Research Report No. 1856, November 1995
Institute of Information Theory and Automation

FINITE PARTIALLY EXCHANGEABLE ARRAYS[1]

F. Matúš[2]

Institute of Information Theory and Automation, Prague

Abstract. Let $(\xi_{i_1,i_2};\ 1 \le i_1 \le n_1,\ 1 \le i_2 \le n_2)$ be an $n_1 \times n_2$ separately exchangeable array of random variables. The distribution of its $m_1 \times m_2$ subarray can be approximated by a mixture of arrays $(f(\alpha_{i_1}, \beta_{i_2});\ 1 \le i_1 \le m_1,\ 1 \le i_2 \le m_2)$ where α_{i_1} and β_{i_2} are mutually independent and identically distributed. In the variation distance our approximation is at most $m_1(m_1-1)/n_1 + m_2(m_2-1)/n_2$ far from the subarray. More-dimensional variants of this result are presented for both jointly and separately exchangeable finite arrays. The bounds on approximations are shown to be sharp. As a consequence, infinite de Finetti theorems for the partially exchangeable arrays with a compact Hausdorff state space are then obtained.[3]

1. Introduction. Results of de Finetti style usually concern infinite random sequences, arrays and other random elements subject to symmetry conditions of exchangeability type. They provide a mixture characterization of all distributions having the chosen symmetry through probability measures on extreme symmetrical distributions. Starting with the classical de Finetti's theorem on exchangeable two-valued sequences, the field of exchangeability has been reviewed regularly, see Hewitt and Savage [11], Koch and Spizzichino [14], Aldous [2], Diaconis [4], Lauritzen [15], Diaconis, Eaton and Lauritzen [7], etc.

One particular line of research in exchangeability, initiated by Diaconis and Freedman [5], focuses on finite sequences with prescribed symmetries and on mixture approximations of their marginal distributions. Results in this direction, e.g. Diaconis and Freedman [6], Zaman [23], Diaconis, Eaton and Lauritzen [7], not only imply easily the corresponding infinite versions of de Finetti theorems but seem to be interesting on its own right.

[1]This research was supported by Grant AV ČR 275105 and Grant EC CIPA 3511CT930053.
[2]*Postal address:* Institute of Information Theory and Automation, Academy of Sciences of the Czech Repubic, Pod vodrenskou v 4, 182 08 Prague, Czech Republic. (matus@utia.cas.cz)
[3]*Key words and phrases:* exchangeable, symmetric, de Finetti's theorem, Hewitt and Savage's theorem, variation distance, random array, Radon measure, Bauer simplex, birthday problem.

AMS 1980 subject classifications. Primary 60G10, Secondary 60J07

The aim of this paper is to study finite separately and jointly exchangeable arrays from this point of view; for references on infinite arrays of this sort see e.g. Aldous [1], [2], Koch and Spizzichino [14], Kallenberg [12], [13] and Matúš [17]. We prove finite de Finetti theorems giving approximations of subarrays of these arrays in terms of mixtures of special dissociated arrays in Section 2 and Section 3. Our bound on the variation distance between a subarray and its approximating mixture is found to be tight in Section 5. Infinite versions of de Finetti theorems for the arrays in a very general topological setting (Radon probability measures on arbitrary compact Hausdorff spaces) are then obtained by compactness and uniqueness arguments. Since we work with d-dimensional arrays, $d \geq 1$, these versions comprehend the generalization of the classical de Finetti's result from Diaconis and Freedman [5], cf. also Ressel [19].

2. Subarrays of finite separately exchangeable arrays. Let d be a positive integer number, $N = (N_1, \ldots, N_d)$ a d-tuple of nonempty sets and $(X, \mathcal{X})$ a measurable space. By the symbol $(X_N, \mathcal{X}_N)$ we denote the Cartesian product of $N_1 \times \ldots \times N_d$ copies of the space $(X, \mathcal{X})$. Typically, the elements of X_N will be $x_N = (x_N^{i_1,\ldots,i_d}; i_1 \in N_1, \ldots, i_d \in N_d)$ each $x_N^{i_1,\ldots,i_d}$ belonging to X. The set $\mathcal{P}_N$ of all probability measures P_N on $(X_N, \mathcal{X}_N)$, when endowed with the smallest σ-algebra making all mappings $P_N \to P_N(A_N)$, $A_N \in \mathcal{X}_N$, measurable, is taken as a measurable space, too.

Where $M = (M_1, \ldots, M_d)$ and $N = (N_1, \ldots, N_d)$ are two d-tuples of nonempty sets and $\varphi = (\varphi_1, \ldots, \varphi_d)$ is a d-tuple of mappings, φ_c defined on M_c and ranging in N_c, $1 \leq c \leq d$, we fix the notation $\widetilde{\varphi}$ for the mapping on X_N into X_M given by

$$\widetilde{\varphi}\, x_N = (x_N^{\varphi_1 j_1,\ldots,\varphi_d j_d}; j_1 \in M_1, \ldots, j_d \in M_d), \quad x_N \in X_N.$$

Observing that $\widetilde{\varphi}$ is measurable, let the image of a measure $P_N \in \mathcal{P}_N$ by $\widetilde{\varphi}$ be $\approx\!\!\!\!\!\varphi\, P_N$, i.e. $\overset{\approx}{\varphi} P_N(A_M) = P_N(\widetilde{\varphi}^{\,-1} A_M)$, $A_M \in \mathcal{X}_M$. The mapping $\overset{\approx}{\varphi}$ on $\mathcal{P}_N$ with values in $\mathcal{P}_M$ is easily recognized to be measurable, too. In the special case when $M_c \subset N_c$ and π_{M_c,N_c} is the natural embedding of M_c into N_c, $1 \leq c \leq d$, the d-tuple of mappings $\pi_{M,N} = (\pi_{M_1,N_1}, \ldots, \pi_{M_d,N_d})$ gives rise to the canonical projection $\widetilde{\pi}_{M,N}$ of X_N on X_M and $\overset{\approx}{\pi}_{M,N} P_N$, for the sake of convenience shortened to P_N^M, is a marginal measure of P_N.

