The Annals of Probability
1980, Vol. 8, No. 4, 745–764

FINITE EXCHANGEABLE SEQUENCES

BY P. DIACONIS[1] AND D. FREEDMAN[2]

Bell Laboratories and Stanford University, and University of California at Berkeley

Let $X_1, X_2, \cdots, X_k, X_{k+1}, \cdots, X_n$ be exchangeable random variables taking values in the set S. The variation distance between the distribution of $X_1, X_2, \cdots, X_k$ and the closest mixture of independent, identically distributed random variables is shown to be at most $2ck/n$, where c is the cardinality of S. If c is infinite, the bound $k(k-1)/n$ is obtained. These results imply the most general known forms of de Finetti's theorem. Examples are given to show that the rates k/n and $k(k-1)/n$ cannot be improved.

The main tool is a bound on the variation distance between sampling with and without replacement. For instance, suppose an urn contains n balls, each marked with some element of the set S, whose cardinality c is finite. Now k draws are made at random from this urn, either with or without replacement. This generates two probability distributions on the set of k-tuples, and the variation distance between them is at most $2ck/n$.

1. Introduction. de Finetti's theorem involves an infinite sequence $X_1, X_2, \cdots$ of exchangeable 0-1 valued random variables. de Finetti showed that there is a unique probability measure μ on the Borel sets of [0, 1] such that

(1) $$P\{X_i = e_i \text{ for } i = 1, \cdots, k\} = \int_{[0,1]} p^j(1-p)^{k-j}\mu(dp), \text{ where } j = \Sigma e_i.$$

The main results of this paper concern finite exchangeable sequences $X_1, X_2, \cdots, X_k$. As is well known, the representation (1) need not hold exactly for finite exchangeable sequences. For example, let

(2) $$P(X_1 = 0, X_2 = 1) = P(X_1 = 1, X_2 = 0) = \tfrac{1}{2},$$
$$P(X_1 = 0, X_2 = 0) = P(X_1 = 1, X_2 = 1) = 0.$$

Now X_1 and X_2 are exchangeable; but if a representation like (1) held,

$$0 = \int_0^1 p^2\mu(dp) = \int_0^1 (1-p)^2\mu(dp).$$

This implies that μ puts mass one both at 0 and at 1, which is impossible.

However, suppose k is much smaller than n and $X_1, \cdots, X_k$ is the beginning of a long exchangeable sequence $X_1, \cdots, X_k, X_{k+1}, \cdots, X_n$. Then (1) should be approximately true. Our main theorem (3) makes this precise, with the universal error bound $4k/n$. And, as we will indicate in Section 4, there is an example where the error is essentially $(2/\pi e)^{\frac{1}{2}}k/n$, so k/n is the right order of magnitude.

Received June 12, 1978; revised May 22, 1979.

[1] This author's research was partially supported by National Science Foundation Grant MPS 74-21416 and by the Energy Research and Development Administration under Contract EY-76-C-03-0515.

[2] This author's research was partially supported by National Science Foundation Grant MCS-77-01665.

AMS 1970 *subject classifications.* Primary 60G10; secondary 60J05.

Key words and phrases. de Finetti's theorem, exchangeable, symmetric, extreme points, presentable, representable, variation distance, sampling with and without replacement.

So far, we have discussed only 0–1 valued random variables. However, similar results apply to variables with values in any finite set. To state theorem (3), let S be a finite set of cardinality c. Let S^k be the set of k-tuples of elements of S. A probability P on S^k is said to be *exchangeable* provided it is invariant under permutations. More precisely, if π is a permutation of $1, \cdots, k$, then

$$P\{(s_1, \cdots, s_k)\} = P\{(s_{\pi(1)}, \cdots, s_{\pi(k)})\}.$$

To state the analog of (1), let S^* be the set of probabilities on S; geometrically, S^* is the unit simplex in R^c. For $p \in S^*$, let p^k be the distribution of k independent picks from p. Formally, p^k is the power probability on S^k:

$$p^k\{(s_1, \cdots, s_k)\} = \Pi_{j=1}^k p\{s_j\}.$$

If μ is a probability of the Borel subsets of S^*, we define the probability $P_{\mu k}$ on S^k as follows: choose p at random from μ, then make k independent picks from p. Formally,

$$P_{\mu k}(A) = \int_{S^*} p^k(A)\mu(dp).$$

If Q is a probability on S^n and $k \leqq n$, let Q_k be the projection of Q onto S^k. More formally, Q_k is the distribution of $(s_1, \cdots, s_k)$ when the n-tuple $(s_1, \cdots, s_k, s_{k+1}, \cdots, s_n)$ is distributed according to Q. Clearly, $(P_{\mu n})_k = P_{\mu k}$.

The variation distance $\|P - Q\|$ is defined as usual:

$$\|P - Q\| = 2 \sup_A |P(A) - Q(A)|.$$

(3) THEOREM. *Let S be a finite set of cardinality c. Let P be an exchangeable probability on S^n. Then there exists a probability μ on the Borel subsets of S^* such that*

$$\|P_k - P_{\mu k}\| \leqq 2ck/n \qquad \text{for all } k \leqq n.$$

The measure μ depends on n and P, but not on k; notice that the same constant $2c$ applies for all k, n and P. The rate k/n is sharp, as will be seen in Section 4.

The result is almost immediate from the following estimate.

(4) THEOREM. *Suppose an urn U contains n balls, each marked by one or another element of the set S, whose cardinality c is finite. Let H_{Uk} be the distribution of k draws made at random without replacement from U, and M_{Uk} be the distribution of k draws made at random with replacement. (H stands for hypergeometric; M for multinomial.) Thus, H_{Uk} and M_{Uk} are two probabilities on S^k. Then*

$$\|H_{Uk} - M_{Uk}\| \leqq 2ck/n.$$

This result will be proved in Section 2.

PROOF OF THEOREM (3). Each extreme exchangeable probability on S^n is of the form H_{Un} for some urn U, and the exchangeable probability P on S^n is a unique mixture of extreme points:

$$P = \Sigma_U w_U H_{Un}. \tag{5}$$

The sum runs over the finite set of all possible urns U of the type considered in

Theorem (4). The w_U's are nonnegative weights which sum to 1. The representation (5) may be proved by conditioning on the order statistics: see Kendall (1967), de Finetti (1969), Crisma (1971), or Diaconis (1977). Clearly,

$$P_k = \Sigma w_U H_{Uk}.$$

Now use theorem (4):

$$\|P_k - \Sigma w_U M_{Uk}\| \leqq \Sigma w_U \|H_{Uk} - M_{Uk}\| \leqq 2ck/n.$$

This proves Theorem (3), because $M_{Uk} = p_U^k$, where $p_U \varepsilon S^*$ is the distribution of one pick from U: the probability measure μ can be constructed by assigning weight w_U to p_U. □

Remark. We have stated our results for finite exchangeable sequences by assuming a measure P on S^n and working with the projection P_k. Another way to formulate the results starts with an exchangeable probability Q on S^k and asks if there exists an exchangeable probability P on S^n such that $P_k = Q$. Then we say that Q can be *extended* to P. The probability defined by equation (2), for instance, cannot be extended at all. Our results show that if an exchangeable probability Q can be extended from k-tuples to n-tuples, where n is much larger than k, then Q is nearly a mixture of power probabilities.

To pursue this a bit, suppose Q on S^k can be extended to P on S^n; that is, P is exchangeable on S^n and $Q = P_k$.

The extreme exchangeable probabilities on S^n are the urn measures H_{Un} constructed in the proof of (5). And if P is exchangeable on S^n then

$$P = \Sigma_U w_U H_{Un}$$

so

$$Q = P_k = \Sigma_U w_U H_{Uk}.$$

This gives an algorithm for deciding if Q is extendable to S^n; check if Q is a mixture of H_{Uk}. When S is finite, this can be done via linear programming. Further discussion is in Diaconis (1977).

Theorem (3) can be modified to handle S of infinite cardinality. If $(S, \mathfrak{B})$ is an abstract measurable space and P is an exchangeable probability on $(S^n, \mathfrak{B}^n)$, then Theorem (13) shows there is a probability μ on S^* such that

$$\|P_k - P_{\mu k}\| \leqslant k(k-1)/n.$$

Here, S^* is the set of probabilities p on $(S, \mathfrak{B})$. The set S^* is equipped with the σ-field $\mathfrak{B}^*$ generated by the functions $p \to p(A)$ as A ranges over $\mathfrak{B}$. And

$$P_{\mu k} = \int_{S^*} p^k \mu(dp)$$

is a probability on S^k. The rate $k(k-1)/n$ is sharp, as will be seen in Section 4.

