The Annals of Probability
1992, Vol. 20, No. 3, 1167–1173

BINOMIAL MIXTURES AND FINITE EXCHANGEABILITY

By G. R. Wood

University of Canterbury

We answer two questions: "What is the probability that a randomly chosen distribution function on $\{0, 1, \dots, n\}$ is a mixture of binomial distributions?" and "What is the probability that an n-exchangeable sequence is the initial segment of an infinite exchangeable sequence?" Curiously, the answers are the same.

1. Introduction. Suppose that we choose at random a probability distribution function on $\{0, 1, \dots, n\}$ for some fixed natural number n. In the first half of this paper, we calculate the probability that the distribution chosen will be a mixture of binomial distributions. We set the problem geometrically, and so introduce a probability model which makes the problem precise. Results are summarized in Theorem 1 and explicit numerical values are given in Table 1.

A closely related problem is discussed in the second half of the paper. Here we choose at random an n-exchangeable sequence of 0–1 random variables and calculate the probability that it is the initial segment of an infinite exchangeable sequence of 0–1 random variables. We review the geometry of this problem, as introduced in Diaconis (1977), and present the main results in Theorem 2. This is followed by Table 2 which shows the probabilities as limits of values presented in Crisma (1982).

2. Binomial mixtures. Let $h(i|p)$ be the binomial probability distribution function, with parameters n, a natural number and p, $0 \le p \le 1$. A probability distribution $\mathbf{p} = (p_0, p_1, \dots, p_n)$ on $\{0, 1, \dots, n\}$ is a mixture of binomial distributions if

$$p_i = \int_0^1 h(i|p)\, dG(p) \quad \text{for } i = 0, 1, \dots, n, \tag{1}$$

for some cumulative distribution function G on $[0, 1]$.

For fixed n and p, we may view $h(i|p)$ as a point in R^{n+1}, the $(n+1)$-tuple

$$(h(0|p), h(1|p), \dots, h(n|p)).$$

If we now allow p to move from 0 to 1, these points trace out a "binomial curve" B_n in the simplex $T_n = \{x = (x_0, x_1, \dots, x_n)\colon \sum_{i=0}^n x_i = 1 \text{ and } x_i \ge 0$ for all $i\}$ of all probability distribution functions on $\{0, 1, \dots, n\}$. Mixtures $\mathbf{p}$, as defined in (1), are then seen geometrically as elements of the convex hull of the

Received May 1990; revised April 1991.

AMS 1980 *subject classifications*. Primary 60E05, 60G09; secondary 60D05.

Key words and phrases. Binomial, mixture, finite exchangeability, infinite exchangeability, de Finetti, moment curve, simplex, volume.

