Open Science Framework DOI:10.31219/osf.io/***

A formula for partial and conditional infinite exchangeability

P.G.L. Porta Mana
Kavli Institute, Trondheim <pgl@portamana.org>

14 May 2020; updated 1 June 2020

[draft] A formula is given for conditionally, infinitely exchangeable probability distributions.

1 Full, partial, and conditional exchangeability

De Finetti's theorem for the representation of infinitely, fully exchangeable probability distributions yields one of the formulae, derived from the probability calculus, with the richest practical and philosophical consequences. It belongs to family of theorems whose members are still under exploration. Its closest relative is the theorem for infinitely, *partially* exchangeable probability distributions.

The present note has two purposes. The first is to show an alternative representation of the theorem for partial exchangeability. This representation emphasizes the role of *conditional* exchangeability and of the conditional character of the limit distributions that appear in the usual representation.

The second purpose is to give a representation formula for distributions that satisfy a combination of conditional and full exchangeability for the marginals. I find the representation interesting because it expresses the combination of exchangeability condition as the *factorization* of the density that appears in de Finetti's representation.

The next section explains some notation and gives a reminder of the theorem for full exchangeability. The subsequent two sections show the alternative representation for partial exchangeability and the representation for conditional & marginal exchangeability. The final section offer some thoughts about these representations.

For the details about exchangeable distributions I refer to Bernardo & Smith[1] and Diaconis & Freedman[2].

[1] Bernardo & Smith 2000 §§ 4.3, 4.6. [2] Diaconis & Freedman 1980a,b.

This document is designed for screen reading and two-up printing on A4 or Letter paper

2 Notation and summary of representation for full exchangeability

Most of this section can be skimmed through by readers familiar with the exchangeability theorems, to grasp the notation I use.

Our domain of discourse consists of a countably infinite set of atomic statements (in the logical sense)

$$\big\{X_i = x_i \,\big|\, i \in \mathbf{N},\ \forall i\ x_i \in \mathfrak{X}\big\} \tag{1}$$

where $\mathfrak{X}$ is a finite set. For each i the statements $\{X_i = x \mid x \in \mathfrak{X}\}$ are assumed mutually exclusive on information I. (The theorem also holds for statements not involving equalities.)

A probability distribution over these atomic statements is called fully (infinitely) exchangeable if

for every N, every set $\{i_1, \dots, i_N\} \subset \mathbf{N}$, and every permutation π thereof,

$$\begin{aligned} &\mathrm{P}(X_{i_1} = x_{i_1}, X_{i_2} = x_{i_2}, \dots, X_{i_N} = x_{i_N} \mid I) = \\ &\qquad \mathrm{P}(X_{i_1} = x_{\pi(i_1)}, X_{i_2} = x_{\pi(i_2)}, \dots, X_{i_N} = x_{\pi(i_N)} \mid I), \end{aligned} \tag{2}$$

and if all such probabilities are consistently related by marginalization. This property is equivalent to declaring the empirical frequencies of the values x to be sufficient statistics. The commas between statements denote logical conjunction ('$\land$'), so the order of the statements is immaterial.

In the following I let '$\{1, 2, \dots, N\}$' denote any subset of $\mathbf{N}$, to avoid a proliferation of subscripts. They should be read as '$\{i_1, i_2, \dots, i_N\}$'.

Denote by $\boldsymbol{f_x} \coloneqq (f_x)$ a normalized distribution over the values $x \in \mathfrak{X}$. The set of all such distributions is a simplex of dimension $|\mathfrak{X}| - 1$.

For each $x \in \mathfrak{X}$, denote by F_x the empirical relative frequency of x in the set $\{x_1, \dots, x_N\}$:

$$NF_x \coloneqq \textstyle\sum_i \delta(x, x_i), \quad x \in \mathfrak{X}\,. \tag{3}$$

De Finetti's theorem states that a fully exchangeable distribution can be written as follows:

$$\begin{aligned} &\mathrm{P}(X_1 = x_1,\ \dots,\ X_N = x_N \mid I) = \\ &\qquad \int \prod_i f_{x_i}\ \mathrm{p}(\boldsymbol{f_x} \mid I)\, \mathrm{d}\boldsymbol{f_x} \equiv \int \prod_x f_x{}^{NF_x}\ \mathrm{p}(\boldsymbol{f_x} \mid I)\, \mathrm{d}\boldsymbol{f_x}\,, \end{aligned} \tag{4}$$

where the integral is over the simplex of distributions $\{\boldsymbol{f_x}\}$.