In Section 3, these results are shown to imply the most generally known versions of de Finetti's theorem for infinite exchangeable sequences. For a discussion of the role of these results in the foundations of Bayesian inference, see Diaconis and

Freedman (1978a). For an application of finite exchangeability, see Chapter 3 of Galambos (1978).

2. Variation bounds on the difference between sampling with and without replacement. Our first object in this section is to prove Theorem (4). Since U and k may be taken as fixed, we abbreviate H for H_{Uk} and M for M_{Uk}. We label S as $\{1, \cdots, c\}$. For $1 \leqq i \leqq c$, let n_i be the number of balls in U which are labelled i, so $\Sigma_{i=1}^{c} n_i = n$. Without loss of generality, we may suppose $n_i \geqslant 1$ for all i. Fix $s \in S^k$; for $i \in S$, let ν_i be the number of indices $j \leqq k$ with $s_j = i$, so $\Sigma_{i=1}^{c} \nu_i = k$. Then

$$M\{s\} = \Pi_{i=1}^{c}(n_i/n)^{\nu_i} = n^{-k}\Pi_{i=1}^{c} n_i^{\nu_i}$$
$$H\{s\} = [(n-k)!/n!]\Pi_{i=1}^{c} n_i!/(n_i-\nu_i)!$$

The main step in proving (4) is

(6) LEMMA.

$$H\{s\}\Pi_{i=1}^{c}[1-(\nu_i/n_i)] \leqq M\{s\}.$$

PROOF. The left side vanishes unless $\nu_i < n_i$ for all i, so assume this condition. Now $M\{s\} > 0$, and

$$\text{(7)} \quad \frac{H\{s\}}{M\{s\}}\Pi_{i=1}^{c}[1-(\nu_i/n_i)] = \frac{\Pi_{i=1}^{c}[1-(\nu_i/n_i)](n_i^{-\nu_i} n_i!/(n_i-\nu_i)!)}{n^{-k}n!/(n-k)!}.$$

For $0 \leqq x < 1$, let

$$f(x) = -(1-x)\log(1-x) - x = -\Sigma_{i=1}^{\infty} x^{i+1}/i(i+1).$$

Write $\exp(x)$ for e^x. Then we claim

$$\text{(8)} \quad n^{-k}n!/(n-k)! \geqq \exp[nf(k/n)].$$

Indeed, the logarithm of the left side of (8) is

$$\Sigma_{j=1}^{k-1}\log(1-j/n) = -\Sigma_{j=1}^{k-1}\Sigma_{i=1}^{\infty}(1/i)(j/n)^i$$
$$= -\Sigma_{i=1}^{\infty}(1/in^i)\Sigma_{j=1}^{k-1} j^i \geqq nf(k/n),$$

using the elementary estimate

$$\text{(9)} \quad \Sigma_{j=1}^{k-1} j^i \leqq (1/i+1)k^{i+1}.$$

Similarly, with n_i for n and ν_i for k, it can be shown that

$$\text{(10)} \quad [1-(\nu_i/n_i)](n_i^{-\nu_i} n_i!/(n_i-\nu_i)!) \leqq \exp[n_i f(\nu_i/n_i)],$$

replacing (9) by

$$\text{(11)} \quad \Sigma_{j=1}^{\nu_i} j^i \geqq (1/i+1)\nu_i^{i+1}.$$

The factor $[1-(\nu_i/n_i)]$ was put in precisely to extend the range of summation from $\nu_i - 1$ to ν_i.

To complete the proof, notice that f is concave. By Jensen's inequality,

$$\Sigma_{i=1}^{c} n_i f(\nu_i/n_i) \leqq nf(k/n).$$

On the right side of (7), the denominator has been bounded below by $\exp[nf(k/n)]$; the numerator has been bounded above by the same quantity. □

PROOF OF THEOREM (4). Lemma (6) implies

$$M\{s\} - H\{s\} \geqq -H\{s\}\Sigma_{i=1}^{c}\nu_i/n_i.$$

Writing $x^{-} = \max\{-x, 0\}$, we get

$$(12) \qquad (M\{s\} - H\{s\})^{-} \leqq H\{s\}\Sigma_{i=1}^{c}\nu_i/n_i.$$

Now

$$\|M - H\| = 2\Sigma_s(M\{s\} - H\{s\})^{-}.$$

Recall that ν_i is the number of $j \leqq k$ with $s_j = i$, so summing the right side of (12) over s is tantamount to computing a certain expectation relative to H. Since the H-expectation of ν_i/n_i is k/n, the sum of the right side of (12) over s is just ck/n. □

3. Results for general range spaces. In this section we give finite forms of de Finetti's theorem for variables taking values in an arbitrary space. The most general known infinite versions of de Finetti's theorem are then derived by taking limits.

Let $(S, \mathfrak{B})$ be an abstract measurable space and write S^* for the set of probabilities on $(S, \mathfrak{B})$. Endow S^* with the smallest σ-field $\mathfrak{B}^*$ making $p \to p(A)$ measurable for all $A \in \mathfrak{B}$. If $p \in S^*$, then p^k is the power probability on $(S^k, \mathfrak{B}^k)$. Clearly, $p \to p^k(A)$ is $\mathfrak{B}^*$-measurable for any $A \in \mathfrak{B}^k$. If μ is a probability on $(S^*, \mathfrak{B}^*)$, define the probability $P_{\mu k}$ on $(S^k, \mathfrak{B}^k)$ as before:

$$P_{\mu k}(A) = \int_{S^*} p^k(A)\mu(dp).$$

The main result of this section can be stated as follows, with P_k the projection of P onto $(S^k, \mathfrak{B}^k)$, and $\beta(n, k)$ defined as follows:

$$1 - \beta(n,k) = n^{-k}n!/(n-k)!$$

(13) THEOREM. *Let P be an exchangeable probability on $(S^n, \mathfrak{B}^n)$. Then there exists a probability μ on $(S^*, \mathfrak{B}^*)$ such that*

$$\|P_k - P_{\mu k}\| \leqq 2\beta(n,k) \qquad \text{for all } k \leqq n.$$

Note. $\beta(n,k) \leqq \frac{1}{2}k(k-1)/n$.

The probability μ depends on n and P, but not on k. The bound is sharp, as will be seen in Section 4.

PROOF. For $\omega \in S^n$, let $U(\omega)$ be the urn containing n balls, marked $\omega_1, \cdots, \omega_n$ respectively. Let $H_{U(\omega)k}$ be the distribution of k draws made at random without replacement from $U(\omega)$, while $M_{U(\omega)k}$ is the distribution of k draws made with replacement. Thus, $H_{U(\omega)k}$ and $M_{U(\omega)k}$ are probabilities on $(S^k, \mathfrak{B}^k)$. The maps

$\omega \to H_{U(\omega)k}(A)$ and $\omega \to M_{U(\omega)k}(A)$ are measurable on $(S^n, \mathfrak{B}^n)$ for each $A \in \mathfrak{B}^k$.

As is easily verified, the exchangeability of P entails

$$P(A) = \int_{S^n} H_{U(\omega)n}(A) P(d\omega) \qquad \text{for } A \in \mathfrak{B}^n.$$

So

$$P_k(A) = \int_{S^n} H_{U(\omega)k}(A) P(d\omega) \qquad \text{for } A \in \mathfrak{B}^k.$$

Clearly,

$$M_{U(\omega)k} = (M_{U(\omega)1})^k.$$

Now $\omega \to M_{U(\omega)1}$ is a measurable map from $(S^n, \mathfrak{B}^n)$ to $(S^*, \mathfrak{B}^*)$. Let μ be the image of P under this map. If $A \in \mathfrak{B}^k$, then

$$P_{\mu k}(A) = \int_{S^n} M_{U(\omega)k}(A) P(d\omega)$$

and

$$P_k(A) - P_{\mu k}(A) = \int_{S^n} \left[H_{U(\omega)k}(A) - M_{U(\omega)k}(A) \right] P(d\omega).$$

But Freedman (1977) shows

$$|H_{U(\omega)k}(A) - M_{U(\omega)k}(A)| \leqq \beta(n,k). \qquad \square$$

The balance of this section is somewhat technical. Readers who want to see that the rates in Theorems (3) and (13) are sharp can skip to Section 4, which does not depend on later results in this section.

