

## Handout 4: Exchangeability and the Bayesian model

Lecturer &amp; author: Georgios P. Karagiannis

georgios.karagiannis@durham.ac.uk

### Aim

Get familiar with the concept of exchangeability, and its relation to Subjective probability and Bayesian paradigm.

### Reading list:

- Bernardo, J. M., & Smith, A. F. (2009, Section 4.3). Bayesian theory (Vol. 405). John Wiley & Sons.
- Berger, J. O. (2013, Section 3.5.7). Statistical decision theory and Bayesian analysis. Springer Science & Business Media.

## 1 Exchangeability

As mentioned in Handout 3, one way to specify the Bayesian model is by subjectively specifying the probability distributions  $F(y|\theta)$  and  $\Pi(\theta)$ , or the joint distribution  $P(y, \theta)$ , that enables You to derive the rest distributions.

Alternatively, You can specify a probability distribution on the data generating process  $G(y_{1:n})$  describing the actual sequence of the data  $y_{1:n} = (y_1, \dots, y_n)$ . One approach is to subjectively set certain invariance assumptions on the observables  $y$  involving probabilistic believes of invariant with respect to some aspect of the observable quantities.

A reasonable invariance assumption about  $y_{1:n} = (y_1, \dots, y_n)$  (to specify the data generating model  $G(y_{1:n})$ ) is the Exchangeability: The 'labels' identifying the individual observable quantities are 'uninformative', in the sense the information that the  $y_i$ 's provide is independent of the order in which they are collected. Exchangeability, although a simple assumption, it accurately describes a large class of experimental setups.

**Definition 1.** A sequence of random quantities  $y_{1:n} = (y_1, \dots, y_n)$  is finitely exchangeable under a probability distribution  $G$  if all permutations of  $\{y_1, \dots, y_n\}$  have the same joint distribution  $G$ . Namely; if

$$G(y_1, \dots, y_n) = G(y_{p(1)}, \dots, y_{p(n)})$$

for all permutations  $p$  defined on the set  $\{1, \dots, n\}$ .

**Definition 2.** An infinite sequence of random quantities  $y_1, y_2, \dots$  is infinitely exchangeable under a probability distribution  $G$  if every finite sub-sequence is finitely exchangeable under  $G$ .

**Example 3.** If a sequence of random quantities  $y_{1:n} = (y_1, \dots, y_n)$  is mutually independent, then it is exchangeable.

**Solution.** It is straightforward, since  $G(y_1, \dots, y_n) = \prod_{i=1}^n G(y_i)$  which is invariant to permutations of the indexes.

## 2 The representation theorem

*Note 4.* The assumption of exhcengeability leads to the following development, which theoretically justifies (to some extend) the existence of the Prior distribution, and the Bayesian paradigm.

**Theorem 5. (General representation theorem)** *If  $y_1, y_2, \dots$  is an infinitely exchangeable sequence of random quantities with probability distribution  $F$ , there exists a probability measure  $\Pi$  over  $\mathcal{F}$ , the space of all distribution functions on*

$\mathcal{Y}^n \subseteq \mathbb{R}^n$  for  $n \geq 1$ , such that the joint distribution function of  $y_{1:n} = (y_1, \dots, y_n)$  has CDF

$$G(y_1, \dots, y_n) = \int_{\mathcal{F}} \prod_{i=1}^n F(y_i) d\Pi(F) \quad (1)$$

where  $F$  is an unknown/unobservable distribution function,  $\Pi(F) = \lim_{n \rightarrow \infty} P(\hat{F}_n)$  is a probability distribution on the space of functions  $\mathcal{F}$ , which is defined as a limit distribution on the empirical distribution function  $\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n 1(y_i \leq x)$  defined by  $y_1, \dots, y_n$  (as  $n \rightarrow \infty$ ), and  $F(x) = \lim_{n \rightarrow \infty} \hat{F}_n(x)$ .

**Remark 6.** The Representation Theorem shows that if  $y_1, y_2, \dots$  is infinitely exchangeable, then the elements of  $y_{1:n} = (y_1, \dots, y_n)$  are i.i.d. conditional on the empirical distribution of  $y_{1:n}$ .

**Remark 7.** (Interpretation) The general representation Theorem 5 says

- $y_{1:n}$  are considered to be an i.i.d. sample generated from an unknown (i.e., random) distribution function  $F$ , (conditional on  $F$ ); i.e.  $y_i | F \sim F(\cdot)$ .
- $F$  is an the unknown CDF, which follows itself a probability distribution  $\Pi$  representing (prior) believes about  $F$
- and  $F$  has the operational role of what You believe the empirical distribution function would look like for a large sample.