In the first integral form, the product is over the set of instances $1, \dots, N$. In the second, equivalent integral form, the product is over the set of values x. This form shows that the empirical frequency distribution (F_x) is a sufficient statistic; it also hint at the important role played in the theorem by the relative entropy of (F_x) with respect to (f_x).

For enough large N, the probability of observing an empirical frequency distribution $\boldsymbol{F_x}$ within a small volume v centred around the distribution $\boldsymbol{f_x}$ is approximately given by the density $\mathrm{p}(\boldsymbol{f_x} \mid I)\,\mathrm{d}\boldsymbol{f_x}$:

$$\mathrm{P}(\boldsymbol{F_x} \in v \mid N \text{ large}, I) \approx \mathrm{p}(\boldsymbol{f_x} \mid I)\, v \;. \tag{5}$$

For this reason the parameter $\boldsymbol{f_x}$ can be interpreted as a long-run frequency distribution. I will therefore call it so sometimes, but without the intention to force such interpretation on you.

3 Partial exchangeability: alternative form

In the theorem for partially exchangeable distributions, the set $\{X_i\}$ is divided into two or more categories represented by subsets $\{Y_j\}, \{Z_k\}, \dots$. Partial exchangeability of the distribution means that permutations are allowed within each subset but not necessarily across subsets. The usual representation in this case, after a suitable re-indexing $\{1, 2, \dots\} \mapsto \{1', 2', \dots, 1'', 2'', \dots\}$, has the form

$$\begin{aligned}\mathrm{P}(Y_{1'} = y_{1'},\ Y_{2'} = y_{2'},\ \dots,\ Z_{1''} = z_{1''},\ Z_{2''} = z_{2''},\ \dots \mid J) = \\ \iint \prod_j g_{y_j} \prod_k h_{z_k}\, \mathrm{p}(\boldsymbol{g}, \boldsymbol{h} \mid J)\, \mathrm{d}\boldsymbol{g}\, \mathrm{d}\boldsymbol{h} \;,\end{aligned} \tag{6}$$

with distinct normalized distributions $\boldsymbol{g}$, $\boldsymbol{h}$ for each category. If the density $\mathrm{p}(\boldsymbol{g}, \boldsymbol{h} \mid J)\, \mathrm{d}\boldsymbol{g}\, \mathrm{d}\boldsymbol{h}$ is diagonal, that is, if it contains a term $\delta(\boldsymbol{g} - \boldsymbol{h})$, the fully exchangeable form (4) is recovered.

With a little reflection we realize that if we know the quantities X to belong to category Y in instances $1', 2', \dots$, and to category Z in instances $1'', 2'', \dots$, then (a) there is some other quantity C that allows us to distinguish the two categories, and (b) the values of this quantity *are known* for all instances.

Let us say, for example, that the quantities X_i are the results of animal treatments, with values 'S'uccess and 'F'ailure. Y refers to the results for treatments on Yaks, and Z on Zebras. If we write

$$\mathrm{P}(Y_3 = \mathsf{S},\ Z_5 = \mathsf{F} \,|\, J) = 0.2 \ ,$$

then we must already know that animal number 3 is a yak, $C_3 = \mathsf{Y}$, and animal number 5 is a zebra, $C_5 = \mathsf{Z}$. This is clear from our very notation, otherwise we would not have known whether to use the symbol Y or Z for those instances. This information is evidently implicit in our background information J.

We now make the dependence upon the category information explicit. We thus obtain a slightly different definition of partial exchangeability and a slightly different form of its representation theorem.

Besides the statements $\{X_i = x_i\}$, we introduce an additional, similar set of atomic statements

$$\left\{C_i = c_i \,\middle|\, i \in \mathbf{N},\ \forall i\ c_i \in \mathfrak{C}\right\} . \tag{7}$$

For each i the statements $\{C_i = c \,|\, c \in \mathfrak{C}\}$ are mutually exclusive on information I.

These statements allow us to identify each instance $1, 2, \dots$ as belonging to one or another category out of the finite set $\mathfrak{C}$.

A probability distribution over the $X_i = x_i$ atomic statements is called partially exchangeable if

for every N, every set of indices $\{1, \dots, N\} \subset \mathbf{N}$, and every permutation π thereof such that $\pi(i) = j \ \Rightarrow\ c_i = c_j$,

$$\begin{aligned}&\mathrm{P}(X_1 = x_1, \dots, X_N = x_N \,|\, C_1 = c_1, \dots, C_N = c_N,\ I) = \\ &\qquad\qquad \mathrm{P}(X_{i_1} = x_{\pi(1)}, \dots, X_{i_N} = x_{\pi(N)} \,|\, C_1 = c_1, \dots, C_N = c_N,\ I) \ .\end{aligned} \tag{8}$$

that is, the only allowed permutations are those *which exchange indices having the same c value*.

