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 30 May 2020

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

1 Full, partial, and conditional exchangeability

De Finetti's theorem for infinitely exchangeable probability distributions is one of the formulae derived from the probability calculus having the richest practical and philosophical consequences. Leaving for the moment the definition of symbols to intuition, the theorem rewrites a joint probability distribution as a law of total probability:

$$\mathrm{P}(X_1 = x_1,\ X_2 = x_2,\ \dots \mid J) = \int \prod_i f_{x_i}\ \mathrm{p}(\boldsymbol{f} \mid J)\, \mathrm{d}\boldsymbol{f}\ , \tag{1}$$

where $\boldsymbol{f} \coloneqq (f_x)$ is a distribution over the values that can be assumed by each quantity X_i, and the integral is over the simplex of such distributions.

The condition for the theorem to hold is that the joint distribution be infinitely *fully exchangeable*, that is, symmetric with respect to permutations of the x_i, for any set of indices $\{i\}$. This formula is the infinite limit of the sampling formula from an urn of unknown content. From now on 'exchangeable' shall be understood as 'infinitely exchangeable'.

A more general version of the theorem holds if the joint distribution is *partially exchangeable*, that is, if the set $\{X_i\}$ is divided into two or more categories represented by subsets $\{Y_j\}, \{Z_k\}, \dots,$

and permutations are allowed within each subset but not necessarily across subsets. The formula then becomes, with a suitable re-indexing $\{1,2,\dots\} \mapsto \{1',2',\dots,1'',2'',\dots\}$,

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

with distinct 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 (1) is recovered.

With a little reflection we see that if we know that the quantities X 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 $C_i=c_i$ 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.

Let us make the information about the C_i explicit. Their possible values are 'Y' and 'Z'. We rewrite the probability in eq. (2) as

$$\begin{aligned}&\mathrm{P}(Y_{1'}=y_{1'},\ \dots,\ Z_{1''}=z_{1''},\ \dots \mid J) \equiv \\ &\mathrm{P}(X_{1'}=x_{1'},\ \dots,\ X_{1''}=x_{1''},\ \dots \mid \mathsf{C}_{1'}=\mathsf{Y},\ \dots,\ \mathsf{C}_{1''}=\mathsf{Z},\ \dots,\ I)\ .\end{aligned} \tag{3}$$

Then it is clear that the partially exchangeable probability distribution (2) or (3) can also be called *conditionally exchangeable*.

The present work gives a representation formula for pairs of quantities (X_i, C_i) such that

1. the distribution for the $\{X_i\}$ is conditionally exchangeable, given $\{C_i\}$,

2. the distribution for the $\{C_i\}$ is fully exchangeable.

The formula is:

$$\mathrm{P}\bigl[(X_1 = x_1, C_1 = c_1),\ (X_2 = x_2, C_2 = c_2),\ \dots \mid I\bigr] = \int \prod_i f_{x_i, c_i}\ \mathrm{p}(\boldsymbol{f} \mid I)\,\mathrm{d}\boldsymbol{f} \tag{4a}$$

$$\text{with} \qquad \boxed{\mathrm{p}(\boldsymbol{f} \mid I)\,\mathrm{d}\boldsymbol{f} = \mathrm{p}[(f_{x|c}) \mid I]\,\mathrm{d}(f_{x|c}) \times \mathrm{p}[(f_{,c}) \mid I]\,\mathrm{d}(f_{,c})} \tag{4b}$$

where

- $\boldsymbol{f} \coloneqq (f_{x,c})$ is a joint distribution over the set of values that can be assumed by each (X_i, C_i) pair,
- $\bigl(f_{,c} \coloneqq \sum_x f_{x,c}\bigr)$ is the related *marginal* frequency distribution for the c values,
- $\bigl(f_{x|c} \coloneqq f_{x,c}/f_{,c}\bigr)$ are the related *conditional* frequency distributions of x given c.

The formula (4) shows that the pairs (X_i, C_i) have a fully exchangeable distribution, with a representation analogous to eq. (1). The noteworthy feature of this formula is that *the density* $\mathrm{p}(\boldsymbol{f} \mid I)\,\mathrm{d}\boldsymbol{f}$ *is factorizable into the product of a density for the conditional frequencies* $(f_{x|c})$ *and a density for the marginal frequencies* $(f_{,c})$, according to eq. (4b). This factorization expresses the conditional exchangeability for $\{X_i\}$ given $\{C_i\}$.

