The Annals of Mathematical Statistics
1972, Vol. 43, No. 1, 260-272

A CLASS OF RANDOM CONVEX POLYTOPES

BY A. P. DEMPSTER[1]

Harvard University

0. Summary. In Section 1 it is shown that n interior points of the $(k-1)$-dimensional simplex S_k define a partition of S_k into $\binom{n+k-1}{k-1}$ convex polytopes $R(\mathbf{n})$ which are in one-one correspondence with the partitions of n into a sum of k nonnegative integers. If the n points are uniformly and independently distributed over S_k, then $R(\mathbf{n})$ becomes a random polytope. Basic properties of the random $R(\mathbf{n})$ are given in Section 2. Section 3 presents an algorithm which can be used to compute the distribution of any extremal vertex of $R(\mathbf{n})$.

1. Introduction. The polytopes defined below arise in the context of a model for multinomial sampling which is used as the main example of a system of generalized Bayesian inference [1]–[6]. The parameter space in multinomial sampling with k categories consists of the points

$$\mathbf{x} = [x_1, x_2, \cdots, x_k] \tag{1.1}$$

with

$$x_i \geqq 0 \quad \text{for} \quad i = 1, 2, \cdots, k \qquad \text{and} \qquad \sum_{i=1}^{k} x_i = 1\,, \tag{1.2}$$

i.e., the points $\mathbf{x}$ of a simplex S_k. For a given stochastic vector $\mathbf{x}$ defining multinomial probabilities, a random drawing from one of the k categories may be created by drawing a second stochastic vector $\mathbf{x}^*$ from a uniform distribution over S_k, and declaring the outcome to be in category i for some i on $1 \leqq i \leqq k$ if

$$x_h/x_h^* \leqq x_i/x_i^* \qquad \forall h \neq i\,. \tag{1.3}$$

The validity of the model may by checked as follows. The simplex S_k may be characterized by its k vertices $[1, 0, \cdots, 0], [0, 1, \cdots, 0], \cdots, [0, 0, \cdots, 1]$. The region in S_k corresponding to category i, as defined by (1.3), consists of the subsimplex of S_k formed by substituting $\mathbf{x}$ for the ith vertex of S_k while leaving the remaining $k-1$ vertices unchanged. The subsimplices formed in this way for $i = 1, 2, \cdots, k$ cover all of S_k and they intersect only in their common faces. Furthermore, it is easily checked that the $(k-1)$-dimensional volumes of these subsimplices are proportional to the corresponding elements of $\mathbf{x}$. Thus the sampling model based on (1.3) is unambiguous with probability one, and produces category i with probability x_i, for $i = 1, 2, \cdots, k$, as desired.

In the inference situation, the vector of multinomial probabilities $\mathbf{x}$ is

Received October 2, 1969.

[1] This work was facilitated by a grant from the National Science Foundation (GP-8774).

unknown, and n drawings are made, producing a vector of observations

$$\mathbf{n} = [n_1, n_2, \cdots, n_k], \tag{1.4}$$

where n_i is the number of observations in category i,

$$n_i \geqq 0 \quad \text{for} \quad i = 1, 2, \cdots, k \qquad \text{and} \qquad \textstyle\sum_{i=1}^{k} n_i = n. \tag{1.5}$$

The observations are supposed to have come from n independent drawings $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \cdots, \mathbf{x}^{(n)}$ each distributed uniformly over S_k like $\mathbf{x}^*$ above, and each producing one count in some category i by the rule (1.3). Thus, for given $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \cdots, \mathbf{x}^{(n)}$, the unknown probability vector $\mathbf{x}$ is known to lie in the region $R(\mathbf{n})$ given by:

DEFINITION. *The region $R(\mathbf{n})$ consists of points $\mathbf{x}$ such that $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \cdots, \mathbf{x}^{(n)}$ can be placed into k cells, with n_i points in cell i, where a point $\mathbf{x}^{(j)}$ in cell i must satisfy*

$$x_h/x_h^{(j)} \leqq x_i/x_i^{(j)} \qquad \forall h \neq i. \tag{1.6}$$

The model assumes that only $\mathbf{n}$ is observed, not $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \cdots, \mathbf{x}^{(n)}$, so that the region $R(\mathbf{n})$ remains unknown and random. Thus the inferences about $\mathbf{x}$ as detailed in [1]–[6] say in essence that $\mathbf{x}$ lies in a random region $R(\mathbf{n})$ with a specified distribution depending only on $\mathbf{n}$. In Appendix A of [5] certain distributions associated with random $R(\mathbf{n})$ are derived for $k = 3$. Properties of random $R(\mathbf{n})$ for general k are taken up beginning in Section 2 below. In Section 1 the objective is to explore properties of the regions $R(\mathbf{n})$ defined by fixed $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \cdots, \mathbf{x}^{(n)}$.

The notation $R(\mathbf{n})$ suppresses for convenience the dependence of $R(\mathbf{n})$ on the particular points $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \cdots, \mathbf{x}^{(n)}$. It is not obvious from the definition that $R(\mathbf{n})$ exists for any given $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \cdots, \mathbf{x}^{(n)}$ or that $R(\mathbf{n})$ is a convex polytope.