Now let us rewrite the representation formula accordingly.

For each category $c \in \mathfrak{C}$, introduce a normalized distribution $\{f_{x|c} \,|\, x \in \mathfrak{X}\}$ over the values x. As the notation suggests, it can be considered as a *conditional* distribution over x given c.

Denote (with some abuse of the symbols) by $\boldsymbol{f_{x|c}} \coloneqq (f_{x|c})$ the set of all such conditional distributions. This set is the Cartesian product of $|\mathfrak{C}|$ simplices, each of dimension $|\mathfrak{X}| - 1$.

Denote by $F_{x,c}$ the empirical relative *joint* frequency of the pair of values (x, c) occurring in the set of pairs $\{x_1, c_1), \dots, (x_N, c_N)\}$:

$$NF_{x,c} \coloneqq \textstyle\sum_i \delta(x, x_i)\, \delta(c, c_i), \quad x \in \mathfrak{X},\ c \in \mathfrak{C}\,. \tag{9}$$

Thus $NF_{x,c}$ is the total number of times value x appears among the indices with $c_i = c$.

De Finetti's theorem states that the partially exchangeable distribution (8) can be written as follows:

$$\begin{aligned}\mathrm{P}(X_1 = x_1, \dots, X_N = x_N \mid C_1 = c_1, \dots, C_N = c_N,\ I) = \\ \int \prod_{c,x} f_{x|c}{}^{NF_{x,c}}\ \mathrm{p}\bigl(\boldsymbol{f}_{x|c} \bigm| I\bigr)\, \mathrm{d}\boldsymbol{f}_{x|c}\,.\end{aligned} \tag{10}$$

Scrutiny of this formula shows that this form is equivalent to the more familiar representation. The integral contains one product of $f_{\dots|c}$ terms for every category c. In each such product, $f_{x_i|c}$ terms are multiplied together for those i such that $c_i = c$. There are exactly $NF_{x,c}$ such terms.

The alternative formulation (8) of partial exchangeability shows that this symmetry could also be called 'conditional' exchangeability instead. The role of conditional distributions is clear in the representation (10).

4 Representation for joint distributions with conditional exchangeability symmetries

Suppose that you would assign a partially or conditionally exchangeable distribution of probability to the statements $\{X_i = x_i\}$, if you knew the true $\{C_i = c_i\}$. But you do not know the latter. What kind of properties does the joint probability distribution of these statements have? And the marginal distribution for $\{X_i = x_i\}$?

The joint probability distribution can be rewritten

$$\begin{aligned}\mathrm{P}(X_1 = x_1, C_1 = c_1,\ \dots,\ X_N = x_N, C_N = c_N \mid I) = \\ \mathrm{P}(X_1 = x_1, \dots, X_N = x_N \mid C_1 = c_1, \dots, C_N = c_N,\ I) \times \\ \mathrm{P}(C_1 = c_1, \dots, C_N = c_N \mid I)\,,\end{aligned} \tag{11}$$

where the first factor, partially or conditionally exchangeable, can be represented by the integral of eq. (10).

Let us suppose that our uncertainty about the statements $\{C_i = c_i\}$ is expressed by a fully exchangeable marginal probability distribution. An integral representation analogous to (4) then holds:

$$\mathrm{P}(C_1 = c_1,\ \dots,\ C_N = c_N \mid I) = \int \prod_c f_{,c}{}^{N F_{,c}}\ \mathrm{p}(\boldsymbol{f_c} \mid I)\,\mathrm{d}\boldsymbol{f_c}\ , \tag{12}$$

where $f_{,c} \coloneqq \sum_x f_{x,c}$ and $F_{,c} \coloneqq \sum_x F_{x,c}$ are marginal distributions.

