J. Math. Anal. Appl. • • • (• • • •) • • •–• • •

Contents lists available at ScienceDirect

Journal of Mathematical Analysis and Applications

www.elsevier.com/locate/jmaa

On a representation theorem for finitely exchangeable random vectors

Svante Janson [a,1], Takis Konstantopoulos [a,*,2], Linglong Yuan [b,3]

[a] *Department of Mathematics, Uppsala University, P.O. Box 480, SE-75106 Uppsala, Sweden*
[b] *Institut für Mathematik, Johannes-Gutenberg-Universität Mainz, Staudingerweg 9, 55099 Mainz, Germany*

ARTICLE INFO

Article history:
Received 15 May 2015
Available online xxxx
Submitted by V. Pozdnyakov

Keywords:
Signed measure
Measurable space
Point measure
Exchangeable
Symmetric
Homogeneous polynomial

ABSTRACT

A random vector $X = (X_1, \dots, X_n)$ with the X_i taking values in an arbitrary measurable space $(S, \mathscr{S})$ is exchangeable if its law is the same as that of $(X_{\sigma(1)}, \dots, X_{\sigma(n)})$ for any permutation σ. We give an alternative and shorter proof of the representation result (Jaynes [6] and Kerns and Székely [9]) stating that the law of X is a mixture of product probability measures with respect to a signed mixing measure. The result is "finitistic" in nature meaning that it is a matter of linear algebra for finite S. The passing from finite S to an arbitrary one may pose some measure-theoretic difficulties which are avoided by our proof. The mixing signed measure is not unique (examples are given), but we pay more attention to the one constructed in the proof ("canonical mixing measure") by pointing out some of its characteristics. The mixing measure is, in general, defined on the space of probability measures on S; but for $S = \mathbb{R}$, one can choose a mixing measure on $\mathbb{R}^n$.

1. Introduction

The first result that comes to mind when talking about exchangeability is de Finetti's theorem concerning sequences $X = (X_1, X_2, \dots)$ of random variables with values in some space S and which are invariant under permutations of finitely many coordinates. This remarkable theorem [7, Theorem 11.10] states that the law of such a sequence is a mixture of product measures: let S^∞ be the product of countably many copies of S and let π^∞ be the product measure on S^∞ with marginals $\pi \in \mathscr{P}(S)$ (the space of probability measures on S); then

* Corresponding author.
E-mail addresses: svante.janson@math.uu.se (S. Janson), takiskonst@gmail.com (T. Konstantopoulos), yuanlinglongcn@gmail.com (L. Yuan).
URLs: http://www2.math.uu.se/~svante (S. Janson), http://www2.math.uu.se/~takis (T. Konstantopoulos), http://linglongyuan.weebly.com (L. Yuan).

[1] Partly supported by the Knut and Alice Wallenberg Foundation.
[2] Supported by Swedish Research Council grant 2013-4688.
[3] The work of this author was done while he was a postdoctoral researcher at Uppsala University.

http://dx.doi.org/10.1016/j.jmaa.2016.04.070
0022-247X/© 2016 Published by Elsevier Inc.

$$\mathbb{P}(X \in \cdot) = \int_{\mathscr{P}(S)} \pi^\infty(\cdot)\,\nu(\mathrm{d}\pi),$$

for a uniquely defined probability measure ν which we call a mixing (or directing) measure.

In Bayesian language, this says that any exchangeable random sequence is obtained by first picking a probability distribution π from some prior (probability distribution on the space of probability distributions) and then letting the X_i to be i.i.d. with common law π. As Dubins and Freedman [5] show, de Finetti's theorem does not hold for an arbitrary measurable space S. Restrictions are required. One of the most general cases for which the theorem does hold is that of a Borel space S, i.e., a space which is isomorphic (in the sense of existence of a measurable bijection with measurable inverse) to a Borel subset of $\mathbb{R}$. Indeed, one of the most elegant proofs of the theorem can be found in Kallenberg [8, Section 1.1] from which it is evident that the main ingredient is the ergodic theorem and that the Borel space is responsible for the existence of regular conditional distributions.

For finite dimension n, however, things are different. Let S be a set together with a σ-algebra $\mathscr{S}$, and let $X_1, \dots, X_n$ be measurable functions from a measure space $(\Omega, \mathscr{F})$ into $(S, \mathscr{S})$. Under a probability measure $\mathbb{P}$ on $(\Omega, \mathscr{F})$, assume that $X = (X_1, \dots, X_n)$ is such that $\sigma X := (X_{\sigma(1)}, \dots, X_{\sigma(n)})$ has the same law as $(X_1, \dots, X_n)$ for any permutation σ of $\{1, \dots, n\}$, i.e., that $\mathbb{P}(\sigma X \in B) = \mathbb{P}(X \in B)$ for all $B \in \mathscr{S}^n$, where $\mathscr{S}^n$ is the product σ-algebra on S^n. In such a case, we say that X is n-exchangeable (or simply exchangeable).

Example 1. Simple examples show that a finitely exchangeable random vector may not be a mixture of product measures. For instance, take $S = \{1, \dots, n\}$, with $n \geq 2$, and let $X = (X_1, \dots, X_n)$ take values in S^n such that $\mathbb{P}(X = x) = 1/n!$ when $x = (x_1, \dots, x_n)$ is a permutation of $(1, \dots, n)$, and $\mathbb{P}(X = x) = 0$ otherwise. Clearly, X is n-exchangeable. Suppose that the law of X is a mixture of product measures. Since the space of probability measures $\mathscr{P}(S)$ can naturally be identified with the set $\Sigma_n := \{(p_1, \dots, p_n) \in \mathbb{R}^n : p_1, \dots, p_n \geq 0,\ p_1 + \cdots + p_n = 1\}$, the assumption that the law of X is a mixture of product measures is equivalent to the following: there is a random variable $p = (p_1, \dots, p_n)$ with values in Σ_n such that $\mathbb{P}(X = x) = \mathbb{E}[\mathbb{P}(X = x|p)]$, where $\mathbb{P}(X = x|p) = p_{x_1} \cdots p_{x_n}$ for all $x_1, \dots, x_n \in S$. But then, for all $i \in S$, $0 = \mathbb{P}(X_1 = \cdots = X_n = i) = \mathbb{E}[p_i^n]$, implying that $p_i = 0$, almost surely, for all $i \in S$, an obvious contradiction.