With N as above and $\tau = (\tau_1, \ldots, \tau_d)$ a d-tuple of permutations, τ_c acting on N_c and moving only finite number of its elements, $1 \leq c \leq d$, a probability measure P_N is said to be τ-invariant if $\overset{\approx}{\tau} P_N = P_N$. Let G_N be the group of all d-tuples of these permutations. If $P_N \in \mathcal{P}_N$ is τ-invariant for any $\tau \in G_N$ then it is called *separately exchangeable*. The convex set of all separately exchangeable probabilities on $(X_N, \mathcal{X}_N)$ is denoted by $\mathcal{P}_N^{s.e.}$.

Let now d-tuples $N = (N_1, \ldots, N_d)$ and $M = (M_1, \ldots, M_d)$ consist exclusively from finite nonempty sets. For fixed $x_N \in X_N$ the formula

$$Q_{x_N}^M(A_M) = \frac{|\{\varphi = (\varphi_1, \ldots, \varphi_d) \in \Pi_{c=1}^d N_c^{M_c};\ \widetilde{\varphi}\, x_N \in A_M\}|}{|\Pi_{c=1}^d N_c^{M_c}|}, \quad A_M \in \mathcal{X}_M,$$

defines a probability measure $Q_{x_N}^M$ on $(X_M, \mathcal{X}_M)$. To have another insight, let us consider random variables $\eta_{j_c}^{(c)}$, $j_c \in M_c$, $1 \le c \le d$, that are mutually independent and let $\eta_{j_c}^{(c)}$ be uniformly distributed on the set N_c. Then the distribution of the array

$$(x_N(\eta_{j_1}^{(1)}, \ldots, \eta_{j_d}^{(d)});\ j_1 \in M_1, \ldots, j_d \in M_d),$$

where we sink with the superindices of x_N for visual plausibility, is exactly $Q_{x_N}^M$. In fact, thinking about the collection $\eta^{(c)} = (\eta_{j_c}^{(c)};\ j_c \in M_c)$ as if it were a random mapping from M_c into N_c, $1 \le c \le d$, the d-tuple $\eta = (\eta^{(1)}, \ldots, \eta^{(d)})$ of random mappings provides the above array $\widetilde{\eta}\, x_N$. We see immediately that $Q_{x_N}^M \in \mathcal{P}_M^{s.e.}$ and that the mapping $x_N \to Q_{x_N}^M$ is measurable, or equivalently, all mappings $x_N \to Q_{x_N}^M(A_M)$, $A_M \in \mathcal{X}_M$, are measurable.

Theorem $1^{s.e.}$. *Let sets $M_c \subset N_c$ have finite cardinalities, $1 \le m_c \le n_c$, $1 \le c \le d$. If P_N for $N = (N_1, \ldots, N_d)$ is any separately exchangeable probability measure on $(X_N, \mathcal{X}_N)$ then its marginal measure P_N^M for $M = (M_1, \ldots, M_d)$ can be approximated by the integral expression as follows.*

$$\sup_{A_M \in \mathcal{X}_M} \left| P_N^M(A_M) - \int_{X_N} Q_{x_N}^M(A_M)\, P_N(dx_N) \right| \le 1 - \prod_{c=1}^d \frac{n_c!}{n_c^{m_c}(n_c - m_c)!}$$

Proof. Let $A_M \in \mathcal{X}_M$ and $B_N = \widetilde{\pi}_{M,N}^{-1} A_M$. The two indices at $\pi_{M,N}$ will be omitted below. Since $P_N \in \mathcal{P}_N^{s.e.}$ one can write

$$P_N^M(A_M) = \widetilde{\pi}\, P_N(A_M) = P_N(B_N) = |G_N|^{-1} \sum_{\tau \in G_N} P_N(\widetilde{\tau}^{-1} B_N)\ .$$

Employing the notation

$$f_{A_M}(x_N) = |G_N|^{-1} \cdot |\{\tau \in G_N\, ;\, \widetilde{\pi}\widetilde{\tau} x_N \in A_M\}|, \quad x_N \in X_N,$$

we obtain

$$P_N^M(A_M) = |G_N|^{-1} \int_{X_N} \sum_{\tau \in G_N} \mathbf{1}_{B_N}(\widetilde{\tau} x_N)\, P_N(dx_N) = \int_{X_N} f_{A_M}(x_N)\, P_N(dx_N)$$

where $\mathbf{1}_{B_N}$ is the indicator function of the measurable set $B_N \in \mathcal{X}_N$. Hence, the function f_{A_M} is measurable.

Similarly, the expression

$$\int_{X_N} Q^M_{x_N}(A_M)\, P_N(dx_N) = \int_{X_N} g_{A_M}(x_N)\, P_N(dx_N)$$

is rewritten using

$$g_{A_M}(x_N) = |N^M|^{-1} \cdot |\{\varphi \in N^M;\ \widetilde{\varphi}\, x_N \in A_M\}|, \quad x_N \in X_N\ .$$

By N^M we denoted here the set of all d-tuples $(\varphi_1, \ldots, \varphi_d)$ of mappings, $\varphi_c \in N_c^{M_c}$, $1 \leq c \leq d$. The subset of all d-tuples consisting exclusively from injections will be denoted by N^{*M}.

To compare f_{A_M} and g_{A_M} we need two counting arguments. First, let us consider the binary relation $\diamond$ between the group G_N and the orbit $\widetilde{G}_N x_N = \{\widetilde{\tau} x_N; \tau \in G_N\}$ of some $x_N \in X_N$

$$\tau \diamond z_N \quad \Leftrightarrow \quad \widetilde{\pi}\widetilde{\tau} z_N = y_M \quad , \quad \tau \in G_N\, , \quad z_N \in \widetilde{G}_N x_N\, ;$$

the elements $x_N \in X_N$ and $y_M \in X_M$ are to be taken as fixed. Since the number of z_N incident to a given τ is not dependent on τ, and vice versa, counting the number of all incident pairs we get

$$|G_N| \cdot |\{z_N \in \widetilde{G}_N x_N;\ \widetilde{\pi} z_N = y_M\}| = |\widetilde{G}_N x_N| \cdot |\{\tau \in G_N;\ \widetilde{\pi}\widetilde{\tau} x_N = y_M\}|\ .$$