We can now replace the integral representations (10) and (12) into (11). The products within their integrals can be combined considering that

$$f_{,c}{}^{N F_{,c}} = f_{,c}{}^{N \sum_x F_{x,c}} = \prod_x f_{,c}{}^{N F_{x,c}}\ . \tag{13}$$

We obtain

$$\mathrm{P}(X_1 = x_1, C_1 = c_1,\ \dots,\ X_N = x_N, C_N = c_N \mid I) = \\ \int \prod_{c,x} f_{x,c}{}^{N F_{x,c}}\ \mathrm{p}(\boldsymbol{f_{xc}} \mid I)\,\mathrm{d}\boldsymbol{f_{xc}} \tag{14a}$$

$$\text{with}\qquad \boxed{\mathrm{p}(\boldsymbol{f_{xc}} \mid I)\,\mathrm{d}\boldsymbol{f_{xc}} = \mathrm{p}(\boldsymbol{f_{x|c}} \mid I)\,\mathrm{p}(\boldsymbol{f_c} \mid I)\,\mathrm{d}\boldsymbol{f_{x|c}}\,\mathrm{d}\boldsymbol{f_c}} \tag{14b}$$

The last equality comes from the one-one correspondence between the variables $(\boldsymbol{f_{x|c}}, \boldsymbol{f_c})$ and $\boldsymbol{f_{xc}}$, so that the product of density functions for $\boldsymbol{f_{x|c}}$ and $\boldsymbol{f_c}$ is just a specific case of a density function for $\boldsymbol{f_{xc}}$, apart from a Jacobian factor.

The integral expression (14) is the representation of a fully exchangeable distribution. Thus the joint distribution for the set of *pairs* of statements $\{(X_i = x_i, C_i = c_i)\}$ is fully exchangeable.

The noteworthy feature of the integral expression (14) is that *the density for the joint distribution $\boldsymbol{f_{xc}}$ is factorizable into the product of a density for the conditional distribution $\boldsymbol{f_{x|c}}$ and a density for the marginal distribution $\boldsymbol{f_c}$*. This factorization expresses the partial or conditional exchangeability for the statements $\{X_i = x_i\}$ given the $\{C_i = c_i\}$.

It is easy to show that the reverse also holds: if the density of an integral representation is factorizable as in (14b), then the corresponding probability distributions enjoy a symmetry of partial or conditional exchangeability.

The factorization (14b) is not trivial. With a change of variables the following identities hold:

$$\mathrm{p}(\boldsymbol{f}_{xc} \,|\, I)\, \mathrm{d}\boldsymbol{f}_{xc} \equiv \mathrm{p}[(\boldsymbol{f}_{x|c}, \boldsymbol{f}_{c}) \,|\, I]\, \mathrm{d}\boldsymbol{f}_{x|c}\, \mathrm{d}\boldsymbol{f}_{c} \equiv \\ \mathrm{p}(\boldsymbol{f}_{x|c} \,|\, \boldsymbol{f}_{c}, I)\, \mathrm{p}(\boldsymbol{f}_{c} \,|\, I)\, \mathrm{d}\boldsymbol{f}_{x|c}\, \mathrm{d}\boldsymbol{f}_{c} \,. \quad (15)$$

The factorization condition is thus equivalent to conditional independence:

$$\mathrm{p}(\boldsymbol{f}_{x|c} \,|\, \boldsymbol{f}_{c}, I)\, \mathrm{d}\boldsymbol{f}_{x|c} = \mathrm{p}(\boldsymbol{f}_{x|c} \,|\, I)\, \mathrm{d}\boldsymbol{f}_{x|c} \,. \quad (16)$$

5 Discussion and possible connections with Bayesian networks

The result of the previous section is thus summarized: Given infinitely countable sets of statements $\{X_i = x_i\}$ and $\{C_i = c_i\}$, and assuming that

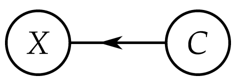
1. the marginal probability distribution for the C statements is fully exchangeable,

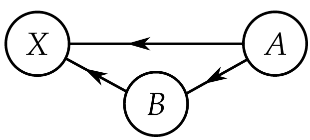
2. the probability distribution for the X statements is partially (or conditionally) exchangeable given the C,

Then the joint distribution for both sets is fully exchangeable, and the density within its integral representation *factorizes* into a density for a conditional long-run frequency distribution, and a density for a marginal long-run frequency distribution, eq. (14b).

This factorization can be expressed in the guise of a Bayesian network,

Using the reasoning and integral representations of § 4 it is possible to generalize this result to more





Bibliography

('de *X*' is listed under D, 'van *X*' under V, and so on, regardless of national conventions.)

Bernardo, J.-M., Smith, A. F. (2000): *Bayesian Theory*, repr. (Wiley, New York). First publ. 1994.

Diaconis, P., Freedman, D. (1980a): *Finite exchangeable sequences*. Ann. Prob. **8**^{4}, 745–764.

— (1980b): *De Finetti's generalizations of exchangeability*. In: Jeffrey (1980), 233–249.

Jeffrey, R. C., ed. (1980): *Studies in inductive logic and probability. Vol. II*. (University of California Press, Berkeley).