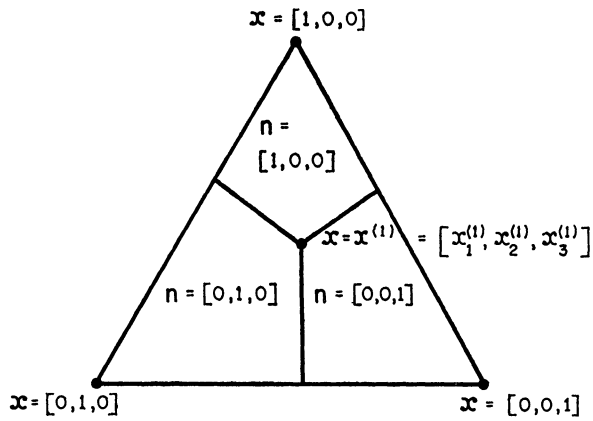

FIG. 1. An illustration of the 3 regions $R(\mathbf{n})$ in the case $k = 3$, $n = 1$ labelled by their associated $\mathbf{n} = [1, 0, 0], [0, 1, 0], [0, 0, 1]$.

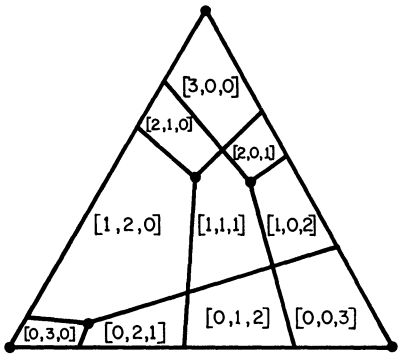

FIG. 2. An illustration of the 10 regions $R(\mathbf{n})$ in the case $k = 3$, $n = 3$ labelled by their associated $\mathbf{n}$.

The reader may begin to develop an intuition by checking out the definition for the regions shown in Figures 1 and 2 which illustrate respectively the cases $k = 3, n = 1$ and $k = 3, n = 3$. The pictures also suggest the conjecture, whose validity is established below, that for given $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \cdots, \mathbf{x}^{(n)}$ the regions $R(\mathbf{n})$ partition the simplex S_k into convex polytopes which correspond one-one to the partitions $\mathbf{n}$ of n. The net result is a generalization of the intervals and orderings produced by n points on a finite line segment. In the special case $k = 2$ there are $n + 1$ intervals on the line segment whereas in the general case there are $\binom{n+k-1}{k-1}$ convex polyhedra. In the special case $k = 2$ each decomposition of n into $n_1 + n_2$ determines a corresponding classification of the n points on a line segment into the n_1 closest to one end and the n_2 closest to the other end, whereas in the general case a set of n points in S_k may be classified into k subsets of size n_j for $1 \leqq j \leqq k$, such that the n_j in the jth subset are closest in some sense to face j of S_k, given any partition $\mathbf{n}$ of n into k parts.

The strategy in the succeeding paragraphs is to define the partition of the simplex anew, without reference to the above definition of $R(\mathbf{n})$, but in such a way that it is obviously a partition into convex polytopes. Then it will be shown that every such polytope is in fact a region $R(\mathbf{n})$ as defined above.

By a partition of S_k into closed convex polytopes will be meant a finite set $\mathscr{P}$ of closed convex polytopes whose union is S_k and whose interiors are mutually exclusive. By the product of the partitions $\mathscr{P}_j$ for $j = 1, 2, \cdots, n$ will be meant the partition of S_k which consists of all different nonempty convex polytopes of the form $\bigcap_1^n P_j$ where $P_j \in \mathscr{P}_j$ for $j = 1, 2, \cdots, n$.

Given a fixed $\mathbf{x}^{(j)} \in S_k$ and a fixed i on $1 \leqq i \leqq k$, a convex polytope may be defined by the $(k - 1)$ linear inequalities (1.6). Any $\mathbf{x} \in S_k$ belongs to at least one of the k polytopes so defined as i ranges over $1, 2, \cdots, k$ while $\mathbf{x}^{(j)}$ remains fixed, namely one corresponding to i satisfying

$$x_i/x_i^{(j)} = \max_{1 \leqq h \leqq k} \{x_h/x_h^{(j)}\} . \tag{1.7}$$

Moreover, if $\mathbf{x}$ is interior to this polytope then the inequalities (1.6) must be strict inequalities so that all can hold only for a single i. It follows that the k polytopes defined in this paragraph by a fixed $\mathbf{x}^{(j)}$ constitute a partition $\mathscr{P}_j$ of S_k into k convex polytopes.

To simplify the following discussion it will be assumed that each $\mathbf{x}^{(j)}$ is interior to S_k so that each of the k polytopes in $\mathscr{P}_j$ has positive content. It will be assumed further that the points $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \cdots, \mathbf{x}^{(n)}$ are arranged such that every nonempty polytope in the product $\mathscr{P}$ of $\mathscr{P}_1, \mathscr{P}_2, \cdots, \mathscr{P}_n$ has positive content. These assumptions clearly hold for the examples illustrated in Figures 1 and 2. To show that they pose no difficulty for the subsequent distribution theory of this paper, it will now be shown that the assumptions hold with probability one when $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \cdots, \mathbf{x}^{(n)}$ are continuously distributed over the n-fold product of S_k with itself. The result clearly holds for $n = 1$. Assume it holds for some $n \geqq 1$, so that $\mathscr{P}$ as defined consists with probability one of a finite number of polytopes each having positive content. The product of $\mathscr{P}$ and $\mathscr{P}_{n+1}$ can fail to retain this property only if at least one of the ratios $x_i^{(n+1)}/x_h^{(n+1)}$ coincides with a corresponding ratio $x_i^{(j)}/x_h^{(j)}$ for some j on $1, 2, \cdots, n$, because these ratios define the bounding hyperplanes of the polytopes. Since such a coincidence occurs with probability zero, the product of $\mathscr{P}$ and $\mathscr{P}_{n+1}$ has, like $\mathscr{P}$, the desired property with probability one.

