Chapter 12
Backwards Martingales and Exchangeability

With many data acquisitions, such as telephone surveys, the order in which the data come does not matter. Mathematically, we say that a family of random variables is *exchangeable* if the joint distribution does not change under finite permutations. De Finetti's structural theorem says that an infinite family of E-valued exchangeable random variables can be described by a two-stage experiment. At the first stage, a probability distribution Ξ on E is drawn at random. At the second stage, i.i.d. random variables with distribution Ξ are implemented.

We first define the notion of exchangeability. Then we consider backwards martingales and prove the convergence theorem for them. This is the cornerstone for the proof of de Finetti's theorem.

12.1 Exchangeable Families of Random Variables

Definition 12.1 Let I be an arbitrary index set and let E be a Polish space. A family $(X_i)_{i\in I}$ of random variables with values in E is called *exchangeable* if

$$\mathcal{L}\big[(X_{\varrho(i)})_{i\in I}\big] = \mathcal{L}\big[(X_i)_{i\in I}\big]$$

for any finite permutation $\varrho : I \to I$.

Recall that a finite permutation is a bijection $\varrho : I \to I$ that leaves all but finitely many points unchanged.

Remark 12.2 Clearly, the following are equivalent.

(i) $(X_i)_{i\in I}$ is exchangeable.
(ii) Let $n \in \mathbb{N}$ and assume $i_1, \ldots, i_n \in I$ are pairwise distinct and $j_1, \ldots, j_n \in I$ are pairwise distinct. Then we have $\mathcal{L}[(X_{i_1}, \ldots, X_{i_n})] = \mathcal{L}[(X_{j_1}, \ldots, X_{j_n})]$.

In particular ($n = 1$), exchangeable random variables are identically distributed. ◊

A. Klenke, *Probability Theory*, Universitext, DOI 10.1007/978-1-4471-5361-0_12,

Example 12.3

(i) If $(X_i)_{i\in I}$ is i.i.d., then $(X_i)_{i\in I}$ is exchangeable.

(ii) Consider an urn with N balls, M of which are black. Successively draw without replacement all of the balls and define

$$X_n := \begin{cases} 1, & \text{if the } n\text{th ball is black,} \\ 0, & \text{else.} \end{cases}$$

Then $(X_n)_{n=1,\ldots,N}$ is exchangeable. Indeed, this follows by elementary combinatorics since for any choice $x_1,\ldots,x_N \in \{0,1\}$ with $x_1+\ldots+x_N = M$, we have

$$\mathbf{P}[X_1 = x_1,\ldots,X_N = x_N] = \frac{1}{\binom{N}{M}}.$$

This formula can be derived formally via a small computation with conditional probabilities. As we will need a similar computation for Pólya's urn model in Example 12.29, we give the details here. Let $s_k = x_1+\ldots+x_k$ for $k = 0,\ldots,N$ and let

$$g_k(x) = \begin{cases} M - s_k, & \text{if } x = 1, \\ N - M + s_k - k, & \text{if } x = 0. \end{cases}$$

Then $\mathbf{P}[X_1 = x_1] = g_0(x_1)/N$ and

$$\mathbf{P}[X_{k+1} = x_{k+1} \mid X_1 = x_1,\ldots,X_k = x_k] = \frac{g_k(x_{k+1})}{N-k} \quad \text{for } k = 1,\ldots,N-1.$$

Clearly, $g_k(0) = N - M - l$, where $l = \#\{i \le k : x_i = 0\}$. Therefore,

$$\begin{aligned} &\mathbf{P}[X_1 = x_1,\ldots,X_N = x_N] \\ &= \mathbf{P}[X_1 = x_1] \prod_{k=1}^{N-1} \mathbf{P}[X_{k+1} = x_{k+1} \mid X_1 = x_1,\ldots,X_k = x_k] \\ &= \frac{1}{N!} \prod_{k=0}^{N-1} g_k(x_{k+1}) = \frac{1}{N!} \prod_{k:x_k=1} g_k(1) \prod_{k:x_k=0} g_k(0) \\ &= \frac{1}{N!} \prod_{l=0}^{M-1} (M-l) \prod_{l=0}^{N-1} (N-M-l) = \frac{M!(N-M)!}{N!}. \end{aligned}$$