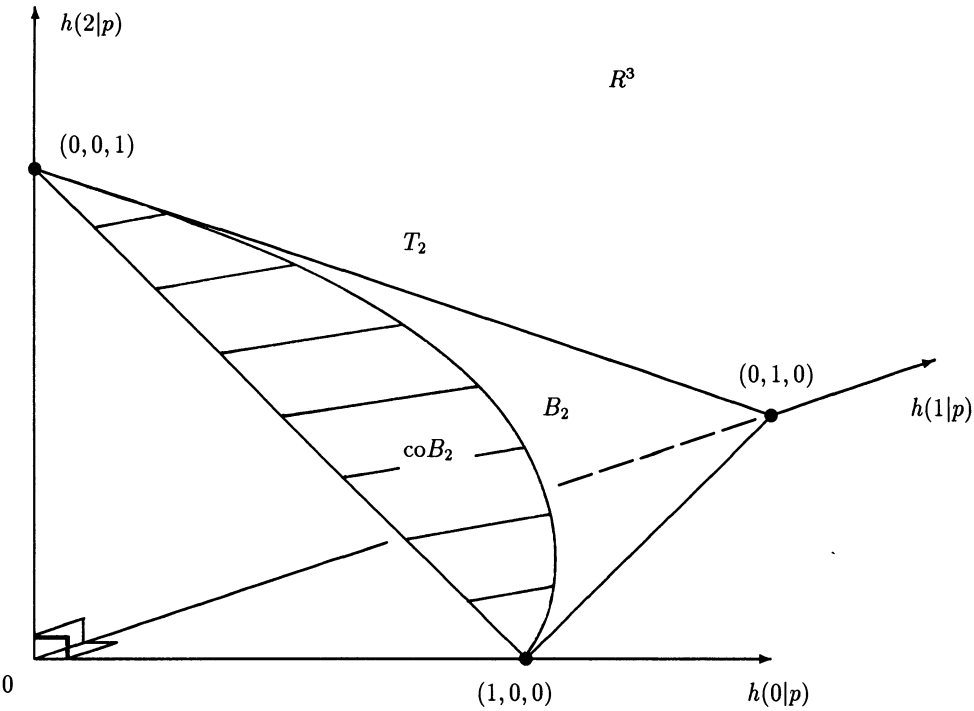

FIG. 1. *The simplex T_2 of probability distribution functions on* $\{0,1,2\}$, *the curve of binomial probability distributions B_2 and the mixtures of binomial distributions,* co B_2, *shaded.*

binomial curve, denoted co B_n. The sets T_2, B_2 and co B_2 are pictured in Figure 1. Here

$$B_2 = \left\{\left((1-p)^2, 2p(1-p), p^2\right): 0 \le p \le 1\right\}.$$

We consider T_n equipped with Lebesgue measure and assume that a distribution $\mathbf{p}$ is drawn from T_n according to this measure. In order to find the probability that $\mathbf{p}$ is a mixture of binomial distributions, we determine the n-dimensional volume of co B_n, denoted $V(\text{co } B_n)$, and compare it with the volume of the containing simplex, $V(T_n)$. The results are presented in the following theorem:

THEOREM 1.

(i) $$V(\text{co } B_n) = \sqrt{n+1}\,(n!)^{n-1} \prod_{k=1}^{n} \frac{1}{(2k-1)!}.$$

(ii) $$\frac{V(\text{co } B_n)}{V(T_n)} = (n!)^n \prod_{k=1}^{n} \frac{1}{(2k-1)!}.$$

(iii) *Given* $0 < r < 1$, *there exists an* n_0 *such that for* $n \geq n_0$,

$$\frac{V(\text{co } B_{n+1})}{V(T_{n+1})} < r\frac{V(\text{co } B_n)}{V(T_n)}.$$

Thus, a fortiori, $V(\text{co } B_n)/V(T_n) \to 0$ *as* $n \to \infty$.

Table 1 shows some values of the ratio $V(\text{co } B_n)/V(T_n)$ for low n and points to their rapid convergence to 0.

TABLE 1
The probability that a randomly chosen distribution on $\{0, 1, \dots, n\}$ *is a mixture of binomials*

n	$V(\text{co } B_n) / V(T_n)$
2	$0.\dot{6}$
3	0.3
4	9.14×10^{-2}
5	1.89×10^{-2}
6	2.65×10^{-3}
7	2.52×10^{-4}
8	1.63×10^{-5}
9	7.17×10^{-7}
10	2.14×10^{-8}
12	5.96×10^{-12}
15	1.53×10^{-18}

PROOF OF THEOREM 1. The proof rests on a single critical observation: The binomial curve B_n is an affine transform of the moment curve M_n in R^n, where

$$M_n = \{(p, p^2, \dots, p^n): 0 \leq p \leq 1\}.$$

An affine transform [see, e.g., McMullen and Shephard (1971), page 14] is the composition of a linear transformation and a translation. We show that $B_n = A(M_n)$, where the affine transformation A: $R^n \to R^{n+1}$ has the form

$$Ax = Lx + e_1 \quad \text{for each } x \in R^n,$$

with $e_1 = (1, 0, \dots, 0)$ the first standard basis vector in R^{n+1} and L the map taking R^n linearly onto the hyperplane H in R^{n+1} orthogonal to the equiangular vector, $(1, 1, \dots, 1)$. Specifically, L has matrix (l_{ij}) for $i = 0, \dots, n$, and $j = 1, \dots, n$, where

$$l_{ij} = \begin{cases} (-1)^{j-i}\binom{n}{i}\binom{n-i}{j-i}, & j \geq i, \\ 0, & j < i. \end{cases}$$

It is straightforward to confirm that $A(M_n) = B_n$, whence it follows that $A(\text{co } M_n) = \text{co } B_n$.

