ETH Zurich

Semester Thesis

De Finetti's Representation Theorem

Author:
Peter Hinz

Professor:
Prof. Sara van de Geer

October 26, 2015

Contents

1 Introduction

This thesis will introduce the notion of exchangeability and Theorem of De Finetti. De Finetti's Theorem gives a full characterization of the joint distribution of finite and infinite sequences of exchangeable random variables. Useful mathematical notions can be found in the Appendix.

The main reference for this thesis is [1]. However, some notation and ideas come from [2].

2 Exchangeability

The notion of exchangeability describes a symmetry property of sequences of random variables. It is the weakest possible symmetry assumption in the sense that relabelling does not change the joint distribution.

Definition 1 (Exchangeability for finitely many random variables)**.** A finite number of random variables $(X_1, \ldots, X_n)$ is said to be *exchangeable* if every permutation[1] $(X_{\rho(1)}, \ldots, X_{\rho(n)})$ of $(X_1, \ldots, X_n)$ has the same joint distribution as $(X_1, \ldots, X_n)$ itself.

Definition 2 (Exchangeability for countably many random variables)**.** A sequence of random variables $X = (X_1, X_2, \ldots)$ is *exchangeable* if $(X_j)_{j \in J}$ is exchangeable for every finite $J \subset \mathbb{N}$.

The following lemma unifies the notion of exchangeability for finite and countably infinite sets of random variables. The proof is obvious.

Lemma 3. *A finite or countable set S of random variables is exchangeable if and only if for every $n \in \{1, \ldots, |S|\}$ every n-tuple of distinct elements of S has the same joint distribution as any other n-tuple of distinct elements of S.*

Example 4. *Any sequence $(X_1, X_2, \ldots)$ of independent identically distributed random variables is exchangeable.*

However, there exist random variables which are exchangeable but not i.i.d.:

Example 5. *Let X_1, X_2 be Bernoulli random variables with joint distribution*

$$\mathbb{P}[\{X_1 = 0, X_2 = 0\}] = \mathbb{P}[\{X_1 = 1, X_2 = 1\}] = \frac{1}{3}$$
$$\mathbb{P}[\{X_1 = 1, X_2 = 0\}] = \mathbb{P}[\{X_1 = 0, X_2 = 1\}] = \frac{1}{6}.$$

Then they are exchangeable but certainly not i.i.d.

De Finetti's Theorem gives a characterization of all possible forms of exchangeability and it will reveal that one has to distinguish between the case of finitely and the case of infinitely many exchangeable random variables.

The Backward Martingale convergence theorem allows to prove a strong law of large numbers for a countably infinite number exchangeable random variables. The theorem and proof can be found in [2] for example.

[1] Let ρ be an element of the symmetric group $S(n)$

Theorem 6 (backward martingale convergence theorem)**.** *Let* $(X_n)_{n\in-\mathbb{N}_0}$ *be a martingale with respect to a filtration* $\mathbb{F} = (\mathcal{F}_n)_{n\in-\mathbb{N}_0}$*. Then there exists* $X_{-\infty} = \lim_{n\to-\infty} X_n$ *almost surely and in* L^1*. Furthermore,* $X_{-\infty} = \mathbb{E}\left[X_0|\mathcal{F}_{-\infty}\right]$*, with* $\mathcal{F}_{-\infty} = \bigcap_{n=1}^{\infty} \mathcal{F}_{-n}$*.*

The notion of the exchangeable σ-algebra $\mathcal{E}$ is useful to make use of this theorem.

Definition 7 (symmetric map)**.** Let $n \in \mathbb{N}$ and E, E' two sets. A function $f : E^n \mapsto E'$ is called *symmetric* if $f = f^\rho$ for each $\rho \in S(n)$, where $S(n)$ is the symmetric group permuting elements $(1, \dots, n)$ and f^ρ is defined as

$$f^\rho : \begin{cases} E^n & \to E' \\ (e_1, \dots, e_n) & \mapsto f(e_{\rho(1)}, \dots, e_{\rho(n)}). \end{cases}$$

A function $F : E^{\mathbb{N}} \to E'$ is called *n-symmetric* if $F = F^\rho$ for each $\rho \in S(n)$, where F^ρ is defined as

$$F^\rho : \begin{cases} E^{\mathbb{N}} & \to E' \\ (e_i)_{i\in\mathbb{N}} & \mapsto F((e_{\rho(i)})_{i\in\mathbb{N}}) \end{cases} \qquad \text{(with } \rho(i) := i \text{ for } i > n\text{).}$$

F is called *symmetric* if it is n-symmetric for every $n \in \mathbb{N}$.

Definition 8 (exchangeable σ-algebra)**.** Let $X = (X_1, X_2, \dots)$ be a sequence of random variables from a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to some measurable space $(E, \mathcal{A})$. For $n \in \mathbb{N}$ let

$$\begin{aligned} \mathcal{E}'_n &:= \sigma(F : E^{\mathbb{N}} \to \mathbb{R} \quad (\mathcal{A}^{\mathbb{N}}/\mathcal{B}(\mathbb{R}))\text{-measurable and } n\text{-symmetric}) \\ \mathcal{E}_n &:= X^{-1}(\mathcal{E}'_n) \end{aligned}$$

and

$$\begin{aligned} \mathcal{E}' &:= \bigcap_{n=1}^{\infty} \mathcal{E}'_n \\ \mathcal{E} &:= \bigcap_{n=1}^{\infty} \mathcal{E}_n = X^{-1}(E') \end{aligned}$$

Call $\mathcal{E}$ the *exchangeable* σ*-algebra* for X.

In other words a function from a finite power E^n of E to E' is symmetric if permuting the arguments does not change the function value and a function from $E^{\mathbb{N}}$ to E' is symmetric if permuting a finite number of arguments does not change the function value.