However, Jaynes [6] showed that (for the $|S| = 2$ case) there is mixing provided that signed measures are allowed; see equation (1) below. Kerns and Székely [9] observed that the Jaynes result can be generalized to *an arbitrary* measurable space S, but the proof in [9] requires some further explicit arguments. In addition, [9] uses a non-trivial algebraic result without a proof. Our purpose in this note is to give an alternative, shorter, and rigorous proof of the representation result (see Theorem 1 below) but also to briefly discuss some consequences and open problems (Theorem 2 and Section 4). An independent proof of an algebraic result needed in the proof of Theorem 1 is presented in Appendix A as Theorem 3. To the best of our knowledge, the proof is new and, possibly, of independent interest.

Theorem 1 *(Finite exchangeability representation theorem). Let $X_1, \dots, X_n$ be random variables on some probability space $(\Omega, \mathscr{F}, \mathbb{P})$ with values in a measurable space $(S, \mathscr{S})$. Suppose that the law of $X = (X_1, \dots, X_n)$ is exchangeable. Then there is a signed measure ξ on $\mathscr{P}(S)$*

$$\mathbb{P}(X \in A) = \int_{\mathscr{P}(S)} \pi^n(A)\,\xi(\mathrm{d}\pi), \quad A \in \mathscr{S}^n, \tag{1}$$

where π^n is the product of n copies of $\pi \in \mathscr{P}(S)$.

Please cite this article in press as: S. Janson et al., On a representation theorem for finitely exchangeable random vectors, J. Math. Anal. Appl. (2016), http://dx.doi.org/10.1016/j.jmaa.2016.04.070

We stress that the theorem does not ensure uniqueness of ξ.

Example 2. To see this in an example, consider Example 1 with $n = 2$, that is, let $S = \{1, 2\}$ and let $X = (X_1, X_2)$ take values $(1, 2)$, $(2, 1)$, $(1, 1)$, $(2, 2)$ with probabilities $1/2$, $1/2$, 0, 0, respectively. We identify $\mathscr{P}(S)$ with the interval $[0, 1]$, via $\pi\{1\} = p$, $\pi\{2\} = 1 - p$, for $\pi \in \mathscr{P}(S)$. We give three different signed measures that can be used in the representation.

(i) Let ξ be the signed measure on $[0, 1]$ defined by

$$\xi = -\frac{1}{2}\delta_0 - \frac{1}{2}\delta_1 + 2\delta_{1/2}.$$

Then $\mathbb{P}(X_1 = 1, X_2 = 2) = \int_{[0,1]} p(1 - p)\, \xi(\mathrm{d}p) = 1/2 = \mathbb{P}(X_1 = 2, X_1 = 1)$, while $\mathbb{P}(X_1 = 1, X_2 = 1) = \int_{[0,1]} p^2\, \xi(\mathrm{d}p) = 0 = \mathbb{P}(X_1 = 2, X_2 = 2)$.

(ii) Let

$$\xi = -\frac{5}{8}\delta_0 - \frac{5}{8}\delta_1 + \frac{9}{8}\delta_{1/3} + \frac{9}{8}\delta_{2/3}.$$

Again, $\int_{[0,1]} p(1 - p)\, \xi(\mathrm{d}p) = 1/2$, $\int_{[0,1]} p^2\, \xi(\mathrm{d}p) = \int_{[0,1]} (1 - p)^2\, \xi(\mathrm{d}p) = 0$.

(iii) Let ξ be a signed measure with density

$$f(p) := -\frac{7}{2} \cdot \mathbf{1}_{p \le 1/3 \text{ or } p \ge 2/3} + 10 \cdot \mathbf{1}_{1/3 < p < 2/3}.$$

We can easily see that $\int_0^1 f(p)\mathrm{d}p = 1$, $\int_0^1 p^2 f(p)\mathrm{d}p = 0$, $\int_0^1 p(1 - p)f(p)\mathrm{d}p = 1/2$.

Remark 1. The difference between this situation and the one in de Finetti's setup is that a finitely exchangeable random vector $(X_1, \dots, X_n)$ is not necessarily extendible to an infinite sequence $(X_1, \dots, X_n, X_{n+1}, \dots)$ that is exchangeable. (See Examples 3 and 4 below.) If it were, then the signed measure ξ could have been chosen as a probability measure (and would then have been unique). The question of extendibility of an n-exchangeable $(X_1, \dots, X_n)$ to an N-exchangeable $(X_1, \dots, X_N)$, for some $N > n$ (possibly $N = \infty$) is treated in the sequel paper [10] that strongly uses the framework and results of the present paper. Assuming such extendibility, Diaconis and Freedman [3,4] show that the total variation distance of an n-exchangeable probability measure on S^n from the set of mixtures of product probability measures is at most $n(n - 1)/N$ when S is an infinite set (and at most $2|S|n/N$ if S is finite).

When $S = \mathbb{R}$, it is possible to say more than in Theorem 1:

Theorem 2. *Let $(X_1, \dots, X_n)$ be an n-exchangeable random vector in $\mathbb{R}^n$, endowed with the Borel σ-algebra. Then there is a bounded signed measure $\eta(\mathrm{d}\theta_1, \dots, \mathrm{d}\theta_n)$, such that $\mathbb{P}(X \in A) = \int_{\mathbb{R}^n} \pi^n_{\theta_1,\dots,\theta_n}(A)\, \eta(\mathrm{d}\theta_1, \dots, \mathrm{d}\theta_n)$, where $\pi_{\theta_1,\dots,\theta_n}$ is an element of $\mathscr{P}(S)$ depending measurably on the n parameters $(\theta_1, \dots, \theta_n)$.*