Theorem (13) implies the most general known representation theorem for an infinite exchangeable sequence. To begin with, we derive a result of Hewitt and Savage (1955). Suppose S is a compact Hausdorff space, and $\mathfrak{B}$ is the σ-field making the real continuous functions measurable. The σ-field $\mathfrak{B}$ is called the *Baire σ-field*. The infinite product $\mathfrak{B}^\infty$ is the Baire σ-field of the product S^∞.

As before, S^* is the set of probabilities on $(S, \mathfrak{B})$. In the weak * topology, S^* is compact Hausdorff; the σ-field $\mathfrak{B}^*$ is generated by the functions $p \to p(A)$ as A ranges over $\mathfrak{B}$, and $\mathfrak{B}^*$ is the Baire σ-field in S^*.

If μ is a probability on S^*, define the probability P_μ on $(S^\infty, \mathfrak{B}^\infty)$ as usual, by

$$P_\mu = \int_{S^*} p^\infty \mu(dp).$$

If Q is a probability on $(S^\infty, \mathfrak{B}^\infty)$, let Q_k be its projection onto the first k coordinates. Clearly, $P_{\mu k} = \int_{S^*} p^k \mu(dp)$, so there is no conflict with previous notation.

(14) THEOREM (Hewitt-Savage). *Let S be a compact Hausdorff space, and P an exchangeable probability on the Baire field $\mathfrak{B}^\infty$ of S^∞. Then there exists a unique measure μ on $\mathfrak{B}^*$ such that $P = P_\mu$.*

PROOF. For each n, Theorem (13) yields a measure μ_n on $(S^*, \mathfrak{B}^*)$ such that

(15) $$\|P_k - P_{\mu_n k}\| \leqq k(k-1)/n \qquad \text{for } k \leqq n.$$

The set of probabilities on $(S^*, \mathfrak{B}^*)$ is compact in the weak* topology. So there is a probability μ on $(S^*, \mathfrak{B}^*)$ and a net $n(\alpha)$ of positive integers such that $n(\alpha) \to \infty$ and $\mu_{n(\alpha)} \to \mu$. The map $\nu \to P_{\nu k}$ is weak* continuous, so $P_{\mu_{n(\alpha)}k} \to P_{\mu k}$ in the weak* topology. But $P_{\mu_{n(\alpha)}k} \to P_k$ in variation norm, by (15). So $P_{\mu k} = P_k$ for all k, and $P_\mu = P$. This completes the existence proof. We omit the uniqueness proof: see, for instance, (3.4) of Dubins and Freedman (1979). ∎

Note added in proof. In (13-14-15), μ_n is a probability on S^*. Indeed, it is (in disguise) the P-law of the empirical measure

$$\theta_{n,\omega} = \frac{1}{n}\Sigma_{i=1}^n \delta_{\omega_i}$$

where δ_x is point mass as x. A relatively elementary argument shows that μ_n converges weak-star; that is, for any continuous function ϕ on S^*, $\int_{S^*}\phi(p)\mu_n(dp)$ converges as $n \to \infty$. First, suppose

$$\phi(p) = \Pi_{j=1}^k \int_S f_j(x)p(dx),$$

where the f_j are continuous functions on S. Then

$$\begin{aligned} \int_{S^*}\phi(p)\mu_n(dp) &= \int_{\Omega^n}\phi(\theta_{n,\omega})P(d\omega) \\ &= \int_{\Omega^n}\Pi_{j=1}^k\left[\frac{1}{n}\Sigma_{i=1}^n f_j(\omega_i)\right]P(d\omega) \\ &\to \int_{\Omega^k}\Pi_{j=1}^k f_j(\omega_j)P(d\omega); \end{aligned}$$

because, after expanding the product, there will be $n^k - O(n^{k-1})$ terms in which each of the k factors $f_j(\omega_i)$ corresponds to a different index i. By exchangeability, all such terms are equal. ∎

Theorem (14) implies de Finetti's theorem for exchangeable probabilities in standard spaces, which are isomorphic to Borel subsets of the unit interval. As shown in Dubins and Freedman (1979), de Finetti's theorem can fail for separable, metrizable spaces which are not standard.

The compact Hausdorff space S also admits the *Borel* σ-field $\tilde{\mathfrak{B}}$, which is larger than the Baire σ-field $\mathfrak{B}$ when S is not metrizable. For our purposes, $\tilde{\mathfrak{B}}$ may be defined as the smallest σ-field relative to which all lower semicontinuous functions are measurable. A probability p on $\mathfrak{B}$ has a unique regular extension $\tilde{p}$ to $\tilde{\mathfrak{B}}$. We propose to show that the formula $P = \int p^\infty \mu(dp)$ can be extended regularly from the Baire to the Borel σ-fields.

It may be helpful to review very briefly the procedure for extending the Baire probability p on $\mathfrak{B}$ to the regular Borel probability $\tilde{p}$ on $\tilde{\mathfrak{B}}$. Let f_α be a net of continuous functions, with $f_\alpha \uparrow f$. Then $\tilde{p}(f) = \lim_\alpha p(f_\alpha)$. The value $\tilde{p}(f)$ does not depend on which net of continuous functions is used to approximate the lower semicontinuous function f. The extension of $\tilde{p}$ from the lower semicontinuous functions to all Borel functions is done by the usual procedures. For a careful discussion of these matters, see Choquet (1969). We point out that the Borel sets usually do not turn up in the measure-theoretic completion of $\mathfrak{B}$. To keep the notation straight: $p \in S^*$ is a Baire probability, defined only on $\mathfrak{B}$; its regular extension to $\tilde{\mathfrak{B}}$ is $\tilde{p}$.

(16) Let f be a nonnegative Borel function on S. Then $p \to \tilde{p}(f)$ is a Borel function on S^*.

Indeed, suppose f were lower semicontinuous. Choose a net f_α of continuous functions with $f_\alpha \uparrow f$. Now $p \to p(f_\alpha)$ is a continuous function of p, and this net increases to $p \to \tilde{p}(f)$. The latter function is lower semicontinuous, and hence Borel. Consider the class of Borel f for which $p \to \tilde{p}(0 \vee f \wedge k)$ is Borel. This class includes the lower semicontinuous functions and is closed under sequential limits, so must comprise all Borel f's. Letting $k \uparrow \infty$ proves (16).

(17) The map $p \to \overset{\frown}{p^\infty}$ is weak* continuous from S^* to $(S^\infty)^*$.

(18) Let $\mathfrak{B}^\infty$ be the Borel σ-field of S^∞. Usually, $\overset{\frown}{\mathfrak{B}^\infty}$ is larger than $\tilde{\mathfrak{B}}^\infty$.

For $p \in S^*$, let $\overset{\frown}{p^\infty}$ be the regular Borel extension of p^∞ to $\overset{\frown}{\mathfrak{B}^\infty}$.

(19) Let f be a nonnegative Borel function on S^∞. Then $p \to \overset{\frown}{p^\infty}(f)$ is a Borel function on S^*.

(20) THEOREM. *Let S be compact Hausdorff, and P an exchangeable Baire probability on $(S^\infty, \mathfrak{B}^\infty)$. According to* (14), *there is a unique Baire probability μ on $(S^*, \mathfrak{B}^*)$ such that*

$$P(A) = \int_{(S^*, \mathfrak{B}^*)} p^\infty(A)\mu(dp) \qquad \textit{for all } A \in \mathfrak{B}^\infty. \tag{21}$$

This representation extends regularly to the Borel σ-field:

$$\tilde{P}(A) = \int_{(S^*, \overset{\frown}{\mathfrak{B}^*})} \overset{\frown}{p^\infty}(A)\tilde{\mu}(dp) \qquad \textit{for all } A \in \tilde{\mathfrak{B}}^\infty. \tag{22}$$

Notation. In (21), the space S^* of probabilities p on $(S, \mathfrak{B})$ has been equipped with the Baire σ-field $\mathfrak{B}^*$. In (22), the space S^* is the same, but is equipped with the larger Borel σ-field $\tilde{\mathfrak{B}}^*$.