Second, the binary relation

$$z_N \circ \varphi \quad \Leftrightarrow \quad \widetilde{\varphi}\, z_N = y_M \quad , \quad z_N \in \widetilde{G}_N x_N\, , \quad \varphi \in N^{*M}\, ,$$

has

$$|\widetilde{G}_N x_N| \cdot |\{\varphi \in N^{*M};\ \widetilde{\varphi}\, x_N = y_M\}| = |N^{*M}| \cdot |\{z_N \in \widetilde{G}_N x_N;\ \widetilde{\pi} z_N = y_M\}|$$

couples of incidence arguing as before. A comparision provides

$$f_{A_M}(x_N) = |N^{*M}|^{-1} \cdot |\{\varphi \in N^{*M};\ \widetilde{\varphi}\, x_N \in A_M\}|\ .$$

However, the difference $g_{A_M}(x_N) - f_{A_M}(x_N)$ is equal to

$$|N^M|^{-1} \cdot \Big|\{\varphi \in N^M - N^{*M};\ \widetilde{\varphi}\, x_N \in A_M\}\Big| \ -\ f_{A_M}(x_N)\left(1 - |N^{*M}| \cdot |N^M|^{-1}\right)$$

and thus

$$-\, f_{A_M}\left(1 - |N^{*M}| \cdot |N^M|^{-1}\right) \ \leq\ g_{A_M} - f_{A_M} \ \leq\ (1 - f_{A_M})\left(1 - |N^{*M}| \cdot |N^M|^{-1}\right)\ .$$

The desired supremum can be now estimated from above by

$$\sup_{A_M \in \mathcal{X}_M} \int_{X_N} |f_{A_M} - g_{A_M}|\, dP_N \;\leq\; 1 - |N^{*M}| \cdot |N^M|^{-1}$$

since $0 \leq f_{A_M} \leq 1$. ∎

3. Subarrays of finite jointly exchangeable arrays. In this section we employ formally the same notation as in the previous one but we change its meaning substantially. The letter N will denote here any nonempty set and $(X_N, \mathcal{X}_N)$ the Cartesian product of N^d copies of a space $(X, \mathcal{X})$. If φ is a mapping from a nonempty set M into N then $\widetilde{\varphi}$ maps X_N into X_M according to

$$\widetilde{\varphi}\, x_N = (x_N^{\varphi j_1, \ldots, \varphi j_d};\, j_1 \in M, \ldots, j_d \in M), \quad x_N \in X_N,$$

and $\widetilde{\widetilde{\varphi}}$ the probability measures P_N on $(X_N, \mathcal{X}_N)$, we write $P_N \in \mathcal{P}_N$, into $\mathcal{P}_M$. In the special case when $M \subset N$ and $\pi_{M,N}$ is the natural embedding of M into N, again, $\widetilde{\pi}_{M,N}$ is a canonical projection and $\widetilde{\widetilde{\pi}}_{M,N} P_N = P_N^M$ a marginal measure.

With N as above let G_N denote the set of all permutations τ on N that move only finite number of its elements. If $P_N \in \mathcal{P}_N$ and $\widetilde{\widetilde{\tau}} P_N = P_N$ for all all $\tau \in G_N$ then the probability measure P_N is called *jointly exchangeable* and we reserve the symbol $\mathcal{P}_N^{j.e.}$ to the convex set of all these measures.

Where N and M are finite nonempty sets and N^M the set of all mappings of M into N, every $x_N \in X_N$ defines the probability measure

$$Q_{x_N}^M(A_M) = \frac{|\{\varphi \in N^M;\, \widetilde{\varphi}\, x_N \in A_M\}|}{|N^M|}, \quad A_M \in \mathcal{X}_M,$$

which is nothing else as the distribution of the random array

$$\widetilde{\eta}\, x_N = (x_N(\eta_{j_1}, \ldots, \eta_{j_d});\, j_1 \in M, \ldots, j_d \in M)$$

constructed from random variables $\eta = (\eta_j; j \in M)$ that are mutually independent and uniformly distributed on N. Clearly, $Q_{x_N}^M \in \mathcal{P}_M^{j.e.}$ and the mapping $x_N \to Q_{x_N}^M$ is measurable.

Theorem $1^{j.e.}$. *Let nonempty sets $M \subset N$ have finite cardinalities $m \leq n$. Then the marginal measure P_N^M of every jointly exchangeable probability measure P_N on $(X_N, \mathcal{X}_N)$ can be approximated by*

$$\sup_{A_M \in \mathcal{X}_M} \left| P_N^M(A_M) - \int_{X_N} Q_{x_N}^M(A_M) P_N(dx_N) \right| \;\leq\; 1 - \frac{n!}{n^m (n-m)!} \;.$$

The reader is now invited to read literaly the previous proof (with two exceptions: $\mathcal{P}_N^{j.e.}$ replaces $\mathcal{P}_N^{s.e.}$ and N^{*M} denotes all injections from M into N) and to imagine under that notation the current meaning. Let us point out that the bound of this theorem is not dependent on d at all.

4. Remarks. The assertions of Theorem $1^{s.e.}$ and Theorem $1^{j.e.}$ can be rephrased in the language of mixtures of probability measures. We want to comment both situations simultaneously so that the notation will have double meaning up to $\mathcal{P}_N^{s.e.}$ and $\mathcal{P}_N^{j.e.}$ which are written as $\mathcal{P}_N^{*.e.}$.

Having a probability measure μ on $\mathcal{P}_N$ the prescription

$$A_N \to \int_{\mathcal{P}_N} R_N(A_N)\,\mu(dR_N)\,, \quad A_N \in \mathcal{X}_N\,,$$

defines a probability measure on X_N denoted by $\int_{\mathcal{P}_N} R_N\,d\mu$ and called μ-mixture. If μ is concentrated on $\mathcal{P}_N^{*.e.}$ in the sense $\mu(\mathcal{R}_N) = 0$ for any measurable subset $\mathcal{R}_N$ of $\mathcal{P}_N - \mathcal{P}_N^{*.e.}$ then it is not difficult to see that the μ-mixture belongs to $\mathcal{P}_N^{*.e.}$.