2. Preliminaries and notation

We make use of the following notations and terminology in the paper. If S is a set with a σ-algebra $\mathscr{S}$, then $\mathscr{P}(S)$ is the set of probability measures on $(S, \mathscr{S})$. The space $\mathscr{P}(S)$ is equipped with the σ-algebra generated by sets of the form $\{\pi \in \mathscr{P}(S) : \pi(B) \le t\}$, $B \in \mathscr{S}$, $t \in \mathbb{R}$. We shall write $\mathscr{P}(S, \mathscr{S})$ if we wish to emphasize the role of the σ-algebra.[4] Similarly, $\mathscr{M}(S)$ or $\mathscr{M}(S, \mathscr{S})$ will be the space of bounded signed measures, equipped with a σ-algebra as above. In particular, $\mathscr{M}(\mathscr{P}(S))$ is the space of bounded signed measures on $\mathscr{P}(S, \mathscr{S})$. A random measure on S is a measurable mapping from Ω into $\mathscr{P}(S)$ and a random signed measure on $\mathscr{P}(S)$ is a measurable mapping from Ω into $\mathscr{M}(\mathscr{P}(S))$. The measure ξ in Theorem 1 is an element of $\mathscr{M}(\mathscr{P}(S))$. The delta measure δ_a at a point $a \in S$ is, as usual, the set function $\delta_a(B) := \mathbf{1}_{a \in B}$, $B \subset S$. A finite point measure is a finite linear combination of delta measures where the coefficients are nonnegative integers. We let $\mathscr{N}(S)$ be the set of finite point measures on S and $\mathscr{N}_n(S)$ the set of point measures ν such that $\nu(S) = n$. The symbol $(\nu)!$ is defined as

$$(\nu)! := \prod_{a \in S} \nu\{a\}!$$

where $\nu\{a\}$ is the value of ν at the singleton $\{a\}$ and where the product is over the support of ν $(0! := 1)$. The symbol $\mathscr{S}^n$ stands for the product σ-algebra on S^n. If $\pi \in \mathscr{P}(S)$ then $\pi^n \in \mathscr{P}(S^n)$ is the product measure of π with itself, n times. If $x = (x_1, \dots, x_n) \in S^n$ then the *type* of x is the element ε_x of $\mathscr{N}_n(S)$ defined by

$$\varepsilon_x := \sum_{i=1}^{n} \delta_{x_i}.$$

The set $S^n(\nu) \subset S^n$ is defined by, for $\nu \in \mathscr{N}_n(S)$,

$$S^n(\nu) := \{y \in S^n : \varepsilon_y = \nu\}.$$

It is a finite set with cardinality

$$\binom{n}{\nu} := \frac{n!}{(\nu)!}.$$

We let $\boldsymbol{u}_\nu$ be the uniform probability measure on $S^n(\nu)$, that is,

$$\boldsymbol{u}_\nu = \binom{n}{\nu}^{-1} \sum_{z \in S^n(\nu)} \delta_z.$$

If $\mathscr{S}$ is too coarse, then $S^n(\nu)$ may not belong to $\mathscr{S}^n$. This is not a problem when $\mathscr{S}$ is, say, the Borel σ-algebra of a Hausdorff space, but we wish to prove the result without any topological assumptions. Moreover, notice that

$$S^n = \bigcup_{\nu \in \mathscr{N}_n(S)} S^n(\nu), \tag{2}$$

since $y \in S^n(\varepsilon_y)$ for all $y \in S^n$. The sets in the union are pairwise disjoint because $S^n(\nu) \cap S^n(\nu') = \varnothing$ if ν and ν' are distinct elements of $\mathscr{N}_n(S)$. If σ is a permutation of $\{1, \dots, n\}$ and $x \in S^n$, then $\sigma x := (x_{\sigma(1)}, \dots, x_{\sigma(n)})$.

[4] In other words, when we write $\mathscr{P}(S, \mathscr{S})$, we mean that $\mathscr{P}(S)$ is given the σ-algebra generated by sets $\{\pi \in \mathscr{P}(S) : \pi(B) \le t\}$, $B \in \mathscr{S}$, $t \in \mathbb{R}$.

Please cite this article in press as: S. Janson et al., On a representation theorem for finitely exchangeable random vectors, J. Math. Anal. Appl. (2016), http://dx.doi.org/10.1016/j.jmaa.2016.04.070

3. Proof of the finite exchangeability representation theorem

By exchangeability, for any $B \in \mathscr{S}^n$,

$$\mathbb{P}(X \in B) = \frac{1}{n!}\sum_{\sigma} \mathbb{P}(\sigma X \in B) = \mathbb{E}\frac{1}{n!}\sum_{\sigma} \delta_{\sigma X}(B) = \mathbb{E}\boldsymbol{U}_X(B)$$

where the sum is taken over all permutations σ of $\{1, \dots, n\}$, and where

$$\boldsymbol{U}_x := \frac{1}{n!}\sum_{\sigma} \delta_{\sigma x}, \quad x \in S^n.$$

Notice that the map $x \mapsto \boldsymbol{U}_x$ is a measurable function from $(S^n, \mathscr{S}^n)$ into $\mathscr{P}(S^n, \mathscr{S}^n)$, and, since X is a measurable function from $(\Omega, \mathscr{F})$ into $(S^n, \mathscr{S}^n)$, we have that $\boldsymbol{U}_X$ is a random element of $\mathscr{P}(S^n, \mathscr{S}^n)$. The mean measure $\mathbb{E}\boldsymbol{U}_X$ is the probability law of X.

Forgetting temporarily that our original space is S, consider a finite set T and let Q be an exchangeable probability measure on T. Then, for all $\nu \in \mathscr{N}_n(T)$, Q assigns the same value to every singleton of $T^n(\nu)$. Hence

$$Q = \sum_{\nu \in \mathscr{N}_n(T)} Q(T^n(\nu))\, \boldsymbol{u}_\nu, \tag{3}$$

where $\boldsymbol{u}_\nu$ is the uniform probability measure on $T^n(\nu)$. In particular, let $Q = \pi^n$, where $\pi \in \mathscr{P}(S)$. It is easy to see (multinomial distribution) that

$$\pi^n(T^n(\nu)) = \binom{n}{\nu} \pi^\nu,$$

where $\pi^\nu := \prod_{a \in T} \pi\{a\}^{\nu\{a\}}$ (adopting the convention $0^0 = 1$). Specialize further by letting $\pi = \lambda/n$ where $\lambda \in \mathscr{N}_n(T)$. Canceling a factor, (3) gives

$$\lambda^n = \sum_{\nu \in \mathscr{N}_n(T)} \binom{n}{\nu} \lambda^\nu\, \boldsymbol{u}_\nu. \tag{4}$$

Let W be a matrix with entries $W(\lambda, \nu) := \binom{n}{\nu}\lambda^\nu$, $\lambda, \nu \in \mathscr{N}_n(T)$. This is essentially the multinomial Dyson matrix; see (A.1) in Appendix A and the discussion therein. Let M be the inverse of W; see (A.2) in Appendix A. From (A.2) and (4) we have

$$\boldsymbol{u}_\nu = \sum_{\lambda \in \mathscr{N}_n(T)} M(\nu, \lambda)\, \lambda^n. \tag{5}$$

This is an equality between measures on T^n.