Theorem 9. *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space and* $(E, \mathcal{A})$ *be a measurable space. Let* $X = (X_n)_{n\in\mathbb{N}}$ *be an exchangeable sequence of random variables from* $(\Omega, \mathcal{F})$ *to* $(E, \mathcal{A})$*. If* $\varphi : E^{\mathbb{N}} \to \mathbb{R}$ *is* $(\mathcal{A}^{\mathbb{N}}/\mathcal{B}(\mathbb{R}))$*-measurable and* $\varphi \circ X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$*, then for every* $n \in \mathbb{N}$ *and every* $\rho \in S(n)$

$$\mathbb{E}\left[\varphi(X)|\mathcal{E}_n\right] = \mathbb{E}\left[\varphi^\rho(X)|\mathcal{E}_n\right]$$

Proof. By definition of $\mathcal{E}_n$, for each $A \in \mathcal{E}_n$ there exists a $B_A \in \mathcal{E}'_n$ such that $A = X^{-1}(B_A)$. It follows for all $A \in \mathcal{E}_n$

$$\begin{aligned} \mathbb{E}\left[\varphi(X)\mathbb{1}_A\right] &= \mathbb{E}\left[\varphi(X)\mathbb{1}_{B_A}(X)\right] = \mathbb{E}\left[(\varphi \cdot \mathbb{1}_{B_A})(X)\right] \\ &= \mathbb{E}\left[(\varphi \cdot \mathbb{1}_{B_A})^\rho(X)\right] = \mathbb{E}\left[\varphi^\rho(X)\mathbb{1}^\rho_{B_A}(X)\right] \\ &= \mathbb{E}\left[\varphi^\rho(X)\mathbb{1}_{B_A}(X)\right] = \mathbb{E}\left[\varphi^\rho(X)\mathbb{1}_A\right]. \end{aligned}$$

The step from the first to the second line follows from exchangeability of X and the step from the second to the third line from n-symmetry of the function $\mathbb{1}_{B_A}$. $\square$

From this theorem, it follows for φ and X as before that

$$\mathbb{E}\left[\varphi(X)|\mathcal{E}(n)\right] = \mathbb{E}\left[\frac{1}{n!}\sum_{\rho\in S(n)}\varphi^\rho(X)|\mathcal{E}_n\right] = \frac{1}{n!}\sum_{\rho\in S(n)}\varphi^\rho(X) \tag{1}$$

because $\frac{1}{n!}\sum_{\rho\in S(n)}\varphi^\rho(X)$ is already $\mathcal{E}_n$-measurable. Now one can make use of the backward martingale convergence theorem. Let $X = (X_1, X_2, \dots)$ be a sequence of exchangeable and integrable random variables from a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. For $n \in \mathbb{N}$ set

$$\begin{aligned} \mathcal{F}_{-n} &:= \mathcal{E}_n \\ Y_{-n} &:= \frac{1}{n}\sum_{k=1}^{n} X_k \end{aligned}$$

Then $(\mathcal{F}_n)_{n\in-\mathbb{N}}$ is a filtration since for all $n \in \mathbb{N}$ one has $\mathcal{F}_{-n-1} = \mathcal{E}_{n+1} \subset \mathcal{E}_n = \mathcal{F}_{-n}$ and for all $n \in \mathbb{N}$, Y_{-n} is $\mathcal{F}_{-n}$-measurable by definition. Furthermore, for all $n \in \mathbb{N}$, it holds by equation (1)

$$\begin{aligned} \mathbb{E}\left[Y_{-n}|\mathcal{F}_{-n-1}\right] &= \mathbb{E}\left[\frac{1}{n}\sum_{k=1}^{n} X_k|\mathcal{E}_{n+1}\right] = \frac{1}{(n+1)!}\sum_{\rho\in S(n+1)}\frac{1}{n}\sum_{k=1}^{n} X_{\rho(k)} \\ &= \frac{1}{(n+1)!}\frac{1}{n}\sum_{k=1}^{n}\sum_{\rho\in S(n+1)} X_{\rho(k)} \\ &= \frac{1}{(n+1)!}\frac{1}{n}\sum_{k=1}^{n} n!X_1 + \dots + n!X_{n+1} = Y_{-n-1} \end{aligned}$$

Therefore $(Y_n)_{n\in-\mathbb{N}}$ is a $(\mathcal{F}_n)_{n\in-\mathbb{N}}$ backward martingale and Theorem 6 implies that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n} X_k = \lim_{n\to-\infty} Y_n = \mathbb{E}\left[X_1|\mathcal{E}\right] \quad \text{almost surely.}$$

This shows the following theorem:

Theorem 10 (strong law of large numbers for exchangeable random variables)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $X = (X_1, X_2, \dots)$ be a sequence of exchangeable and integrable random variables. Then the sequence $(\frac{1}{n}\sum_{k=1}^{n} X_k)$ converges almost surely to $\mathbb{E}[X_1|\mathcal{E}]$.*

Note that this theorem cannot be used if only finitely many exchangeable random variables are given.

3 De Finetti's Representation Theorem

To be precise, there is not a single Representation Theorem of De Finetti but two. One for the case of countably infinitely many exchangeable random variables and one for the case of only finitely many such random variables. The second case allows only a weaker conclusion. Since the Theorems are not so easy to understand and since additional notions need to be introduced, the special case of Bernoulli random variables will be treated first.

3.1 De Finetti's Representation Theorem for Bernoulli random variables

In the sequel, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(\mathcal{X}, \mathcal{B})$ be a Borel space.

3.1.1 Finite Version

It is easy to determine all possible distributions of finitely many Bernoulli random variables such that they are exchangeable.

