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 2 June 2020

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

0 Introduction and notation

De Finetti's theorem for the representation of infinitely exchangeable probability distributions yields some of the formulae, derivable from the probability calculus, with the richest practical and philosophical consequences.

In this work I present three results; each may interest a different audience.

The first result, derived in § **??**, is a reformulation of partial exchangeability and of its representation theorem in a slightly unfamiliar form. Though not remarkable, this reformulation gives some insights into the connection between partial and conditional exchangeability, and their connection to regression.

The second result, derived is § **??**, is an integral representation for joint (predictive) probability distributions with particular symmetries: some of the conditional distributions obtained from them satisfy partial exchangeability. This integral representation has the usual de Finetti form and, remarkably, its density must factorize in a specific way. This factorization is implied by, and implies, the conditional-exchangeability symmetries.

The third result, § **??**, brings together exchangeability and Bayesian networks. A Bayesian network expresses assumptions of conditional dependence and independence. Exchangeability expresses assumptions about the mutual informational relevance of a set of phenomena or experiments, deemed to be 'similar'. When we combine these two kinds

of assumptions we obtain (predictive) probability distributions with a particular integral representation: a mixture of copies of the same Bayesian network. This result has some analogies with de Finetti's usual representation, which is a mixture of independent probability distributions ('i.i.d.').

Each of these three results builds on the preceding one(s). The final section discusses them further and hints at possible applications for them.

The remainder of this section introduces some notation and summarizes de Finetti's theorem for full exchangeability. It can be skimmed through by readers familiar with exchangeability theorems, just to grasp the notation I use.

0.1 Notation and summary of representation for full exchangeability

For the details about exchangeable distributions I refer to Bernardo & Smith[1], Diaconis & Freedman[2], and Dawid's[3] review.

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

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

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 holds for any set of statements with these properties, even if the statements are not of the form '$X = x$'.)

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,

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

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 ('$\wedge$'), so the order of the statements is immaterial.

[1] Bernardo & Smith 2000 §§ 4.3, 4.6. [2] Diaconis & Freedman 1980a,b. [3] Dawid 2013.

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:

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

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[4]. I will therefore call it so sometimes, but without the intention to force such interpretation on you.

1 Partial exchangeability: alternative form

In de Finetti's theorem for partially exchangeable distributions, the set $\{X_i = x_i\}$ is divided into two or more categories represented by subsets $\{Y_j = y_j\}, \{Z_k = z_k\}, \dots$. Partial exchangeability of the distribution means

[4] 'But this *long run* is a misleading guide to current affairs. *In the long run* we are all dead' (Keynes 2013 § 3.I, p. 65).

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

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

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 (**??**) 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

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

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

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

For each i the statements $\{C_i = c \mid c \in \mathfrak{C}\}$ 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 $\pi(i) = j \;\Rightarrow\; c_i = c_j$,

$$\mathrm{P}(X_1 = x_1, \dots, X_N = x_N \mid C_1 = c_1, \dots, C_N = c_N,\ I) = \mathrm{P}(X_{i_1} = x_{\pi(1)}, \dots, X_{i_N} = x_{\pi(N)} \mid C_1 = c_1, \dots, C_N = c_N,\ I)\ . \quad (8)$$

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 $\{f_{x|c} \mid x \in \mathfrak{X}\}$ 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 $\{x_1, c_1), \dots, (x_N, c_N)\}$:

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

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 (**??**) can be written as follows:

$$\mathrm{P}(X_1 = x_1, \dots, X_N = x_N \mid C_1 = c_1, \dots, C_N = c_N,\ I) = \int \prod_{c,x} f_{x|c}{}^{NF_{x,c}}\ \mathrm{p}(\boldsymbol{f_{x|c}} \mid I)\, \mathrm{d}\boldsymbol{f_{x|c}}\ . \quad (10)$$

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

2 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. (**??**).

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 (**??**) then holds:

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

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 (**??**) and (**??**) into (**??**). The products within their integrals can be combined considering that

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

We obtain

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

with

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

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

$f_{x|c}$ and f_c is just a specific case of a density function for f_{xc}, apart from a Jacobian factor.

The integral expression (**??**) 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 (**??**) is that *the density for the joint distribution f_{xc} is factorizable into the product of a density for the conditional distribution $f_{x|c}$ and a density for the marginal distribution 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 (**??**), then the corresponding probability distributions enjoy a symmetry of partial or conditional exchangeability.