Specialize further by letting $T = [n] := \{1, \dots, n\}$. Fix $x = (x_1, \dots, x_n) \in S^n$. Define $\varphi_x : [n] \to S$ by $\varphi_x(i) = x_i$, $i = 1, \dots, n$. This induces a linear map $\mathscr{M}([n]) \to \mathscr{M}(S)$, also denoted by φ_x, by the formula $\varphi_x(\delta_i) = \delta_{\varphi_x(i)} = \delta_{x_i}$, $i \in [n]$, and extended by linearity: $\varphi_x(\sum_{i=1}^n c_i\delta_i) = \sum_{i=1}^n c_i\varphi_x(\delta_i)$. Define $\varphi_x^n : [n]^n \to S^n$ by $\varphi_x^n(i_1, \dots, i_n) = (\varphi_x(i_1), \dots, \varphi_x(i_n))$. This again induces a linear map $\mathscr{M}([n]^n) \to \mathscr{M}(S^n)$, also denoted by φ_x^n, by the formula $\varphi_x^n(\delta_j) = \delta_{\varphi_x^n(j)}$, $j \in [n]^n$, and extended by linearity. We can then easily show that $\varphi_x^n(\mu_1 \times \cdots \times \mu_n) = \varphi_x(\mu_1) \times \cdots \times \varphi_x(\mu_n)$ for any $\mu_1, \dots, \mu_n \in \mathscr{M}([n])$. Let now ν_n be the measure on $[n]$ with $\nu_n\{i\} = 1$, $i = 1, \dots, n$. Let $\mathscr{N}_n(n) := \mathscr{N}_n([n])$. Then (5) yields

$$\boldsymbol{u}_{\nu_n} = \sum_{\lambda \in \mathscr{N}_n(n)} M_n(\lambda)\lambda^n,$$

where $M_n(\lambda) := M(\nu_n, \lambda)$. It is easy to see that

$$\boldsymbol{u}_{\nu_n} = \frac{1}{n!}\sum_{\sigma} \delta_{\sigma\iota} = \boldsymbol{U}_\iota,$$

where $\iota := (1, \ldots, n)$ and where the sum is taken over all permutations σ of $[n]$. The last two displays are equalities between measures on $\{1, \ldots, n\}^n$.

For each $x \in S^n$ define

$$\psi_x := \sum_{\lambda \in \mathscr{N}_n(n)} n^n M_n(\lambda)\delta_{\varphi_x(\lambda/n)}, \tag{6}$$

a signed measure on $\mathscr{P}(S)$. We are going to show that

(i) $\int_{\mathscr{P}(S)} \tau^n \psi_x(\mathrm{d}\tau) = \boldsymbol{U}_x$,
(ii) $x \mapsto \psi_x$ is a measurable map $S \to \mathscr{M}(\mathscr{P}(S))$.

To show (i), observe, directly from the definition of ψ_x, that

$$\int_{\mathscr{P}(S)} \tau^n \psi_x(\mathrm{d}\tau) = \sum_{\lambda \in \mathscr{N}_n(n)} n^n M_n(\lambda)\varphi_x(\lambda/n)^n. \tag{7}$$

But $\varphi_x(\lambda/n)^n = \varphi_x^n((\lambda/n)^n) = n^{-n}\varphi_x^n(\lambda^n)$ and so

$$\begin{aligned} \int_{\mathscr{P}(S)} \tau^n \psi_x(\mathrm{d}\tau) &= \sum_{\lambda \in \mathscr{N}_n(n)} M_n(\lambda)\varphi_x^n(\lambda^n) = \varphi_x^n\left(\sum_{\lambda \in \mathscr{N}_n(n)} M_n(\lambda)\lambda^n\right) = \varphi_x^n(\boldsymbol{u}_{\nu_n}) = \varphi_x^n(\boldsymbol{U}_\iota) \\ &= \frac{1}{n!}\sum_\sigma \varphi_x^n(\delta_{\sigma\iota}) = \frac{1}{n!}\sum_\sigma \delta_{\varphi_x^n(\sigma\iota)} = \frac{1}{n!}\sum_\sigma \delta_{\sigma x} = \boldsymbol{U}_x. \end{aligned} \tag{8}$$

To show (ii), we first observe that $\varphi_x(\lambda/n) = \sum_{i=1}^n \frac{\lambda_i}{n}\delta_{x_i}$ and that the maps $x \mapsto \delta_{x_i}$, $S^n \to \mathscr{P}(S)$, are measurable. It then follows that $x \mapsto \varphi_x(\lambda/n)$, $S^n \to \mathscr{P}(S)$, is measurable. Also, the map $\mu \mapsto \delta_\mu$, $\mathscr{P}(S) \to \mathscr{M}(\mathscr{P}(S))$, is measurable. Since composition of measurable functions is measurable, we have that $x \mapsto \delta_{\varphi_x(\lambda/n)}$ is a measurable function from S into $\mathscr{M}(\mathscr{P}(S))$.

Since $x \mapsto \psi_x$ is measurable we have that ψ_X is a random element of $\mathscr{M}(\mathscr{P}(S))$ and thus $\xi := \mathbb{E}\psi_X$ is a well-defined element of $\mathscr{M}(\mathscr{P}(S))$. Note that ψ_X is also bounded so there is no problem with taking the expectation. On the other hand, since $\boldsymbol{U}_X = \int_{\mathscr{P}(S)} \tau^n \psi_X(\mathrm{d}\tau)$, a.s., and since $\mathbb{E}\boldsymbol{U}_X$ is the probability distribution of X, the assertion (1) follows with $\xi = \mathbb{E}\psi_X$.

