_1)}, \cdots$, $U_{(k_1+\cdots+k_s)} - U_{(k_1+\cdots+k_{s-1})}$, for k_i's as in (1.1), is $D(k_1, k_2, \ldots k_s; n - \sum_1^s k_i + 1)$.

Let $p_1, p_2, \cdots p_{k-1}, p_k$ represent the cell probabilities for a multinomial population with k categories. For a Bayesian analysis it is common to assume a conjugate prior of the form $D(\eta_1, \eta_2, \cdots; \eta_k)$ for $p_1, \cdots p_{k-1}$. Suppose $\eta_1, \eta_2, \cdots \eta_k$ are integers. Let n_i, $i = 1, \cdots k$, be the observed frequency for the ith category and $n = \sum_{i=1}^k n_i$. Then the posterior distribution of $p_1, \cdots p_{k-1}$ is $D(n_1 + \eta_1, \cdots; n_k + \eta_k)$.

Suppose we wish to make posterior probability statements about events of the form $\{\sum_1^k a_i p_i \leqq x\}$ for real numbers $a_1, \cdots a_k$ and x. Let $v_j = \sum_{i=1}^j (n_i + \eta_i)$, $j = 1, \cdots k$. Then we have

$$P\{\textstyle\sum_1^k a_i p_i \leqq x\} = P\{\sum_1^k a_i[U_{(v_i)} - U_{(v_{i-1})}] \leqq x\} \tag{3.1}$$
$$= P\{a_k + \textstyle\sum_1^{k-1} (a_i - a_{i+1}) U_{(v_i)} \leqq x\},$$

where $U_{(j)}$ is the jth smallest observation from a sample of size $(v_k - 1)$ from the uniform distribution on [0, 1]. Thus the algorithm of Section 2 can be applied.

For example, suppose we have $k = 5$, and we assume the improper prior $D(0, 0, 0, 0; 0)$ suggested by Lindley (1964) for (p_1, p_2, p_3, p_4). Suppose also that

$$\begin{aligned} a_1 &= -5 & n_1 &= 10 \\ a_2 &= -2 & n_2 &= 15 \\ a_3 &= 0 & n_3 &= 10 \\ a_4 &= +2 & n_4 &= 10 \\ a_5 &= +5 & n_5 &= 6. \end{aligned}$$

From (3.1)

$$P\{\textstyle\sum_1^5 a_i p_i \leqq x\} = P\{5 - 3U_{(10)} - 2U_{(25)} - 2U_{(35)} - 3U_{(45)} \leqq x\}$$
$$= P\{3U_{(10)} + 2U_{(25)} + 2U_{(35)} + 3U_{(45)} \geqq 5 - x\},$$

where the order statistics are from a sample of size 50.

A computer program to implement the algorithm of Section 2 has been successfully run and used to obtain the following results for this example.

x	$P\{\sum_1^5 a_i p_i \leqq x\}$
1.0	.9998
.8	.9992
.6	.9967
.4	.9885
.2	.9660
0	.9150
− .2	.8196
− .4	.6738
− .6	.4929
− .8	.3119
−1.0	.1669
−1.2	.0741
−1.4	.0269
−1.6	.0079
−1.8	.0018
−2.0	.0003.

REFERENCES

[1] DEMPSTER, A. P. and KLEYLE, R. M. (1968). Distributions determined by cutting a simplex with hyperplanes. *Ann. Math. Statist.* **39** 1473–78.

[2] LINDLEY, D. V. (1964). Bayesian analysis of contingency tables. *Ann. Math. Statist.* **35** 1622–43

[3] RALSTON, A. R. (1965). *A First Course in Numerical Analysis.* McGraw Hill, New York.

[4] WILKS, S. S. (1962). *Mathematical Statistics.* Wiley, New York.