Under the assumptions of Theorem $1^{*.e.}$ let μ be the image of a probability measure $P_N \in \mathcal{P}_N^{*.e.}$ under the measurable mapping $x_N \to Q_{x_N}^M$. The integral expression rewrites as

$$\int_{X_N} Q_{x_N}^M(A_M)\,P_N(dx_N) = \int_{\mathcal{P}_M} R_M(A_M)\,\mu(dR_M)\,, \quad A_M \in \mathcal{X}_M\,,$$

what means that Theorem $1^{*.e.}$ gives an upper bound on the total variation of the difference $P_N^M - \int_{\mathcal{P}_N} R_N\,d\mu$. The bound depends only on N and M, the sizes of arrays and subarrays, and not on individual P_N. This μ-mixture has one remarkable property. Namely, it is concentrated on the set $\mathcal{P}_N^{*.e.d.}$ of all $R_N \in \mathcal{P}_N^{*.e.}$ that are *dissociated* what amounts validity of the equations

$$R_N(\widetilde{\pi}_{K,N}^{-1}A_K \,\cap\, \widetilde{\pi}_{L,N}^{-1}B_L) = R_N^K(A_K)\cdot R_N^L(B_L)\,, \quad A_K \in \mathcal{X}_K,\ B_L \in \mathcal{X}_L\,,$$

for any K and L disjoint subsets of N (thinking in d-tuples this is to be understood coordinatewisely). In other words, the equations express independence of some subarrays. All measures $Q_{x_N}^M$ are dissociated as easily seen from the structure of $\widetilde{\eta}\,x_N$ and this entails the claimed concentration property.

Let us stop at the end on the case $d = 1$. Separate and joint exchangeabilities coincide with the usual exchangeability. The measure $Q_{x_N}^M$, $x_N \in X_N$, is now equal to the distribution of i.i.d. system of random variables $(x_N(\eta_j);\ j \in M)$. Each of these variables is concentrated on the finite set $\{x_N^i;\ i \in N\}$ with its distribution proportional to the numbers of repeatings of x_N^i. Thus, $Q_{x_N}^M$ is the product of identical measures $Q_{x_N}^{\{j\}}$, $j \in M$. Both approximation theorems estimate consonantly the variation distance between P_N^M, a marginal of an exchangeable measure P_N, and a mixture of product measures. The upper bounds are both the same as in Diaconis and Freedman [5], p. 749.

5. Sharpness of the approximation bounds. Let $N = (N_1, \ldots, N_d)$, $M = (M_1, \ldots, M_d)$ and $L = (L_1, \ldots, L_d)$ be three d-tuples of nonempty finite sets where $M_c \subset N_c$, $1 \le c \le d$, and $P_N \in \mathcal{P}_N^{s.e.}$. One might strive to look for a probability measure $S_L \in \mathcal{P}_L$ giving a better approximation for P_N^M than Theorem $1^{s.e.}$

$$\inf_{S_L \in \mathcal{P}_L} \quad \sup_{A_M \in X_M} \quad \left| P_N^M(A_M) - \int_{X_L} Q_{x_L}^M(A_M)\, S_L(dx_L) \right| \ .$$

The infimum is interesting if L is varying as well. The next proposition shows, however, that if the space $(X, \mathcal{X})$ is rich enough then for the "worst" $P_N \in \mathcal{P}_N^{s.e.}$ no improvement of the bound of Theorem $1^{s.e.}$ can be attained.

Proposition. *If the measurable space $(X, \mathcal{X})$ has at least $n_1 \ldots n_d$ nonempty disjoint measurable sets then the maximum below exists and*

$$\max_{P_N \in \mathcal{P}_N^{s.e.}} \inf_{L;\, S_L \in \mathcal{P}_L} \sup_{A_M \in X_M} \left| P_N^M(A_M) - \int_{X_L} Q_{x_L}^M(A_M)\, S_L(dx_L) \right| = 1 - \prod_{c=1}^{d} \frac{n_c!}{n_c^{m_c}(n_c - m_c)!}$$

This section is occupied mainly by a proof of this assertion. We need first a combinatorial inequality; n_c, m_c and l_c, $1 \le c \le d$, denote the cardinalities, respectively.

Lemma 1. *For every function $v : L_1 \times \ldots \times L_d \to N_1 \times \ldots \times N_d$ the set*

$$\{(\psi, \varphi) \in L^{*M} \times N^{*M};\ v(\psi_1 j_1, \ldots, \psi_d j_d) \ = \ (\varphi_1 j_1, \ldots, \varphi_d j_d) \ , j_1 \in M_1, \ldots, j_d \in M_d\}$$

has cardinality smaller than or equal to

$$\prod_{c=1}^{d} \frac{n_c!}{(n_c - m_c)!} \left(\frac{l_c}{n_c} \right)^{m_c} \ .$$

Proof. At the beginning we are going to prove an auxiliary inequality that implies the case $d = 1$ and will be used repeatedly thereafter.

Let N, M and L be for a while nonempty finite sets, $|N| = n \ge |M| = m$, and $\psi \colon M \to L$, $\varphi \colon M \to N$ be injections. For every collection of nonnegative numbers w_k^i, $i \in N$, $k \in L$ the inequality

$$\sum_{\psi} \sum_{\varphi} \prod_{j \in M} \ w_{\psi j}^{\varphi j} \ \le \ \frac{n!}{(n-m)!} \left(\frac{1}{n} \sum_{i \in N} \sum_{k \in L} \ w_k^i \right)^m$$

is satisfied. In fact, since for each $k \in L$ separately the left-hand side is an affine function of the variables w_k^i, $i \in N$, it can be estimated from above by

$$\sum_{\psi} \sum_{\varphi} \prod_{j \in M} \underline{w}_{\psi j}^{\varphi j}$$

where $\underline{w}_k^i = 0$ for $k \notin L^i$ and $\underline{w}_k^i = w_k^\bullet$ for $k \in L^i$. Here, $L = \sum_{i \in N} L^i$ is a partition of the set L into not necessarily nonempty blocks and $w_k^\bullet = \sum_{i \in N} w_k^i$. The product is nonzero only if $\psi j \in L^{\varphi j}$ for all $j \in M$ and thus the above expression can be rewritten to

$$\sum_{\varphi} \prod_{j \in M} \sum_{k \in L^{\varphi j}} w_k^\bullet \ .$$

This is $m!$ multiple of the value of the m-th order elementary symmetric polynomial in n arguments $\sum_{k \in L^i} w_k^\bullet$, $i \in N$. Substitution of the n identical arguments $\sum_{k \in L} w_k^\bullet / n$ does not decrease the value and leads to the desired right-hand side.

