2.3. Characterising Implicitly Bayesian Prediction Rules

In the case of assumed *i.i.d.* data, De Finetti's theorem gives a simple condition for a prediction rule to be implicitly Bayesian. As mentioned before, a prediction rule implies a unique joint distribution over the sequence of random variables $(X_1, X_2, \dots)$. By a version of the De Finetti's theorem, under some mild assumptions, a prediction rule is implicitly Bayesian if and only if the joint distribution it implies over $(X_1, X_2, \dots)$ is *exchangeable* (Hewitt and Savage, 1955):

Definition 1 (Exchangeable Sequence of Random Variables) *A finite sequence of n random variables $(X_1, \dots, X_n)$ is said to be exchangeable if for any permutation $\pi : \{1, \dots, n\} \to \{1, \dots, n\}$ the joint distribution of $(X_1, \dots, X_n)$ is the same as the joint distribution of $(X_{\pi(1)}, \dots, X_{\pi(n)})$.*

An infinite sequence of random variables $(X_1, X_2, \dots)$ is said to be exchangeable if, for any n, the finite sequence $(X_1, \dots, X_n)$ is exchangeable.

By De Finetti's theorem (see Appendix **??** for a more thorough explanation), we know that a sequence of random variables $(X_1, X_2, \dots)$ is exchangeable if and only if there exists a (unique) prior probability π on the space $\mathcal{P}$ of probability measures on $\mathcal{X}$ such that:

$$P[X_1 \in A_1, \dots, X_n \in A_n] = \int_{\mathcal{P}} \prod_{i=1}^{n} p(A_i) d\pi(p) \qquad \forall A_1, \dots, A_n \forall n \in \mathbb{N},$$

where $p \in \mathcal{P}$ is a probability measure on $\mathcal{X}$. In other words, only if the there exists a likelihood/prior construction that defines the same joint distribution as the prediction rule. Hence, exchangeability is the defining characteristic of implicitly Bayesian prediction rules on *i.i.d.* data. This suggests one direct way of checking whether a prediction rule is implicitly Bayesian: check whether the joint distribution on $(X_1, X_2, \dots)$ implied by the prediction rule is exchangeable.

Conditionally Identically Distributed Another desirable coherence property that we might expect of a prediction rule is that the future observations are identically distributed given the past. For example, under the *i.i.d.*assumption, if we were to observe $(x_1, \dots, x_n)$, the prediction for the next observation X_{n+1} *surely* shouldn't be different from the prediction for the observation after that. After all, we know they are identically distributed, we just don't know what the distribution is; it'd be irrational to make different predictions for X_{n+1}, X_{n+2} given the same data.

This property can be formalised as follows:

Definition 2 (Conditionally Identically Distributed) *We say that a sequence of random variables $(X_1, X_2, \dots)$ is conditionally identically distributed (c.i.d.) if for any n:*

$$P[X_{n+1} \in \cdot | X_1 = x_1, \dots, X_n = x_n] = P[X_{n+k} \in \cdot | X_1 = x_1, \dots, X_n = x_n] \quad \forall k > n$$

holds almost surely.

A prediction rule is then *c.i.d.* if the joint distribution it implies over $(X_1, X_2, \dots)$ is *c.i.d.*. Conditionally Identically Distributed sequences have been introduced in (Kallenberg, 1988) and studied and applied in a range of works (Berti et al., 2004, 2013; Fong et al., 2021a).

Appendix A. De Finetti's theorem

Consider a probability space $(\Omega, \mathcal{F}_\Omega, P)$ and a sequence of random variables $(X_1, X_2, \dots)$, $X_i : \Omega \to \mathcal{X}$, where X_i takes values in some measurable space $(\mathcal{X}, \mathcal{F})$. Then, by a version of the De Finetti's theorem given by Hewitt and Savage (1955), under fairly general conditions, if $(X_1, X_2, \dots)$ is exchangeable as per Definition 1, then the distribution of $(X_1, X_2, \dots)$ can be represented as a mixture of *i.i.d.* distributions.

