Journal of Theoretical Probability, Vol. 19, No. 3, July 2006 (© 2006)
DOI: 10.1007/s10959-006-0028-z

Definetti's Theorem for Abstract Finite Exchangeable Sequences

G. Jay Kerns[1] **and Gábor J. Székely**[2,3,4]

Received September 27, 2004; revised April 14, 2005

We show that a finite collection of exchangeable random variables on an arbitrary measurable space is a signed mixture of i.i.d. random variables. Two applications of this idea are examined, one concerning Bayesian consistency, in which it is established that a sequence of posterior distributions continues to converge to the true value of a parameter θ under much wider assumptions than are ordinarily supposed, the next pertaining to Statistical Physics where it is demonstrated that the quantum statistics of Fermi-Dirac may be derived from the statistics of classical (i.e. independent) particles by means of a signed mixture of multinomial distributions.

KEY WORDS: Exchangeable variables; de Finetti's theorem; finite exchangeable sequence; signed measure; extreme points.

2000 MATHEMATICAL SUBJECT CLASSIFICATION. Primary: 60B05; Secondary: 60E99, 62E99.

1. INTRODUCTION

The roots of the idea of finite exchangeability of events can be traced back to the new "Algebra" in Ref. 11 where the game "Rencontre" ("Matches") was analyzed (Problem XXXV. and Problem XXXVI.) In the

[1] Department of Mathematics and Statistics, Youngstown State University, Youngstown, OH 44555-0002, USA. E-mail: gjkerns@math.ysu.edu
[2] Department of Mathematics and Statistics, Bowling Green State University, Bowling Green, OH 43403-0221, USA
[3] Renyi Institute of Mathematics, Hungarian Academy of Sciences, P.O.B. 127, Budapest 1364, Hungary. E-mail: gabors@bgnet.bgsu.edu
[4] To whom correspondence should be addressed.

Preface de Moivre writes enthusiastically, “In the 35th and 36th Problems I explain a new sort of Algebra,... I assure the Reader, that the Method I have followed has a degree of Simplicity, not to say of Generality,...”.

Infinite sequences of exchangeable events seem to have first been discussed by Haag.[25] Not long after, de Finetti[12] independently introduced exchangeable random variables and proved his famous representation theorem for the 2-valued case. More precisely: if $X_1, X_2, \dots,$ take values in $\{0, 1\}$ and

$$\mathbb{P}(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = \mathbb{P}(X_1 = x_{\pi(1)}, X_2 = x_{\pi(2)}, \dots, X_n = x_{\pi(n)})$$

holds for all $n \in \mathbb{N}$, $x_1, \dots, x_n \in \{0, 1\}$, and all finite permutations π of $\{1, 2, \dots, n\}$, then for some measure μ on $[0, 1]$

$$\mathbb{P}(X_1 = x_1, X_2 = x_2, \dots X_n = x_n) = \int_0^1 p^{\sum x_i} (1-p)^{n-\sum x_i} d\mu(p).$$

In words, $X_1, X_2, \dots,$ are conditionally i.i.d. given the random variable p, which is distributed according to the measure μ. The success and impact of exchangeable random variables is well documented. Recently Dawid[10] wrote, “Perhaps the greatest and most original success of de Finetti’s methodological program is his theory of exchangeability (Ref. 13).”

The theorem has been extended and generalized in various directions. De Finetti himself showed that his representation held for real-valued random variables.[13] Many years later Dynkin[21] replaced $\mathbb{R}$ with more general spaces that are in some sense separable, and Hewitt and Savage[26] soon after extended the result to compact Hausdorff spaces, from which it can readily be extended to, for example, Polish or locally compact spaces. The theorem does not hold without some topological assumptions, however, as was shown by Dubins and Freedman.[20]

It is well known that de Finetti’s theorem also does not hold in general for finite sequences of exchangeable random variables; for an easy example showing failure in case $n = 2$ (see Refs. 16, 17 or 34). It is not difficult to see the trouble: suppose that $X_1, X_2, \dots, X_N$ are conditionally i.i.d. given some random variable θ which is distributed according to some probability measure ν. Then for $1 \leqslant i, j \leqslant N$,

$$\begin{aligned}
\mathbb{E}X_i X_j &= \mathbb{E}X_1 X_2 \quad \text{(by exchangeability)} \\
&= \mathbb{E}_\nu\{\mathbb{E}[X_1 X_2 | \theta]\} \\
&= \mathbb{E}_\nu\{\mathbb{E}[X_1 | \theta]\ \mathbb{E}[X_2 | \theta]\} \quad \text{(conditional independence)} \\
&= \mathbb{E}_\nu\{(\mathbb{E}[X_1 | \theta])^2\} \quad \text{(conditionally identically dist'd).}
\end{aligned}$$

And of course $\mathbb{E}X_i = \mathbb{E}_\nu\{\mathbb{E}[X_1|\theta]\} = \mathbb{E}X_j$, so that

$$\begin{aligned}\mathrm{Cov}(X_i, X_j) &= \mathbb{E}_\nu\{(\mathbb{E}[X_1|\theta])^2\} - (\mathbb{E}_\nu\{\mathbb{E}[X_1|\theta]\})^2 \\ &= \mathrm{Var}_\nu(\mathbb{E}[X_1|\theta]) \\ &\geqslant 0.\end{aligned}$$

This gives immediately the familiar fact that the classical de Finetti representation works only for sequences that are nonnegatively correlated, and instances where this condition fails are often encountered in practice; take, for example, hypergeometric sequences.

In response to this problem there have been several versions and modifications of the theorem developed for the finite case. Kendall(29) (also Ref. 14) showed that every finite system of exchangeable events is equivalent to a random sampling scheme without replacement, where the number of items in the sampling has an arbitrary distribution. Diaconis(16) and Diaconis and Freedman(17) used this to find total variation distances to the closest mixture of i.i.d. random variables, (which turns out to imply de Finetti's theorem in the limit). Gnedin(24) explored conditions on the density of a finite exchangeable sequence and found a criterion for extendability to an infinite sequence. Several other aspects of finite exchangeability are in Refs. 1, 18, 19, 30, 35 give useful surveys of exchangeability (see also Ref. 33).

The present paper reexamines the failure in the finite case of de Finetti's theorem and in the process reveals that while for infinite exchangeable sequences the classical (that is, nonnegative) mixtures of i.i.d. random variables are sufficient, with some finite sequences, we need to consider an extended notion of "mixture" to retain de Finetti's convenient representation. In particular, we have the

Theorem 1.1. Let $(S, \mathcal{B})$ be an abstract measurable space and write S^* for the set of probability measures on $(S, \mathcal{B})$. Endow S^* with the smallest σ-field $\mathcal{B}^*$ making $p \mapsto p(A)$ measurable for all $A \in \mathcal{B}$. If $p \in S^*$, then p^n is the product measure on $(S^n, \mathcal{B}^n)$.

If $\mathbb{P}$ is an exchangeable probability on $(S^n, \mathcal{B}^n)$ then there exists a signed measure ν of bounded variation such that

$$\mathbb{P}(A) = \int_{S^*} p^n(A)\, d\nu(p) \qquad \text{for all } A \in \mathcal{B}^n.$$

Further, ν satisfies $\nu(S^*) = 1$.