Now, it is straighforward that Lemma 1 is valid for $d = 1$, namely

$$|\{(\psi_1, \varphi_1); v(\psi_1 j_1) = (\varphi_1 j_1),\ j_1 \in M_1\}| = \sum_{\psi_1} \sum_{\varphi_1} \prod_{j_1 \in M_1} w_{\psi_1 j_1}^{\varphi_1 j_1}$$

where $w_{k_1}^{i_1} = 1$ if $v(k_1) = i_1$ and $w_{k_1}^{i_1} = 0$ otherwise, $i_1 \in N_1$, $k_1 \in L_1$. Let $d \geq 2$ and, by no loss of generality, $m_1 \leq m_2 \leq \ldots \leq m_d$. The desired number of pairs of d-tuples ψ and φ is not greater than

$$\sum_{\psi_1} \sum_{\varphi_1} \prod_{j_1 \in M_1} \left| \{(\psi_2, \ldots, \psi_d, \varphi_2, \ldots, \varphi_d); v(\psi_1 j_1, \ldots, \psi_d j_d) = (\varphi_1 j_1, \ldots, \varphi_d j_d), j_2 \in M_2, \ldots, j_d \in M_d\} \right|^{1/m_1}$$

Abbreviating $\sum_{\psi_c} \sum_{\varphi_c}$ followed by the geometric mean over $j_c \in M_c$ as $\coprod_c$ for $1 \leq c \leq d$, we continue in this kind of estimation to

$$\coprod_1 \cdots \coprod_d \ w_{\psi_1 j_1, \ldots, \psi_d j_d}^{\varphi_1 j_1, \ldots, \varphi_d j_d}$$

where $w_{k_1, \ldots, k_d}^{i_1, \ldots, i_d}$ equals to 1 of 0 according to $v(k_1, \ldots, k_d)$ equals to $(i_1, \ldots, i_d)$ or not, respectively, $i_c \in N_c$, $k_c \in L_c$, $1 \leq c \leq d$. By the auxiliary inequality this expression is bounded from above by

$$\coprod_1 \cdots \coprod_{d-1} \frac{n_d!}{(n_d - m_d)!} \left(\frac{1}{n_d} \sum_{i_d \in N_d} \sum_{k_d \in L_d} w_{\psi_1 j_1, \ldots, \psi_{d-1} j_{d-1}, k_d}^{\varphi_1 j_1, \ldots, \varphi_{d-1} j_{d-1}, i_d} \right)^{m_d} .$$

With an additional notation

$$\langle n_c, m_c, l_c \rangle = \frac{n_c!}{(n_c - m_c)!} \left(\frac{l_c}{n_c} \right)^{m_c}, \qquad 1 \leq c \leq d,$$

we can rewrite it as

$$\langle n_d, m_d, l_d\rangle \ \coprod_1 \cdots \coprod_{d-2} \sum_{\psi_{d-1}} \sum_{\varphi_{d-1}} \prod_{j_{d-1}\in M_{d-1}} \left(\frac{1}{l_d} \sum_{i_d\in N_d} \sum_{k_d \in L_d} w^{\varphi_1 j_1,\ldots,i_d}_{\psi_1 j_1,\ldots,k_d} \right)^{m_d/m_{d-1}}$$

and omit the exponent $m_d/m_{d-1} \geq 1$. Now this series of tricks is repeated from right to the left. In the last but one step we arrive at

$$\langle n_2, m_2, l_2\rangle \ldots \langle n_d, m_d, l_d\rangle \ \sum_{\psi_1} \sum_{\varphi_1} \prod_{j_1 \in M_1} \left(\frac{1}{l_2 \ldots l_d} \sum_{i_2,k_2} \cdots \sum_{i_d,k_d} w^{\varphi_1 j_1, i_2,\ldots,i_d}_{\psi_1 j_1, k_2 \ldots, k_d} \right)^{(m_2/m_1)}$$

and again by the auxiliary inequality to the upper bound $\prod_{c=1}^d \langle n_c, m_c, l_c\rangle$. ∎

Proof of Proposition. If the maximum is replaced by supremum the inequality $\leq$ follows from Theorem $1^{s.e.}$ (chose $L = N$ and $S_L = P_L$).

To verify the opposite inequality we can assume that $X = N_1 \times \ldots \times N_d$ and that every subset of X is in $\mathcal{X}$. We will show that the maximum is attained for the uniform probability measure R_N on the orbit $\{\widetilde{\tau} z_N;\ \tau \in G_N\}$ of

$$z_N = ((i_1, \ldots, i_d);\ \ i_1 \in N_1, \ldots, i_d \in N_d) \in X_N.$$

Note that z_N is diagonal of $N_1 \times \ldots \times N_d$ and its orbit has $n_1! \ldots n_d!$ elements. To estimate inf sup from below we substitute for $A_M \in \mathcal{X}_M$ the set

$$\begin{aligned} B_M &= \{\widetilde{\varphi}\, z_N;\ \varphi \in N^{*M}\} = \\ &= \{((\varphi_1 j_1, \ldots, \varphi_d j_d);\ j_1 \in M_1, \ldots, j_d \in M_d);\ \ (\varphi_1, \ldots, \varphi_d) \in N^{*M}\} \end{aligned}$$

what gives

$$\inf_{L;\, S_L \in \mathcal{P}_L} \quad \int_{X_L} \left[1 - Q^M_{x_L}(B_M)\right]\ S_L(dx_L)$$

because $R^M_N(B_M) = 1$. But

$$Q^M_{x_L}(B_M) = \sum_{\varphi} \left|\{\psi \in L^M; \widetilde{\psi}\, x_L = \widetilde{\varphi}\, z_N\}\right| \cdot \prod_{c=1}^{d} l_c^{-m_c} \ \leq \ \prod_{c=1}^{d} \frac{n_c!}{n_c^{m_c}(n_c - m_c)!}$$

by use of Lemma 1 for $v = x_L$. Thus, R_N is "worst" approximable. ∎

In the case $d = 1$ the measure R^M_N used in the previous proof can be interpreted by an urn scheme without replacement: the urn should have $|N_1|$ different balls and $|M_1|$ draws are to be performed. The integral approximation of R^M_N from Theorem $1^{s.e.}$ is nothing else as the same scheme with replacement. Our proposition gives thus also variation distance between sampling with and without replacement in accordance with Freedman [10]; cf. Stam [20], [21]. On the other hand, the set B_M means an event when different balls are drawn out. This is always the situation under no replacement in our urn ($R_N^M(B_M) = 1$) while under replacement computing probabilities of B_M has occured in the "birthday problem", see Munford [18].