Theorem 11. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $X_1, \dots, X_N$ be Bernoulli random variables. The random variables $(X_1, \dots, X_N)$ are exchangeable if and only there exists numbers $p_0, \dots, p_N \in [0,1]$ with $\sum_{k=0}^N p_k = 1$ such that for any $(b_1, \dots, b_N) \in \{0,1\}^N$*

$$\mathbb{P}\left[\{X_1 = b_1, \dots, X_N = b_N\}\right] = p_{\sum_{k=1}^N b_k} \frac{1}{\binom{N}{\sum_{k=1}^N b_k}}.$$

Proof. The "if" direction is clear since the above probability distribution is invariant under permutation of the $(X_1, \dots, X_N)$. For the "only if" direction assume that the $(X_1, \dots, X_N)$ are exchangeable and let $b_1, \dots, b_N \in \{0,1\}$. Then by exchangeability, $\mathbb{P}\left[\{X_1 = b_1, \dots, X_N = b_N\}\right]$ can be expressed as a function of $\sum_{k=1}^N b_k$. Renormalizing leads to the above formula. □

3.1.2 Infinite Version

Theorem 12. *A sequence of Bernoulli random variables $X = (X_1, X_2, \dots)$ from $(\Omega, \mathcal{F})$ to $(\mathcal{X}, \mathcal{B})$ is exchangeable if and only if there exists a random variable $\Theta : \Omega \to [0,1]$ such that conditional on $\Theta = \theta$, the random variables $X_1, \dots, X_n$ are i.i.d. Bernoulli(θ) distributed. In this case the distribution of Θ is unique and it holds*

$$\frac{1}{n} \sum_{k=1}^n X_k \to \Theta \text{ almost surely.}$$

Proof. We first prove the "if" direction. Let μ_Θ be the distribution of such a Θ. By assumption it holds for every $n \in \mathbb{N}$, every $B = (b_1, \dots, b_n) \in \{0,1\}^n$ and every choice of pairwise distinct natural numbers $i_1, \dots, i_n$ that

$$\begin{aligned} \mathbb{P}\left[\{X_{i_1} = b_1, \dots, X_{i_n} = b_n\}\right] &= \int_{[0,1]} \theta^{\sum_{k=1}^n b_k} (1-\theta)^{n - \sum_{k=1}^n b_k} d\mu_\Theta(\theta) \\ &= \mathbb{P}\left[\{X_1 = b_1, \dots, X_n = b_n\}\right] \end{aligned}$$

This proves that X is an exchangeable sequence. Now assume that X is an exchangeable sequence. For $m \in \mathbb{N}$, let $Y_m := \frac{1}{m} \sum_{k=1}^m X_k$. The strong law of large numbers for exchangeable random variables shows (the X_i are certainly integrable) that Y_n converges almost surely to some random variable Θ. Let μ_Θ be the distribution of Θ and let

$$k_{X,\Theta} : \mathbb{R} \times \mathcal{B}(\mathbb{R}) \to [0,1]$$

be a regular conditional distribution of X given Θ. The goal is to show that for μ_Θ-almost every $\theta \in [0,1]$[2] it holds that

$$k_{X,\Theta}(\theta, .) = \otimes_{k=1}^\infty ((1-\theta)\delta_0 + \theta \delta_1) \quad \text{(the product measure of } \theta\text{-Bernoulli measures)}$$

[2]We can take $[0,1]$ instead of $\mathbb{R}$ because $\mu_\Theta(\mathbb{R} \setminus [0,1]) = 0$.

In other words for μ_Θ-almost every $\theta \in [0,1]$, every $n \in \mathbb{N}$ and every $B = (b_1, \dots, b_n) \in \{0,1\}^n$

$$k_{X,\Theta}(\theta, \{X_1 = b_1, \dots, X_n = b_n\}) = \theta^{\sum_{k=1}^n b_k}(1-\theta)^{1-\sum_{k=1}^n b_k}.$$

This is equivalent to

$$\begin{aligned}
&\forall n \in \mathbb{N}\ \forall C \in \mathcal{B}([0,1])\ \forall B = (b_1, \dots, b_n) \in \{0,1\}^n : \\
&\mathbb{P}\left[\{X_1 = b_1, \dots, X_n = b_n, \Theta \in C\}\right] = \int_C \theta^{\sum_{k=1}^n b_k}(1-\theta)^{n-\sum_{k=1}^n b_i} d\mu_\Theta(\theta)
\end{aligned} \tag{2}$$

In order to prove this last equation define for n, C and B as before for each $m \in \mathbb{N}$

$$Z_m := \mathbb{1}_C(\Theta) Y_m^{\sum_{k=1}^n b_k}(1-Y_m)^{n-\sum_{k=1}^n b_i}$$

and

$$Z := \mathbb{1}_C(\Theta)\Theta^{\sum_{k=1}^n b_k}(1-\Theta)^{n-\sum_{k=1}^n b_k}.$$

Then $Z_m \to Z$ almost surely and therefore in distribution and in L^1 (by dominated convergence). Since $\mathbb{E}[Z]$ is equal to the right-hand side of (2), we only have to prove that $\mathbb{E}[Z_m]$ converges to the left-hand side as $m \to \infty$. For that purpose define for $k \in \{1, \dots, n\}$, $j \in \mathbb{N}$

$$W_{j,k} = \mathbb{1}_{\{b_k\}}(X_j).$$

It follows for all $m \geq n$, $k \in \{1, \dots, n\}$:

$$\frac{1}{m}\sum_{j=1}^m W_{j,k} = \begin{cases} Y_m & \text{if } b_k = 1 \\ 1 - Y_m & \text{if } b_k = 0 \end{cases}.$$