It is noted that in the book by Dellacherie and Meyer(15) there is presented a sketch of how a proof of this result would proceed (credited to P. Cartier) in the more restrictive scenario of an exchangeable law on $\mathbb{R}^n$. Their treatment is concise, and they remark, "...we shall not give any details, since for us this result is a luxury, which we shall not use."

Some years later, it appears that Jaynes[27] independently discovered this theorem in the case where the random variables take only the values $\{0, 1\}$. In that essay he alludes to the extension of the theorem to the abstract case, saying, "...a more powerful and abstract approach, which does not require us to go into all that detail, was discovered by Dr. Eric Mjolsness while he was a student of [mine]. We hope that, with its publication, the useful results of this representation will become more readily obtainable...". Alas, to the knowledge of the present authors that publication never appeared.

In any event, this research is distinct from the above papers in a number of ways. First, this result holds for an arbitrary measurable space without any topological assumptions. As we have stated, Jaynes proved it for $\{0, 1\}$, while Dellacherie and Meyer's argument[15] works for real-valued random variables. The great generality of this paper is afforded in part because, in contrast to Ref. 20, one deals with finite sequences which avoids some of the pathologies that arise with infinite product spaces and limits, and it is also due in part to clever devices used by Diaconis and Freedman.[17]

Second, the method of proof of the representation is of a different character and may have some independent interest. For example, de Finetti[14] shows that finite exchangeable sequences are mixtures of hypergeometric processes, and then takes a weak limit as the sample size increases without bound. For us, there is no limit whatever; we have only finitely many exchangeable random variables to use. Similarly, Kendall[29] used the (Reversed) Martingale Convergence Theorem combined with the hypergeometric mixture representation to show the existence of a sigma algebra conditional on which $X_1, X_2, \dots$, are i.i.d., again a limiting argument. Hewitt and Savage[26] start the analysis with an infinite sequence and derive its mixture representation; we can get no help from there. And while the reasoning in Ref. 27 applies to finite sequences, nevertheless his proof uses Bernstein and Legendre polynomials to find a continuous mixture. Ours instead manipulates discrete sums of urn measures in the spirit of Refs. 16 and 17. The argument in Ref. 15 is more of an algebraic approach, in which they compare coefficients in certain expansions of point masses. Their idea easily generalizes to $\mathbb{R}^n$, however, the method presented here is perhaps more natural, since it capitalizes on the intuition provided by the urn process analogy.

Third, we present two applications of the idea of finite exchangeability, that is, results that specifically use the signed mixture representation to provide insight in the areas of Bayesian theory and Statistical Physics, both of which seem to be new.

And lastly, the conclusion of this paper usefully complements what is currently known regarding bounds for abstract finite exchangeable sequences. Indeed, Diaconis and Freedman,[17] and Freedman[23] showed that for a finite exchangeable real-valued sequence of length k that can be embedded in a sequence of length $N \geqslant k$, a classical mixture of an i.i.d. sequence is at most $2(1 - N_k/N^k)$ away (in terms of total variation distance), where $N_k = N(N-1)\ldots(N-k+1)$. This bound is sharp.

Now, suppose we have an exchangeable sequence that is only *slightly* negatively correlated, say $\rho = -0.01$. Then it is easy to see (in Refs. 30 and 34, for example) that N can be at most 100. One can see the reason why by considering exchangeable $X_1, X_2, \ldots, X_N$ with X_1 having finite nonzero variance σ^2 and correlation coefficient with X_2 being ρ. Then the variance of $\sum_{i=1}^N X_i$ is

$$\sum_{i=1}^{N} \sigma^2 + \sum_{i \neq j} \rho\sigma^2 = N\sigma^2[1 + (N-1)\rho]$$

and since the above quantity is positive it follows

$$\rho > -(N-1)^{-1}$$

or in other words, the integer N is at most 100.

The bound $2(1 - 100_k/100^k)$ monotonically increases as a function of k, and by the time $k = 12$, it has already passed the value 1, the maximum total variation distance between any two distributions being of course 2. From a practical standpoint this means that for even moderately sized samples, the present bound gives little insight. It is completely unclear how close the closest classical mixture is, and the situation only worsens as the variables become more correlated. Again, with an infinite sequence the above correlation inequality would hold for all $N \geqslant 1$, and in such a case ρ would therefore be nonnegative.

This paper guarantees an exact representation, for any finite k, and regardless of the underlying correlation structure. In an application to Bayesian theory, we show that by allowing the mixture to stand in for a Prior distribution one can formally justify common practical procedures; in addition, the resulting Posteriors continue to converge to degenerate distributions under the usual regularity conditions. Thus any nonclassical behavior in the Posterior would become negligible in the limit.

We present a second application in the area of Statistical Physics. We demonstrate that the quantum statistics of Fermi-Dirac may be derived from the statistics of classical (i.e. independent) particles by means of a signed mixture of multinomial distributions, or Maxwell–Boltzmann statistics. This work continues in the vein of work by Bach *et al.*,[4] who performed a similar derivation of Bose–Einstein Statistics using the classical de Finetti theorem. In many ways the resulting asymmetry for Fermi-Dirac statistics is eliminated.

As a final remark, it should be pointed out that the use of signed measures in probability theory is by no means new; it has claimed even proponents such as Refs. 5 and 22. The existing literature is voluminous. For the purposes of this paper those results are not needed, but the interested reader could begin by consulting the survey by Mückenheim[32].

2. FINITE EXCHANGEABLE SEQUENCES

We start by extending some earlier results of Ref. 16 from the space $\{0, 1\}$ to the more general $\{0, 1, \dots, n\}$. Let $\mathcal{P}_n$ represent all probabilities on $\prod_{i=1}^{n} S_i$, where $S_i = \{s_0, s_1, \dots, s_{n-1}\}$ for each i. Then $\mathcal{P}_n$ is a $n^n - 1$ dimensional simplex which is naturally embedded in Euclidean n^n space.

$\mathcal{P}_n$ can be coordinatized by writing $\mathbf{p} = (p_0, p_1, \dots, p_{n^n-1})$ where p_j represents the probability of the outcome j, and $j = 0, 1, \dots, n^n - 1$ is thought of as having its n-ary representation, written with n n-ary digits. Thus if $n = 3$, then $j = 5$ refers to the point 012. Let $\Omega_n(\mathbf{k}) = \Omega_n(k_0, k_1, \dots, k_{n-1})$ be the set of $j, 0 \leqslant j < n^n$, with exactly k_0 digits 0, k_1 digits 1, $\dots, k_{n-1}$ digits $n-1$. The number of elements in $\Omega_n(\mathbf{k})$ is $n!/(k_0!k_1! \cdots k_{n-1}!)$.

Let $\mathcal{E}_n$ be the exchangeable measures in $\mathcal{P}_n$; then $\mathcal{E}_n$ is convex as a subset of $\mathcal{P}_n$.