4. Additional results and applications

4.1. Proof of Theorem 2

By (6), the integration on the left-hand side of (8) actually takes place over the set $\{\varphi_x(\lambda/n) : \lambda \in \mathscr{N}_n(n)\}$; since $\varphi_x(\lambda/n) = \frac{1}{n}\varphi_x(\lambda)$ and $\varphi_x(\lambda) \in \mathscr{N}_n(S)$ when $\lambda \in \mathscr{N}_n(n)$, it follows that it suffices to integrate over

$\frac{1}{n}\mathscr{N}_n(S) := \{\frac{1}{n}\lambda : \lambda \in \mathscr{N}_n(S)\} \subset \mathscr{P}(S)$. Let $S = \mathbb{R}$. Then $\frac{1}{n}\mathscr{N}_n(\mathbb{R})$ is a measurable subset of $\mathscr{P}(\mathbb{R})$. Then, for any Borel subset B of $\mathbb{R}^n$,

$$\mathbb{P}(X \in B) = \int_{\frac{1}{n}\mathscr{N}_n(\mathbb{R})} \tau^n(B)\,\xi(\mathrm{d}\tau).$$

We can write $\mathscr{N}_n(\mathbb{R}) = \bigcup_{d=1}^n \mathscr{N}_{n,d}(\mathbb{R})$ where $\mathscr{N}_{n,d}(\mathbb{R})$ is the set of all point measures ν on $\mathbb{R}$ with total mass equal to n and support of size d. The set $\mathscr{N}_{n,n}(\mathbb{R})$ can be identified with all $(x_1, \dots, x_n) \in \mathbb{R}^n$ such that $x_1 < \cdots < x_n$. This is an open cone C. Each of the other sets, $\mathscr{N}_{n,d}(\mathbb{R})$, $d = 1, \dots, n-1$, corresponds to a particular subset of the boundary of C. Therefore, we can replace the integration by integration on a cone of $\mathbb{R}^n$. □

This result tells us that in order to represent an n-exchangeable random vector in $\mathbb{R}^n$ as an integral against an unknown signed measure we may as well search for a signed measure on a space of dimension n rather than on the space $\mathscr{P}(\mathbb{R})$. Of course, we chose $S = \mathbb{R}$ in Theorem 2 as a matter of convenience. A similar result can be formulated more generally.

4.2. A more explicit formula for the mixing measure ψ_x

We claim that ψ_x, defined by (6), also is given by

$$\psi_x = \sum_{\lambda \in \mathscr{N}_n(T)} n^n M(\varepsilon_x, \lambda)\,\delta_{\lambda/n}, \tag{9}$$

for any finite set T such that $x_1, \dots, x_n \in T$.

Indeed, if we temporarily denote the right-hand side of (9) by ψ'_x, then, using (5),

$$\int_{\mathscr{P}(S)} \tau^n\,\psi'_x(\mathrm{d}\tau) = \sum_{\lambda \in \mathscr{N}_n(T)} n^n M(\varepsilon_x, \lambda)\,(\lambda/n)^n = \boldsymbol{u}_{\varepsilon_x} = \boldsymbol{U}_x, \tag{10}$$

where the last equality follows from the definitions. By (10) and (8),

$$\int_{\mathscr{P}(S)} \tau^n\,\psi'_x(\mathrm{d}\tau) = \int_{\mathscr{P}(S)} \tau^n\,\psi_x(\mathrm{d}\tau). \tag{11}$$

Now note that if $\lambda \in \mathscr{N}(n)$, then $\varphi_x(\lambda/n) \in \frac{1}{n}\mathscr{N}(T)$, cf. Section 4.1, and thus the measures ψ_x and ψ'_x both are supported on $\frac{1}{n}\mathscr{N}(T)$. Moreover, the measures $\boldsymbol{u}_\nu$, $\nu \in \mathscr{N}(T)$, are linearly dependent and form thus a basis in a linear space of dimension $|\mathscr{N}(T)|$. By (4) and (5), the measures λ^n, $\lambda \in \mathscr{N}(T)$, span the same space, so they form another basis and are therefore linearly independent. Hence, the measures τ^n, $\tau \in \frac{1}{n}\mathscr{N}(T)$, are linearly independent. Consequently, the equality (11) implies that $\psi'_x = \psi_x$, as claimed.

In particular, we can in (9) always choose $T = T_x := \{x_1, \dots, x_n\}$.

4.3. The canonical mixing measure

This is the particular signed measure ξ constructed as the mean measure of the random signed measure ψ_X, given by (6).

For example, let $n = 2$. An easy computation of the matrix W and its inverse M when $T = \{1, 2\}$ shows that $M_2(\lambda) := M(\nu_2, \lambda) = -1/8,\ 1/2,\ -1/8$ when $\lambda = 2\delta_1,\ \delta_1 + \delta_2,\ 2\delta_2$, respectively. So, we have

$$\psi_{(X_1,X_2)} = -\frac{1}{2}\delta_{\delta_{X_1}} - \frac{1}{2}\delta_{\delta_{X_2}} + 2\delta_{(\delta_{X_1}+\delta_{X_2})/2}. \tag{12}$$

Hence,

$$\xi = 2\mathbb{P}((\delta_{X_1} + \delta_{X_2})/2 \in \cdot) - \mathbb{P}(\delta_{X_1} \in \cdot). \tag{13}$$

We can also use the formula (9) for ψ_X, taking $T = T_X$. Let $d(X)$ be the cardinality of the set $\{X_1, \dots, X_n\}$. On the event $\{d(X) = d\}$, for some $d \in \{1, \dots, n\}$, the variable λ in the summation in (9) ranges over a set of cardinality $\binom{n+d-1}{n}$.