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

$$\mathrm{p}(f_{xc} \mid I)\, \mathrm{d}f_{xc} \equiv \mathrm{p}[(f_{x|c}, f_c) \mid I]\, \mathrm{d}f_{x|c}\, \mathrm{d}f_c \equiv \mathrm{p}(f_{x|c} \mid f_c, I)\, \mathrm{p}(f_c \mid I)\, \mathrm{d}f_{x|c}\, \mathrm{d}f_c \,. \quad (15)$$

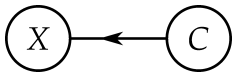
The factorization condition is thus equivalent to conditional independence:

$$\mathrm{p}(f_{x|c} \mid f_c, I)\, \mathrm{d}f_{x|c} = \mathrm{p}(f_{x|c} \mid I)\, \mathrm{d}f_{x|c} \,. \quad (16)$$

3 Exchangeable Bayesian networks

3.1 Graphical representation of conditional exchangeability

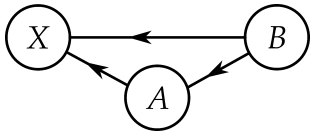
The partial (or conditional) exchangeability (**??**) of the probabilities of the X statements given the C, the full exchangeability (**??**) of the probabilities of the latter, and the final integral representation (**??**) for the join probability distribution, can be together expressed in the guise of a Bayesian network:



The two nodes represent the two sets of statements. The arrow from the C-node to the X-node represent the conditional exchangeability

of the probabilities for the latter set of statements conditional on the former set. The absence of incoming arrows to the C-node represents the full exchangeability for its probability distribution. The final integral representation for the joint probability of full network has a factorizable density, with one factor per node.

Using the reasoning and integral representations of §§ **??**–**??** it is indeed possible to generalize these rules to more complex networks of statements. The following network, for example:



has the representation

$$\mathrm{P}(X_1 = x_1, A_1 = a_1, B_1 = b_1, \ldots, X_N = x_N, A_N = a_N, B_N = b_N \mid I) = \int \prod_{x,a,b} {f_{x,a,b}}^{NF_{x,a,b}} \, \mathrm{p}(\boldsymbol{f}_{\boldsymbol{x}|\boldsymbol{ab}} \mid I)\, \mathrm{p}(\boldsymbol{f}_{\boldsymbol{a}|\boldsymbol{b}} \mid I)\, \mathrm{p}(\boldsymbol{f}_{\boldsymbol{b}} \mid I)\, \mathrm{d}\boldsymbol{f}_{\boldsymbol{x}|\boldsymbol{ab}}\, \mathrm{d}\boldsymbol{f}_{\boldsymbol{a}|\boldsymbol{b}}\, \mathrm{d}\boldsymbol{f}_{\boldsymbol{b}} \,. \quad (17)$$

3.2 Additional independence assumptions

In the last two examples the joint probability density for all the statements is decomposed in full accord with the product rule. For example,

$$\begin{aligned}\mathrm{P}(\{X_i = x_i\}, \{A_i = a_i\}, \{B_i = b_i\} \mid I) \equiv \\ \mathrm{P}(\{X_i = x_i\} \mid \{A_i = a_i\}, \{B_i = b_i\}, I) \times \\ \mathrm{P}(\{A_i = a_i\} \mid \{B_i = b_i\}, I) \times \mathrm{P}(\{B_i = b_i\} \mid I)\,, \end{aligned} \quad (18)$$

which is an identity in the probability-calculus. In other words, no special properties of conditional independence hold. In this case the integral representation involves an integration over a set of conditional or marginal long-run frequencies which is equivalent to the set of joint frequencies. For example, in the representation (**??**) the two sets

$$\{f_{x|a,b},\, f_{,a|b},\, f_{,,b} \mid x \in \mathfrak{X}, a \in \mathfrak{A}, b \in \mathfrak{B}\} \leftrightarrow \{f_{x,a,b} \mid x \in \mathfrak{X}, a \in \mathfrak{A}, b \in \mathfrak{B}\} \quad (19)$$

are in one-one correspondence.

The factorizability of the density therefore only implies, and is implied by, the exchangeability symmetries of conditional probability distributions. It does not imply additional independences.