**Note 8.** For convenience, hereafter, we will consider the concept of exchangeability by using the parametric form 2 and 3 in Fact 9.

**Fact 9.** Given some additional problem specific invariance assumptions (see Bernardo & Smith (2009, Section 4.3)) via subjunctive judgments regarding the generating process of  $y_1, y_2, \dots$ , the unknown sampling distribution  $G(y_i)$  in (1) can be written as a parametric model  $F(y_i|\theta)$  labeled by an unknown parameter  $\theta \in \Theta$ , which is the limit of some function of  $y_{1:n}$  (as  $n \rightarrow \infty$ ), and there exists a probability distribution  $\Pi$  for  $\theta$  such that

$$G(y_1, \dots, y_n) = \int_{\Theta} \prod_{i=1}^n F(y_i|\theta) d\Pi(\theta) \quad (2)$$

In PDF/PMF, (2) is written as

$$g(y_1, \dots, y_n) = \int_{\Theta} \prod_{i=1}^n f(y_i|\theta) d\Pi(\theta) \quad (3)$$

**Remark 10.** Regarding the Bayesian paradigm, Fact 9 provides a rational for the consideration of the uncertain parameter  $\theta$  as a random variable and the subjective prior  $\Pi(\theta)$ . If  $y_1, y_2, \dots$  is an exchangeable sequence of real-valued random quantities, then any finite subset of them is an i.i.d. random sample from parametric model  $F(\cdot|\theta)$  labeled by some uncertain parameter  $\theta \in \Theta$ , and there exists a (prior) probability distribution  $\Pi(\theta)$  for  $\theta$  which has to describe the initially available information about the parameter which labels the model; Hence a rational for the parametric Bayesian model:

$$\begin{cases} y_i | \theta & \stackrel{\text{iid}}{\sim} F(\cdot|\theta) \quad \forall i = 1, \dots, n \\ \theta & \sim \Pi(\cdot) \end{cases}$$

**Note 11.** The following Example 12 of the ‘representation Theorem with 0 – 1 quantities’ presents a special case.

**Example 12.** (Representation Theorem with 0 – 1 quantities). If  $y_1, y_2, \dots$  is an infinitely exchangeable sequence of 0 – 1 random quantities with probability measure  $G$ , there exists a distribution function  $\Pi$  such that the joint mass function  $g(y_1, \dots, y_n)$  for  $y_1, \dots, y_n$  has the form

$$g(y_1, \dots, y_n) = \int_0^1 \prod_{i=1}^n \underbrace{\theta^{y_i} (1 - \theta)^{1-y_i}}_{f_{\text{Br}(\theta)}(y_i|\theta)} d\Pi(\theta)$$

where

$$\Pi(t) = \lim_{n \rightarrow \infty} P\left(\frac{1}{n} \sum_{i=1}^n y_i \leq t\right) \quad \text{and} \quad \theta \stackrel{\text{as}}{=} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n y_i$$

aka  $\theta$  is the limiting relative frequency of 1s, by SLLN.

**Solution.** It is given in the Exercise sheet to read; see Exercise 34.

**Example 13.** (cont. Example 12) The representation of exchangeable sequence of 0 – 1 random quantities  $y_1, \dots, y_n$  can be interpreted as follows:

- $y_i$  are considered to be conditionally independent and identically distributed Bernoulli random quantities given the random quantity  $\theta$ ; i.e.  $y_i|\theta \stackrel{iid}{\sim} \text{Br}(\theta)$ .
- $\theta$  is itself assigned a probability distribution  $\Pi$  which can be interpreted as its prior distribution,
- by the SLLN,  $\theta$  is defined as  $\theta = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n y_i$ , and hence  $\Pi$  can be interpreted as beliefs about the limiting relative frequency of 1's.

*Note 14.* The following Example 15 can justify the notion of prior, posterior and predictive distribution in the context of the exchangeability.