For example, let again $n = 2$, assume that the event $\{d(X) = 2\} = \{X_1 \neq X_2\}$ is measurable and let $p := \mathbb{P}(d(X) = 2)$. On the event $\{d(X) = 1\}$, we have $T_X = \{X_1\}$ and so $\mathscr{N}_2(T_X) = \{2\delta_{X_1}\}$. Hence $\psi_X = \delta_{\delta_{X_1}}$. On the event $\{d(X) = 2\}$, we obtain (12). This yields the formula, obviously equivalent to (13),

$$\begin{aligned}\xi = (1-p)\mathbb{P}(\delta_{X_1} \in \cdot \mid d(X) = 1) - p\,\mathbb{P}(\delta_{X_1} \in \cdot \mid d(X) = 2)& \\ + 2p\mathbb{P}((\delta_{X_1} + \delta_{X_2})/2 \in \cdot \mid d(X) = 2).&\end{aligned}$$

4.4. Moment functional

The k-th moment functional of a mixing signed measure ξ is defined by

$$C_k(B_1, \dots, B_k) := \int_{\mathscr{P}(S)} \pi(B_1) \cdots \pi(B_k)\, \xi(\mathrm{d}\pi).$$

If $k \le n$, then, from (1),

$$C_k(B_1, \dots, B_k) = \mathbb{P}(X \in B_1 \times \cdots \times B_k).$$

This means that any mixing measure ξ will have the same C_k for all $k \le n$. But if $k > n$, then $C_k(B_1, \dots, B_k)$ may be negative and will depend on the choice of ξ. For the canonical ξ, we have, using (6),

$$\begin{aligned}C_k(B_1, \dots, B_k) &= \mathbb{E} \sum_{\lambda \in \mathscr{N}_n(T_X)} n^n M(\varepsilon_X, \lambda)(\lambda/n)(B_1) \cdots (\lambda/n)(B_k) \\ &= \mathbb{E} \sum_{\lambda \in \mathscr{N}_n(T_X)} M(\varepsilon_X, \lambda)\lambda(B_1) \cdots \lambda(B_k).\end{aligned}$$

4.5. Laplace functional

Define next the Laplace functional of the canonical mixing measure ξ by

$$\Lambda(f) := \xi\Big[\exp\Big(-\int_S f(a)\pi(\mathrm{d}a)\Big)\Big] = \int_{\mathscr{P}(S)} \mathrm{e}^{-\int_S f(a)\pi(\mathrm{d}a)}\, \xi(\mathrm{d}\pi),$$

for $f : S \to \mathbb{R}_+$ measurable. We obtain

Please cite this article in press as: S. Janson et al., On a representation theorem for finitely exchangeable random vectors, J. Math. Anal. Appl. (2016), http://dx.doi.org/10.1016/j.jmaa.2016.04.070

$$\begin{aligned}\Lambda(f) &= \mathbb{E}\int_{\mathscr{P}(S)} \mathrm{e}^{-\int_S f(a)\pi(\mathrm{d}a)}\,\psi_X(\mathrm{d}\pi) = \mathbb{E}\sum_{\lambda\in\mathscr{N}_n(n)} n^n M_n(\lambda)\,\mathrm{e}^{-\frac{1}{n}\int_S f\mathrm{d}\varphi_x(\lambda/n)} \\ &= \mathbb{E}\sum_{\lambda\in\mathscr{N}_n(n)} n^n M_n(\lambda)\,\mathrm{e}^{-\frac{1}{n}\sum_{i=1}^n \lambda_i f(x_i)}.\end{aligned} \tag{14}$$

For example, if $n = 2$, by the values of $M_2(\lambda)$ in Section 4.3 and symmetry,

$$\Lambda(f) = 2\mathbb{E}\left[\mathrm{e}^{-f(X_1)/2}\mathrm{e}^{-f(X_2)/2}\right] - \mathbb{E}\mathrm{e}^{-f(X_1)}.$$

4.6. Extendibility

It is easy to see that an n-exchangeable random vectors may not be extendible to an N-exchangeable random vector (see Remark 1). Here are two easy examples.

Example 3. As in Example 2, with $S = \{1, 2\}$, the random variable (X_1, X_2) taking values in S^2, such that $\mathbb{P}(X = (1, 2)) = \mathbb{P}(X = (2, 1)) = 1/2$, cannot be extended to an exchangeable random variable (X_1, X_2, X_3) with values in S^3.

Example 4. Let (X_1, X_2) be a Gaussian vector with $\mathbb{E}X_1 = \mathbb{E}X_2 = 0$, $\mathbb{E}X_1^2 = \mathbb{E}X_2^2 = 1 + \varepsilon$, $\varepsilon \in (0, 1)$, $\mathbb{E}X_1X_2 = -1$. If this were extendible to an exchangeable vector (X_1, X_2, X_3), then we would have had $\mathbb{E}X_1X_3 = \mathbb{E}X_2X_3 = -1$. An easy calculation shows that the matrix $\begin{pmatrix} 1+\varepsilon & -1 & -1 \\ -1 & 1+\varepsilon & -1 \\ -1 & -1 & 1+\varepsilon \end{pmatrix}$ is not positive definite when $\varepsilon < 1$, and thus fails to be the covariance matrix of (X_1, X_2, X_3).

In [10] we give a necessary and sufficient condition for extendibility.

4.7. Some applications

Consider the following statement.

Lemma 1. *Let $(S, \mathscr{S})$ be a measurable space, n a positive integer, and $f : S^n \to \mathbb{R}$ a bounded measurable function such that $\int_{S^n} f(x_1, \dots, x_n)P(\mathrm{d}x_1)\cdots P(\mathrm{d}x_n) = 0$ for any probability measure P on $(S, \mathscr{S})$. Then $\int_{S^n} f\mathrm{d}Q = 0$ for any exchangeable probability measure Q on S^n.*

Although this can be proven by other methods, it follows immediately from Theorem 1.

For a more practical application, we refer to the paper of Kerns and Székely [9] for an application of Theorem 1 to the Bayesian consistency problem. In situations where one has a fixed number n of unordered samples, one can refer to de Finetti's theorem in order to prove consistency of standard Bayesian estimators. The theorem assumes that the samples come from random vectors that are infinitely extendible (otherwise, de Finetti's theorem does not hold). As pointed out in [9], the result of Theorem 1 still allows proving Bayesian consistency.

For a practical application of the representation result to the Bayesian properties of normalized maximum likelihood, see Barron, Ross and Watanabe [1].

4.8. Open problems

Estimate the size (in terms of total variation) of the signed measure ξ in the representation (1). How does this behave as a function of the dimension n? What is the best bound?