Additional independence properties, such as

$$\begin{aligned}\mathrm{P}(\{X_i = x_i\}, \{A_i = a_i\}, \{B_i = b_i\} \mid I) = \\ \mathrm{P}(\{X_i = x_i\} \mid \{A_i = a_i\}, \{B_i = b_i\}, I) \times \\ \mathrm{P}(\{A_i = a_i\} \mid I) \times \mathrm{P}(\{B_i = b_i\} \mid I)\ ,\end{aligned} \tag{20}$$

which is not an identity of the probability-calculus, are instead expressed by a reduction in the number of conditional or marginal long-run frequencies in the representation (similarly to what happens when partial exchangeability reduces to full exchangeability; see § **??**).

For example, when the independence (**??**) hold, the representation (**??**) becomes

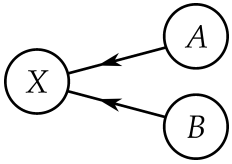
$$\begin{aligned}\mathrm{P}(X_1 = x_1, A_1 = a_1, B_1 = b_1, \dots, X_N = x_N, A_N = a_N, B_N = b_N \mid I) = \\ \int \prod_{x,a,b} (f_{x|a,b}\, f_{,a}\, f_{,,b})^{N f_{x,a,b}}\ \mathrm{p}(\boldsymbol{f_{x|ab}} \mid I)\, \mathrm{p}(\boldsymbol{f_a} \mid I)\, \mathrm{p}(\boldsymbol{f_b} \mid I)\ \mathrm{d}\boldsymbol{f_{x|ab}}\, \mathrm{d}\boldsymbol{f_a}\, \mathrm{d}\boldsymbol{f_b}\end{aligned} \tag{21}$$

with $f_{,a} \coloneqq \sum_{x,b} f_{x,a,b}$ and so on. We see that the integration is now over the reduced set of long-run distributions

$$\{f_{x|a,b},\ f_{,a},\ f_{,,b} \mid x \in \mathfrak{X}, a \in \mathfrak{A}, b \in \mathfrak{B}\}\ . \tag{22}$$

Equivalently this reduction can be expressed as $f_{,a|b} = f_{,,a}$ for all b, implying the presence of a delta term in the corresponding density.

The independence condition (**??**) and integral representation (**??**) can be expressed by the network



3.3 Generalization and possible graphical rules

It seems possible to generalize the example presented thus far into a consistent set of rules.

We have

i. several sets of N statements, $\{X_i = x_i\}$, $\{A_i = a_i\}$, ... To each set is associated a set of values $x \in \mathfrak{X}$, $a \in \mathfrak{A}$, ...

ii. several conditionally exchangeable probabilities for some groups of such statements conditional on some other groups, such as eq. (**??**);

iii. several assumptions of conditional independence, such as eq. (**??**).

The integral representation of the joint probability distribution for all the statements involves several sets of integration variables. Each set contains either conditional or marginal long-run frequencies in the quantities $x, a, \ldots$.

– and a factorizable density over such variables. The density is multiplied by the product of these variables, raised to the joint empirical frequencies

can then be represented by an acyclic directed network with these properties:

1. nodes represent sets of statements;
2. there is a set of integration variables for each node. These variables are the long-run frequencies for the quantity of the node conditional on the quantities of the parent nodes. Nodes without parents have marginal, unconditional long-run frequencies;
3. each set of integration variables has its own density. The densities are multiplied in the integral representation.

Moreover, s explained in § **??**, independence assumptions are expressed by the set of integration variables, whereas conditional-exchangeability assumption are expressed by the factorization of the density over the integration variables.

indicates the conditional exchangeability of the latter set conditional on the former set. Nodes with no incoming arrows are fully exchangeable. The integral representation for the full network, that is, for the all the statements, has a factorizable density, one factor per node. The arguments of these factors are conditional or marginal long-run frequencies accordingly to the arrows incoming to their nodes.

4 Discussion

*** first result can be useful for infinite limits, leading to regression

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. (**??**).

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.

Damien, P., Dellaportas, P., Polson, N. G., Stephens, D. A., eds. (2013): *Bayesian Theory and Applications*. (Oxford University Press, Oxford).

Dawid, A. P. (2013): *Exchangeability and its ramifications*. In: Damien, Dellaportas, Polson, Stephens (2013), ch. 2, 19–29.

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).

Keynes, J. M. (2013): *A Tract on Monetary Reform*, repr. of 2nd ed. (Cambridge University Press, Cambridge). First publ. 1923.