The n-dimensional volumes of co M_n and co B_n will be related as

$$V(\text{co } B_n) = |\det L'| V(\text{co } M_n),$$

where $\det L'$ is the determinant of L', the $(n+1) \times (n+1)$ matrix whose first n columns are those of L and whose last column is the normalized equiangular vector, namely $(1, \dots, 1)/\sqrt{n+1}$. Karlin and Shapley [(1953), Theorem 15.2] obtained the remarkable result that $V(\text{co } M_n) = \prod_{k=1}^{n}((k-1)!(k-1)!)/(2k-1)!$, so it remains to compute the determinant of L'. Fortunately, L' can readily be reduced to upper triangular form as follows. Replace the second row with the sum of the first and second rows, then replace the third row with the sum of the new second and third rows and so on. For $k = 2, \dots, n$, we find that $k-1$ of these elementary row operations brings

$$\sum_{i=0}^{k-1} (-1)^{k-i} \binom{n}{i} \binom{n-i}{k-i} = -\binom{n}{k}$$

to the kth diagonal position, which at the next row operation removes the value of $\binom{n}{k}$ lying directly beneath that diagonal element. The nth row operation produces an upper triangular matrix, with $(n+1)/\sqrt{n+1}$ in the $(n+1)$st diagonal position. Thus

$$\det L' = (-1)^n \sqrt{n+1} \prod_{k=1}^{n} \binom{n}{k}.$$

Multiplying the absolute value of this determinant and the expression for the volume of co M_n yields

$$\begin{aligned} V(\text{co } B_n) &= \sqrt{n+1} \prod_{k=1}^{n} \left[\frac{(k-1)!(k-1)!}{(2k-1)!} \frac{n!}{(n-k)!k!} \right] \\ &= \sqrt{n+1}\,(n!)^n \prod_{k=1}^{n} \frac{(k-1)!}{(n-k)!} \prod_{k=1}^{n} \frac{1}{k} \prod_{k=1}^{n} \frac{1}{(2k-1)!} \\ &= \sqrt{n+1}\,(n!)^{n-1} \prod_{k=1}^{n} \frac{1}{(2k-1)!}, \quad \text{giving (i)}. \end{aligned}$$

Further, $V(T_n) = \sqrt{n+1}/n!$, so $V(\text{co } B_n)/V(T_n) = (n!)^n \prod_{k=1}^{n} 1/(2k-1)!$, giving (ii).

Finally, let $a_n = (n!)^n \prod_{k=1}^{n} 1/(2k-1)!$. Then

$$\frac{a_{n+1}}{a_n} = \frac{n+1}{2n+1} \cdot \frac{n+1}{2n} \cdots \frac{n+1}{n+2} = b_n,$$

say. Now b_{n+1}/b_n can be shown to converge to $1/2$ as $n \to \infty$. Thus $b_n \downarrow 0$, whence (iii) follows. □

3. Extending finite exchangeable sequences. Let X_i, $i = 1, \dots, n$, be random variables taking only the values 0 or 1. Then $\{X_i\}_{i=1}^n$ are said to be n-exchangeable if for every fixed sequence of 0's and 1's $\{e_i\}_{i=1}^n$ and every permutation π of $\{1, 2, \dots, n\}$ we have

$$P(X_1 = e_1, \dots, X_n = e_n) = P(X_1 = e_{\pi(1)}, \dots, X_n = e_{\pi(n)}).$$