The proof of formula is given in the next section. In the final section I discuss possible generalizations and connections with Bayesian belief networks.

2 Setup and proof

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

The domain of discourse consists of a countably infinite set of statements

$$\bigl\{X_i = x_i \bigm| i \in \mathbf{N},\ \forall i\ x_i \in \mathfrak{X}\bigr\} \tag{5}$$

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

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

A probability distribution over these atomic statements (in the logical sense) is called fully (infinitely) exchangeable if

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

$$\begin{aligned}\mathrm{P}(X_{i_1} = x_{i_1}, X_{i_2} = x_{i_2}, \dots, X_{i_N} = x_{i_N} \mid I) &= \\ &\mathrm{P}(X_{i_1} = x_{\pi(i_1)}, X_{i_2} = x_{\pi(i_2)}, \dots, X_{i_N} = x_{\pi(i_N)} \mid I),\end{aligned} \tag{6}$$

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.

In the following I let $\{1, 2, \dots\}$ denote any subset of $\mathbf{N}$, to avoid a proliferation of subscripts.

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

$$\begin{aligned}\mathrm{P}(X_1 = x_1,\ \dots,\ X_N = x_N \mid I) &= \int \prod_i f_{x_i}\ \mathrm{p}(\boldsymbol{f} \mid I)\,\mathrm{d}\boldsymbol{f} \\ &= \int \prod_x {f_x}^{NF_x}\ \mathrm{p}(\boldsymbol{f} \mid I)\,\mathrm{d}\boldsymbol{f}\end{aligned} \tag{7}$$

where (F_x) are the empirical relative frequencies of the values x in the set $\{x_1, \dots, x_N\}$:

$$NF_x \coloneqq \textstyle\sum_i \delta(x, x_i), \quad x \in \mathfrak{X}\,. \tag{8}$$

I will state the definition of partial exchangeability and its theorem in a slightly unfamiliar form, according to the discussion of § 1. The role of the additional information about the category to which each instance belongs will be explicit.

We now have an additional set of atomic statements

$$\left\{C_i = c_i \,\middle|\, i \in \mathbf{N},\ \forall i\ c_i \in \mathfrak{C}\right\} \tag{9}$$

identifying each instance $1, 2, \dots$ as belonging to one or another category from a finite set $\mathfrak{C}$. For each i the statements $\{C_i = c \mid c \in \mathfrak{C}\}$ are mutually exclusive on information I.

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

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

for every N, every set of indices $\{1, \dots, N\} \subset \mathbf{N}$, and every permutation π thereof such that $\pi(i) = j \implies c_i = c_j$ (10)

that is, for every permutation π *which only exchanges indices having the same c value*.

As remarked in § 1, partial exchangeability is usually written by using different symbols for indices belonging to different categories, and leaving the information implied by such notation implicit in the context.

I will also write de Finetti's theorem for partial exchangeability in a slightly unfamiliar form.

For each value $c \in \mathfrak{C}$, 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 by $\boldsymbol{f}_{|} \coloneqq (f_{x|c})$ 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 distribution over the values (x, c) in the set of pairs $\big\{(x_i, c_i), \dots, (x_N, c_N)\big\}$:

$$NF_{x,c} \coloneqq \textstyle\sum_i \delta(x, x_i)\, \delta(c, c_i), \quad x \in \mathfrak{X},\ c \in \mathfrak{C}\,. \tag{11}$$

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

$$\mathrm{P}(X_1 = x_1, \dots, X_N = x_N \mid C_1 = c_1, \dots, C_N = c_N,\ I) = \\ \int \prod_{c,x} {f_{x|c}}^{NF_{x,c}}\ \mathrm{p}\big(\boldsymbol{f}_{|} \,\big|\, I\big)\, \mathrm{d}\boldsymbol{f}_{|}\,. \tag{12}$$

Scrutiny of this formula should show that this form is equivalent to the usual representation. There is 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.

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: **jeffrey1980**, 233–249.