Lemma 2.1. $\mathcal{E}_n$ has $\binom{2n-1}{n}$ extreme points $\mathbf{h_0}, \mathbf{h_1}, \dots, \mathbf{h}_{\binom{2n-1}{n}-1}$, where $\mathbf{h_k}$ is the measure putting mass $k_0!k_1! \cdots k_{n-1}!/n!$ at each of the coordinates $i \in \Omega_n(\mathbf{k})$, and mass 0 elsewhere. Here, the index vector $\mathbf{k} = (k_0, k_1, \dots, k_{n-1})$ runs over all possible distinct "urns", as described in the proof below.

Each exchangeable probability $\mathbf{p} \in \mathcal{E}_n$ has a unique representation as a mixture of the $\binom{2n-1}{n}$ extreme points.

Proof. This is a direct generalization of Ref. 16 results to a higher dimensional setting. Intuitively, $\mathbf{h_k}$ is a column vector representing the measure associated with n drawings without replacement from the unique urn containing n balls, k_0 marked with s_0, k_1 marked with $s_1, \dots,$ k_{n-1} marked with s_{n-1}; $\mathbf{h}$ stands for "hypergeometric". Each $\mathbf{h_k}$ is exchangeable, and finding the total number of distinct urns amounts to finding the number of ways to distribute n indistinguishable balls into n boxes marked $s_0, s_1, \dots, s_{n-1}$, respectively. It is well known that the number of ways to distribute r balls into m (ordered) boxes is just $\binom{m+r-1}{r}$. In this case, $m=r=n$.

Note that when $S=\{s_0, s_1\}=\{0, 1\}$, this matches Diaconis[16] result where $\mathcal{E}_n$ had $n+1$ extreme points, because in that case $m=2$ and $r=n$, so that

$$\binom{2+n-1}{n}=\binom{n+1}{n}=n+1.$$

Now suppose that $\mathbf{h_k}$ is a mixture of two other exchangeable measures, that is $\mathbf{h_k}=p\mathbf{a}+(1-p)\mathbf{b}$, where $\mathbf{a}$ and $\mathbf{b}$ are exchangeable and $0<p<1$. Then $\mathbf{a}$ and $\mathbf{b}$ put mass zero at every point where $\mathbf{h_k}$ puts mass zero, namely, on the complement of $\Omega_n(\mathbf{k})$.

By exchangeability, outcomes with the same number of s_0's, s_1's, $\dots$, s_{n-1}'s have the same probability. Therefore, the entries of both $\mathbf{a}$ and $\mathbf{b}$ must be equal at all coordinates $j\in\Omega_n(\mathbf{k})$. But the sum of the entries is 1 in each vector, so the mass at each $j\in\Omega_n(\mathbf{k})$ must be $k_0!k_1!\cdots k_{n-1}!/n!$. This implies $\mathbf{a}=\mathbf{b}$, and $\mathbf{h_k}$ is in fact an extreme point of $\mathcal{E}_n$.

It is well known that every point of a simplex has a unique representation as a mixture of extreme points. □

Now we see that $\mathcal{E}_n$ is a $\binom{2n-1}{n}$ sided polyhedron. The extreme points $\mathbf{h_0}, \mathbf{h_1}, \dots, \mathbf{h}_{\binom{\mathbf{2n-1}}{\mathbf{n}}-\mathbf{1}}$ are linearly independent because they are supported on the disjoint sets $\Omega_n(\mathbf{j}), \mathbf{j}=\mathbf{0}, \dots, \binom{\mathbf{2n-1}}{\mathbf{n}}-\mathbf{1}$. And, a probability $\mathbf{p}$ is in $\mathcal{E}_n$ if and only if it is constant on the sets $\Omega_n(\mathbf{k})$.

An interesting subclass is the class $\mathcal{M}_n\subset\mathcal{E}_n$ of i.i.d. probabilities on $\mathcal{P}_n$. $\mathcal{M}_n$ can be parameterized as an nth degree polynomial of $n-1$ variables:

$$\mathbf{m}=(p_0^n, p_0^{n-1}p_1, \dots, p_{n-2}p_{n-1}^{n-1}, p_{n-1}^n).$$

In general, for $j\in\Omega_n(\mathbf{k})$, we have $\mathbf{m}_j=\prod_{i=0}^{n-1}p_i^{k_i}$, where $p_i\geqslant 0$, $i=0, \dots, n-1$, and $p_0+p_1+\cdots+p_{n-1}=1$. The set $\mathcal{M}_n$ is a smooth surface that twists through $\mathcal{E}_n$.

The class of mixtures of i.i.d. probabilities (with which de Finetti's theorem is associated) is a convex set lying in $\mathcal{E}_n$. Measures in this set can be represented in uncountably many ways.

The above has similarly been a many-dimensional analogue of Ref. 16. This next proposition may be viewed as an alternative proof of the claim for $\mathbb{R}^n$ made in Ref. 15.

Proposition 2.1. Each exchangeable probability $\mathbf{p}\in\mathcal{E}_n$ can be written as a (possibly) signed mixture of measures $\{\mathbf{v}_i\}\subset\mathcal{M}_n$.

Proof. Consider the $\binom{2n-1}{n}-1$ dimensional hyperplane $\mathcal{H}$ of Euclidean n^n space that is determined by the linearly independent hypergeometric vectors $\{\mathbf{h_k}\}$, and consider the surface $\mathcal{M}_n \subset \mathcal{H}$. We choose $\binom{2n-1}{n}$ linearly independent vectors $\{\mathbf{v}_i : i=1,\ldots,\binom{2n-1}{n}\}$ in the uncountable collection $\mathcal{M}_n$ to construct a basis for $\mathcal{H}$. Then any $\mathbf{p}\in\mathcal{E}_n\subset\mathcal{H}$ can be written as a (possibly) signed linear combination $\mathbf{p}=\sum_i a_i\mathbf{v}_i$. The fact that $\sum_i a_i=1$ follows from the observation that $\mathbf{v}_i(\prod_{j=1}^n S_j)=1$ for all i, so that $\mathbf{p}(\prod_{j=1}^n S_j)=\sum_i a_i=1$. □

In symmetry to the above we may, as Diaconis[17] did, define the column vector $\mathbf{m_k}$ to represent the measure associated with n drawings with replacement from the urn containing k_0 balls marked with s_0, k_1 marked with $s_1,\ldots,$ k_{n-1} marked with s_{n-1}; $\mathbf{m}$ stands for "multinomial". From our Proposition 2.1 follows the corollary.

Corollary 2.1. Each extreme point $\mathbf{h_k}$ can be written as a unique signed mixture of the multinomial measures $\{\mathbf{m_j}\}$.