The following three lemmas are sufficient to establish a one-one correspondence among partitions $\mathbf{n}$ of n, regions $R(\mathbf{n})$ as defined, and polytopes in the product $\mathscr{P}$ of $\mathscr{P}_1, \mathscr{P}_2, \cdots, \mathscr{P}_n$.

LEMMA 1.1. *Any polytope in $\mathscr{P}$ is entirely contained in some $R(\mathbf{n})$.*

LEMMA 1.2. *Two different polytopes in $\mathscr{P}$ are not contained in the same $R(\mathbf{n})$.*

LEMMA 1.3. *The region $R(\mathbf{n})$ is nonempty for every $\mathbf{n}$.*

The proof of Lemma 1.1 is immediate, for the definition of any region of $\mathscr{P}$ as $\bigcap_1^n P_j$ and the definition of P_j from (1.7) together ensure that all of the points $\bigcap_1^n P_j$ belong to the same $R(\mathbf{n})$ *by virtue even of the same partition of* $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \cdots, \mathbf{x}^{(n)}$ *into k cells.*

Suppose that $\mathbf{y}$ and $\mathbf{z}$ are interior points of two different polytopes in $\mathscr{P}$. Accordingly, there must exist a point $\mathbf{x}^{(j)}$ and cells i and h such that $\mathbf{x}^{(j)}$ is in cell i relative to $\mathbf{y}$ and in cell h relative to $\mathbf{z}$, so that $y_h/x_h^{(j)} > y_i/x_i^{(j)}$ and $z_h/x_h^{(j)} < z_i/x_i^{(j)}$ from which

$$y_h/z_h > y_i/z_i \,. \tag{1.8}$$

If the points $\mathbf{x}^{(j)}$ distribute themselves according to $\mathbf{n}$ relative to $\mathbf{y}$, and redistribute themselves according to $\mathbf{n}^*$ relative to $\mathbf{z}$, then (1.8) asserts that every $\mathbf{x}^{(j)}$ which shifts cells does so monotonely relative to the ordering of the cells

defined by the values y_i/z_i. Such a monotone shifting is not consistent with $\mathbf{n} = \mathbf{n}^*$, thus proving Lemma 1.2.

To prove Lemma 1.3 the following construction process can be used. Write

$$
\begin{aligned}
\mathbf{x} &= [x_1, x_2, \cdots, x_k] \\
&= [x_1, (1 - x_1)x_{22}, (1 - x_1)x_{23}, \cdots, (1 - x_1)x_{2k}] \\
&= [x_1, (1 - x_1)x_{22}, (1 - x_1)(1 - x_{22})x_{33}, \cdots, (1 - x_1)(1 - x_{22})x_{3k}] \\
&\ \ \vdots \\
&= [x_1, (1 - x_1)x_{22}, (1 - x_1)(1 - x_{22})x_{33}, (1 - x_1)(1 - x_{22})(1 - x_{33})x_{44}, \cdots, \\
&\qquad (1 - x_1)(1 - x_{22}) \cdots (1 - x_{k-1,k-1})x_{kk}] \,.
\end{aligned} \tag{1.9}
$$

Note that $x_{22} + x_{23} + \cdots + x_{2k} = 1$, $x_{33} + \cdots + x_{3k} = 1$, $\cdots$, $x_{kk} = 1$. For fixed $x_{22}, \cdots, x_{2k}$ the curve of points $\mathbf{x}$ as x_1 ranges from 0 to 1 passes through regions $R(\mathbf{n})$ where n_1 varies monotonely from 0 to n, with any given n_1 achieved for some x_1. Define $x_1^* = x_1^*(x_{22}, \cdots, x_{2k})$ to be the smallest x_1 for which $R(\mathbf{n})$ with the given n_1 is encountered. For fixed $x_{33}, \cdots, x_{3k}$ a curve of points is defined as x_2 ranges from 0 to 1 and x_1 is taken to be $x_1^*(x_2, (1 - x_2)x_{33}, \cdots, (1 - x_2)x_{3k})$. This curve passes through regions $R(\mathbf{n})$ for which n_1 has the desired value while n_2 varies monotonely from 0 to $n - n_1$. Define $x_2^{**} = x_2^{**}(x_{33}, \cdots, x_{3k})$ to be the smallest x_2 for which $R(\mathbf{n})$ with the desired n_1 and n_2 is achieved. Continuing in this way, one eventually constructs a point in $R(\mathbf{n})$ for given $\mathbf{n}$.

For certain purposes it may be convenient to replace the coordinates $\mathbf{x}$ of any interior point of S_k by coordinates

$$\mathbf{t} = [t_1, t_2, \cdots, t_k] \tag{1.10}$$

satisfying

$$-\infty < t_i < \infty \quad \text{for} \quad i = 1, 2, \cdots, k \qquad \text{and} \qquad \textstyle\sum_{i=1}^{k} t_i = 0 \,, \tag{1.11}$$

where the relations between $\mathbf{x}$ and $\mathbf{t}$ are given by

$$t_i = \log x_i - k^{-1} \textstyle\sum_{h=1}^{k} \log x_h \tag{1.12}$$

and

$$x_i = e^{t_i} / \textstyle\sum_{h=1}^{k} e^{t_h} \,, \tag{1.13}$$

for $i = 1, 2, \cdots, k$. The relations (1.12) or (1.13) thus define a one-one correspondence between interior points $\mathbf{x}$ of S_k and points $\mathbf{t}$ in an ordinary $(k - 1)$-space, say E_k.