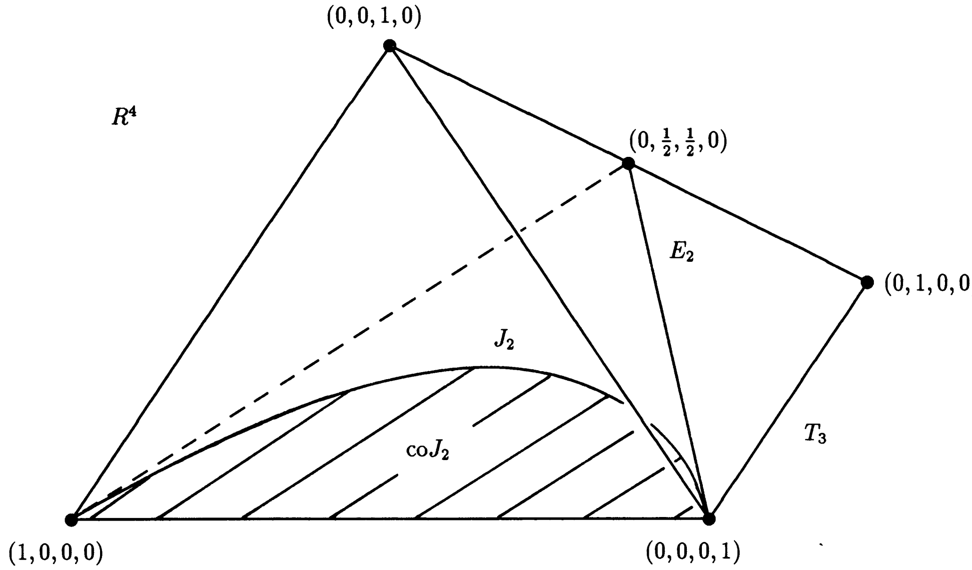

FIG. 2. *The 2-exchangeable sequences, E_2, and those which are infinitely extendible, the convex hull of the curve J_2 in E_2. All are contained in the simplex T_3 of all probability distributions on four points.*

An infinite sequence of 0–1 random variables $\{X_i\}_{i=1}^{\infty}$ is termed an infinite exchangeable sequence if $\{X_i\}_{i=1}^{n}$ is exchangeable for each n. The initial segment, $\{X_i\}_{i=1}^{n}$, of an infinite exchangeable sequence is certainly n-exchangeable. How likely is it that an n-exchangeable sequence is the initial segment of an infinite exchangeable sequence?

In Diaconis (1977), one view of the geometry behind this question is elucidated. A 2-exchangeable sequence, for example, corresponds to an assignment of probabilities to the four events $(X_1 = 0, X_2 = 0)$, $(X_1 = 0, X_2 = 1)$, $(X_1 = 1, X_2 = 0)$ and $(X_1 = 1, X_2 = 1)$, say p_1, p_2, p_3 and p_4, such that $p_2 = p_3$. Such 4-tuples may be viewed as the slice E_2 through the standard simplex T_3 in R^4, shown in Figure 2.

A 2-exchangeable sequence which can be extended to an infinite exchangeable sequence corresponds to a point in the convex hull of the curve J_2 in E_2, given by

$$J_2 = \left\{\left((1-p)^2, p(1-p), p(1-p), p^2\right): 0 \leq p \leq 1\right\},$$

as described in Diaconis [(1977), page 275] and displayed in his Figure 5.

The geometry underlying n-exchangeable sequences can be described in a similar fashion. Probabilities on $\{0, 1\}^n$ are viewed as points in the standard simplex T_{2^n-1} in R^{2^n}, the n-exchangeable sequences E_n as the intersection of T_{2^n-1} with a suitably chosen n-dimensional flat and the infinitely extendible sequences as the convex hull of the curve J_n in E_n, given by

$$J_n = \left\{\left((1-p)^n, p(1-p)^{n-1}, \dots, p(1-p)^{n-1}, \dots, p^n\right) : 0 \le p \le 1\right\}.$$

Details are given in Diaconis [(1977), page 275].

The connection between the two questions of this paper can now be made. Each distribution $\mathbf{p}$ in T_n, as described in Section 2, determines an n-exchangeable distribution $\mu_{\mathbf{p}}$ in E_n via the procedure: (i) pick X at random according to $\mathbf{p}$; and (ii) put X 1's and $n - X$ 0's in random order, uniformly over all $\binom{n}{X}$ possibilities. Moreover, each n-exchangeable distribution arises in this way. By de Finetti's theorem, $\mu_{\mathbf{p}}$ is infinitely extendible if and only if it is a mixture of Bernoulli processes, hence if and only if $\mathbf{p}$ is a mixture of binomials.

If the probability model for random n-exchangeable sequences is "choose $\mathbf{p}$ using Lebesgue measure on T_n and then use $\mu_{\mathbf{p}}$", then certainly the questions have the same answer. That this probability model for random n-exchangeable sequences is equivalent to the direct use of Lebesgue measure on E_n follows from the fact that the map taking $\mathbf{p}$ to $\mu_{\mathbf{p}}$ can be expressed as the restriction of a linear map. Specifically, it occurs as a composition $D \circ U \circ E$, where

$$E: R^{n+1} \to R^{2^n} \quad \text{is an embedding,}$$
$$U: R^{2^n} \to R^{2^n} \quad \text{is an isometry,}$$
$$D: R^{2^n} \to R^{2^n} \quad \text{is a dilation.}$$

By a dilation we mean a mapping which has a positive diagonal matrix with respect to some orthogonal basis. The construction is routine, so is omitted. It reveals that the volume of E_n is $\prod_{k=1}^n 1/\sqrt{\binom{n}{k}}$ times that of T_n. We can sum matters up as follows:

THEOREM 2.