With this notation

$$\begin{aligned}
Z_m &= \mathbb{1}_C(\Theta)\prod_{k=1}^n \left(\frac{1}{m}\sum_{i_k=1}^m W_{i_k,k}\right) = \frac{\mathbb{1}_C(\Theta)}{m^n}\sum_{i_1=1}^m \cdots \sum_{i_n=1}^m \prod_{k=1}^n W_{i_k,k} \\
&= \frac{\mathbb{1}_C(\Theta)}{m^n}\sum_{\text{all } i_k \text{ distinct}} \prod_{k=1}^n W_{i_k,k} + \underbrace{\frac{\mathbb{1}_C(\Theta)}{m^n}\sum_{\text{at least two } i_k \text{ equal}} \prod_{k=1}^n W_{i_k,k}}_{:=e_m}
\end{aligned}$$

and therefore

$$\begin{aligned}
\mathbb{E}[Z_m] &= \frac{1}{m^n}\sum_{\text{all } i_k \text{ distinct}} \mathbb{E}\left[\mathbb{1}_C(\Theta)\prod_{k=1}^n W_{i_k,k}\right] + \mathbb{E}[e_m] \\
&= \frac{1}{m^n}\sum_{\text{all } i_k \text{ distinct}} \mathbb{P}\left[\{X_{i_1} = b_1, \dots, X_{i_n} = b_n, \Theta \in C\}\right] + \mathbb{E}[e_m] \\
&= \frac{1}{m^n}\frac{m!}{(m-n)!}\mathbb{P}\left[\{X_1 = b_1, \dots, X_n = b_n, \Theta \in C\}\right] + \mathbb{E}[e_m] \\
&\to \mathbb{P}\left[\{X_1 = b_1, \dots, X_n = b_n, \Theta \in C\}\right] \quad \text{as } m \to \infty
\end{aligned}$$

since $|e_m| \leq \frac{1}{m^n}(m^n - \frac{m!}{(m-n)!}) \to 0$ and $\frac{m!}{m^n(m-n)!} \to 1$ as $m \to \infty$. The step from the second to the third line follows because in distribution

$$\begin{aligned}
\text{Law}(X_{i_1}, \dots, X_{i_n}, \Theta) &= \lim_{l\to\infty} \text{Law}(X_{i_1}, \dots, X_{i_n}, Y_l) \\
&= \lim_{l\to\infty} \text{Law}(X_1, \dots, X_n, Y_l) = \text{Law}(X_1, \dots, X_n, \Theta)
\end{aligned}$$

This shows equation (2).
It is still left to show that the distribution of Θ is unique if there exists a Θ such that

$$\begin{aligned}&\forall n \in \mathbb{N}\, \forall C \in \mathcal{B}([0,1])\, \forall B = (b_1, \ldots, b_n) \in \{0,1\}^n :\\ &\mathbb{P}\left[\{X_1 = b_1, \ldots, X_n = b_n, \Theta \in C\}\right] = \int_C \theta^{\sum_{k=1}^n b_k}(1-\theta)^{n-\sum_{k=1}^n b_i} d\mu_\Theta(\theta).\end{aligned}$$

Apply this formula for $b_1 = \cdots = b_n = 1$ and $C = [0,1]$ to see that

$$\forall n \in \mathbb{N} \quad \int_{[0,1]} \theta^n d\mu_\Theta(\theta) = \mathbb{P}\left[\{X_1 = 1, \ldots, X_n = 1\}\right].$$

Thus, the integral of every polynomial with respect to the measure μ_Θ is uniquely determined and because each continuous function on $[0,1]$ is bounded can be uniformly approximated by polynomials (Stone-Weierstrass), the integral over each bounded continuous function is uniquely determined, hence the distribution of Θ. □

If finitely many exchangeable random variables $X_1, \ldots, X_n$ are given it is in general not possible to find infinitely many exchangeable random variables such that every n of them have the same joint distribution as $X_1, \ldots, X_n$. In other words in the finite case there are more ways of being exchangeable than in the infinite case. Finitely many exchangeable Bernoulli random variables need not have a conditional i.i.d. Bernoulli(θ) distribution given $\Theta = \theta$ for some random variable Θ.

Example 13. *Consider* 9 *white balls and* 1 *black ball in an urn. Now draw all of them without replacement and let*

$$X_i = \begin{cases} 1 & \textit{if the i-th ball is black} \\ 0 & \textit{if the i-th ball is white} \end{cases}$$

for $i \in 1, \ldots, 10$. *Under the assumption that all possible permutations are equally likely, the random variables* $X_1, \ldots, X_{10}$ *are exchangeable. However, there does not exist a random variable* Θ *such that the* $X_1, \ldots, X_{10}$ *are conditionally i.i.d. Bernoulli*(θ) *given* $\Theta = \theta$. *To see this, assume it by contradiction. Since* $0 < \frac{1}{10} = \mathbb{P}\left[\{X_1 = 1\}\right] = \mathbb{P}\left[\mathbb{P}\left[\{X_1 = 1\}\,|\Theta\right]\right]$, *it follows that* $\mathbb{P}\left[\{X_1 = 1\}\,|\Theta\right] = 0$ *almost surely is impossible. Therefore*

$$\begin{aligned}\mathbb{P}\left[\{\mathbb{P}\left[\{X_1 = 1, X_2 = 1\}\,|\Theta\right] > 0\}\right] &= \mathbb{P}\left[\{\mathbb{P}\left[\{X_1 = 1\}\,|\Theta\right]^2 > 0\}\right] \\ &= \mathbb{P}\left[\{\mathbb{P}\left[\{X_1 = 1|\Theta\}\right] > 0\}\right] > 0\end{aligned}$$

which implies that $\mathbb{P}\left[\{X_1 = 1, X_2 = 1\}\right] > 0$ *but it is impossible to see two black balls in a sample without replacement if there is only one in the urn.*

3.2 General De Finetti's Representation Theorem

For the general case, additional notions need to be introduced first. In the following, let $(\mathcal{X}, \mathcal{B})$ be a Borel space (see appendix). Let $\mathcal{P}$ be the set of all probability measures on $(\mathcal{X}, \mathcal{B})$.