Results of this section were motivated by Proposition 31 of Diaconis and Freedman [5], p. 756, that concerns usual exchangeability; also our proof of Proposition uses analogical ideas generalizing their assertion.

Sharpness of the bound from Theorem $1^{j.e.}$ need not be speciably commented as it is clearly a consequence of the one-dimensional argumentation.

6. Infinite partially exchangeable arrays on compact Hausdorff spaces. The last part of this paper is devoted to the following two theorems on infinite arrays with a topological structure.

Theorem $2^{s.e.}$ *Let $d \geq 1$, $N = (N_1, \dots N_d)$ be a d-tuple of countable sets and X be a compact Hausdorff space. Then the space $\mathcal{M}_N^{s.e.}$ of all separately exchangeable Radon probability measures on $X_N = X^{N_1 \times \ldots \times N_d}$ is a Bauer simplex, $P_N \in \mathcal{M}_N^{s.e.}$ being its extreme point if and only if it is dissociated.*

Theorem $2^{j.e.}$ *Let $d \geq 1$, N be a countable set and X be as above. The space $\mathcal{M}_N^{j.e.}$ of all jointly exchangeable Radon probability measures on $X_N = X^{N^d}$ is a Bauer simplex the extreme points of which are the dissociated measures of $\mathcal{M}_N^{j.e.}$.*

We work out both proofs in a parallel way, so that N denotes a d-tuple of countable sets or a countable set, respectively. The main arguments are organized in two lemmas that are preceded by an extensive notation and almost trivial remarks.

To begin with, we must admit that there is a slight revision in our measurable setting from Section 2. Namely, having a compact Hausdorff space X and the σ-algebra $\mathcal{B}$ of its Borel subsets the letter X_N denotes Cartesian product of copies of X as before whereas $\mathcal{B}_N$ will be the Borel σ-algebra of X_N (containing, in general properly, the product σ-algebra). The Radon probability measures, $P_N \in \mathcal{M}_N$, live on $(X_N, \mathcal{B}_N)$; all facts about these measures that we shall use below can be found in Chapter 2 of [3]. The space $\mathcal{M}_N$ endowed with the weak topology is compact Hausdorff.

Where φ is a mapping from M to N (or a d-tuple of mappings), the mappings $\widetilde{\varphi} : X_N \to X_M$ and $\approx\!\!\!\!\varphi : \mathcal{M}_N \to \mathcal{M}_M$ are both continuous. The set $\mathcal{M}_N^{s.e.}$ of all $P_N \in \mathcal{M}_N$ that are τ-invariant for every $\tau \in G_N$ ($*$. exchangeable Radon measures) is nonempty, convex and compact. Since marginalization and products of Radon measures are continuous the subset $\mathcal{M}_N^{*.e.d.}$ of $\mathcal{M}_N^{*.e.}$ consisting of the dissociated measures is nonempty and compact.

Knowing that every function $R_N \to R_N(B_N)$, $B_N \in \mathcal{B}_N$, on $\mathcal{M}_N$ is Borel measurable it makes sense to speak about μ-mixtures $\int_{\mathcal{M}_N} R_N \, \mu(dR_N)$ for Borel probability measures μ on $\mathcal{M}_N$. If μ is even Radon then the μ-mixture is Radon as well. These μ-mixtures coincide with the notion of barycentre used in general locally convex Hausdorff topological vector spaces. Making μ-mixtures is a continuous mapping on the space $\mathbf{M}_N$ of Radon probability measures on $\mathcal{M}_N$ with the weak topology. If the support of $\mu \in \mathbf{M}_N$ is contained in $\mathcal{M}_N^{*.e.}$ then the μ-mixture is $*$. exchangeable.

Next, let $M \subset N$ and K be both finite and nonempty (sets or d-tuples of sets with the inclusions and properties in each coordinate). For $x_K \in X_K$ fixed, the probability measure $Q^M_{x_K}$ is taken as the average of Dirac measures sitting in the points $\widetilde{\varphi}\, x_K$, $\varphi \in K^M$; this extends consistently the earlier definition to a Radon measure from $\mathcal{M}_M^{*.e.d.}$. There exists exactly one Radon probability measure $Q^N_{x_K}$ on X_N such that $\widetilde{\widetilde{\pi}}_{M,N} Q^N_{x_K} = Q^M_{x_K}$ for every finite $M \subset N$. In addition, $Q^N_{x_K} \in \mathcal{M}_N^{*.e.d.}$. The mapping $x_K \to Q^N_{x_K}$ is continuous. Making images of measures $P_K \in \mathcal{M}_K$ by this mapping is again a continuous mapping ranging among those $\mu \in \mathbf{M}_N$ that are concentrated on $\mathcal{M}_N^{*.e.d.}$.

Lemma 2. *To every $P_N \in \mathcal{M}_N^{*.e.}$ there exists a Radon probability measure μ on $\mathcal{M}_N$ such that P_N equals the μ-mixture and $\mu(\mathcal{M}_N^{*.e.d.}) = 1$.*

Proof. We can suppose N to be the set of all (d-tuples of) positive integers. Let $K = \{1, 2, \ldots, k\}^{(d)}$, $k \geq 1$, and $M = \{1, 2, \ldots, m\}^{(d)}$, $m \geq 1$, and let $P_N \in \mathcal{M}_N^{*.e.}$ and $\mu_k \in \mathbf{M}_N$ be the image of P_N^K by the mapping $x_K \to Q^N_{x_K}$. The desired measure μ is chosen as the limit of arbitrary convergent subsequence of the sequence μ_k, $k \geq 1$. Clearly, $\mu(\mathcal{M}_N^{*.e.d.}) = 1$.