Proof. In essence, we have only chosen $\{\mathbf{m_j}\}$ as a particular basis for the hyperplane $\mathcal{H}$. For each $\mathbf{k}$, the measure $\mathbf{m_k}$ is exchangeable, so it is (from Lemma 2.1) a unique mixture of the measures $\{\mathbf{h_j}\}$. That is, there exist nonnegative weights $w_0, w_1,\ldots, w_{\binom{2n-1}{n}-1}$ that sum to one satisfying

$$\mathbf{m_k}=\sum_{j=0}^{\binom{2n-1}{n}-1} w_j\,\mathbf{h_j}.$$

A moment's reflection will convince us that the weights are *exactly* the ordinary multinomial distribution; in fact, we can display them explicitly in the form

$$\mathbf{m_k}=\sum_{\mathbf{j}}\binom{n}{j_0\, j_1\cdots j_{n-1}}\left(\frac{k_0}{n}\right)^{j_0}\cdots\left(\frac{k_{n-1}}{n}\right)^{j_{n-1}}\cdot\mathbf{h_j},$$

where we interpret $0^0=1$ and the summation runs over the $\binom{2n-1}{n}$ indices $\mathbf{j}$.

Letting $\mathbf{w_k}$ be the column vector containing the weights associated with $\mathbf{m_k}$ we can write in matrix notation $\mathbf{M}=\mathbf{HW}$, where

$$\mathbf{M}=\left[\ \mathbf{m_0}\ \mathbf{m_1}\ \cdots\ \mathbf{m}_{\binom{2n-1}{n}}-1\ \right],\quad \mathbf{H}=\left[\ \mathbf{h_0}\ \mathbf{h_1}\ \cdots\ \mathbf{h}_{\binom{2n-1}{n}-1}\ \right],\ \text{and}$$

$$\mathbf{W}=\left[\ \mathbf{w_0}\ \mathbf{w_1}\ \cdots\ \mathbf{w}_{\binom{2n-1}{n}-1}\ \right].$$

Now, the matrix $\mathbf{W}$ is invertible; indeed, $\mathbf{W}$ is merely the change of basis matrix from the basis $\{\mathbf{h}_j\}$ to the basis $\{\mathbf{m}_j\}$. Linear independence of the $\mathbf{w}_j$'s follows from the observation that any column vector $\boldsymbol{\beta}$ satisfying $\mathbf{W}\boldsymbol{\beta}=\mathbf{0}$ also satisfies $\mathbf{M}\boldsymbol{\beta}=\mathbf{H}(\mathbf{W}\boldsymbol{\beta})=\mathbf{0}$. And the vectors $\mathbf{m_k}$, each corresponding to distinct urns $\mathbf{k}$, are easily verified to be linearly independent, hence $\boldsymbol{\beta}$ must be $\mathbf{0}$ and invertibility of $\mathbf{W}$ follows.

The required representation is obtained by inverting $\mathbf{W}$ to yield $\mathbf{H}=\mathbf{M}\mathbf{W}^{-1}$, which gives each extreme point $\mathbf{h_k}$ to be the mixture of the vectors $\{\mathbf{m_j}\}$, with the weights being the corresponding $\mathbf{k}$-th column of $\mathbf{W}^{-1}$. □

Note that in the above proof no claim was made either way regarding the sign of the weights of $\mathbf{W}^{-1}$. The fact of the matter is that except in the degenerate cases an extreme point $\mathbf{h_k}$ will be a signed mixture of the $\mathbf{m_j}$'s. For an arbitrary exchangeable measure nothing can in general be said. The deciding factor is the vector's location in $\mathcal{E}_n$; if it happens to fall in the convex set of mixtures of i.i.d. vectors in $\mathcal{M}_n$, then its representation will be classical. If it falls too close to the extreme points $\mathbf{h_k}$, then its existing mixture representation is necessarily signed.

Example. Now we go to make precise that which has been mentioned above. An illuminating example occurs already in case $n=2$. Consider sampling without replacement from an urn containing two balls, one marked 0 and the other marked 1. Here the exchangeable random variables $X_1,\ X_2$ satisfy

$$\mathbb{P}(X_1=1, X_2=0)=\mathbb{P}(X_1=0, X_2=1)=\frac{1}{2},$$

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

Suppose for the moment that there existed a nonnegative mixing measure μ for this case. Then one would have

$$0=\mathbb{P}(X_1=1, X_2=1)=\int_0^1 p^2 d\mu(p)$$

implying that μ puts mass 1 at the point $p=0$, but on the other hand,

$$0=\mathbb{P}(X_1=0, X_2=0)=\int_0^1 (1-p)^2 d\mu(p),$$

which implies that μ puts mass 1 at the point $p=1$, which is impossible. This particular example is the one Diaconis[16] and Diaconis and

Freedman[17] used to display the inadequacy of de Finetti's theorem when applied to certain finite sequences of exchangeable random variables.

Now consider the space $\mathcal{P}_2$ of all possible assignments of probabilities to the four events $\{X_1 = x_1, X_2 = x_2\}, x_i = 0, 1$, that is, all probabilities on $\{0, 1\}^2$.$\mathcal{P}_2$ is a three dimensional simplex which may be embedded in $\mathbb{R}^4$.

These probabilities are represented as the set of all points $\mathbf{p} = (p_0, p_1, p_2, p_3)$, where $p_i \geqslant 0$ and $p_0 + p_1 + p_2 + p_3 = 1$. Set

$$\begin{aligned} p_0 &= \mathbb{P}(X_1 = 0, X_2 = 0), \\ p_1 &= \mathbb{P}(X_1 = 0, X_2 = 1), \\ p_2 &= \mathbb{P}(X_1 = 1, X_2 = 0), \\ p_3 &= \mathbb{P}(X_1 = 1, X_2 = 1). \end{aligned}$$

(Notice that p_j represents the probability of the outcome $j, j = 0, 1, 2, 3$, written in binary notation.) Here we have the sets

$$\Omega_2(2, 0) = \{0\}, \qquad \Omega_2(1, 1) = \{1, 2\},$$

$$\Omega_2(0, 2) = \{3\}.$$

The subclass $\mathcal{E}_2$ of exchangeable probabilities is the set of all $\mathbf{p}$ where $p_1 = p_2$, that is, the set of all $\mathbf{p}$ which are constant on the sets $\Omega_2(\mathbf{k}), \mathbf{k} = (2, 0), (1, 1), (0, 2)$.$\mathcal{E}$ is convex as a subset of $\mathcal{P}_2$.

In particular, the hypergeometric vectors (extreme points) $\mathbf{h}_\mathbf{k}$ are given by

$$\mathbf{h}_{(2,0)} = (1, 0, 0, 0)^T, \qquad \mathbf{h}_{(1,1)} = (0, 1/2, 1/2, 0)^T, \text{ and}$$

$$\mathbf{h}_{(0,2)} = (0, 0, 0, 1)^T.$$

It is clear that the $\mathbf{h}_\mathbf{k}$ are linearly independent, since they are supported on the disjoint sets $\Omega_2(\mathbf{k})$.

Moving to the class $\mathcal{M}_2$, we recognize it as the set of vectors $((1 - p)^2, (1 - p)p, p(1 - p), p^2)$ parametrized in p for the values $0 \leqslant p \leqslant 1$, where of course p would represent $\mathbb{P}(X_1 = 1)$. As special cases we identify the multinomial vectors

$$\mathbf{m}_{(2,0)} = (1, 0, 0, 0)^T, \qquad \mathbf{m}_{(1,1)} = (1/4, 1/4, 1/4, 1/4)^T, \text{ and}$$

$$\mathbf{m}_{(0,2)} = (0, 0, 0, 1)^T.$$

Notice that $\mathbf{m}_{(2,0)} = \mathbf{h}_{(2,0)}$ and $\mathbf{m}_{(0,2)} = \mathbf{h}_{(0,2)}$; in the degenerate cases it does not matter whether one samples with or without replacement.