Since the convex polytopes $R(\mathbf{n})$ in S_k are defined by inequalities of the form $x_i/x_h \leqq c$, the corresponding regions in E_k are defined by inequalities of the form $t_i - t_h \leqq \log c$, and it is clear therefore that the regions are convex polytopes $F(\mathbf{n})$ in E_k. The regions $F(\mathbf{n})$ have the conceptually simplifying property that they are defined by hyperplanes belonging only to one of $k(k - 1)/2$ families

of parallel hyperplanes. Their disadvantage lies mainly in the unfamiliar nature of the representation $\mathbf{t}$ of stochastic vectors $\mathbf{x}$.

2. Simple properties of a random $R(\mathbf{n})$. Suppose that $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \cdots, \mathbf{x}^{(n)}$ are independently and uniformly distributed over S_k. For a fixed $\mathbf{n}$, the region $R(\mathbf{n})$ determined by $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \cdots, \mathbf{x}^{(n)}$ becomes a random convex polytope in S_k. From Section 1 it is known that $R(\mathbf{n})$ has positive content with probability one. Several other elementary stochastic properties of $R(\mathbf{n})$ are collected in this section.

The simplest of these is

$$P(\mathbf{x} \in R(\mathbf{n})) = \frac{n!}{n_1!\, n_2! \cdots n_k!}\, x_1^{n_1} x_2^{n_2} \cdots x_k^{n_k}, \tag{2.1}$$

which is a consequence of the remarks following (1.3) together with the definition of $R(\mathbf{n})$.

If the fixed $\mathbf{x}$ is replaced by a random $\mathbf{y}$ distributed independently of $R(\mathbf{n})$, then from (2.1) it follows that

$$P(\mathbf{y} \in R(\mathbf{n})) = \int_{S_k} \frac{n!}{n_1!\, n_2! \cdots n_k!}\, x_1^{n_1} x_2^{n_2} \cdots x_k^{n_k} f(\mathbf{x})\, d\mathbf{x}, \tag{2.2}$$

where $f(\mathbf{x})$ denotes the density of $\mathbf{y}$ over S_k. In particular, if

$$f(\mathbf{x}) = \frac{\Gamma(r)}{\Gamma(r_1)\,\Gamma(r_2) \cdots \Gamma(r_k)}\, x_1^{r_1-1} x_2^{r_2-1} \cdots x_k^{r_k-1}, \tag{2.3}$$

with $r_i > 0$ for $i = 1, 2, \cdots, k$ and $\sum_1^k r_i = r$, then (2.2) becomes

$$P(\mathbf{y} \in R(\mathbf{n})) = \prod_{i=1}^k \frac{\Gamma(n_i + r_i)}{n_1!\,\Gamma(r_i)} \Big/ \frac{\Gamma(n + r)}{n!\,\Gamma(r)}. \tag{2.4}$$

Specializing further to $r_1 = r_2 = \cdots = r_k = 1$, which is a uniform distribution for $\mathbf{y}$, yields

$$P(\mathbf{y} \in R(\mathbf{n})) = \binom{n+k-1}{k-1}^{-1}. \tag{2.5}$$

Formula (2.5) can be interpreted as $E\{P(\mathbf{y} \in R(\mathbf{n}) \mid R(\mathbf{n}))\}$ and therefore gives the expected volume of the random region $R(\mathbf{n})$. It is interesting that each of the regions $R(\mathbf{n})$ as $\mathbf{n}$ ranges over the $\binom{n+k-1}{k-1}$ partitions of n has the same average volume.

Replacing the fixed $\mathbf{x} \in S_k$ by a fixed set $T \subset S_k$, one can generalize (2.1) to

$$\begin{aligned} P(R(\mathbf{n}) \supset T) = \frac{n!}{n_1!\, n_2! \cdots n_k!} &\{1 + b_{12} + b_{13} + \cdots + b_{1k}\}^{-n_1} \\ &\{b_{21} + 1\, b + b_{23} + \cdots + b_{2k}\}^{-n_2} \cdots \\ &\{b_{k1} + b_{k2} + \cdots + 1\}^{-n_k} \end{aligned} \tag{2.6}$$

where

$$b_{ih} = \sup_{\mathbf{x} \in T} \{x_h / x_i\} \tag{2.7}$$

for i and $h = 1, 2, \cdots, k$ and $i \neq h$. From (1.6), $\mathbf{x}^{(j)}$ belongs to cell i for every $\mathbf{x} \in T$ if and only if

$$x_h^{(j)}/x_i^{(j)} \geqq b_{ih} \tag{2.8}$$

for $h = 1, 2, \cdots, k$ and $h \neq i$. The set of points $\mathbf{x}^{(j)}$ satisfying (2.8) consists of the subsimplex of S_k whose vertices are the some as the vertices of S_k except that $[\delta_{i1}, \delta_{i2}, \cdots, \delta_{ik}]$ is replaced by the point such that all of the inequalities (2.8) become equalities, i.e., the point with coordinates proportional to $[b_{i1}, b_{i2}, \cdots, 1, \cdots, b_{ik}]$. The volume of this simplex is $\{b_{i1} + b_{i2} + \cdots + 1 + \cdots + b_{ik}\}^{-1}$, and accordingly this is the probability that $\mathbf{x}^{(j)}$ lies in cell i for all $\mathbf{x} \in T$. Formula (2.6) follows immediately.