**Example 15.** Let  $y_1, y_2, \dots$  be an infinitely exchangeable sequence of random quantities under distribution  $G$  admitting a PDF/PMF  $g$ . Then from the representation theorem in (2), the conditional distribution  $G(y_{n+1:n+m}|y_{1:n})$  has PDF/PMF

$$g(y_{n+1:n+m}|y_{1:n}) = \int_{\Theta} \prod_{i=n+1}^{n+m} f(y_i|\theta) d\Pi(\theta|y_{1:n}) \quad \text{where} \quad d\Pi(\theta|y_{1:n}) = \frac{\prod_{i=1}^n f(y_i|\theta) d\Pi(\theta)}{\int_{\Theta} \prod_{i=1}^n f(y_i|\theta) d\Pi(\theta)}$$

**Solution.** It can be shown that

$$\begin{aligned} g(y_{n+1:n+m}|y_{1:n}) &= \frac{g(y_{1:n}, y_{n+1:n+m})}{g(y_{1:n})} = \frac{\int_{\Theta} \prod_{i=1}^n f(y_i|\theta) \prod_{i=n+1}^{n+m} f(y_i|\theta) d\Pi(\theta)}{\int_{\Theta} \prod_{i=1}^n f(y_i|\theta) d\Pi(\theta)} \\ &= \int_{\Theta} \prod_{i=n+1}^{n+m} f(y_i|\theta) \underbrace{\frac{\prod_{i=1}^n f(y_i|\theta) d\Pi(\theta)}{\int_{\Theta} \prod_{i=1}^n f(y_i|\theta) d\Pi(\theta)}}_{=d\Pi(\theta|y_{1:n})} \end{aligned}$$

*Remark 16.* Subjectively specifying  $G(y_{1:n})$  and then deriving  $F(y_{1:n}|\theta)$  and  $\Pi(\theta)$  is philosophically interesting. It can suggest useful sampling-model & prior ( $F(y_{1:n}|\theta)$ ,  $\Pi(\theta)$ ) decompositions, that allow the design of new meaningful models. For example, see Bernardo, J. M., & Smith, A. F. (2009, Section 4.4). On the other hand, it is often easier to subjectively specify the Bayesian model by  $F(y_{1:n}|\theta)$  and  $\Pi(\theta)$ .

**Example 17.** Consider the parametric form (2) of the general representation theorem. Let  $y_1, y_2, \dots$  be an infinitely exchangeable sequence of real valued random quantities with  $y_i \in \mathbb{R}$  for any  $i$ .

1. Show that  $\text{Corr}(y_i, y_j) \geq 0$ , for  $i \neq j$ .
2. Find the condition under which (i.) I have  $\text{Corr}(y_i, y_j) > 0$ , for  $i \neq j$  (ii.) I have  $\text{Corr}(y_i, y_j) = 0$ , for  $i \neq j$

**Solution.** Since sequence  $y_1, y_2, \dots$  is infinitely exchangeable, I use the general representation theorem (the parametric form for simplicity). Hence, for a given  $\theta$ , it is  $x_i|\theta \stackrel{iid}{\sim} dF(\cdot|\theta)$  for all  $i$ .

1. So

$$\begin{aligned} \text{Cov}_G(y_i, y_j) &= E_G(x_i^\top x_j) - E_G(x_i)^\top E_G(x_j) = E_G(E_F(x_i^\top x_j|\theta)) - E_G(E_F(x_i|\theta))^\top E_G(E_F(x_j|\theta)) \\ &= E_\Pi(E_F(x_i^\top|\theta)E_F(x_j|\theta)) - E_\Pi(E_F(x_i|\theta))^\top E_\Pi(E_F(x_j|\theta)) = \text{Var}_\Pi(\mu(\theta)) \geq 0 \end{aligned}$$

where  $\mu(\theta) = E_F(x_i|\theta)$  for all  $i$ . Also

$$\text{Var}_G(y_i) = \text{Var}_\Pi(E_F(x_j|\theta)) + E_\Pi(\text{Var}_F(x_j|\theta)) = \text{Var}_\Pi(\mu(\theta)) + E_\Pi(\sigma^2(\theta))$$

where  $\sigma^2(\theta) = \text{Var}_F(x_i|\theta)$  for all  $i$ . Lets consider the 1-D case, that is requested;  $y_i \in \mathbb{R}$  for any  $i$ . It is

$$\text{Corr}(y_i, y_j) = \frac{\text{Var}_\Pi(\mu(\theta))}{\text{Var}_\Pi(\mu(\theta)) + E_\Pi(\sigma^2(\theta))} \geq 0$$

2. For  $\text{Var}_\Pi(E_F(x_i|\theta)) > 0$ , I have  $\text{Corr}(y_i, y_j) > 0$ . For  $\text{Var}_\Pi(E_F(x_i|\theta)) = 0$ , I have  $\text{Corr}(y_i, y_j) = 0$ ;

NB: correlation does not necessarily imply independence.

*Remark 18.* Example 17 shows that elements of infinite exchangeable sequence cannot be negatively correlated.

### 3 Practice

**Question 19.** For practice try the Exercises 36, 37, 38, and 39 from the Exercise Sheet.