Now consider the 2-dimensional plane $\mathcal{H}$ of Euclidean 4-space that is spanned by the vectors $\mathbf{h_k}$. Since $\mathcal{E}_2 \subset \mathcal{H}$, then any exchangeable $\mathbf{p}$ may be written as a (possibly signed) linear combination of the $\mathbf{h_k}$'s. For the exchangeable vectors $\mathbf{m_k}$ this mixture takes the form

$$\left(\mathbf{m}_{(2,0)} \ \mathbf{m}_{(1,1)} \ \mathbf{m}_{(0,2)} \right) = \left(\mathbf{h}_{(2,0)} \ \mathbf{h}_{(1,1)} \ \mathbf{h}_{(0,2)} \right) \cdot \mathbf{W},$$

where

$$\mathbf{W} = \begin{pmatrix} 1 & 0 & 1/4 \\ 0 & 1 & 1/4 \\ 0 & 0 & 1/2 \end{pmatrix}.$$

Of course, the multinomial $\mathbf{m_k}$'s lie also in $\mathcal{H}$, and further span the subspace. By a change of basis the $\mathbf{h_k}$'s are signed mixtures of the $\mathbf{m_k}$'s. In other words $\mathbf{H} = \mathbf{M}\mathbf{W}^{-1}$, where

$$\mathbf{W}^{-1} = \begin{pmatrix} 1 & 0 & -1/2 \\ 0 & 1 & -1/2 \\ 0 & 0 & 2 \end{pmatrix}.$$

The two negative entries are evidence that when random variables are negatively correlated we must resort to extended notions of mixtures to retain de Finetti's convenient representation. Going back to the counterexample quoted earlier, from the matrix we may conclude that to resolve the problem it suffices to allow μ to place mass $-1/2$ on the points $x = 0$ and $x = 1$, and place mass 2 on $x = 1/2$.

3. PROOF OF THE THEOREM

Proof. For $\boldsymbol{s} = (s_0, s_1, \ldots, s_{n-1}) \in S^n$, let $U(\boldsymbol{s})$ be the urn consisting of n balls, marked $s_0, s_1, \ldots, s_{n-1}$, respectively. Let $H_{U(\boldsymbol{s})}$ be the distribution of n draws made at random without replacement from $U(\boldsymbol{s})$. Thus, $H_{U(\boldsymbol{s})}$ is a probability on $(S^n, \mathcal{B}^n)$. The map $\boldsymbol{s} \mapsto H_{U(\boldsymbol{s})}(A)$ is measurable on $(S^n, \mathcal{B}^n)$ for each $A \in \mathcal{B}^n$.

We shall see now that exchangeability of $\mathbb{P}$ entails

$$\mathbb{P}(A) = \int_{S^n} H_{U(\boldsymbol{s})}(A)\, \mathbb{P}(d\boldsymbol{s}).$$

This is true because $H_{U(\boldsymbol{s})}$ is the measure placing mass $1/n!$ at the $n!$ points which are permutations of $\boldsymbol{s} = (s_0, s_1, \ldots, s_{n-1})$, in which case $H_{U(\boldsymbol{s})}(A) = 1/n! \sum_\pi 1_A(\pi \boldsymbol{s})$, where π is a permutation of the first n positive integers and the summation extends over the $n!$ such permutations π. Next we calculate

$$\begin{aligned}\int_{S^n} H_{U(\mathbf{s})}(A)\,\mathbb{P}(d\mathbf{s}) &= \int_{S^n} \frac{1}{n!}\sum_{\pi} 1_A(\pi\mathbf{s})\,\mathbb{P}(d\mathbf{s})\\ &= \frac{1}{n!}\sum_{\pi}\int_{S^n} 1_A(\pi\mathbf{s})\,\mathbb{P}(d\mathbf{s})\\ &= \frac{1}{n!}\sum_{\pi}\int_{S^n} 1_A(\mathbf{s})\,\mathbb{P}(d\mathbf{s}) \quad \text{(by exchangeability)}\\ &= \frac{1}{n!}\cdot n!\,\mathbb{P}(A)\\ &= \mathbb{P}(A).\end{aligned}$$

Furthermore, from our Corollary 2.1 we may write

$$H_{U(\mathbf{s})}(A) = \int_{S^*} p^n(A)\,\mu(\mathbf{s}, dp),$$

where $\mu(\mathbf{s}, \cdot)$ is a unique signed measure with finite support on S^*, the space of probability measures on $(S, \mathcal{B})$, and $p^n(A)$ is the probability of getting an outcome in A when doing an n-length i.i.d. experiment based on probability measure $p \in S^*$. We may think of μ as a signed transition kernel. Thus, we may write

$$\begin{aligned}\mathbb{P}(A) &= \int_{S^n} H_{U(\mathbf{s})}(A)\,\mathbb{P}(d\mathbf{s})\\ &= \int_{S^n}\left(\int_{S^*} p^n(A)\,\mu(\mathbf{s}, dp)\right)\mathbb{P}(d\mathbf{s})\\ &= \int_{S^*} p^n(A)\,\nu(dp),\end{aligned}$$

where ν is the signed measure on S^* defined by

$$\nu(B) = \int_{S^n} \mu(\mathbf{s}, B)\,\mathbb{P}(d\mathbf{s}), \qquad B \in \mathcal{B}^*.$$

Of course, $\nu(S^*) = \int_{S^n} \mu(\mathbf{s}, S^*)\,\mathbb{P}(d\mathbf{s}) = \int_{S^n} \mathbb{P}(d\mathbf{s}) = 1$.

There are two remaining details to check. First, μ should have the right measurability properties, and second, ν should be well defined. It is clear on the one hand that for each $\mathbf{s}$ the function $\mu(\mathbf{s}, \cdot)$ is a measure on S^*. On the other hand, we next show that for each $B \in \mathcal{B}^*$, the function $\mu(\cdot, B)$ is a measurable function of $\mathbf{s}$. To see why this is the case it is useful to examine the explicit form of the signed measure μ guaranteed by Corollary 2.1. It is defined by the formula

$$\mu(\mathbf{s}, B) = \sum_{k=1}^{\binom{2n-1}{n}} w_{kj}\, 1_B(p_{\mathbf{s}}^{(k)}),$$

where the (possibly negative) weights w_{kj} are from a jth column of the matrix $\mathbf{W}^{-1}$ and do not depend on $\boldsymbol{s}$, and the measures $p_{\boldsymbol{s}}^{(k)}$ are elements of S^* defined by

$$p_{\boldsymbol{s}}^{(k)}(A) = \sum_{i=0}^{n-1} c_i^{(k)}\, 1_A(s_i) \quad \text{for } A \in \mathcal{B}.$$

The numbers $c_i^{(k)}$ are nonnegative and sum to one.