PROOF. Let f be a nonnegative lower-semicontinuous function on S^∞. We must show

$$\tilde{P}(f) = \int_{(S^*, \overset{\frown}{\mathfrak{B}^*})} \overset{\frown}{p^\infty}(f)\tilde{\mu}(dp). \tag{23}$$

Suppose the net f_α of continuous functions on S^∞ increases to f. Then $p \to p^\infty(f_\alpha)$ is a net of continuous functions on S^* increasing to $p \to \overset{\frown}{p^\infty}(f)$. The latter function is lower-semicontinuous on S^*, and

$$\begin{aligned} \tilde{P}(f) &= \lim_\alpha P(f_\alpha) \\ &= \lim_\alpha \int p^\infty(f_\alpha)\mu(dp) \\ &= \int \overset{\frown}{p^\infty}(f)\tilde{\mu}(dp). \end{aligned}$$

This proves (23). Now (23) can be extended by the usual arguments to all nonnegative Borel functions. □

We turn now to nontopological situations. Let $(S, \mathfrak{B})$ be an abstract measurable space. Let S^* be the set of probabilities on $\mathfrak{B}$, with the weak* σ-field $\mathfrak{B}^*$ generated by the function $p \to p(A)$ as A ranges over $\mathfrak{B}$. Hewitt and Savage call $\mathfrak{B}$ *presentable* if, for every exchangeable P on $(S^\infty, \mathfrak{B}^\infty)$ there is a probability μ on $(S^*, \mathfrak{B}^*)$ with $P = P_\mu$. In this terminology, Theorem (14) says that the Baire σ-field of a compact Hausdorff space S is presentable. We do not know whether the Borel σ-field of S is presentable: (20) only deals with regular probabilities. We also do not know if the Borel σ-field of a locally convex topological vector space is presentable. For a partial result in this direction, see Hulonicki and Phelps (1968).

We propose to show that if a σ-field $\mathfrak{B}$ of subsets of S is presentable, so is the *universal completion* $\overline{\mathfrak{B}}$. For our purposes, $\overline{\mathfrak{B}}$ may be defined as the σ-field of subsets of S which are θ-measurable for any probability θ on $\mathfrak{B}$. Of course, θ has a unique extension $\bar{\theta}$ to $\overline{\mathfrak{B}}$.

(23) THEOREM. *If $\mathfrak{B}$ is a presentable σ-field of subsets of S, and Σ is a σ-field of subsets of S with $\mathfrak{B} \subset \Sigma \subset \overline{\mathfrak{B}}$, then Σ is presentable.*

PROOF. Recall that S^* is the set of probabilities on $(S, \mathfrak{B})$, endowed with the σ-field $\mathfrak{B}^*$ generated by the functions $p \to p(A)$ as A ranges over $\mathfrak{B}$. Let $\overline{\mathfrak{B}^*}$ be the universal completion of $\mathfrak{B}^*$, and $\overline{\mathfrak{B}^\infty}$ the universal completion of $\mathfrak{B}^\infty$. If $A \in \overline{\mathfrak{B}^\infty}$, we claim that $p \to \overline{p^\infty}(A)$ is $\overline{\mathfrak{B}^*}$-measurable on S^*. Indeed, let μ be a probability on $\mathfrak{B}^*$. Choose A_0 and A_1 in $\mathfrak{B}^\infty$ with $A_0 \subset A \subset A_1$ and $P_\mu(A_0) = P_\mu(A_1)$. Then

$$p^\infty(A_0) \leqq \overline{p^\infty}(A) \leqq p^\infty(A_1) \qquad \text{for all } p \in S^*$$

and

$$p^\infty(A_0) = p^\infty(A_1) \qquad \text{for } \mu\text{-almost all } p \in S^*,$$

so $p \to \overline{p^\infty}(A)$ is μ-measurable.

Suppose P is exchangeable on $(S^\infty, \mathfrak{B}^\infty)$. Since $\mathfrak{B}$ is presentable by assumption, there is a probability μ on $(S^*, \mathfrak{B}^*)$ with

$$(24) \qquad P(A) = \int_{(S^*, \mathfrak{B}^*)} p^\infty(A)\mu(dp) \qquad \text{for all } A \in \mathfrak{B}^\infty.$$

We claim

$$(25) \qquad \bar{P}(A) = \int_{(S^*, \overline{\mathfrak{B}^*})} \overline{p^\infty}(A)\bar{\mu}(dp) \qquad \text{for all } A \in \overline{\mathfrak{B}^\infty}.$$

Indeed, let A_0 and A_1 be in $\mathfrak{B}^\infty$ with $A_0 \subset A \subset A_1$ and $P(A_0) = P(A_1)$. Using (24),

$$\begin{aligned} P(A_0) &= \int p^\infty(A_0)\mu(dp) \leqq \int \overline{p^\infty}(A)\bar{\mu}(dp) \\ &\leqq \int p^\infty(A_1)\mu(dp) = P(A_1) = P(A_0) \end{aligned}$$

so equality holds throughout. This completes the proof of (25).

Next, we claim that $\overline{\mathfrak{B}}^\infty \subset \overline{\mathfrak{B}^\infty}$. Indeed, let

$$A = A_1 \times \cdots \times A_n \times S \times S \times \cdots \qquad \text{with } A_i \in \overline{\mathfrak{B}}.$$

Let Q be a probability on $\mathfrak{B}^\infty$, with ith marginal Q_i. Because $A_i \in \overline{\mathfrak{B}}$, there are sets

B_i and C_i in $\mathfrak{B}$ with $B_i \subset A_i \subset C_i$ and $Q_i(C_i - B_i) = 0$. Let

$$B = B_1 \times \cdots \times B_n \times S \times S \times \cdots$$

$$C = C_1 \times \cdots \times C_n \times S \times S \times \cdots .$$

Then B and C are in $\mathfrak{B}^\infty$, and $B \subset A \subset C$, and a moment's thought shows that $Q(C - B) = 0$. Hence $A \in \overline{\mathfrak{B}^\infty}$. Then $\overline{\mathfrak{B}}^\infty \subset \overline{\mathfrak{B}^\infty}$ by a routine argument.

To complete the proof of the theorem, suppose $\mathfrak{B} \subset \Sigma \subset \overline{\mathfrak{B}}$, and Q is an exchangeable probability on Σ^∞. Let P be the restriction of Q to $\mathfrak{B}^\infty$. Confirm that P is exchangeable, and $\bar{P}(A) = Q(A)$ for $A \in \Sigma^\infty$. In (25), confine A to Σ^∞, and verify that $\overline{p^\infty}(A) = \bar{p}^\infty(A)$, so

$$Q(A) = \bar{P}(A) = \int_{(S^*, \overline{\mathfrak{B}^*})} \bar{p}^\infty(A)\bar{\mu}(dp). \tag{26}$$

To review the notation, $p \in S^*$ is a probability on $(S, \mathfrak{B})$, which extends to a probability $\bar{p}$ on $(S, \overline{\mathfrak{B}})$. In the integral, the space S^* has been equipped with the universal completion $\overline{\mathfrak{B}^*}$ of $\mathfrak{B}^*$. Let W^* be the space of probabilities q on (S, Σ), equipped with the weak* σ-field Σ^*. For $p \in S^*$, let ρp be the restriction of $\bar{p}$ from $\overline{\mathfrak{B}}$ to Σ. In particular, $\bar{p}^\infty(A) = (\rho p)^\infty(A)$ for $A \in \Sigma^\infty$. From (26),

$$Q(A) = \int_{(S^*, \overline{\mathfrak{B}^*})} (\rho p)^\infty(A)\bar{\mu}(dp). \tag{27}$$

Verify that $p \to \rho p$ is a $(\overline{\mathfrak{B}^*}, \Sigma^*)$-measurable map of S^* to W^*; let $\nu = \bar{\mu}\rho^{-1}$. From (27),

$$Q(A) = \int_{(W^*, \Sigma^*)} q^\infty(A)\nu(dq). \qquad \square$$

So far, we have considered only countably additive probabilities. There is also a circle of results involving finitely additive probabilities. For example, Hewitt and Savage (1956) showed that any finitely additive probability on the field of cylinder sets of S^∞ is a unique countably additive mixture of finitely additive power probabilities. This can be derived from the case of finite S by a compactness argument.