While this paper deals with a probability measure on S^n that is invariant under all $n!$ permutations of coordinates, it is natural to ask if there is a representation result for measures that are invariant under a subgroup of the symmetric group.

Acknowledgments

We thank Jay Kerns and Gábor Székely for their comments and for pointing reference [11] to us and an anonymous referee for helpful comments and for reference [1].

Appendix A. Invertibility of the multinomial Dyson matrix

Let T be a finite set, say $T = \{1, \dots, d\}$. With the notation established in the introduction,

$$W(\lambda, \nu) := \binom{n}{\nu} \lambda^\nu, \quad \lambda, \nu \in \mathscr{N}_n(T). \tag{A.1}$$

The matrix $[n^{-n} W(\lambda, \nu)]$ on $\mathscr{N}_n(T)$ is referred to as the multinomial Dyson matrix [13]. In fact, $n^{-n} W(\lambda, \nu)$ is the 1-step transition probability of a multitype Wright–Fisher Markov chain with state space $\mathscr{N}_n(T)$. This chain is defined as follows (see, e.g., [2]). There is a population of always constant size n. Individuals in this population are of different types; the set of types is T. Given the vector $\lambda = (\lambda_1, \dots, \lambda_d)$ of type counts of the population currently, select an individual at random and copy its type; do this selection n times, independently. Then the probability that the vector of type counts changes from λ to ν is exactly equal to $n^{-n} W(\lambda, \nu)$. Shelton et al. [13] show that the matrix W is invertible, i.e., that there is a matrix M on $\mathscr{N}_n(T)$ such that

$$\sum_{\lambda \in \mathscr{N}_n(T)} M(\nu, \lambda)\, W(\lambda, \nu') = \mathbf{1}_{\nu = \nu'}, \quad \nu, \nu' \in \mathscr{N}_n(T). \tag{A.2}$$

The inverse matrix M can be expressed explicitly in terms of sums involving binomial coefficients and signed Stirling numbers of the first kind [11, eq. (25)]. If $n = d$ and $\nu = \sum_{i=1}^n \delta_i$, then $M(\nu, \lambda)$ is explicitly known for all $\lambda \in \mathscr{N}_n(T)$ [12, Theorem 4.1].

For a direct proof of the invertibility of W that avoids explicit computations we proceed as follows. The columns of W are linearly independent if and only if the only numbers $c(\lambda), \lambda \in \mathscr{N}_n(T)$, for which

$$\sum_\lambda c(\lambda) \binom{n}{\nu} \lambda^\nu = 0, \quad \text{for all } \nu \in \mathscr{N}_n(T)$$

are zero. But the last display is equivalent to

$$0 = \sum_\nu x^\nu \sum_\lambda c(\lambda) \binom{n}{\nu} \lambda^\nu = \sum_\lambda c(\lambda)(\lambda_1 x_1 + \cdots + \lambda_d x_d)^n, \quad \text{for all } x = (x_1, \dots, x_d) \in \mathbb{R}^d.$$

Invertibility of W thus follows from:

Theorem 3. *Let d, n be positive integers. Let $\mathscr{N}_n(d)$ be the set of all $\lambda = (\lambda_1, \dots, \lambda_d)$ where the λ_i are nonnegative integers such that $\sum_{i=1}^d \lambda_i = n$. Then the polynomials*

$$p_\lambda(x) := (\lambda_1 x_1 + \cdots + \lambda_d x_d)^n, \quad \lambda \in \mathscr{N}_n(d),$$

are linearly independent and form a basis for the space $\mathscr{P}_n(d)$ of homogeneous polynomials of degree n in $x_1, \dots, x_d$.

Please cite this article in press as: S. Janson et al., On a representation theorem for finitely exchangeable random vectors, J. Math. Anal. Appl. (2016), http://dx.doi.org/10.1016/j.jmaa.2016.04.070

Proof. Note that $\{x^\lambda = x_1^{\lambda_1} \cdots x_d^{\lambda_d} : \lambda \in \mathscr{N}_n(d)\}$ is a basis in $\mathcal{P}_n(d)$. Let $\mathcal{Q}_n(d)$ be the linear subspace of $\mathcal{P}_n(d)$ spanned by $\{p_\lambda, \lambda \in \mathscr{N}_n(d)\}$. We will show that $\mathcal{Q}_n(d) = \mathcal{P}_n(d)$. The substitution $x_i \mapsto x_i + x_d$, $i = 1, \dots, d-1$, shows that the $\mathcal{Q}_n(d)$ is isomorphic to the subspace $\widehat{\mathcal{Q}}_n(d) \subset \mathcal{P}_n(d)$ spanned by the polynomials

$$\widehat{p}_\lambda(x_1, \dots, x_d) := (\lambda_1 x_1 + \cdots + \lambda_{d-1} x_{d-1} + n x_d)^n, \quad \lambda \in \mathscr{N}_n(d).$$

For $0 \le j \le n$, denote by $\mathcal{P}_{n,j}(d)$ the subspace of $\mathcal{P}_n(d)$ consisting of polynomials that have degree in x_d at most j. Note that $\mathcal{P}_{n,0}(d) = \mathcal{P}_n(d-1)$ and $\mathcal{P}_{n,n}(d) = \mathcal{P}_n(d)$. Let Δ_h be the difference operator acting on functions $f(t)$ of one real variable t by $\Delta_h f(t) := f(t+h) - f(t)$. It is easy to see that, for all integers $k \ge 1$,

$$\Delta_{h_1} \cdots \Delta_{h_k} f(t) = \sum_{r=0}^{k} (-1)^r \sum_{\substack{I \subset \{1,\dots,k\} \\ |I|=r}} f\big(t + \sum_{\alpha \in I} h_\alpha\big). \tag{A.3}$$