Since $\mu(\cdot, B)$ is a linear combination of indicator functions, it suffices to show that for fixed k the function $1_B(p_{\boldsymbol{s}}^{(k)})$ is a measurable function of $\boldsymbol{s}$, or in other words, we just need to verify that the set $\{\boldsymbol{s} : p_{\boldsymbol{s}}^{(k)} \in B\}$ is $\mathcal{B}^n$ measurable. Further, we remember that we endowed S^* with the weak$^*\sigma$-algebra $\mathcal{B}^*$, generated by the class of sets $\{p : p(A) < t\}$, as A ranges over $\mathcal{B}$ and t over [0, 1]. Thus, it is only necessary to confirm the measurability for such "nice" sets $B \in \mathcal{B}^*$. After these simplifications we find

$$\{\boldsymbol{s} : p_{\boldsymbol{s}}^{(k)} \in B\} = \left\{\boldsymbol{s} : \sum c_i^{(k)}\, 1_A(s_i) < t\right\}$$

and this set is $\mathcal{B}^n$ measurable because the function $g(\boldsymbol{s}) = \sum c_i 1_A(s_i)$ is a measurable function of $\boldsymbol{s}$ when $A \in \mathcal{B}$.

Keeping in mind that all real-valued countably additive set functions are automatically of bounded variation, we finish the proof by verifying that μ is bounded. First fix $\boldsymbol{s}$, let $\mathbf{W}^{-1} = (w_{ij})$ be as in Corollary 2.1 and notice

$$\|\mu(\boldsymbol{s}, \cdot)\|_{var} = \|\sum w^{(k)}\, \delta_{p^{(k)}}\|_{var} \leqslant \sum_{i,j} |w_{ij}|,$$

and this last quantity is a finite constant depending only on n. □

4. AN APPLICATION TO BAYESIAN CONSISTENCY

We turn to Bayesian theory, where de Finetti's Theorem is commonly applied, for example in Predictive Inference. The standard setup involves some unknown (random) parameter θ, conditional on which a sequence of random variables $X_1, X_2, \ldots,$ is distributed according to some family $\{f(\cdot|\theta) : \theta \in \Theta\}$ of p.d.f.'s indexed by θ, called the Likelihood family. In some situations θ may be interpreted as a strong law limit of a sequence of observations. The Bayesian has subjective beliefs about θ, represented by a Prior probability distribution $\pi(\theta)$ on Θ, the parameter space. The goal is to use the information contained in observations $X_1, X_2, \ldots, X_n$ to sequentially update the Prior distribution $\pi(\theta)$ to a Posterior distribution $\pi(\theta|\mathbf{x})$ via Bayes' Rule $\pi(\theta|\mathbf{x}) \propto f(\mathbf{x}|\theta)\,\pi(\theta)$.

One hopes that with more and more information, one would become more confident about the location of θ, which would be reflected in the Posterior $\pi(\theta|\mathbf{x})$ by a concentration as $n \to \infty$ to a degenerate distribution centered at the true value of θ. It is a celebrated fact that under some regularity conditions such convergence does take place (see Refs. 6, 31, and 36). The usual method of proof is to suppose that the Likelihood may be written as a product of identical factors: $f(\mathbf{x}|\theta) = \prod_{i=1}^{n} f(x_i|\theta)$. Then (depending on the particular proof) one takes logarithms, normalizes accordingly, and appeals to the Strong Law of Large Numbers to show that the Posterior does indeed concentrate as $n \to \infty$ to the true value of θ.

In this context the problem arises when one goes to use such a statement in practice; one is always confined by Nature to finite samples. As we have seen in Section 1, under finite exchangeability the Likelihood certainly may *not* be written as a product of identical factors without first supposing that the sequence could be imbedded in an infinite exchangeable sequence and then applying de Finetti's theorem, that is, unless one asserts that the sequence could in principle continue indefinitely. In particular, this immediately restricts the Bayesian to observations which are nonnegatively correlated, and even then it is not guaranteed – in theory or practice – that a given sampling process could continue without end.

The theorem presented in this chapter shows that, with the introduction of an intermediate mixing generalized random variable β one may consider the sequence conditionally i.i.d. given β and θ without concern for the correlation or the conceptual difficulties associated with sampling from a "potentially infinite" exchangeable sequence.

A pointed criticism of this method would be that the resulting Likelihood $f(\mathbf{x}|\theta)$ is decomposed into summands some of which (possibly) take negative values, and this in turn may seem unpleasant or unnatural. In fact, it seems as though we have traded one conceptual difficulty for another! However, it should be realized that the method is merely a mathematical technique meant to justify and explain what is strongly desired and commonly observed in practice: namely, the convergence of the sequence of Posteriors to a distribution degenerate at the true value of the parameter.

We go to demonstrate this fact, and for simplicity we prove it for the case when β and θ take at most a countable number of values by modifying an argument in Ref. 7, a more general setting being much more technically cumbersome (see Ref. 6). Here we suppose that the true parameter θ_{i^*} is distinguishable from the other values in the sense that the logarithmic divergences $\int |f(x|\beta_k, \theta_{i^*})| \log |f(x|\beta_k, \theta_{i^*})/f(x|\beta_l, \theta_i)|\, dx$ are strictly positive for any $k \neq l$ or $i \neq i^*$.

Proposition 4.1. Let $\mathbf{x}=(x_1, x_2, \dots, x_n)$ be exchangeable observations from a Likelihood family $\{f(\cdot|\theta):\theta\in\Theta\}$, where $\Theta=\{\theta_1, \theta_2, \dots,\}$ is countable. Suppose that θ_{i^*} is the true value of θ and the following regularity conditions hold:

(1) The Likelihood $f(\cdot|\beta_k, \theta_i)>0$ w.p.1, for all $k, i\geqslant 1$.
(2) The Prior satisfies $\pi(\theta_{i^*})>0$ and there exists k^* such that $g(\beta_{k^*}|\theta_{i^*})\neq 0$, where g denotes the mass function of β.
(3) The joint mass function $g(\beta, \theta)$ is of bounded variation: $\sum_{k,i}|g(\beta_k, \theta_i)|<\infty$.
(4) For any $k\neq l$ or $i\neq i^*$,

$$\int f(x|\beta_k, \theta_{i^*})\log\left[\frac{f(x|\beta_k, \theta_{i^*})}{f(x|\beta_l, \theta_i)}\right]dx>0.$$

Then

$$\lim_{n\to\infty}\pi(\theta_{i^*}|\mathbf{x})=1, \quad \lim_{n\to\infty}\pi(\theta_i|\mathbf{x})=0, \ i\neq i^*.$$

Proof. By Bayes' rule,

$$\pi(\theta_i|\mathbf{x})=\frac{f(\mathbf{x}|\theta_i)\pi(\theta_i)}{\sum_m f(\mathbf{x}|\theta_m)\pi(\theta_m)}.$$