A variant of Theorem (13) leads to a different kind of representation, in terms of finitely additive mixtures of countably additive powers. To state this variant of (13), let S be a set and $\mathfrak{B}$ a σ-field of subsets of S. Let S_d^* be the set of those countably additive probabilities p on $\mathfrak{B}$ which have finite support. (The "d" stands for "discrete.") In other terms, $p \in S_d^*$ iff for some positive integer n and points $s_1, \cdots, s_n$ in S and nonnegative weights $w_1, \cdots, w_n$ with $w_1 + \cdots + w_n = 1$, we have

$$p(A) = \Sigma_{i=1}^n w_i 1_A(s_i).$$

Let $\mathfrak{B}_d^*$ be the σ-field of subsets of S_d^* spanned by the functions $p \to p(A)$ as A ranges over $\mathfrak{B}$. If $\mathfrak{B}$ is a separable σ-field, then $S_d^* \in \mathfrak{B}^*$: see Dubins and Freedman (1964).

(28) LEMMA. *Let P be a finitely additive probability on $(S^n, \mathfrak{B}^n)$. Then there is a*

finitely additive probability μ_n *on* $(S_d^*, \mathfrak{B}_d^*)$ *such that*

$$\|P_k - P_{\mu_n k}\| \leqq k(k-1)/n \qquad \textit{for all } k \leqq n.$$

The proof is omitted, being identical to that for (13).

REMARK. The same result holds if $\mathfrak{B}$ is only a field: then $\mathfrak{B}^n$ too is only a field. One appropriate definition for $\mathfrak{B}_d^*$ is the smallest field containing all sets of the form $\{p: a \leqq p(A) < b\}$ as a and b range over the rationals and A ranges over $\mathfrak{B}$.

(29) THEOREM. *Let* $\mathfrak{B}$ *be a* σ-field *of subsets of the set* S. *Let* $\mathfrak{B}^f$ *be the field of finite-dimensional product-measurable sets in* $(S^\infty, \mathfrak{B}^\infty)$. *Every finitely additive exchangeable probability* P *on* $(S^\infty, \mathfrak{B}^f)$ *admits the representation*

$$(30) \qquad P(A) = \int_{S^*} p^\infty(A)\mu(dp) \qquad \textit{for all } A \in \mathfrak{B}^f$$

where μ *is a finitely additive probability on* $(S^*, \mathfrak{B}^*)$. *This* μ *is not unique, and need not be countably additive even when* P *is.*

PROOF. Apply (28) to the projection P_n of P onto the first n coordinates. We get a finitely additive probability μ_n on $(S^*, \mathfrak{B}^*)$ such that

$$\|P_k - P_{\mu_n k}\| \leqq k(k-1)/n \qquad \text{for all } k \leqq n.$$

The set of finitely additive probabilities on $(S^*, \mathfrak{B}^*)$ is compact in the topology of setwise convergence, and the rest of the existence proof is like that in (14).

To see that the mixing measure μ is not unique, and need not be countably additive, let S be $[0, 1]$ and $\mathfrak{B}$ the Borel σ-field. Let P make the coordinates of S^∞ independent, with common distribution λ, where λ is Lebesgue measure on $(S, \mathfrak{B})$. Then P is countably additive and exchangeable on $(S^\infty, \mathfrak{B}^f)$. The "obvious" representation (30) uses mixing measure δ_λ—point mass at λ. To construct another representation, use the compactness argument to get a finitely additive probability measure μ on $(S_d^*, \mathfrak{B}_d^*)$ for which (30) holds. Now $\mu \neq \delta_\lambda$: indeed, $\mu(S_d^*) = 1$; and $\delta_\lambda(S_d^*) = 0$ because $\lambda \notin S_d^*$. Likewise, μ cannot be countably additive: if it were, it would have to be δ_λ, by the uniqueness part of the Hewitt-Savage Theorem (14). □

REMARK. In (29), we define $\mathfrak{B}^f$ as the field of subsets of S^∞ of the form $A \times S \times S \times \cdots$, $A \in \mathfrak{B}^n$, $n = 1, 2, \cdots$. The theorem remains true if $\mathfrak{B}^f$ is the (smaller) field generated by the sets

$$A_1 \times \cdots \times A_n \times S \times S \times \cdots, \quad A_i \in \mathfrak{B}, n = 1, 2, \cdots.$$

In this form, the theorem is true even if $\mathfrak{B}$ is only a field. See the remark after (28).

4. The rates are sharp. We begin by showing that the bound $2\beta(n, k)$ in (13) is best possible. The example involves a sample of size k from a population of n elements. Let H_{nk} represent the distribution of k draws made at random without replacement from an urn U_n containing n balls labeled $\{1, 2, \cdots, n\}$. Thus, H_n is an exchangeable probability measure on S_n^k, where $S_n = \{1, 2, \cdots, n\}$. Let M_{nk} be the distribution of k draws made at random with replacement from U_n.

(31) PROPOSITION. *For every probability* μ *on* S_n^*, *and for all* $k \leqq n$,

(32) $$\|H_{nk} - P_{\mu k}\| \geqq \|H_{nk} - M_{nk}\| = 2\beta(n,k).$$

In particular, among all mixtures of power probabilities, the pure power probability M_{nk} *is closest to* H_{nk}: *and its distance to* H_{nk} *is just the bound in* (13). *So* (13) *is sharp —if the state space S is not constrained.*

PROOF. Fix n and k. Consider the set

$$B = \{\omega \in S_n^k : \omega_i = \omega_j \quad \text{for some } i,j \quad \text{with } 1 \leqq i < j \leqq k\}.$$

("B" is for "birthday": in the birthday problem, $n = 365$ and B is the event that at least two people in a sample of size k have the same birthday.) Clearly, $H_{nk}(B) = 0$, and Freedman (1977) shows that

$$\|H_{nk} - M_{nk}\| = 2M_{nk}(B) = 2\beta(n,k).$$

To complete the proof, we use the inequality

(33) $$p^k(B) \geqq M_{nk}(B) \qquad \text{for any probability } p \text{ on } \{1, 2, \cdots, n\}.$$

This can be deduced from the Schur convexity of 1_B, with the help of Rinott (1973, page 68). A direct proof is in Munford (1977). The inequality (33) shows that $P_{\mu k}(B) \geqq M_{nk}(B)$. Thus

$$\|H_{nk} - P_{\mu k}\| \geqq 2M_{nk}(B) = \|H_{nk} - M_{nk}\|.$$

REMARK. As noted in Freedman (1977),

$$1 - \exp\left(-\tfrac{1}{2}k(k-1)/n\right) \leqq \beta(n,k) \leqq \tfrac{1}{2}k(k-1)/n.$$

If $k = 0(n^{\frac{1}{2}})$, then $\beta(n,k)$ is essentially $\frac{1}{2}k(k-1)/n$; otherwise, $\beta(n,k)$ does not tend to 0. That is why the rate $k(k-1)/n$ is sharp for inequalities like the one in (13).

We next show that the rate k/n in Theorem (3) is sharp: the state space S is constrained to have at most $c = 2$ elements. The example involves a sample of size $2k$ from the urn U_{2n} which contains n red balls and n black balls. For the rest of this section we assume $k < n/2$. Let $H_{2n,2k}$ be the distribution of $2k$ draws made at random without replacement from the urn U_{2n}. Let B_{2k} be the distribution of $2k$ draws made at random with replacement from U_{2n}. So $H_{2n,2k}$ and B_{2k} are probabilities on S^{2k}, where $S = \{\text{red}, \text{black}\}$. (The H is for hypergeometric, B for binomial.)

Of course, S^* is isomorphic to [0, 1]: if $0 \leqq \theta \leqq 1$, let

$$p_\theta(\text{red}) = \theta \qquad \text{and} \qquad p_\theta(\text{black}) = 1 - \theta.$$

As usual, if μ is a probability on [0, 1], we define the probability $P_{\mu,2k}$ on S^{2k}:

$$P_{\mu,2k} = \int_{[0,1]} p_\theta^{2k} \mu(d\theta).$$

So $P_{\mu,2k} = B_{2k}$ when $\mu = \delta_{\frac{1}{2}}$—point mass at 1/2.

Later, we will prove

(34) THEOREM. $\|H_{2n,2k} - P_{\mu,2k}\| \geqq \|H_{2n,2k} - B_{2k}\|$, *and the inequality is strict unless* $\mu = \delta_{\frac{1}{2}}$.