Any convex polytope R of the type studied in this paper is completely defined by the $k(k-1)$ coordinates b_{ih} as defined by (2.7) with R in the role of T. Denote by B_{ih} the random coordinates of the random polytope $R(\mathbf{n})$. Then the right side of (2.6) is seen to express the cumulative distribution function

$$\begin{aligned} F(b_{ih}\,;\, i \quad &\text{and} \quad h = 1, 2, \cdots, k \quad \text{and} \quad i \neq h) \\ &= P(B_{ih} \geqq b_{ih}\,;\, i \quad \text{and} \quad h = 1, 2, \cdots, k \quad \text{and} \quad i \neq h) \end{aligned} \tag{2.9}$$

defined over the region in $k(k-1)$-space such that the coordinates b_{ih} define a polytope R, i.e., the region defined by inequalities of the form $b_{i\ell} \leqq b_{ih} b_{h\ell}$.

Unfortunately, the formal characterization of the distribution of $R(\mathbf{n})$ through the cdf of (2.6) and (2.9) does not lead to easy analytical computations of interesting properties of the distribution, such as the distribution of specified vertices of $R(\mathbf{n})$ assayed in Section 3. The reason for the difficulties is that positive probability resides on hyperplanes of dimension less than $k(k-1)$ in the space of the b_{ih}, or in other words some of the inequalities $b_{i\ell} \leqq b_{ih} b_{h\ell}$ become equalities with positive probability. For example, if $k = 3$ the region $R(\mathbf{n})$ may have 6, 5, 4 or 3 sides each with finite probability, as the reader may easily check by experimenting with drawings as in Figure 2. It is a plausible conjecture that the number of hyperlanes required to bound $R(\mathbf{n})$ can be any number between k and $k(k-1)$ with positive probability.

Finally, suppose that $R(\mathbf{m})$ and $R(\mathbf{n})$ are independent random regions based on given partitions $\mathbf{m}$ and $\mathbf{n}$ of m and n. Then

$$P(R(\mathbf{m}) \cap R(\mathbf{n}) \neq \varnothing) = \prod_{i=1}^{k} \binom{m_i + n_i}{m_i}^{-1} \tag{2.10}$$

and the conditional distribution of $R(\mathbf{m}) \cap R(\mathbf{n})$ given that it is not empty is the distribution of $R(\mathbf{m} + \mathbf{n})$. To prove these results, consider first a fixed set of $m + n$ points $\mathbf{z}^{(1)}, \mathbf{z}^{(2)}, \cdots, \mathbf{z}^{(m+n)}$ such that the associated $R(\mathbf{m} + \mathbf{n})$ has positive content, and suppose that $R(\mathbf{m})$ and $R(\mathbf{n})$ are determined by independent uniformly distributed points $\mathbf{y}^{(1)}, \mathbf{y}^{(2)}, \cdots, \mathbf{y}^{(m)}$ and $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \cdots, \mathbf{x}^{(n)}$, respectively. Given that the random points $\mathbf{y}^{(1)}, \mathbf{y}^{(2)}, \cdots, \mathbf{y}^{(m)}, \mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \cdots, \mathbf{x}^{(n)}$ coincide with $\mathbf{z}^{(1)}, \mathbf{z}^{(2)}, \cdots, \mathbf{z}^{(m+n)}$ *in some unspecified order*, the conditional probability that the

$m_i + n_i$ points in cell i in the definition of $R(\mathbf{m} + \mathbf{n})$ consist of m_i points from the $\mathbf{y}$ sequence and n_i points from the $\mathbf{x}$ sequence, for $i = 1, 2, \cdots, k$, is the expression (2.10). But it is easily checked that $R(\mathbf{m})$ and $R(\mathbf{n})$ intersect if and only if they intersect in $R(\mathbf{m} + \mathbf{n})$, so that (2.10) holds conditionally on a set $\{\mathbf{z}^{(1)}, \mathbf{z}^{(2)}, \cdots, \mathbf{z}^{(m+n)}\}$ which has probability one, a result stronger than (2.10). Since the distribution of $\mathbf{z}^{(1)}, \mathbf{z}^{(2)}, \cdots, \mathbf{z}^{(m+n)}$ is independent uniform conditional on any assignment of $\mathbf{x}$'s and $\mathbf{y}$'s to $\mathbf{z}$'s, the distribution of $R(\mathbf{m} + \mathbf{n})$ is the ordinary distribution even conditional on the assignment which makes

$$R(\mathbf{m}) \cap R(\mathbf{n}) = R(\mathbf{m} + \mathbf{n}) ,$$

as required to complete the proof.

3. Random vertices of $R(\mathbf{n})$. Most of the distribution problems associated with applications of $R(\mathbf{n})$ have so far proved intractable. To illustrate the complexities encountered, an attempt will be made in this section to compute the distributions of certain extremal vertices of $R(\mathbf{n})$. Suppose that $\mathbf{x}$ is a fixed point of S_k and $\mathbf{ax}' = 0$ is the equation of a fixed hyperplane through $\mathbf{x}$. (Since $\mathbf{x1}' = 1$, the coefficients $\mathbf{a}$ can be adjusted so that $\mathbf{ax}'$ takes the value zero.) For almost all $\mathbf{a}$, the random $R(\mathbf{n})$ has a unique vertex $\mathbf{v}$ such that

$$\mathbf{av}' = \max_{\mathbf{u} \in R(\mathbf{n})} \{\mathbf{au}'\} . \tag{3.1}$$

The problem posed is to find the probability density function of the random vertex $\mathbf{v}$ at a specific point on the hyperplane $\mathbf{ax}' = 0$.