Concretely, consider the set $\mathcal{P}$ of probability measures $p \in \mathcal{P}$ on $(\mathcal{X}, \mathcal{F})$, and the smallest σ-algebra $\mathcal{S}$ that makes $p \mapsto p(A)$ measurable for all $A \in \mathcal{F}$. For $p \in \mathcal{P}$, denote by p^∞ the product probability measure $\otimes_{i=1}^\infty p$ on the product space $(\prod_{i=1}^\infty \mathcal{X}, \otimes_{i=1}^\infty \mathcal{F})$. Then, if $(X_1, X_2, \dots)$ is exchangeable, there exists a unique probability measure π on $(\mathcal{P}, \mathcal{S})$ such that the joint distribution of $(X_1, X_2, \dots)$ can be represented as:

$$P[(X_1, X_2, \dots) \in A] = \int_{\mathcal{P}} p^\infty(A) d\pi(p) \qquad \forall A \in \otimes_{i=1}^\infty \mathcal{F}, \tag{7}$$

The only condition is that $\mathcal{X}$ must be a Hausdorff space with $\mathcal{F}$ being the σ-algebra of all Baire sets in $\mathcal{X}$. The practically relevant aspect of that condition is that it is satisfied in the typically considered settings of $\mathcal{X}$ being $\mathbb{R}$ or $\mathbb{R}^d$ with the usual Borel σ-algebra.

The above formulation is slightly different than the one given in Section 2.3. Namely, the proposition in (7) was instead replaced with: there exists a unique probability measure π on $(\mathcal{P}, \mathcal{S})$ such that:

$$P[X_1 \in A_1, \dots, X_n \in A_n] = \int_{\mathcal{P}} \prod_{i=1}^n p(A_i) d\pi(p) \qquad \forall A_1, \dots, A_n \in \mathcal{F}, \forall n \in \mathbb{N}, \tag{8}$$

which we think is a little bit more approachable. The propositions in (7) and (8) are equivalent. Clearly, (7) implies (8); if the equality holds for all sets A in the infinite product σ-algebra $\otimes_{i=1}^\infty \mathcal{F}$, then it will hold for rectangles $A = (A_1, \dots, A_n, \mathcal{X}, \mathcal{X}, \mathcal{X}, \dots) \in \otimes_{i=1}^\infty \mathcal{F}$, and so:

$$\begin{aligned}
P[X_1 \in A_1, \dots, X_n \in A_n] &= P[(X_1, X_2, \dots) \in A] \qquad B = (A_1, \dots, A_n, \mathcal{X}, \mathcal{X}, \dots) \\
&= \int_{\mathcal{P}} p^\infty(B) d\pi(p) \\
&= \int_{\mathcal{P}} \prod_{i=1}^n p(A_i) \prod_{i=n+1}^\infty \overbrace{p(\mathcal{X})}^{1} d\pi(p) \qquad \triangle\text{Definition of product measure} \\
&= \int_{\mathcal{P}} \prod_{i=1}^n p(A_i) d\pi(p)
\end{aligned}$$

To go the other way around from (8) to (7), assume that there exists a unique measure π such that (8) holds. We can extend the statement to 'infinite rectangle' sets of the form $B = (A_1, A_2, \dots), A_i \in \mathcal{F} \forall i \in \mathbb{N}$ by noting that since $(A_1, \dots, A_n, \mathcal{X}, \mathcal{X}, \dots) \downarrow B$, we have that $P[X_1 \in A_1, \dots, X_n \in A_n] \downarrow P[(X_1, X_2, \dots) \in B]$. Hence:

$$\begin{aligned}
P[(X_1, X_2, \dots) \in B] &= \lim_{n\to\infty} P[X_1 \in A_1, \dots, X_n \in A_n] \\
&= \lim_{n\to\infty} \int_{\mathcal{P}} \prod_{i=1}^{n} p(A_i) d\pi(p) \\
&= \int_{\mathcal{P}} \lim_{n\to\infty} \prod_{i=1}^{n} p(A_i) d\pi(p) \quad \triangle\text{Dominated Convergence Theorem} \\
&= \int_{\mathcal{P}} p^{\infty}(B) d\pi(p)
\end{aligned}$$

Now, since $P[(X_1, X_2, \dots) \in \cdot]$ and $\int_{\mathcal{P}} p^{\infty}(\cdot) d\pi(p)$ are two measures on $(\prod_{i=1}^{\infty} \mathcal{X}, \otimes_{i=1}^{\infty} \mathcal{F})$ that agree on the π-system of rectangles that generates the product σ-algebra, they must agree on the entire σ-algebra.

It is worth noting that exchangeability of the entire infinite sequence $(X_1, X_2, \dots)$ is required for the result to hold. One might hope that if a finite sequence $(X_1, \dots, X_n)$ is exchangeable for some n, then the joint distribution of these n random variables might also be a mixture distribution. This is not always the case (Kerns and Székely, 2006).