But by Theorem 1.1 we may write

$$f(\mathbf{x}|\theta)=\sum_k\prod_{j=1}^n f(x_j|\beta_k, \theta)\cdot g(\beta_k|\theta),$$

where $f(\cdot|\beta, \theta)\geqslant 0$ and $g(\beta|\theta)$ is (possibly) signed. The above expression then becomes

$$\begin{aligned}\pi(\theta_i|\mathbf{x}) &= \frac{\sum_k\prod_{j=1}^n f(x_j|\beta_k, \theta_i)\,g(\beta_k|\theta_i)\pi(\theta_i)}{\sum_m\sum_l\prod_{j=1}^n f(x_j|\beta_l, \theta_m)\,g(\beta_l|\theta_m)\pi(\theta_m)}\\ &=\sum_k\frac{\prod_{j=1}^n f(x_j|\beta_k, \theta_i)\,g(\beta_k, \theta_i)}{\sum_{m,l}\prod_{j=1}^n f(x_j|\beta_l, \theta_m)\,g(\beta_l, \theta_m)}.\end{aligned}$$

By condition 2, there exists k^* such that $g(\beta_{k^*}, \theta_{i^*})\neq 0$; consequently after using condition 1 and dividing numerator and denominator by $\prod_{j=1}^n f(x_j|\beta_{k^*}, \theta_{i^*})$ there is obtained

$$=\sum_k\frac{\exp\{S_{k,i}\}g(\beta_k, \theta_i)}{\sum_{m,l}\exp\{S_{l,m}\}g(\beta_l, \theta_m)},$$

where

$$S_{k,i} = \log\left(\prod_{j=1}^{n} \frac{f(x_j|\beta_k,\theta_i)}{f(x_j|\beta k^*,\theta_{i^*})}\right) = \sum_{j=1}^{n} \log \frac{f(x_j|\beta_k,\theta_i)}{f(x_j|\beta_{k^*},\theta_{i^*})}.$$

Now, conditional on $(\beta_{k^*},\theta_{i^*})$, $S_{k,i}$ is the sum of n i.i.d. random variables and therefore

$$\lim_{n\to\infty} \frac{1}{n} S_{k,i} = \int f(x|\beta_{k^*},\theta_{i^*}) \log\left[\frac{f(x|\beta_k,\theta_i)}{f(x|\beta_{k^*},\theta_{i^*})}\right] dx,$$

by the Strong Law of Large Numbers. Condition 4 implies that the right-hand side is negative for $k \neq k^*$ or $i \neq i^*$, and it of course equals zero for $k = k^*, i = i^*$; therefore as $n \to \infty$, $S_{k,i} \to -\infty$ for $k \neq k^*$ or $i \neq i^*$ and $S_{k^*,i^*} \to 0$. Proposition 4.1 now follows from Condition 3 and the Dominated Convergence Theorem. □

5. AN APPLICATION TO STATISTICAL PHYSICS

Following Johnson et al.,[28] Bach,[3] Constantini and Garibaldi,[8] consider a physical system comprising a number n of particles of some kind, for example electrons, protons or photons. Suppose that there are d states (energy levels) in which each particle can be. If X_i represents the state of particle $i, i = 1, \ldots, n$, then the overall state of the system is $(X_1, \ldots, X_n)$, and equilibrium is defined as the overall state with the highest probability of occurrence.

If all d^n arrangements are equally likely the system is said to behave according to Maxwell–Boltzmann statistics (MB), where "statistics" is used here in a sense meaningful to physicists. Assumptions on the system that lead to such behavior are as follows:

(1) The particles are identical in terms of physical properties but distinguishable in terms of position. This is equivalent to the statement that the particle size is small compared with the average distance between particles.
(2) There is no theoretical limit on the fraction of the total number of particles in a given energy state, but the density of particles is sufficiently low and the temperature sufficiently high that no more than one particle is likely to be in a given state at the same time.

However, modern experimentation (particularly at low temperatures) has yielded two more plausible sets of hypotheses concerning physical systems; these result in Bose–Einstein statistics (BE) and Fermi–Dirac statistics (FD).

For both, one supposes that the particles are indistinguishable (thereby granting exchangeability). The BE statistics are obtained by retaining the second assumption that there is no limit on the number of particles that may occupy a particular energy state. Particles that are observed to obey BE statistics are called bosons and include photons, alpha particles, and deuterons.

For FD statistics one stipulates instead that only one particle can occupy a particular state at a given time (a condition well known as the Pauli exclusion principle). Particles that obey the principle are called fermions and include protons, neutrons, and electrons. All known elementary particles fall into one of the two above categories.

Much work has been done studying these different models and their consequences for the interpretation of the physical concepts. Constantini *et al.*[9] proposed a new set of ground hypotheses for deriving the three models. Bach *et al.*,[4] using the argument that exchangeable random variables are appropriate for describing indistinguishable particles, used a multivariate de Finetti's theorem to derive BE statistics. Bach[2] explored the quantum properties of indistinguishable particles, while Bach,[3] Constantini and Garibaldi[8] attempted to base the derivation of the different statistics on the correlation structure and an introduced relevance quotient.

This section, in the spirit of Ref. 4, shows that the FD of indistinguishable (exchangeable) particles can be derived from the statistics of classical (*i.e.* independent) particles by means of Theorem 1.1.

More precisely, we are concerned with the statistical problem of distributing n particles into d cells. We introduce a probability space $(\Omega, \mathcal{B}, \mu)$ and random variables $X_i\colon \Omega \to \{1, \dots, d\}$, where the event $\{X_i = j\}$ represents the outcome that particle i is in cell j. As remarked above, there are d^n different configurations, which are characterized by the events $\{\mathbf{X} = \mathbf{j}\}$, where $\mathbf{X} = (X_1, \dots, X_n)$ and $\mathbf{j} \in \{1, \dots, d\}^n$.

For a given configuration $\{\mathbf{X} = \mathbf{j}\}$, we define the occupation numbers $n_k,\ k = 1, \dots, d$, by

$$n_k(\mathbf{j}) = \sum_{i=1}^{n} \delta_{k, j_i}.$$

For the particles of MB statistics we obtain

$$\mathbb{P}_{\mathrm{MB}}(\mathbf{X} = \mathbf{j}) = d^{-n} \quad \mathbf{j} \in \{0, \dots, n\}^d.$$

However, under BE statistics we have

$$\mathbb{P}_{\mathrm{BE}}(\mathbf{X}=\mathbf{j})=\binom{n}{n_1(\mathbf{j})\cdots n_d(\mathbf{j})}^{-1}\binom{n+d-1}{n}^{-1} \quad \mathbf{j}\in\{0,\ldots,n\}^d.$$

For FD statistics, according to the Pauli exclusion principle we must assume $n \leqslant d$. Also, whenever there exists an occupation number n_k larger than 1 we must set the corresponding probability zero, with the remaining configurations being equally likely:

$$\mathbb{P}_{\mathrm{FD}}(\mathbf{X}=\mathbf{j})=\begin{cases} 0, & \text{if } \mathbf{n}(\mathbf{j})\notin\{0,1\}^d,\\ \binom{d}{n}^{-1}(n!)^{-1}, & \text{if } \mathbf{n}(\mathbf{j})\in\{0,1\}^d.\end{cases}$$

It is immediately clear that $\mathbb{P}_{\mathrm{MB}}, \mathbb{P}_{\mathrm{BE}}$, and $\mathbb{P}_{\mathrm{FD}}$ are exchangeable. It was shown in Ref. 4 that $\mathbb{P}_{\mathrm{BE}}$ is a nonnegative Dirichlet mixture of multinomial distributions, i.e., statistics corresponding to classical, independent particles. And in Ref. 3, the correlation structures for the models were found to be

$$\mathrm{Corr}_{\mathrm{BE/FD}}(X_i, X_j)=\pm(d\pm 1)^{-1},$$

where of course $\mathrm{Corr}_{\mathrm{MB}}(X_i, X_j)=0$. The negative correlation in FD statistics (due to the Pauli exclusion principle) shows that a classical mixture representation will not hold, and also suggests why no such derivation of them using classical particles has been done until now. However, by exchangeability we may immediately apply Theorem 1.1 to see that $\mathbb{P}_{\mathrm{FD}}$ is indeed a (necessarily signed) mixture, more precisely a linear combination, of multinomial distributions.