Suppose that $R(\mathbf{n})$ is determined as in Section 1 to be a member of the partition $\mathscr{P}$ defined by $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \cdots, \mathbf{x}^{(n)}$. It is evident that, with probability one, every vertex of $R(\mathbf{n})$ is defined by an intersection of $k - 1$ hyperplanes from the family of hyperplanes defining $\mathscr{P}$. In terms of $R(\mathbf{n})$, this means that the vertex $\mathbf{v}$ is defined, with probability one, to be the intersection of $k - 1$ hyperplanes defining actual faces of $R(\mathbf{n})$, i.e., faces with positive $(k - 2)$-dimensional content. These faces are defined by a subset of $(k - 1)$ of the coordinates b_{ih} of $R(\mathbf{n})$, i.e., $\mathbf{v}$ satisfies

$$v_h/v_i = b_{ih} \tag{3.2}$$

for some set of $(k - 1)$ ordered pairs (i, h). In order that the $(k - 1)$ hyperplanes defined by the $(k - 1)$ equations (3.2) intersect in a single point, it is necessary and sufficient that the $(k - 1)$ subscript pairs (i, h) form the branches of a tree joining nodes labelled $1, 2, \cdots, k$. Consequently, one need only seek among trees for potential vertices.

Moreover, if $\mathscr{T}$ is any tree where the $(k - 1)$ branches are not ordered, then for almost all $\mathbf{a}$ there is a unique way to assign an order to the pair of nodes defining each branch such that the vertex defined by the $(k - 1)$ equations (3.2) using the ordered pairs (i, h) is a possible extremal vertex of the

type (3.1). To be a potential vertex of the desired kind it is necessary and sufficient that every $\mathbf{u}$ which satisfies

$$u_h/u_i < b_{ih} \tag{3.3}$$

for pairs (i, h) in the ordered tree (i.e., $\mathbf{u}$ is potentially inside $R(\mathbf{n})$ according to (2.7)) must also satisfy

$$\mathbf{au}' < \mathbf{av}' , \tag{3.4}$$

where $\mathbf{v}$ is the point of intersection of the $k-1$ hyperplanes (i.e., the vertex defined by (3.1)). The condition can also be stated in terms of the lines of intersection of the $(k-1)$ hyperplanes taken $(k-2)$ at a time. The halves of these lines which bound the orthant of points satisfying (3.3) must all lie on the same side of the hyperplane $\mathbf{ax}' = 0$, specifically must all consist of points $\mathbf{u}$ satisfying (3.4). The condition (3.4) thus implies the particular ordering which must be assigned to every unordered pair (i, h) in a given tree $\mathscr{T}$, to wit: *the pair (i, h) partitions the nodes into subset I and H where I consists of i and the nodes connected to i when the branch (i, h) is dropped, and H consists of h and the nodes connected to h when branch (i, h) is dropped; the particular direction (i, h) must be chosen which makes*

$$\textstyle\sum_I a_i v_i = -\sum_H a_i v_i < 0 \tag{3.5}$$

at the vertex $\mathbf{v}$. The reasoning is that along the half-line corresponding to the desired ordering (i, h) the ratios among coordinates u_ℓ with $\ell \in I$ are fixed, and likewise the ratios among coordinates u_ℓ with $\ell \in H$ are fixed. Thus, as one moves away from the vertex v in the desired direction, the u_ℓ with $\ell \in H$ must decrease in a fixed ratio while the u_ℓ with $\ell \in I$ increase in a fixed ratio, and at the same time the overall $\mathbf{au}'$ is decreasing. The combination of conditions (3.3) and (3.4) thus yields (3.5) as claimed.

The density of $\mathbf{v}$ is the sum of the contributions from each tree. For a particular tree, the density is easily computable, not in terms of the original coordinates $\mathbf{v}$, but rather in terms of a coordinate system special to that tree. An additional step is required to transform back to the coordinates $\mathbf{v}$.

Suppose that $\mathscr{T}$ is a tree and $\mathscr{T}_0$ denotes the set of $(k-1)$ ordered pairs (i, h) which define the appropriate ordered tree as defined above. Suppose that the vertex $\mathbf{v}$ is represented by the $(k-1)$ coordinates

$$b_{ih} = v_h/v_i \qquad \text{for} \quad (i, h) \in \mathscr{T}_0 . \tag{3.6}$$

The contribution to the density of $\mathbf{v}$ from the tree $\mathscr{T}$, expressed in terms of the coordinates (3.6), is found by the following operations on the expression (2.6):

(i) Differentiate with respect to b_{ih} for each pair $(i, h) \in \mathscr{T}_0$,

and

(ii) Replace each b_{ih} for $(i, h) \notin \mathscr{T}_0$ by the expression for v_h/v_i in terms of b_{ih} with $(i, h) \in \mathscr{T}_0$ as defined by the inverse of (3.6). Step (i) here relies on the fact that (2.6) is essentially a cumulative distribution function as in (2.9), so that its derivative with respect to appropiate coordinates defines the desired density. See Appendix A of [5] for a more detailed discussion of this point.

Step (ii) merely expresses the density in a consistent set of coordinates.