(i) $$V(\operatorname{co} J_n) = \sqrt{n+1}\,(n!)^{(n-1)/2} \prod_{k=1}^{n} \frac{(k-1)!}{(2k-1)!}.$$

(ii) $$\frac{V(\operatorname{co} J_n)}{V(E_n)} = (n!)^n \prod_{k=1}^{n} \frac{1}{(2k-1)!}.$$

(iii) $$\frac{V(\operatorname{co} J_n)}{V(E_n)} \to 0 \quad \text{as } n \to \infty.$$

For $n < 6$ and $r > n$, Crisma (1982) was able to compute the probability that an n-exchangeable sequence is extendible to an r-exchangeable sequence. His technique involved a challenging summation of simplex volumes. Theorem 2(ii) provides the limits to which his values tend as $r \to \infty$. Table 2 presents Crisma's results, together with the limiting values.

For $n \ge 6$ and $r > 1$, the values in the table are based on Crisma's conjectured form for the probability that an n-exchangeable sequence is extendible to an r-exchangeable sequence, given in Crisma [(1971), page 20]. Based upon his conjecture, Crisma obtained our limiting numerical values [see

TABLE 2
The probability that an n-exchangeable 0–1 sequence is extendible to an r-exchangeable sequence. Results for finite r are due to Crisma [2]. Entries are given to four decimal place accuracy

	$r-n$									
n	1	2	3	4	5	6	7	8	9	∞
2	0.8889	0.8333	0.8000	0.7778	0.7619	0.7500	0.7407	0.7333	0.7273	0.6667
3	0.7500	0.6300	0.5600	0.5143	0.4821	0.4583	0.4400	0.4255	0.4136	0.3000
4	0.6144	0.4480	0.3583	0.3031	0.2661	0.2397	0.2200	0.2048	0.1928	0.0914
5	0.4938	0.3054	0.2148	0.1638	0.1321	0.1108	0.0958	0.0847	0.0762	0.0189
6	0.3917	0.2019	0.1225	0.0828	0.0603	0.0464	0.0372	0.0308	0.0262	0.0027
7	0.3076	0.1304	0.0673	0.0397	0.0257	0.0179	0.0131	0.0101	0.0080	0.0003
8	0.2398	0.0826	0.0358	0.0182	0.0103	0.0064	0.0043	0.0030	0.0022	0.0000
9	0.1858	0.0516	0.0186	0.0080	0.0040	0.0022	0.0013	0.0008	0.0006	0.0000
10	0.1433	0.0318	0.0094	0.0034	0.0015	0.0007	0.0004	0.0002	0.0001	0.0000

Crisma (1971), page 24]. This paper offers no evidence to refute his conjecture, still outstanding after two decades.

Acknowledgments. The author would like to thank Richard Askey, Lucio Crisma, Persi Diaconis, Frank Lad, Tom Leonard and Jostein Lillestøl for encouragement and assistance during the preparation of these results. Thanks also go to a referee whose constructive criticism led to improvements in the final version of the paper.

REFERENCES

CRISMA, L. (1971). Alcune valutazione quantitative interessanti la proseguibilità di processi aleatori scambiabili. *Rend. Mat. Trieste* **3** 96–124.

CRISMA, L. (1982). Quantitative analysis of exchangeability in alternative processes. In *Exchangeability in Probability and Statistics* (G. Koch and F. Spizzichino, eds.) 207–216. North-Holland, Amsterdam.

DE FINETTI, B. (1969). Sulla proseguibilità di processi aleatori scambiabili. *Rend. Mat. Trieste* **1** 53–67.

DIACONIS, P. (1977). Finite forms of de Finetti's theorem on exchangeability. *Synthese* **36** 271–281.

DIACONIS, P. and FREEDMAN, D. (1980). Finite exchangeable sequences. *Ann. Probab.* **8** 745–764.

KARLIN, S. and SHAPLEY, L. S. (1953). Geometry of moment spaces *Mem. Amer. Math. Soc.* **12** 51–53.

MCMULLEN, P. and SHEPHARD, G. C. (1971). Convex polytopes and the upper bound conjecture. *London Math. Soc. Lecture Note Ser.* **3**. Cambridge Univ. Press.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CANTERBURY
CHRISTCHURCH
NEW ZEALAND