For every open set $A_M \subset X_M$ and $B_N = \widetilde{\pi}^{-1}_{M,N} A_M$ we have

$$\int_{\mathcal{M}_N} R_N(B_N)\, \mu_k(dR_N) = \int_{X_K} Q^N_{x_K}(B_N)\, P_N^K(dx_K) = \int_{X_K} Q^M_{x_K}(A_M)\, P_N^K(dx_K)$$

and if $k \geq m$ we can apply[4] Theorem $1^{*.e.}$ to the measure $P_N^K \in \mathcal{M}_N^{*.e.}$ to obtain

$$\int_{\mathcal{M}_N} R_N(B_N) \mu_k(dR_N) \;\leq\; P_N(B_N) + 1 - \left[\frac{k!}{k^m (k-m)!}\right]^{(d)} .$$

Here a lower semicontinuous function is integrated and thus the portmanteau theorem implies

$$\int_{\mathcal{M}_N} R_N(B_N) \mu(dR_N) \;\leq\; P_N(B_N) \quad .$$

This inequality is then clearly valid for any open $B_N \subset X_N$ what entails that P_N is equal to the μ-mixture. ∎

[4] Frankly, we cannot directly. We need a version of this theorem where $\mathcal{X}_N$ is replaced by $\mathcal{B}_N$ and P_N is in addition Radon. However, this version has the same proof as Theorem $1^{*.e.}$.

The next lemma is an application of general facts on groups of automorphisms of measurable spaces as investigated e.g. in Farell [9], Varadarajan [22] and Maitra [16]. For readers convenience we conserve the self-containing character of this paper and present some details.

Lemma 3. *Every Radon probability measure $\mu \in \mathbf{M}_N$ such that $\mu(\mathcal{M}_N^{*,e.d.}) = 1$ is uniquely determined by its μ-mixture.*

Proof. Let H_k, $k \geq 1$, be an increasing sequence of finite subgroups of G_N such that every $\tau \in G_N$ is in some H_k. Let $\mathcal{A}_{H_k}$ be the σ-algebra of H_k-invariant Borel subsets of X_N, i.e. $B_N \in \mathcal{A}_{H_k}$ if and only if $B_N = \widetilde{\tau} B_N$ for all $\tau \in H_k$ and let $\mathcal{A}_{G_N} = \bigcap_{k\geq 1} \mathcal{A}_{H_k}$ be the σ-algebra of G_N-invariant sets. First auxiliary fact we are going to prove is that for every $B_N \in \mathcal{B}_N$ there exists a real $\mathcal{A}_{G_N}$-measurable function h_{B_N} on X_N satisfying $P_N(B_N) = \int_{X_N} h_{B_N}(x_N)\, P_N(dx_N)$ for all $P_N \in \mathcal{M}_N^{*,e.}$.

This function is constructed for fixed $B_N \in \mathcal{B}_N$ from the sequence of functions $h_k = |H_k|^{-1} \sum_{\tau \in H_k} \mathbf{1}_{\widetilde{\tau} B_N}$, $k \geq 1$. Observing that the Borel set A_N of all points x_N for which the sequence $h_k(x_N)$ converges is G_N-invariant it suffices to take $h_{B_N}(x_N) = \lim_{k\to\infty} h_k(x_N)$ for $x_N \in A_N$ and $h_{B_N}(x_N) = 0$ otherwise. Evidently, $h_{B_N}(x_N) = h_{B_N}(\widetilde{\tau} x_N)$ for any $x_N \in X_N$ and $\tau \in G_N$ whence h_{B_N} is $\mathcal{A}_{G_N}$-measurable. Now, if $P_N \in \mathcal{M}_N^{*,e.}$ then h_k is a version of the conditional probability $P_N(B_N|\mathcal{A}_{H_k})$ and then by the reverse martingale theorem h_k converges P_N a.e. to a version of the conditional probability $P_N(B_N|\mathcal{A}_{G_N})$. This implies $P_N(A_N) = 1$ and since $P_N(B_N) = \int_{A_N} h_k(x_N)\, P_N(dx_N)$, $k \geq 1$, we have the desired equality.

Further we need the fact that the dissociated measures $P_N \in \mathcal{M}_N^{*,e.d.}$ takes on $\mathcal{A}_{G_N}$ only the values 0 or 1. A proper reasoning comes from the standard proof of Hewitt-Savage's 0-1 law: having some $A_N \in \mathcal{A}_{G_N}$ and $a > 1 > b > 0$ one can find open sets $B_N \supset A_N$ and $C_N \subset B_N$ such that $aP_N(A_N) \geq P_N(B_N)$, C_N depends only on finite number of coordinates and $P_N(C_N) \geq P_N(B_N)$. Then for a suitable $\tau \in G_N$ we have

$$\begin{aligned} a^2 P_N(A_N)^2 &\geq P_N(C_N)^2 = P_N(C_N \cap \widetilde{\tau} C_N) \geq 2bP_N(B_N) - P_N(B_N \cap \widetilde{\tau} B_N) \geq \\ &\geq 2bP_N(A_N) - 2P_N(B_N) + P_N(A_N \cap \widetilde{\tau} A_N) \geq (2b - 2a + 1) P_N(A_N) \ . \end{aligned}$$

The crucial point of the proof is introduction of the mapping

$$t(A_N) = \{P_N \in \mathcal{M}_N^{*,e.d.};\ P_N(A_N) = 1\}\ , \quad A_N \in \mathcal{A}_{G_N}\ ,$$

that sends invariant Borel subsets of X_N to Borel subsets of $\mathcal{M}_N^{*,e.d.}$.[5] It is easy to verify that every set $A_N = \{x_N \in X_N; h_{B_N}(x_N) \leq a\}$ where $B_N \in \mathcal{B}_N$ and $0 \leq a \leq 1$

[5] This idea goes back, as far as we were able to trace it, to the work Blum J.P.R. and Hanson D.H. (1960) On invariant probability measures I., *Pacific J. Math.* **10** 1125–1129.