Since the coordinates (3.6) are to be replaced immediately by the coordinates $\mathbf{v}$, step (ii) above need not actually be carried out, and the conversion to $\mathbf{v}$ terms is easy because the expressions

$$\{1 + b_{12} + b_{13} + \cdots + b_{1k}\}^{-1}, \{b_{21} + 1 + b_{23} + \cdots + b_{2k}\}^{-1}, \cdots, \{b_{k1} + b_{k2} + \cdots + 1\}^{-1}$$

reduce simply to $v_1, v_2, \cdots, v_k$. The Jacobian of the transformation (3.6) is easily checked to be

$$J = \prod_{(i,h) \in T_0} \frac{v_h}{v_i} \Big/ \prod_{\ell=1}^{k} v_\ell \,. \tag{3.7}$$

The results of the two preceding paragraphs may be summarized in analytic terms as follows. Associated with the ordered tree $\mathscr{T}_0$ is a pair of vectors

$$\mathbf{r} = [r_1, r_2, \cdots, r_k] \quad \text{and} \tag{3.8}$$
$$\mathbf{s} = [s_1, s_2, \cdots, s_k] \,,$$

where r_i denotes the number of times index i appears as the first member of a pair $(i, h) \in \mathscr{T}_0$ and s_h denotes the number of times h appears as the second member of such a pair. The differentiation in step (i) above has the effect of replacing the exponents $-n_1, -n_2, \cdots, -n_k$ in (2.6) by $-n_1 - r_1, -n_2 - r_2, \cdots, -n_k - r_k$ and of multiplying the constant term by

$$C(\mathbf{n}, \mathbf{r}) = \prod_{\ell=1}^{k} \prod_{m_\ell=0}^{r_\ell - 1} (n_\ell + m_\ell) \,, \tag{3.9}$$

where signs are dropped and factors corresponding to $r_\ell = 1$ are unity. Substituting for $\mathbf{v}$ and multiplying by the Jacobian J yields the formula

$$g(\mathbf{v}) = C(\mathbf{n}, \mathbf{r}) \cdot \frac{n!}{\prod_{\ell=1}^{k} n_\ell!} \cdot \prod_{\ell=1}^{k} v_\ell^{\,n_\ell + s_\ell - 1} \tag{3.10}$$

for the density contribution from the tree $\mathscr{T}_0$.

EXAMPLE. Take $k = 4$ and general $\mathbf{n} = [n_1, n_2, n_3, n_4]$, and ask for the vertex $\mathbf{v}$ such that $v_2 + v_3 + v_4$ is maximum or v_1 is minimum. The canonical form of the hyperplane $\mathbf{av}' = 0$ has three positve terms $a_2 v_2, a_3 v_3, a_4 v_4$ and one negative term $a_1 v_1$, and the rule (3.5) is easily applied to yield for each of the 16 trees $\mathscr{T}$ the directed trees $\mathscr{T}_0$ shown in Table I. Since $\mathbf{s}$ is the same for each tree, each contribution from (3.10) is of the form $K v_1^{n_1 - 1} v_2^{n_2} v_3^{n_3} v_4^{n_4}$ and the

TABLE I

The directed trees $\mathscr{T}_0$ and their associated **r** *and* **s** *for the vertex with maximum $v_2 + v_3 + v_4$ when $k = 4$*

$\mathscr{T}_0$	$[r_1, r_2, r_3, r_4]$	$[s_1, s_2, s_3, s_4]$
(1, 2) (3, 4) (2, 3)	[1, 1, 1, 0]	[0, 1, 1, 1]
(1, 3) (2, 4) (3, 2)	[1, 1, 1, 0]	[0, 1, 1, 1]
(1, 4) (3, 2) (4, 3)	[1, 0, 1, 1]	[0, 1, 1, 1]
(1, 4) (2, 3) (4, 2)	[1, 1, 0, 1]	[0, 1, 1, 1]
(1, 2) (4, 3) (2, 4)	[1, 1, 0, 1]	[0, 1, 1, 1]
(1, 3) (4, 2) (3, 4)	[1, 0, 1, 1]	[0, 1, 1, 1]
(1, 2) (3, 4) (1, 3)	[2, 0, 1, 0]	[0, 1, 1, 1]
(1, 3) (2, 4) (1, 2)	[2, 1, 0, 0]	[0, 1, 1, 1]
(2, 3) (1, 4) (1, 2)	[2, 1, 0, 0]	[0, 1, 1, 1]
(3, 2) (1, 4) (1, 3)	[2, 0, 1, 0]	[0, 1, 1, 1]
(1, 2) (4, 3) (1, 4)	[2, 0, 0, 1]	[0, 1, 1, 1]
(4, 2) (1, 3) (1, 4)	[2, 0, 0, 1]	[0, 1, 1, 1]
(1, 2) (1, 3) (1, 4)	[3, 0, 0, 0]	[0, 1, 1, 1]
(1, 2) (2, 3) (2, 4)	[1, 2, 0, 0]	[0, 1, 1, 1]
(1, 3) (3, 2) (3, 4)	[1, 0, 2, 0]	[0, 1, 1, 1]
(1, 4) (4, 2) (4, 3)	[1, 0, 0, 2]	[0, 1, 1, 1]

sum of the 16 contributions must yield the Dirichlet density

$$(3.11)\qquad D(n_1, n_2 + 1, n_3 + 1, n_4 + 1) = \frac{\Gamma(n + 3)}{\Gamma(n_1)\,\Gamma(n_2 + 1)\,(\Gamma n_3 + 1)\,\Gamma(n_4 + 1)}\, v_1^{n_1 - 1}\, v_2^{n_2}\, v_3^{n_3}\, v_4^{n_4}\,.$$