In particular, among all mixtures of power probabilities, the pure power probability B_{2k} is closest to $H_{2n,2k}$.

We will also prove

(35) THEOREM. *As n tends to infinity,*

(36) $$\|H_{2n,2k} - B_{2k}\| = \gamma k/n + 0(k/n) \quad \text{if } k = o(n)$$

(37) $$\|H_{2n,2k} - B_{2k}\| \to \phi(\alpha) \quad \text{if } k/n \to \alpha \quad \text{with } 0 < \alpha < 1/2,$$

where

(38) $$\gamma = \tfrac{1}{2}(2\pi)^{-\frac{1}{2}}\int|1 - u^2|\exp(-u^2/2)\,du = (2/\pi e)^{\frac{1}{2}}$$

and for $0 < \alpha < 1/2$

(39) $$\phi(\alpha) = (2\pi)^{-\frac{1}{2}}\int|1 - (1-\alpha)^{\frac{1}{2}}\exp(\alpha u^2/2)|\exp(-u^2/2)\,du$$

and

$$\exp(x) = e^x.$$

Theorems (34) and (35) show that k/n is the best rate possible in (3):

(40) THEOREM. *Let n tend to* ∞ *and let* $k = o(n)$. *Then, for any probability* μ *on* [0, 1],

(41) $$\|H_{2n,2k} - P_{\mu,2k}\| \geqq \gamma k/n + o(k/n).$$

Likewise, if $n \to \infty$ *and* $k/n \to \alpha$ *with* $0 < \alpha < 1/2$, *then*

(42) $$\|H_{2n,2k} - P_{\mu,2k}\| \geqq \phi(\alpha) + o(1).$$

NOTE. In (36) and (41), we do not assume $k \to \infty$.

(43) LEMMA. *For* $\omega \in S^{2k}$, *let* $R(\omega)$ *be the number of reds in* ω. *Let P and Q be exchangeable probabilities on* S^{2k}. *Then*

$$\|P - Q\| = \|PR^{-1} - QR^{-1}\|.$$

PROOF. The conditional distribution of P given R coincides with the conditional distribution of Q given R. □

(44) LEMMA. *Let* μ *be a probability measure on the Borel sets of* [0, 1]. *Let* $\tilde{\mu}$ *be the image of* μ *under the map* $x \to 1 - x$. *Let* $\bar{\mu} = \frac{1}{2}(\mu + \tilde{\mu})$. *Then*

$$\|H_{2n,2k} - P_{\bar{\mu},2k}\| \leqq \|H_{2n,2k} - P_{\mu,2k}\|.$$

PROOF. The measure $H_{2n,2k}$ is invariant under the transformation T which switches red and black. But T transforms $P_{\mu k}$ into $P_{\tilde{\mu}k}$. Thus

$$\|H_{2n,2k} - P_{\mu,2k}\| = \|H_{2n,2k} - P_{\tilde{\mu},2k}\|.$$

Since $P_{\tilde{\mu},2k}$ is the average of $P_{\mu,2k}$ and $P_{\tilde{\mu},2k}$, we have

$$\begin{aligned}\|H_{2n,2k} - P_{\bar{\mu},2k}\| &= \|H_{2n,2k} - \tfrac{1}{2}(P_{\mu,2k} + P_{\tilde{\mu},2k})\| \\ &\leqq \tfrac{1}{2}\|H_{2n,2k} - P_{\mu,2k}\| + \tfrac{1}{2}\|H_{2n,2k} - P_{\tilde{\mu},2k}\| \\ &= \|H_{2n,2k} - P_{\mu,2k}\|. \qquad \square\end{aligned}$$

To simplify the writing, let

$$h_{nk}(j) = H_{2n,2k}\{R = k + j\}$$
$$b_k(j) = B_{2k}\{R = k + j\}.$$

So $h_{nk}(j)$ is the chance of getting $k+j$ red balls in $2k$ draws made at random without replacement from U_{2n}; and $b_k(j)$ is the chance when the draws are made with replacement. Notice that $H_{2n,2k}$ and B_{2k} are probabilities on S^{2k}. However, h_{nk} and b_k are probabilities on the integers between $-k$ and k: indeed, h_{nk} is a centered hypergeometric distribution and b_k a centered binomial.

The next result is included to make the relationship between h_{nk} and b_k easier to visualize: it will not be used later.

(45) LEMMA. *There is a unique index* $j^* = j^*(n,k)$ *such that*

$$h_{nk}(j) \geqq b_k(j) \qquad \text{for } |j| < j^*$$
$$h_{nk}(j) < b_k(j) \qquad \text{for } |j| \geqq j^*.$$

PROOF. Let $L_{n,k}(j) = h_{nk}(j)/b_k(j)$. By algebra,

$$L_{n,k}(j) = \binom{2n-2k}{n-k+j}\left(\tfrac{1}{2}\right)^{2n-2k} \Big/ \binom{2n}{n}\left(\tfrac{1}{2}\right)^{2n}.$$

Therefore, $L_{n,k}(j)$ increases strictly to its maximum as j increases from $-k$ to 0 and then strictly decreases. (We have used the assumption $k < n/2$, to ensure $n - k + j > 0$.) $\square$

The next result shows that $j^*(n,k) \leqq (k/2)^{\frac{1}{2}}$.

(46) LEMMA. *If* $|k| \geqq |j| \geqq (k/2)^{\frac{1}{2}}$, *then* $h_{nk}(j) < b_k(j)$.

PROOF. Define $L_{n,k}$ as above. By algebra,

$$\frac{L_{n,k}(j)}{L_{n-1,k}(j)} = \frac{n^2}{(n-k-j)(n-k+j)} \Big/ \frac{2n(2n-1)}{(2n-2k)(2n-2k-1)}.$$

Notice that $n - k \pm j > 0$ because $|j| \leqq k < n/2$. So $L_{n,k}(j)/L_{n-1,k}(j) > 1$ if and only if

$$k^2 - j^2 > (k - 2j^2)n.$$

In particular, fix j and k with $|j| \geqq (k/2)^{\frac{1}{2}}$. Then $L_{n,k}(j) > L_{n-1,k}(j)$. So $L_{n,k}(j)$ is strictly increasing with n. But $\lim_{n\to\infty} L_{n,k}(j) = 1$ since the hypergeometric merges with the binomial as the urn U_{2n} becomes infinite. Thus $L_{n,k}(j) < 1$ and $h_{nk}(j) < b_k(j)$ if $|j| \geqq (k/2)^{\frac{1}{2}}$. $\square$

We next consider symmetric two-point mixtures of binomials. Let

$$b_{k,p}(j) = \binom{2k}{k+j} p^{k+j}(1-p)^{k-j}.$$

For $-1 \leqq x \leqq 1$, define

$$m_{k,x}(j) = \tfrac{1}{2}\left[b_{k,(1+x)/2}(j) + b_{k,(1-x)/2}(j)\right]. \tag{47}$$

Clearly,

$$m_{k,-x}(j) = m_{k,x}(j) = m_{k,x}(-j). \tag{48}$$

(49) LEMMA. *If* $|j| \leqq (k/2)^{\frac{1}{2}}$ *and* $|x| < |y|$, *then* $m_{k,y}(j) < m_{k,x}(j)$.