We will need a random variable with values in $\mathcal{P}$. Therefore it is necessary to introduce a sigma algebra $\mathcal{C}_\mathcal{P}$ on $\mathcal{P}$ such that $(\mathcal{P}, \mathcal{C}_\mathcal{P})$ is a measurable space. For $B \in \mathcal{B}$ and $t \in [0,1]$ define

$$A_{B,t} := \{\mathbb{P} \in \mathcal{P} \mid \mathbb{P}(B) \leq t\}$$

and

$$\mathcal{C}_{\mathcal{P}} := \sigma\left(\left\{A_{b,t} \in 2^{\mathcal{P}} | B \in \mathcal{B}, t \in [0,1]\right\}\right),$$

where $2^{\mathcal{P}}$ is the power set of $\mathcal{P}$. Thus, $\mathcal{C}_{\mathcal{P}}$ is the σ-algebra generated by all of the $A_{B,t}$ for $B \in \mathcal{B}$ and $t \in [0,1]$. One can now introduce the notion of a random probability measure.

Definition 14 (random probability measure)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $(\mathcal{X}, \mathcal{B})$ a Borel space and $(\mathcal{P}, \mathcal{C}_{\mathcal{P}})$ defined as above. A random variable $\mathfrak{P}$ from $(\Omega, \mathcal{F})$ to $(\mathcal{P}, \mathcal{C}_{\mathcal{P}})$ is called *random probability measure.*

As an example, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space with random variables $X_1, \ldots, X_N$ mapping to the measurable space $(\mathcal{X}, \mathcal{B})$. Define the map

$$\hat{\mathfrak{P}}_N : \begin{cases} \Omega & \to \mathcal{P} \\ \omega & \mapsto \frac{1}{N} \sum_{k=1}^{N} \delta_{X_k(\omega)} \end{cases} \tag{3}$$

Then $\hat{\mathfrak{P}}_N$ is a random probability measure because for each $t \in [0,1], B \in \mathcal{B}$ it holds that

$$\begin{aligned} \hat{\mathfrak{P}}_N^{-1}(A_{B,t}) &= \hat{\mathfrak{P}}_N^{-1}(\{\mathbb{P} \in \mathcal{P} | \mathbb{P}(B) \leq t\}) \\ &= \left\{\omega \in \Omega | \quad \frac{1}{N} \sum_{k=1}^{N} \mathbb{1}_B(X_k(\omega)) \leq t\right\} \in \mathcal{F}, \end{aligned}$$

since $\frac{1}{N} \sum_{k=1}^{N} \mathbb{1}_B \circ X_k$ is $(\mathcal{F}/\mathcal{B}(\mathbb{R}))$-measurable. Since $\hat{\mathfrak{P}}_N$ is a random variable, it is possible to condition on it. This will be done in the general versions of the De Finetti Representation Theorem.

3.2.1 Finite Version

Let $N \in \mathbb{N}$ and define

$$\mathcal{Q}_N := \left\{P \in \mathcal{P} | P = \frac{1}{N} \sum_{k=1}^{N} \delta_{x_k} \text{ for } x_1, \ldots, x_N \in \mathcal{X}\right\} \subset \mathcal{P}$$

$\hat{\mathfrak{P}}_N$ maps to $\mathcal{Q}_N$. For each $P \in \mathcal{Q}_N$ there exist by definition $x^P = (x_1^P, \ldots, x_N^P) \in \mathcal{X}^N$ such that $P = \sum_{k=1}^{n} \delta_{x_k^P}$. Of course these $(x_1^P, \ldots, x_N^P)$ are not unique but they are unique up to reordering. This allows us to define the distribution of a random sample without replacement from an element of $\mathcal{Q}_N$ for $N \in \mathbb{N}$.

Definition 15. For $N \in \mathbb{N}$, $n \in \{1, \ldots, N\}$, $x = (x_1, \ldots, x_N) \in \mathcal{X}^N$ and $B \in \mathcal{B}^n$ define $H_n^N(B|x)$ to be the probability that n draws without replacement from an urn containing balls labeled $x_1, \ldots, x_N$ form a point in B, i.e.

$$H_n^N(.|x) : \begin{cases} \mathcal{B}^n & \to [0,1] \\ B & \mapsto \frac{1}{\binom{N}{n}} \sum_{\substack{\text{distinct} \\ j_1,\ldots,j_n \in \{1,\ldots,N\}}} \mathbb{1}_B(x_{j_1}, \ldots, x_{j_n}) \end{cases}$$

Remark For N, n, x as above, the projections

$$Y_i : \begin{cases} \mathcal{X}^n & \to \mathcal{X} \\ (a_1, \dots, a_n) & \mapsto a_i \end{cases}, \quad i \in \{1, \dots, n\}$$

are obviously exchangeable random variables from the probability space $(\mathcal{X}^n, \mathcal{B}^n, H_n^N(.|x))$ to $(\mathcal{X}, \mathcal{B})$. In particular for all $B_1, \dots, B_n \in \mathcal{B}$ and all permutations ρ in the symmetric group $S(n)$

$$H_n^N(\{Y_1 \in B_1, \dots, Y_n \in B_n\} |x) = H_n^N(\{Y_{\rho(1)} \in B_1, \dots, Y_{\rho(n)} \in B_n\} |x). \tag{4}$$

Definition 16 (simple random sample without replacement)**.** Let $N \in \mathbb{N}$, $P \in \mathcal{Q}_N$ and $n \in \{1, \dots, N\}$. The *distribution of a simple random sample of length* n from P is the distribution of the labels of an ordered draw of n balls out of N balls labeled $x_1^P, \dots, x_N^P$ where each of the $\frac{N!}{(N-n)!}$ possibilities has equal probability. It is a distribution on B^N and nothing more than $H_n^N(.|x^P)$.

The following lemma from [1] will be used in the proof of the general finite version of the De Finetti Theorem 18.