is mapped by t to the set $\{P_N \in \mathcal{M}_N^{*.e.d.};\ P_N(B_N) \le a\}$. Namely, if $P_N(A_N) = 1$ for $P_N \in \mathcal{M}_N^{*.e.d.}$ and this set A_N then $P_N(B_N) = \int_{A_N} h_{B_N}(x_N)\, P_N(dx_N) \le a$ and in the opposite case $P_N(A_N) = 0$ similarly $P_N(B_N) = \int_{X_N - A_N} h_{B_N}(x_N)\, P_N(dx_N) > a$. Finally, if $\mu \in \mathbf{M}_N$ is sitting on $\mathcal{M}_N^{*.e.d.}$ then owing to

$$\mu(t(A_N)) = \int_{t(A_N)} P_N(A_N)\, \mu(dP_N) = \int_{\mathcal{M}_N^{*.e.d.}} P_N(A_N)\, \mu(dP_N)\ , \quad A_N \in \mathcal{A}_{G_N}\ ,$$

the measure μ can be recovered from its μ-mixture uniquely. ∎

Proof of Theorem 2$^{*.e.}$. We can summarize: $\mathcal{M}_N^{*.e.}$ is a convex compact set and by Lemma 2 its extreme points ex $\mathcal{M}_N^{*.e.}$ are contained in $\mathcal{M}_N^{*.e.d.}$. Lemma 3 implies that the dissociated $P_N \in \mathcal{M}_N^{*.e.}$ is extreme in $\mathcal{M}_N^{*.e.}$, hence ex $\mathcal{M}_N^{*.e.} = \mathcal{M}_N^{*.e.d.}$ is compact. Every $P_N \in \mathcal{M}_N^{*.e.}$ is a mixture, equivalently, a barycentre, of exactly one $\mu \in \mathbf{M}_N$ with its support in ex $\mathcal{M}_N^{*.e.}$ again by Lemma 2. (note that the sequence μ_k in its proof always converges). This set of μ's is homeomorphic to $\mathcal{M}_N^{*.e.}$, the homeomorphism being making of mixtures. ∎

Having fixed $d \ge 1$ and the kind $(*)$ of exchangeability, let us say that a Hausdorff space X is *Radon presentable* if every $P_N \in \mathcal{M}_N^{*.e.}$ can be expressed as the μ-mixture for unique Radon probability measure μ on $\mathcal{M}_N^{*.e.d.}$. Theorem $2^{*.e.}$ implies that compact Hausdorff spaces are Radon presentable. A direct modification of the proof of Theorem 3 by Ressel(1985) brings in that every completely regular Hausdorff space is Radon presentable, too. Especially, for polish spaces Lemma 2 (and then also Radon presentability) follows from the representation theorem of Aldous(1981).

References

[1] Aldous D.J. (1981) Representations for partially exchangeable arrays of random variables. *J. Multivariate Anal.* **11**, 581–598.

[2] Aldous D.J. (1985) Exchangeability and related topics. In: *Springer Lecture Notes in Math. 1117*, 1–198, Springer-Verlag, Berlin.

[3] Berg Ch., Christensen J.P.R. and Ressel P. (1984) *Harmonic Analysis on Semigroups.* Springer-Verlag, New York.

[4] Diaconis P. (1988) Recent progress on de Finetti's notions of exchangeability. In: *Bayesian Statistics 3*, 111–125, (eds. Bernardo J.B., de Groot M.H., Lindley D.V. and Smith A.F.M.), Oxford University Press, Oxford.

[5] Diaconis P.W. and Freedman D.A. (1980) Finite exchangeable sequences. *Ann. Probab.* **8**, 745–764.

[6] Diaconis P.W. and Freedman D.A. (1987) A dozen de Finetti-style results in search of a theory. *Ann. Inst. H. Poincaré*, Sup. au no. 2, **23**, 397-423.

[7] Diaconis P.W., Eaton M.L. and Lauritzen S.L. (1992) Finite de Finetti theorems in linear models and multivariate analysis. *Scand. J. Statist.* **19**, 289–315.

[8] Dubins L.E. and Freedman D.A. (1979) Exchangeable processes need not be mixtures of independent identically distributed random variables. *Z. Warsch. verw. Gebiete* **48**, 115–132.

[9] Farell R.H. (1962) Representation of invariant measures. *Illinois J. Math.* **6**, 447–467.

[10] Freedman D. (1977) A remark on the difference between sampling with and without replacement. *J. Amer. Statist. Assoc.* **32**, 681.

[11] Hewitt E. and Savage L.J. (1955) Symmetric measures on Cartesian products. *Trans. Amer. Math. Soc.* **80**, 470–501.

[12] Kallenberg O. (1988) Some new representations in bivariate exchangeability. *Probab. Th. Rel. Fields* **77**, 415–455.

[13] Kallenberg O. (1993) Symmetries on random arrays and set-indexed processes. *J. Theor. Probab.* **6**, 727–765.

[14] Koch G. and Spizzichino F., Eds. (1982) *Exchangeability in Probability and Statistics.* North-Holland Publ. Co., Amsterdam.

[15] Lauritzen S.L. (1988) *Extremal Families and Systems of Sufficient Statistics.* Lecture Notes in Statistics 49, Springer-Verlag, New York.

[16] Maitra A. (1977) Integral representations of invariant measures. *Trans. Amer. Math. Soc.* **229**, 209–225.

[17] Matúš F. (1994) On two-block-factor sequences and one-dependence. (to appear in *Proc. Amer. Math. Soc.*)

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

[19] Ressel P. (1985) De Finetti-type theorems: an analytical approach. *Ann. Probab.* **13**, 898–922.

[20] Stam A.J. (1978) Distance between sampling with and without replacement. *Stat. Neerl.* **32**, 81–91.

[21] Stam A.J. (1986) A note on sampling with and without replacement. *Stat. Neerl.* **40**, 35–38.

[22] Varadarajan V.S. (1963) Groups of automorphisms of Borel spaces. *Trans. Amer. Math. Soc.* **109**, 191–220.

[23] Zaman A. (1986) A finite form of de Finetti's theorem for stationary Markov exchangeability. *Ann. Probab.* **14**, 1418–1427.