PROOF. Put $j = 0$. Then

$$m_{k,x}(0) = \binom{2k}{k}(1-x^2)^k/2^{2k},$$

and the monotonicity is easy. In view of (48), we may suppose $1 \leqq j \leqq (k/2)^{\frac{1}{2}}$ and $0 \leqq x < y \leqq 1$. The case $x = 0$ follows by taking limits, so we may even suppose $x > 0$. Fix j and k. The derivative of $m_{k,x}(j)$ with respect to x is a constant times

$$(1-x^2)^{k-1}\left[\left(\frac{1+x}{1-x}\right)^j (j-kx) - \left(\frac{1-x}{1+x}\right)^j (j+kx)\right].$$

To show that this derivative is strictly negative, we must argue that

$$\left(\frac{1+x}{1-x}\right)^j j < \left(\frac{1+x}{1-x}\right)^j kx + \left(\frac{1-x}{1+x}\right)^j (j+kx). \tag{50}$$

The right side of (50) is made smaller if k is replaced by $2j^2$, because $|j| \leqq (k/2)^{\frac{1}{2}}$ by assumption. Make this substitution and divide the resulting inequality by j, which is positive by assumption. So (50) will be proved if we demonstrate

$$\left(\frac{1+x}{1-x}\right)^j < \left(\frac{1+x}{1-x}\right)^j 2jx + \left(\frac{1-x}{1+x}\right)^j (1+2jx). \tag{51}$$

But (51) is equivalent to each of the following three inequalities.

$$1 < 2jx + \left(\frac{1-x}{1+x}\right)^{2j}(1+2jx), \tag{52}$$

$$(1+x)^{2j} < (1+x)^{2j}2jx + (1-x)^{2j}(1+2jx), \tag{53}$$

$$(1+x)^{2j} - (1-x)^{2j} < 2jx\left[(1+x)^{2j} + (1-x)^{2j}\right]. \tag{54}$$

Clearly,

$$\binom{2j}{2i-1} < 2j\binom{2j}{2i-2} \qquad \text{for } i = 2,\cdots,j,$$

with equality for $i = 1$. Now (54) follows:

$$(1+x)^{2j} - (1-x)^{2j} = 2\Sigma_{i=1}^{j}\binom{2j}{2i-1}x^{2i-1}$$
$$< 4jx\Sigma_{i=1}^{j+1}\binom{2j}{2i-2}x^{2i-2}$$
$$= 2jx\left[(1+x)^{2j} + (1-x)^{2j}\right]. \qquad \square$$

PROOF OF THEOREM (34). Fix n and k. Recall that $H_{2n,2k}$ is the distribution of $2k$ draws at random without replacement from the urn U_{2n} which has n reds and n blacks; B_{2k} is the distribution with replacement; and R is the number of reds. Abbreviate

$$h(j) = h_{nk}(j) = H_{2n,2k}\{R = k+j\}$$
$$b(j) = b_k(j) = B_{2k}\{R = k+j\}.$$

Thus, h is the centered hypergeometric distribution and b is the centered binomial.

Let J be the set of j's with $h(j) \geqq b(j)$. As (46) implies, $|j| < (k/2)^{\frac{1}{2}}$ for $j \in J$. Now (49) shows that for $j \in J$ and $x \neq 0$,

(55) $$m_{k,x}(j) < m_{k,0}(j) = b(j).$$

Suppose μ is a probability on $[0,1]$, symmetric around $1/2$, with $\mu\{1/2\} < 1$. Abbreviate

$$\phi(j) = P_{\mu,2k}(R = k+j).$$

Of course,

$$\mu = \int \tfrac{1}{2}(\delta_{\frac{1}{2}-y} + \delta_{\frac{1}{2}+y})\mu(dy)$$

so

$$\phi(j) = \int m_{k,2y-1}(j)\mu(dy).$$

Relationship (55) implies

(56) $$\phi(j) < b(j) \qquad \text{for } j \in J.$$

Now

$$\tfrac{1}{2}\|H_{2n,2k} - P_{\mu,2k}\| \geqq H_{2n,2k}\{R-k \in J\} - P_{\mu,2k}\{R-k \in J\}$$
$$= \Sigma_{j\in J} h(j) - \phi(j)$$
$$> \Sigma_{j\in J} h(j) - b(j) \qquad \text{by (56)}$$
$$= \tfrac{1}{2}\|h-b\|$$
$$= \tfrac{1}{2}\|H_{2n,2k} - M_{2k}\| \qquad \text{by (43)}.$$

This completes the argument for symmetric μ. The essence of it is that $h(j) > b(j)$ implies $b(j) > \phi(j)$, where $\phi(j)$ is any symmetric mixture of binomial probabilities.

Unsymmetric μ's are handled by (44). $\square$

PROOF OF THEOREM (35). Define h and b as for the proof of (34). As (43) implies, $\|H_{2n,2k} - B_{2k}\| = \|h - b\|$. We propose to estimate $\|h-b\|$ by the central limit theorem. The argument is a bit technical, and the following heuristic discussion may help explain the results. The probability distributions h and b are both essentially normal with the same mean 0. The variance of b is $2k\cdot(\frac{1}{4})$. The variance of h differs by the finite population correction factor, which is essentially $1-\alpha$ where $\alpha = k/n$. Thus, $\|h-b\|$ is essentially the variation distance between two normal distributions with the same means, and variances differing by the factor $1-\alpha$. This distance is $\phi(\alpha)$. When α is small, $\phi(\alpha) \doteq \gamma\alpha$.

We begin by proving (37). As noted in the proof of (45), $h(j)/b(j) = N/D$, where

(57) $$N = \binom{2n-2k}{n-k+j}\left(\tfrac{1}{2}\right)^{2n-2k},$$

(58) $$D = \binom{2n}{n}\left(\tfrac{1}{2}\right)^{2n}.$$

We are assuming $k < n/2$. The binomial probabilities N and D may be estimated using the local Berry-Esseen theorem, as on page 197 of Petrov (1975):

(59) $$D = (\pi n)^{-\frac{1}{2}} + o(n^{-\frac{1}{2}})$$

(60) $$N = [\pi(n-k)]^{-\frac{1}{2}}\exp(-x^2/2) + o(n^{-\frac{1}{2}})$$

where

(61) $$x = [2/(n-k)]^{\frac{1}{2}}j.$$

The error term in (60) is uniform in j. Thus

(62) $$\frac{h(j)}{b(j)} = \frac{N}{D} = \left(1-\frac{k}{n}\right)^{-\frac{1}{2}}\exp\left(-\frac{1}{2}x^2\right) + o(1),$$

the error term being uniform in j. Now

(63) $$\|h-b\| = \Sigma_j\left|\frac{h(j)}{b(j)} - 1\right|b(j)$$

so

(64) $$\|h-b\| \to \mathrm{E}\left|(1-\alpha)^{-\frac{1}{2}}\exp\left[-\frac{\alpha Z^2}{2(1-\alpha)}\right] - 1\right|$$

where Z is normal with mean 0 and variance 1. This follows by an elementary argument from de Moivre's central limit theorem: relative to b, the random variable defined by (61) is almost normal, with mean 0 and variance $k/n-k \to \alpha/1-\alpha$. The limit in (64) is the $\phi(\alpha)$ of (39), as one sees by the change of variables $x/(1-\alpha)^{\frac{1}{2}} = u$. This completes the proof of (37).

The argument for (36) is similar but more delicate. We are assuming that $n \to \infty$ and $k = o(n)$; however, k need not tend to ∞. We must estimate to within $o(k/n)$. Let $C(k, n)$ be the set of j's with

(65) $$|j| \leqq 2[k\log(n/k)]^{\frac{1}{2}}.$$

Clearly

(66) $$\begin{cases} \|h-b\| = T_1 + T_2, \quad \text{where} \\ T_1 = \Sigma_{j\notin C(k,n)}|h(j)-b(j)| \\ T_2 = \Sigma_{j\in C(k,n)}|h(j)-b(j)|. \end{cases}$$

Now

$$
\begin{aligned}
T_1 &\leqq \Sigma_{j \notin C(k,n)} h(j) + b(j) \\
&= H_{2k,2n}\{|R-k| > 2[k\log(n/k)]^{\frac{1}{2}}\} \\
&\qquad + B_{2k}\{|R-k| > 2[k\log(n/k)]^{\frac{1}{2}}\} \\
&\leqq 2B_{2k}\{|R-k| > 2[k\log(n/k)]^{\frac{1}{2}}\} \\
&\leqq 4(k/n)^4 \\
&= o(k/n).
\end{aligned} \tag{67}
$$

The third step in (67) follows from (46); for more general results in this direction, see Hoeffding (1963). The fourth step follows from Bernstein's inequality, as will now be explained in more detail. Bernstein's inequality appears, for instance, as Theorem 15 on page 52 of Petrov (1975). In our case, the X's of that theorem are $\pm\frac{1}{2}$ with probability $\frac{1}{2}$ each, so his constants g may all be taken as $\frac{1}{4}$, and T as ∞. More explicitly,

$$E[\exp(tX)] \leqq \exp\left[\tfrac{1}{8}t^2\right] \qquad \text{for } 0 \leqq t < \infty \tag{68}$$

because

$$(2m)!/m! > 2^m \qquad \text{for } m = 2, 3, \cdots. \tag{69}$$

Now $R - k$ is the sum of $2k$ of these X's and $G = \Sigma_1^{2k} g = k/2$, so

$$B_{2k}\{R - k \geqq x\} \leqq \exp\{-x^2/k\}.$$

Put $x = 2[k\log(n/k)]^{\frac{1}{2}}$ and use symmetry to complete the argument for (67).