Two remarks are in order. First, it is unnecessary to repeat all of the hard work present in the above papers to assert the mixture representation; it is a natural corollary of our current study. Secondly, an obvious (and perhaps tempting) question would concern the physical *interpretation* of the mixing generalized random variable and its signed distribution. Let it be clear that no such explanation is made or implied here. In all of the above mentioned papers there was no attempt to understand the physical significance of the mixture; throughout the goal was to find some way to mathematically base the modern, quantum particles on the more familiar classical particles of Maxwell–Boltzmann.

It seems that in the 1980s an asymmetry was created in the literature by Bach and his colleagues since such a basis was possible only for BE, while the status of FD remained unresolved. It is hoped that the theorem of this paper and the representation of this section may help to restore the symmetry to this long standing problem.

REFERENCES

1. Aldous, D. J. (1985). *Exchangeability and Related Topics*, Lecture Notes in Mathematics, Vol. 117, Springer, New York.
2. Bach, A. (1985). On the quantum properties of indistinguishable classical particles. *Lett. Nuovo Cimento*. **43**, 383–387.
3. Bach, A. (1988). The concept of indistinguishable particles in classical and quantum physics. *Found. Phy*. **18**, 639–649.
4. Bach, A., Blank, H., and Francke, H. (1985). Bose-Einstein statistics derived from the statistics of classical particles. *Lett. Nuovo Cimento*. **43**, 195–198.
5. Bartlett, M. S. (1945) Negative probability. *Proc. Cambridge Philos. Soc.* **41**, 71–73.
6. Berk, R. (1970). Consistency a posteriori. *Ann. Math. Stat.* **41**, 894–906.
7. Bernardo, J. M., and Smith, A. F. M. (1993). *Bayesian Theory*, Wiley, New York.
8. Constantini, D., and Garibaldi, U. (1989). Classical and quantum statistics as finite random processes. *Found. Phy*. **19**, 743–754.
9. Constantini, D., Galavotti, M., and Rosa, R. (1983). A set of ground hypotheses for elementary-particle statistics. *Il Nuovo Cimento* **74**, 151–158.
10. Dawid, A. P (2004). Probability, causality and the empirical world: a Bayes – de Finetti – Popper – Borel synthesis. *Stat. Sci.* **19**, 44–57.
11. De Moivre, A. (1718). *The Doctrine of Chances*, London.
12. De Finetti, B. (1931). Funzione caratteristica di un fenomeno allatorio. *Atti della R. Accademia Nazionale dei Lincii Ser. 6, Memorie, Classe di Scienze, Fisiche, Matematiche e Naturali*. **4**, 251–299.
13. De Finetti, B. (1937). La prévision: ses logiques, ses sources subjectives. *Ann. l'Inst. Henri Poincaré* **7**, 1–68.
14. De Finetti, B. (1975). *Theory of Probability*, Vol. 2, Wiley, New York.
15. Dellacherie, C., and Meyer, P. A. (1982). *Probabilities and Potential B.*, North-Holland, New York. pp. 46–47.
16. Diaconis, P. (1977). Finite forms of de finetti's theorem on exchangeability. *Synthese* **36**, 271–281.
17. Diaconis, P., and Freedman, D. (1980). *Finite Exchangeable Sequences*. *Ann. Probab.* **8**, 745–764.
18. Diaconis, P., and Freedman, D. (1987). *A dozen de Finitti style results in search of a theory*. *Ann. l'Inst. Henri Poincaré* **23**, 397–423.
19. Diaconis, P. W., Eaton, M. L., and Lauritzen, S. L. (1992). *Finite de Finetti theorems in linear models and multivariate analysis. Scand. J. Stat*. **19**, 289–315.
20. Dubins, L. E., and Freedman, D. (1979). Exchangeable processes need not be distributed mixtures of independent, identically distributed random variables. *Z. Wahrsch. Verw. Gebiete* **48**, 115–132.
21. Dynkin, E. B. (1953). Klassy ekvivalentnyh Slučaĭnyh veličin. *Uspehi Matematičeskih Nauk 8* **54**, 125–134.
22. Feynman, R.P. (1987). Negative probability. *Quantum Implications*, Routledge and Kegan Paul, London, 235–248.
23. Freedman, D. (1977). A remark on the difference between sampling with and without replacement. *J. Am. Stat. Assoc.* **72**, 681.
24. Gnedin, A. V. (1996). On a class of exchangeable sequences. *Stat. Probab. Lett.* **28**, 159–164.
25. Haag, J. (1928). Sur un problème général de probabilitès et ses diverses applications. *Proc. Int. Congr. Mathematicians*, Toronto, 1924, pp. 659–674.

26. Hewitt, E., and Savage, L. J. (1955). Symmetric measures on Cartesian products. *Trans. Am. Math. Soc.* **80**, 470–501.
27. Jaynes, E. (1986). Some applications and extensions of the de finetti representation theorem. *Bayesian Inference and Decision Techniques*, Elsevier, Amsterdam, pp. 31–42.
28. Johnson, K., Kotz, S., and Kemp, A. (1993). *Univariate Discrete Distributions*, 2nd ed., Wiley, New York.
29. Kendall, D.K. (1967). On finite and infinite sequences of exchangeable events. *Stud. Sci. Math. Hung.* **2**, 319–327.
30. Kingman, J. (1978). Uses of exchangeability. *Ann. Probab.* **6**, 183–197.
31. Le Cam, L. (1953). On some asymptotic properties of maximum likelihood estimates and related Bayes estimates. *Univ. Calif. Publ. Stat.* **1**, 277–300.
32. Mückenheim, W. (1986). A review of extended probabilities. *Phys. Rep.* **6**, 337–401.
33. Ressel, P. (1985). de Finetti-type theorems: an analytical approach. *Ann. Probab.* **13**, 898–922.
34. Spizzichino, F. (2001). *Subjective Probability Models for Lifetimes*, Monographs on Statistics and Applied Probability, Vol. 91, Chapman and Hall, New York.
35. von Plato, J. (1991). Finite partial exchangeability. *Stat. Probab. Lett.* **11**, 99–102.
36. Wald, A. (1949). Note on the consistency of the maximum likelihood estimate. *Ann. Math. Stat.* **20**, 595–601.