As a check, the 16 normalizing factors $C(\mathbf{n}, \mathbf{r})$ may be computed and summed, yielding

$$\begin{aligned} &n_1n_2n_3 + n_1n_2n_3 + n_1n_3n_4 + n_1n_2n_4 \\ &\quad + n_1n_2n_4 + n_1n_3n_4 + n_1(n_1 + 1)n_3 + n_1(n_1 + 1)n_2 \\ &\quad + n_1(n_1 + 1)n_2 + n_1(n_1 + 1)n_3 + n_1(n_1 + 1)n_4 + n_1(n_1 + 1)n_4 \\ &\quad + n_1(n_1 + 1)(n_1 + 2) + n_1n_2(n_2 + 1) + n_1n_3(n_3 + 1) + n_1n_4(n_4 + 1) \\ &\qquad = n_1\{(n_1 + n_2 + n_3 + n_4)^2 + 3(n_1 + n_2 + n_3 + n_4) + 2\} \\ &\qquad = n_1(n + 1)(n + 2)\,. \end{aligned}$$

Multiplying this sum by the combinations factor $n!/\prod_1^4 n_i!$ yields the normalizing factor in (3.11) as desired. The density of the random vertex with maximum coordinate v_1 may also be found from Table I, for the revised table appropriate to maximizing v_1 instead of minimizing v_1 merely replaces every branch (i, h) with (h, i) and so interchanges **r** and **s**. The resulting density is

$$(3.12)\qquad D(n_1 + 1, n_2, n_3, n_4) = \frac{\Gamma(n + 1))}{\Gamma(n_1 + 1)\,\Gamma(n_2)\,\Gamma(n_3)\,\Gamma(n_4)}\, v_1^{n_1}\, v_2^{2-1}\, v_3^{n_3 - 1}\, v^{n_4 - 1}\,.$$

The reader may easily check that the result just given for $k = 4$ extends immediately to general k. Specifically, the densities of the vertices with minimum and maximum v_1 are $D(n_1, n_2 + 1, n_3 + 1, \cdots, n_k + 1)$ and $D(n_1 + 1, n_2, n_3, \cdots, n_k)$, respectively. The key to this result is the validity of the obvious generalization of the last column of Table I.

The author has not found any other instances of vertices with Dirichlet densities, although it is obvious that any vertex must have a distribution which is a mixture of Dirichlet densities, at least locally.

One difficulty in computing the density of the vertex associated with general $\mathbf{a}$ is that the condition (3.5) changes as $\mathbf{v}$ moves among subregions of S_k, even subregions within the same hyperplane $\mathbf{av}' = 0$. Consequently, the ordered tree $\mathscr{T}_0$ corresponding to a given nonordered tree varies, and the analytic expression for the density varies likewise. Under the exponential coordinates defined in (1.10), (1.11) and (1.12), however, the difficulty does not arise. To see this, consider the vertex with maximum $\mathbf{ct}'$ where $\mathbf{c}$ is a vector of constants. Since $\sum t_i = 0$, adding a constant to each component of $\mathbf{c}$ does not affect $\mathbf{ct}'$, and $\mathbf{c}$ may therefore be chosen such that $\sum c_i = 0$. In terms of the original coordinates, the surface $\mathbf{ct}' = \mathbf{ct}_0'$ is a curved surface which passes through the point $\mathbf{x}_0$ with exponential coordinates $\mathbf{t}_0$ and whose tangent hyperplane at $\mathbf{x}_0$ has the equation

$$\sum_{i=1}^{k} \{c_i / x_{i0}\} x_i = 0 . \tag{3.13}$$

The desired extremal vertex at x_0 is therefore also extremal relative to the hyperplane $\mathbf{ax}' = 0$ where $a_i = c_i / x_{i0}$ for $i = 1, 2, \cdots, k$. Thus the theory developed above can be used, and in particular the condition (3.5) reduces to

$$\sum_I c_i = -\sum_H c_i < 0 , \tag{3.14}$$

so that the same $\mathscr{T}_0$ and the same analytic expression for the density holds, given $\mathbf{c}$, at all points $\mathbf{v}$ in S_k.

Since the exponential coordinates lead to a simplification, it is natural to seek to employ them in applications. In terms of the application to estimating parameters of frequency distributions, the corresponding approach is to use parameters linear in the log frequency. Thus, one may expect the theory to be relatively simple for estimating the parameters of exponential families of distributions such as are already widely used in statistical theory. Further progress in such applications will require, however, an ability to compute probabilities that the random polytopes intersect hyperplanes and subregions of hyperplanes defined in terms of the exponential coordinates $\mathbf{t}$. The present work is but an introduction to these problems.

REFERENCES

[1] DEMPSTER, A. P. (1966). New methods for reasoning towards posterior distributions based on sample data. *Ann. Math. Statist.* **37** 355–374.

[2] DEMPSTER, A. P. (1967 a). Upper and lower probabilities induced by a multivalued mapping. *Ann. Math. Statist.* **38** 325–339.

[3] DEMPSTER, A. P. (1967 b). Upper and lower probability inferences based on a sample from a finite univariate population. *Biometrika* **54** 515–528.

[4] DEMPSTER, A. P. (1968 a). Upper and lower probabilities generated by a random closed interval. *Ann. Math. Statist.* **39** 957–966.

[5] DEMPSTER, A. P. (1968 b). A generalization of Bayesian inference (with discussion). *J. Roy. Statist. Soc. Ser. B* **30** 205–247.

[6] DEMPSTER, A. P. (1969). Upper and lower probability inferences for families of hypotheses with monotone density ratios. *Ann. Math. Statist.* **40** 953–969.