Using (A.3) with $f(t) = t^n$ and induction on k we easily obtain

$$\Delta_{h_1} \cdots \Delta_{h_k} \{t^n\} = (n)_k \, h_1 \cdots h_k t^{n-k} + r_{h_1,\dots,h_k}(t), \quad k = 1, \dots, n, \tag{A.4}$$

where $t \mapsto r_{h_1,\dots,h_k}(t)$ is a polynomial of degree $\le n-k-1$, whereas $(h_1, \dots, h_k, t) \mapsto r_{h_1,\dots,h_k}(t)$ is a homogeneous polynomial of degree n. The meaning of (A.4) for $k = n$ is that

$$\Delta_{h_1} \cdots \Delta_{h_n} \{t^n\} = n! \, h_1 \cdots h_n. \tag{A.5}$$

Let $(i_1, \dots, i_k)$ be a sequence with values in $\{1, \dots, d-1\}$. Using (A.3) with $f(t) = t^n$ and then setting $t = n x_d$ and $h_1 = x_{i_1}, \dots, h_k = x_{i_k}$ we obtain

$$\Delta_{x_{i_1}} \cdots \Delta_{x_{i_k}} \{t^n\}\big|_{t=nx_d} = \sum_{r=0}^{k} (-1)^r \sum_{\substack{I \subset \{1,\dots,k\} \\ |I|=r}} (n x_d + \sum_{\alpha \in I} x_{i_\alpha})^n, \quad 1 \le k \le n.$$

For $I \subset \{1, \dots, k\}$, we have $(n x_d + \sum_{\alpha \in I} x_{i_\alpha})^n = \widehat{p}_\lambda(x_1, \dots, x_d)$ where λ_j is the number of terms of the sequence $(i_1, \dots, i_k)$ that are equal to j, $j = 1, \dots, d$. This implies that (recall that $\widehat{\mathcal{Q}}_n(d)$ is spanned by $\{\widehat{p}_\lambda, \lambda \in \mathscr{N}_n(d)\}$) the function $(x_1, \dots, x_d) \mapsto \Delta_{x_{i_1}} \cdots \Delta_{x_{i_k}} \{t^n\}\big|_{t=nx_d}$ is a polynomial in d variables that belongs to $\widehat{\mathcal{Q}}_n(d)$. Using this observation in (A.4) and (A.5) we obtain

$$\begin{aligned} n^{n-k} \, (n)_k \, x_{i_1} \cdots x_{i_k} x_d^{n-k} &= \Delta_{x_{i_1}} \cdots \Delta_{x_{i_k}} \{t^n\}\big|_{t=nx_d} - r_{x_{i_1},\dots,x_{i_k}}(n x_d) \\ &\in \widehat{\mathcal{Q}}_n(d) + \mathcal{P}_{n,n-k-1}(d), \qquad 1 \le k \le n-1, \end{aligned} \tag{A.6}$$

$$n! \, x_{i_1} \dots x_{i_n} = \Delta_{x_{i_1}} \cdots \Delta_{x_{i_n}} \{t\}\big|_{t=nx_d} \in \widehat{\mathcal{Q}}_n(d). \tag{A.7}$$

Since every polynomial on $\mathcal{P}_{n,n-k}(d)$ is a linear combination of the monomials appearing in the left-hand side of (A.6), (A.6) implies

$$\mathcal{P}_{n,n-k}(d) \subset \widehat{\mathcal{Q}}_n(d) + \mathcal{P}_{n,n-k-1}(d), \quad 1 \le k \le n-1.$$

Similarly, every polynomial on $\mathcal{P}_{n,0}(d)$ is a linear combination of monomials as in the left-hand side of (A.7), and thus (A.7) implies

$$\mathcal{P}_{n,0}(d) \subset \widehat{\mathcal{Q}}_n(d).$$

The last two displays imply that $\mathcal{P}_{n,n-1}(d) \subset \widehat{\mathcal{Q}}_n(d)$. Since every polynomial in $\mathcal{P}_n(d)$ is a linear combination of the monomial x_d^n and a polynomial in $\mathcal{P}_{n,n-1}(d)$, it follows that $\mathcal{P}_n(d) \subset \widehat{\mathcal{Q}}_n(d)$ and so the proof is completed. □

References

[1] Andrew Barron, Teemu Roos, Kazuho Watanabe, Bayesian properties of normalized maximum likelihood and its fast computation, in: Proc. IEEE International Symposium on Information Theory, ISIT-2014, IEEE Press, 2014, pp. 1667–1671.

[2] Donald Dawson, Introductory Lectures on Stochastic Population Systems, Tech. Report of the Lab. for Research in Statistics and Probability No. 451, 2010.

[3] Persi Diaconis, Finite forms of de Finetti's theorem on exchangeability. Foundations of probability and statistics, II, Synthese 36 (1977) 271–281.

[4] Persi Diaconis, David Freedman, Finite exchangeable sequences, Ann. Probab. 8 (1980) 745–764.

[5] Lester E. Dubins, David A. Freedman, Exchangeable processes need not be mixtures of independent, identically distributed random variables, Z. Wahrsch. Verw. Gebiete 48 (1979) 115–132.

[6] E.T. Jaynes, Some applications and extensions of the de Finetti representation theorem, in: Bayesian Inference and Decision Techniques, 1986, p. 31.

[7] Olav Kallenberg, Foundations of Modern Probability, Springer, New York, 2002.

[8] Olav Kallenberg, Probabilistic Symmetries and Invariance Principles, Springer, 2005.

[9] G. Jay Kerns, Gábor J. Székely, De Finetti's theorem for abstract finite exchangeable sequences, J. Theoret. Probab. 19 (2006) 589–608.

[10] Takis Konstantopoulos, Linglong Yuan, On the extendibility of finitely exchangeable probability measures, arXiv:1501.06188 [math.PR].

[11] Daniel S. Moak, Combinatorial multinomial matrices and multinomial Stirling numbers, Proc. Amer. Math. Soc. 108 (1990) 1–8.

[12] Daniel Moak, Konrad J. Heuvers, K.P.S. Bhaskara Rao, Karen Collins, An inversion relation of multinomial type, Discrete Math. 1 (31(1)) (1986) 195–204.

[13] Robert Shelton, Konrad J. Heuvers, Daniel Moak, K.P.S. Bhaskara Rao, Multinomial matrices, Discrete Math. 61 (1) (1986) 107–114.

Please cite this article in press as: S. Janson et al., On a representation theorem for finitely exchangeable random vectors, J. Math. Anal. Appl. (2016), http://dx.doi.org/10.1016/j.jmaa.2016.04.070