Lemma 17. *Let $X = (X_1, \dots, X_N)$ be a vector of exchangeable random variables from $(\Omega, \mathcal{F}, \mathbb{P})$ to $(\mathcal{X}, \mathcal{B})$. Let μ_X be the probability measure induced by X on $(\mathcal{X}^N, \mathcal{B}^N)$. For each $B \in \mathcal{B}^n$, $C \in \mathcal{B}^N$ and each choice of distinct $i_1, \dots, i_n \in \{1, \dots, N\}$ it holds*

$$\mathbb{P}\left[\{(X_{i_1}, \dots, X_{i_n}) \in B, X \in C\}\right] = \int_C H_n^N(B|x)\, d\mu_X(x)$$

Proof. By exchangeability

$$\begin{aligned} &\mathbb{P}\left[\{(X_{i_1}, \dots, X_{i_n}) \in B, X \in C\}\right] \\ = \;& \frac{1}{\binom{N}{n}} \sum_{\substack{\text{distinct} \\ (j_1,\dots,j_n) \in \{1,\dots,N\}}} \mathbb{P}\left[\{(X_{j_1}, \dots, X_{j_n}) \in B, X \in C\}\right] \\ = \;& \frac{1}{\binom{N}{n}} \sum_{\substack{\text{distinct} \\ (j_1,\dots,j_n) \in \{1,\dots,N\}}} \int_C \mathbb{1}_B(x_{j_1}, \dots, x_{j_n}) d\mu_X(x) \\ = \;& \int_C H_n^N(B|x)\, d\mu_X(x) \end{aligned}$$

□

Theorem 18. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $(\mathcal{X}, \mathcal{B})$ a Borel space and let $X_1, \dots, X_N$ be random variables from Ω to $\mathcal{X}$. Let $\hat{\mathfrak{P}}_N$ be defined as in (3). The random variables $(X_1, \dots, X_N)$ are exchangeable if and only if for every ordered n-tuple $(i_1, \dots, i_n)$ of distinct elements of $\{1, \dots, N\}$, the joint distribution of $(X_{i_i}, \dots, X_{i_n})$ conditional on $\hat{\mathfrak{P}}_N = P$ is that of a simple random sample without replacement from the distribution of P.*

Proof. For the "if" direction assume that for every ordered n-tuple $(i_1, \dots, i_n)$ of distinct elements of $\{1, \dots, N\}$, the joint distribution of $(X_{i_1}, \dots, X_{i_n})$ conditional on $\hat{\mathfrak{P}}_n = P$ is that of a simple random sample without replacement from the distribution P. We want to show that $(X_1, \dots, X_N)$ are exchangeable. Let $\mu_{\hat{\mathfrak{P}}_N}$ be the distribution of $\hat{\mathfrak{P}}_N$ on $(\mathcal{Q}_N, \mathcal{C}_{\mathcal{P}}|_{\mathcal{Q}})$[3]. The assumptions imply that for $n = N$ and an N-tuple $(\rho(1), \dots, \rho(N))$ of distinct elements $\{1, \dots, N\}$ with $\rho \in S(n)$ we have therefore for $B_1, \dots, B_N \in \mathcal{B}$:

$$\begin{aligned} &\mathbb{P}\left[\{X_1 \in B_1, \dots, X_N \in B_N\}\right] = \mathbb{P}\left[\mathbb{P}\left[\{X_1 \in B_1, \dots, X_N \in B_N\} | \hat{\mathfrak{P}}_N\right]\right] \\ = \; &\int_{\mathcal{Q}_n} H_N^N(\{X_1 \in B_1, \dots, X_N \in B_N\} | x^P) d\mu_{\hat{\mathfrak{P}}_N}(P) \\ = \; &\int_{\mathcal{Q}_n} H_N^N(\{X_{\rho(1)} \in B_1, \dots, X_{\rho(N)} \in B_N\} | x^P) d\mu_{\hat{\mathfrak{P}}_N}(P) \\ = \; &\mathbb{P}\left[\{X_{\rho(1)} \in B_1, \dots, X_{\rho(N)} \in B_N\}\right] \end{aligned}$$

Here, the notation x^P from above is used and the step from the second to the third line follows from equation (4). But this implies that $(X_1, \dots, X_N)$ are exchangeable.
For the "only if" direction assume that the sequence $(X_1, \dots, X_N)$ is exchangeable. We want to show that for all $n \in \mathbb{N}$, $B \in \mathcal{B}^n$, $C \in \mathcal{C}_{\mathcal{P}}|_{\mathcal{Q}_n}$ and all distinct $i_1, \dots, i_n \in \{1, \dots, N\}$ it holds

$$\mathbb{P}\left[\left\{(X_{i_1}, \dots, X_{i_n}) \in B, \hat{\mathfrak{P}}_N \in C\right\}\right] = \int_C H_n^N(B|x^P) d\mu_{\hat{\mathfrak{P}}_N}(P)$$

For that purpose we factorize the map $\hat{\mathfrak{P}}_N : \Omega \to \mathcal{Q}_N$ as $\hat{\mathfrak{P}}_N = h \circ X$ with

$$h : \begin{cases} \mathcal{X}^N & \to \mathcal{Q}_n \\ (x_1, \dots, x_N) & \mapsto \frac{1}{N}\sum_{k=1}^N \delta_{x_k} \end{cases}.$$

One can see as for $\hat{\mathfrak{P}}_N$ that h is $(\mathcal{B}^N / \mathcal{C}_{\mathcal{P}}|_{\mathcal{Q}_N})$-measurable. Now it follows

$$\begin{aligned} \mathbb{P}\left[\left\{(X_{i_1}, \dots, X_{i_n}) \in B, \hat{\mathfrak{P}}_N \in C\right\}\right] &= \mathbb{P}\left[\{(X_{i_1}, \dots, X_{i_n}) \in B, X \in h^{-1}(C)\}\right] \\ &= \int_{h^{-1}(C)} H_n^N(B|x) d\mu_X(x) \\ &= \int_C H_n^N(B|x^P) d\mu_{\hat{\mathfrak{P}}_N}(P) \end{aligned}$$