This disposes of T_1, and we turn to T_2. Again, $h(j)/b(j) = N/D$, where the probabilities N and D in (57–58) must be estimated to within $o(k/n)$. This can be done using the Edgeworth expansion, as on page 508 of Feller (1966) or pages 205–206 of Petrov (1975):

$$D = (\pi n)^{-\frac{1}{2}}\left(1 - \frac{1}{8n}\right) + o(n^{-\frac{3}{2}}) \tag{70}$$

$$N = [\pi(n-k)]^{-\frac{1}{2}}[\exp(-x^2/2)]\left[1 - \frac{H_4(x)}{24(n-k)}\right] + o[(n-k)^{-\frac{3}{2}}] \tag{71}$$

where the error term in (71) is uniform in j, the variable x was defined in (61), and H_4 is the fourth Hermite polynomial:

$$H_4(x) = x^4 - 6x^2 + 3. \tag{72}$$

We assume $j \in C(k,n)$, that is, j satisfies (65). All our estimates below will be uniform over j's in this region.

We claim

$$h(j)/b(j) = 1 + \tfrac{1}{2}(k - 2j^2)/n + o(k/n). \tag{73}$$

Indeed, $h(j)/b(j) = N/D$, where N and D are estimated in (70–71). Clearly

$$\frac{1}{D} = (\pi n)^{\frac{1}{2}}\left(1 + \frac{1}{8n}\right) + o(n^{-\frac{3}{2}}).$$

Since $k = o(n)$, the error term in (71) is $o(n^{-\frac{3}{2}})$, and

$$\frac{N}{D} = \left(1 - \frac{k}{n}\right)^{-\frac{1}{2}}\left[\exp\left(-\tfrac{1}{2}x^2\right)\right]\left[1 - \frac{H_4(x)}{24(n-k)}\right]\left(1 + \frac{1}{8n}\right) + o(n^{-\frac{3}{2}}).$$

Clearly,

$$\left(1 - \frac{k}{n}\right)^{-\frac{1}{2}} = 1 + \tfrac{1}{2}\frac{k}{n} + o\left(\frac{k}{n}\right).$$

Relation (65) implies $j^2/n - k = j^2/n + o(k/n)$; relations (61) and (65) imply that

(74) $$x^2 = 0\left(\frac{k}{n}\log\frac{n}{k}\right)$$

so $x^4 = o(k/n)$ and

$$\exp\left(-\tfrac{1}{2}x^2\right) = 1 - \frac{j^2}{n} + o\left(\frac{k}{n}\right).$$

Likewise, (72) and (74) imply

$$\frac{H_4(x)}{24(n-k)} = \frac{1}{8n} + o\left(\frac{k}{n}\right).$$

Putting these estimates together proves (73): note that $1/n^2 = o(k/n)$.

We can now estimate T_2 in (66):

(75) $$\frac{n}{k}T_2 = \Sigma_{j \in C(k,n)}\frac{n}{k}\cdot\left|\frac{h(j)}{b(j)} - 1\right|\cdot b(j)$$

so

(76) $$\frac{n}{k}T_2 = \tfrac{1}{2}\Sigma_{j \in C(k,n)}\left|1 - \frac{2j^2}{k}\right|\cdot b(j) + o(1),$$

because estimate (73) is uniform over j in $C(k, n)$.

Recall that B_{2k} is the distribution of $2k$ draws made at random with replacement from the urn U_{2n} which contains n red balls and n blacks: R is the number of red balls among the draws.

Let $Z_k = (R - k)/(k/2)^{\frac{1}{2}}$. Now $\int Z_k^4\, dB_{2k}$ is uniformly bounded, so $|1 - Z_k^2|$ is uniformly B_{2k}-integrable. The sum on the right side of (76) is the B_{2k}-integral of $|1 - Z_k^2|$ over the region $|Z_k| \leqq [8\log(n/k)]^{\frac{1}{2}}$. By de Moivre's central limit theorem, Z_k is almost normal with mean 0 and variance 1, so our B_{2k}-integral converges to the γ of (38). □

REMARK. An argument similar to the one for Theorem (40) gives results for the nonsymmetric case. Let the urn U_n contain r_n red balls and b_n black balls, where $r_n + b_n = n$. Suppose there is an $\varepsilon > 0$ for which $\varepsilon \leqq r_n/n \leqq 1 - \varepsilon$. Suppose too that $k_n \to \infty$ with $k_n/n \to 0$. Let H_{n,k_n} be the distribution of k_n draws made at random from U_n without replacement, and B_{n,k_n} the distribution with replacement. Then

(77) $$\|H_{n,k_n} - P_{\mu,k_n}\| \geqq \|H_{n,k_n} - B_{n,k_n}\| + o(k_n/n)$$

(78) $$\|H_{n,k_n} - B_{n,k_n}\| = \gamma k_n/n + o(k_n/n).$$

The error terms are uniform over the indicated region. The assumption that r_n/n be bounded away from 0 and 1 is essential. Indeed, suppose $r_n = 1$ and $k_n = n - 1$. The variation distance on the left side of (78) is of order $(k_n/n)^2$. And the closest binomial measure to H_{n,k_n} is at variation distance of order $(k_n/n)^3$, so (77) and (78) are false.

Acknowledgments. We thank Andre Cabannes and Michael Perlman for their comments. We also thank an extremely helpful referee.

Note added in proof. Sandy Zabell called to our attention a paper by Stam with results which overlap ours.

REFERENCES

CHOQUET, G. (1969). *Lectures on Analysis*, I. Benjamin, Reading, Mass.

CRISMA, L (1971). Sulla prosequibilità di processi scambiabili. *Rend. Ist. Mat. Univ. Trieste* **3** 96–124.

DE FINETTI, B. (1969). Sulla prosequibilità di processi aleatori scambiabili. *Rend. Ist. 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. (1979). de Finetti's generalizations of exchangeability. In *Foundations of Subjective Probabilities*, (R. Jeffrey, ed.). Univ. Calif. Press, Los Angeles.

DUBINS, L. E. and FREEDMAN, D. (1964). Measurable sets of measures. *Pacific J. Math.* **14** 1211–1222.

DUBINS, L. E. and FREEDMAN, D. (1979). Exchangeable processes need not be distributed mixtures of independent, identically distributed random variables. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **48** 115–132.

FELLER, W. (1966). *An Introduction to Probability Theory and its Applications*, II. Wiley, New York.

FELLER, W. (1968). *An Introduction to Probability Theory and its Applications*, I, 3rd ed. Wiley, New York.

FREEDMAN, D. (1977). A remark on the difference between sampling with and without replacement. *J. Amer. Statist. Assoc.* **73** 681.

GALAMBOS, J. (1978). *The Asymptotic Theory of Extreme Order Statistics*. Wiley, New York.

HEWITT, E. and SAVAGE, L. J. (1955). Symmetric measures on Cartesian products. *Trans. Amer. Math. Soc.* **80** 470–501.

HOEFFDING, W. (1963). Probability inequalities for sums of random variables. *J. Amer. Statist. Assoc.* **58** 13–30.

HULONICKI, A. and PHELPS, R. (1968). Some applications of tensor products of partially ordered linear spaces. *J. Functional Analysis* **2** 177–201.

KENDALL, D. G. (1967). On finite and infinite sequences of exchangeable events. *Studia Sci. Math. Hungar.* **2** 319–327.

MUNFORD, A. G. (1977). A note on the uniformity assumption in the birthday problem. *Amer. Statist.* **31** 119.

OLSHEN, R. (1973). A note on exchangeable sequences. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **28** 317–321.

PETROV, V. V. (1975). *Sums of Independent Random Variables*. Springer, New York.

RINOTT, J. (1973). Multivariate majorization and rearrangement inequalities with some applications to probability and statistics. *Israel J. Math.* **15** 60–77.

STAM, A. J. (1978). Distance between sampling with and without replacement. *Statistica Neerlandica* **32** 81–91.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA, BERKELEY
BERKELEY, CALIFORNIA, 94720