The step from the first to the second line follows from Lemma 17 and the step from the second to the third line follows because $\mu_{\hat{\mathfrak{P}}_N}$ is the image of the measure μ_X under h.
□

3.2.2 Infinite Version

The infinite version for the general case will not be proved here. The reader may find a proof in [2] and [1]

Theorem 19. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(\mathcal{X}, \mathcal{B})$ be a Borel space. A sequence $(X_1, X_2, \dots)$ of random variables from Ω to $\mathcal{X}$ is exchangeable if and only if there is a random probability measure $\mathfrak{P} : \Omega \mapsto \mathcal{P}$ such that conditional on $\mathfrak{P} = P$, the $X_i, i \in \mathbb{N}_{>0}$ are i.i.d. with distribution P. In this case the distribution of $\mathfrak{P}$ is unique and $\hat{\mathfrak{P}}_n(B)$ converges almost surely to $\mathfrak{P}(B)$ for each $B \in \mathcal{B}$.*

[3] Let $\mathcal{C}_{\mathcal{P}}|_{\mathcal{Q}} := \{S \cap \mathcal{Q} | S \in \mathcal{C}_{\mathcal{P}}\}$

3.3 Example

This example is due to Bayes (1764) and taken from [1]. More examples can be found there.

Example 20. *Suppose that a sequence of Bernoulli random variables $(X_n)_{n\in\mathbb{N}}$ are distributed in the following way*

$$\forall n \in \mathbb{N}: \ \forall k \in \{0,\dots,n\}: \quad \mathbb{P}\left[\textit{exactly } k \textit{ out of } n \textit{ variables are } 1\right] = \frac{1}{n+1}$$

The first step is to show that with this definition induces a consistent family of probability measures. Then one can see with the Kolmogorov Existence Theorem that the sequence $(X_n)_{n\in\mathbb{N}}$ exists and it is obvious that it is exchangeable. Consistency is implied by

$$\begin{aligned} & \forall n \in \mathbb{N}: \forall (x_1,\dots,x_n) \in \{0,1\}^n: \\ & \mathbb{P}\left[\{X_1 = x_1,\dots,X_n = X_n\}\right] = \frac{1}{(n+1)\binom{n}{\sum_{k=1}^n x_k}} = \\ = \quad & \frac{1}{(n+2)\binom{n+1}{1+\sum_{k=1}^n x_k}} + \frac{1}{(n+2)\binom{n+1}{\sum_{k=1}^n x_k}} \\ = \quad & \mathbb{P}\left[\{X_1 = x_1,\dots,X_n = x_n, X_{n+1} = 1\}\right] + \mathbb{P}\left[\{X_1 = x_1,\dots,X_n = x_n, X_{n+1} = 0\}\right]. \end{aligned}$$

We want to find out what the random probability measure $\mathfrak{P}$ is in this case. From Theorem 19 and Theorem 12, we know that

$$\begin{aligned} \mathfrak{P}(\{1\}) &= \lim_{n\to\infty} \hat{\mathfrak{P}}_n\left[\{1\}\right] = \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^n X_k = \Theta \\ \mathfrak{P}(\{0\}) &= \lim_{n\to\infty} \hat{\mathfrak{P}}_n\left[\{0\}\right] = \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^n (1 - X_k) = 1 - \Theta \end{aligned}$$

and that the limits exist almost surely. This means

$$\mathfrak{P} = \delta_0 + \left(\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^n X_k\right)\delta_1,$$

which may not be defined on a nullset. One can also determine the distribution of $\mathfrak{P}$, which is equivalent to determining the distribution of Θ. For notational convenience, the latter will be done. For $t \in [0,1]$, it holds

$$\begin{aligned} \lim_{n\to\infty} \mathbb{P}\left[\frac{1}{n}\sum_{k=1}^n X_k \le t\right] &= \lim_{n\to\infty} \mathbb{P}\left[\textit{at most } \lfloor nt \rfloor \textit{ out of } n \textit{ variables are } 1\right] \\ &= \lim_{n\to\infty} \frac{\lfloor nt \rfloor + 1}{n+1} = t \end{aligned}$$

This implies that the $\Theta \sim$ Uniformly$(0,1)$ and we know that the $(X_1, X_2, \dots)$ are conditional i.i.d. Bernoulli(θ) given $\Theta = \theta$.

A Appendix

Definition 21 (Borel Space)**.** A measurable space $(X, \mathcal{A})$ is called *Borel space* if there exists a bimeasurable map $\varphi : X \mapsto (C, \mathcal{C})$, where C is a Borel subset of $\mathbb{R}$ endowed with the σ-algebra $\mathcal{C}$ induced by $\mathcal{B}(\mathbb{R})$.

Definition 22 (Polish Space)**.** A complete and separable metric space is called *Polish space.*

Examples are $\mathbb{R}$ or $\mathbb{R}^n$ for $n \in \mathbb{N}$.

Lemma 23. *If X is Polish space with Borel σ-algebra $\mathcal{B}$ then $(X, \mathcal{B})$ is a Borel space.*

Theorem 24 (Conditional Expectation)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and X a random variable in $\mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$. Let $\mathcal{C}$ be a sub-σ-algebra of $\mathcal{F}$. Then there exists a $\mathcal{C}$-measurable random variable $g \in \mathcal{L}^1(\Omega, \mathcal{G}, \mathbb{P})$ which satisfies*

$$\forall C \in \mathcal{C} : \quad \mathbb{E}\left[X \mathbb{1}_C\right] = \mathbb{E}\left[g \mathbb{1}_C\right] \tag{5}$$

Every such $\mathcal{C}$-measurable function satisfying (5) is called a version of the conditional expectation and will be denoted by denoted by $\mathbb{E}\left[X|\mathcal{G}\right]$.

Lemma 25 (Factorization Lemma)**.** *Let $(\Omega_1, \mathcal{F}_1)$, $(\Omega_2, \mathcal{F}_2)$ and $(\Omega_3, \mathcal{F}_3)$ be measurable spaces. Assume that $\mathcal{F}_3$ contains all singletons:*

$$\forall \omega \in \Omega_3 : \quad \{\omega\} \in \mathcal{F}_3$$

For a $(\mathcal{F}_1/\mathcal{F}_2)$-measurable function $f : \Omega_1 \to \Omega_2$ and a $(\sigma(f)/\mathcal{F}_3)$-measurable function $g : \Omega_1 \to \Omega_3$ there exists a factorization of g in the following sense:

$$\exists \ (\mathcal{F}_2/\mathcal{F}_3)\text{-measurable function } h : \Omega_2 \to \Omega_3 \text{ such that} \quad g = h \circ f.$$

The factorization lemma can be applied to conditional expectation.

Corollary 26. *Let $(\Omega_1, \mathcal{F}_1, \mathbb{P})$ be a probability space and let $(\Omega_2, \mathcal{F}_2)$ be a measurable space. For a random variable $X \in \mathcal{L}^1(\Omega_1, \mathcal{F}_1, \mathbb{P})$ and a $(\mathcal{F}_1/\mathcal{F}_2)$-measurable function $Y : \Omega_1 \to \Omega_2$ there exists a $(\mathcal{F}_2/\mathcal{B}(\mathbb{R}))$-measurable function h such that for a version g of the conditional $\mathbb{E}[X|\sigma(Y)]$ it holds*

$$g = h \circ Y$$

In particular, one has[4]

$$\forall B \in \mathcal{F}_2 : \quad \mathbb{E}\left[X \mathbb{1}_{Y^{-1}(B)}\right] = \int_B h d(Y_* \mathbb{P}). \tag{6}$$

Denote any such function h which is $(\mathcal{F}_2/\mathcal{B}(\mathbb{R}))$-measurable and satisfies (6) by

$$\mathbb{E}\left[X|Y = \cdot\right] : \begin{cases} \Omega_2 & \to \mathbb{R} \\ y & \mapsto \mathbb{E}\left[X|Y = y\right] := h(y). \end{cases}$$

[4] Let $(Y_* P)$ be the image measure defined by

$$(Y_* \mathbb{P}) : \begin{cases} \mathcal{F}_2 & \to [0, 1] \\ F & \mapsto \mathbb{P}(Y^{-1}(B)). \end{cases}$$

Definition 27 (Stochastik Kernel)**.** Let $(\Omega_1, \mathcal{F}_1)$, $(\Omega_2, \mathcal{F}_2)$ be measurable spaces. A function $k : \Omega_1 \times \mathcal{F}_2 \to [0,1]$ is called *stochastic kernel* from $(\Omega_1, \mathcal{F}_1)$ to $(\Omega_2, \mathcal{F}_2)$ if it satisfies the following two conditions:

- $\forall F_2 \in \mathcal{F}_2$, the function $\begin{cases} \Omega_1 & \to \mathbb{R} \\ \omega_1 & \mapsto k(\omega_1, F_2) \end{cases}$ is $(\mathcal{F}_1, \mathcal{B}(\mathbb{R}))$-measurable.
- $\forall \omega_1 \in \Omega_1$, the function $\begin{cases} \mathcal{F}_2 & \to [0,1] \\ F_2 & \mapsto k(\omega_1, F_2) \end{cases}$ is a probability measure on $(\Omega_2, \mathcal{F}_2)$.

Definition 28 (regular conditional distribution)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(E, \mathcal{E})$ be a measurable space. Further, let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. A stochastic kernel $k_{X,\mathcal{G}}$ from $(\Omega, \mathcal{F})$ to $(E, \mathcal{E})$ is called *regular conditional distribution* if it satisfies

$$\forall B \in \mathcal{E} : \quad \mathbb{E}\left[\mathbb{1}_{X^{-1}(B)} | \mathcal{G}\right] = k_{X,\mathcal{G}}(\cdot, B) \ \mathbb{P}\text{-almost surely.}$$

One can show that if $(E, \mathcal{E})$ from the previous definition is a Borel space, then a regular conditional distribution of X exists.

Definition 29 (Conditional Independence)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $\mathcal{A}$ a sub-σ-algebra of $\mathcal{F}$ and $(\mathcal{A}_i)_{i \in I}$ a collection of sub-σ-algebras of $\mathcal{F}$. The collection $(A_i)_{i \in \mathbb{I}}$ is called *conditionally independent given* $\mathcal{A}$ if for each finite $J \subset I$ and each choice of $A_j \in \mathcal{A}_j$ for $j \in J$, it holds

$$\mathbb{P}\left[\bigcap_{j \in J} A_j | \mathcal{A}\right] = \prod_{j \in J} \mathbb{P}\left[A_j | \mathcal{A}\right] \ \mathbb{P}\text{-almost surely.}$$

A family $(X_i)_i \in I$ of random variables from $(\Omega, \mathcal{F}, \mathbb{P})$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is called conditionally i.i.d. given $\mathcal{A}$ if the σ-algebras $(\sigma(X_i))_{i \in I}$ are conditionally independent given $\mathcal{A}$ and the conditional distributions $\mathbb{P}[X_i \in \cdot | \mathcal{A}]$ coincide pairwise almost surely.

References

[1] Mark J. Schervish. Theory of Statistics. Springer Series in Statistics, 1995.

[2] Achim Klenke. Probability Theory *A comprehensive course*. Springer, 2008.