(iii) Let Y be a random variable with values in $[0,1]$. Assume that, given Y, the random variables $(X_i)_{i\in I}$ are independent and Ber_Y-distributed. That is, for any finite $J \subset I$,

$$\mathbf{P}[X_j = 1 \text{ for all } j \in J \mid Y] = Y^{\#J}.$$

Then $(X_i)_{i\in I}$ is exchangeable. ◊

Let $X = (X_n)_{n\in\mathbb{N}}$ be a stochastic process with values in a Polish space E. Let $S(n)$ be the set of permutations $\varrho : \{1,\ldots,n\} \to \{1,\ldots,n\}$. We consider ϱ also as a map $\mathbb{N} \to \mathbb{N}$ by defining $\varrho(k) = k$ for $k > n$. For $\varrho \in S(n)$ and $x = (x_1,\ldots,x_n) \in E^n$, denote $x^\varrho = (x_{\varrho(1)},\ldots,x_{\varrho(n)})$. Similarly, for $x \in E^{\mathbb{N}}$, denote $x^\varrho = (x_{\varrho(1)}, x_{\varrho(2)},\ldots) \in E^{\mathbb{N}}$. Let E' be another Polish space. For measurable maps $f : E^n \to E'$ and $F : E^{\mathbb{N}} \to E'$, define the maps f^ϱ and F^ϱ by $f^\varrho(x) = f(x^\varrho)$ and $F^\varrho(x) = F(x^\varrho)$. Further, we write $f(x) = f(x_1,\ldots,x_n)$ for $x \in E^n$ and for $x \in E^{\mathbb{N}}$.

Definition 12.4

(i) A map $f : E^n \to E'$ is called *symmetric* if $f^\varrho = f$ for all $\varrho \in S(n)$.
(ii) A map $F : E^{\mathbb{N}} \to E'$ is called n-symmetric if $F^\varrho = F$ for all $\varrho \in S(n)$. F is called symmetric if F is n-symmetric for all $n \in \mathbb{N}$.

Example 12.5

(i) For $x \in \mathbb{R}^{\mathbb{N}}$, define the nth arithmetic mean by $a_n(x) = \frac{1}{n}\sum_{i=1}^n x_i$. Clearly, a_n is an n-symmetric map (but not m-symmetric for any $m > n$). Furthermore, $\bar{a}(x) := \limsup_{n\to\infty} a_n(x)$ defines a symmetric map $\mathbb{R}^{\mathbb{N}} \to \mathbb{R} \cup \{-\infty, +\infty\}$.
(ii) The map $s : \mathbb{R}^{\mathbb{N}} \to [0,\infty]$, $x \mapsto \sum_{i=1}^\infty |x_i|$ is symmetric. Unlike $\bar{a}$, the value of s depends on every coordinate if it is finite.
(iii) For $x \in E^{\mathbb{N}}$, define the nth empirical distribution by $\xi_n(x) = \frac{1}{n}\sum_{i=1}^n \delta_{x_i}$ (recall that δx_i is the Dirac measure at the point x_i). Clearly, ξ_n is an n-symmetric map.
(iv) Let $k \in \mathbb{N}$ and let $\varphi : E^k \to \mathbb{R}$ be a map. The nth symmetrized average

$$A_n(\varphi) : E^{\mathbb{N}} \to \mathbb{R}, \quad x \mapsto \frac{1}{n!}\sum_{\varrho\in S(n)} \varphi\big(x^\varrho\big) \tag{12.1}$$

is an n-symmetric map. ◊

Definition 12.6 Let $X = (X_n)_{n\in\mathbb{N}}$ be a stochastic process with values in E. For $n \in \mathbb{N}$, define

$$\mathcal{E}'_n := \sigma\big(F : F : E^{\mathbb{N}} \to \mathbb{R} \text{ is measurable and } n\text{-symmetric}\big)$$

and let $\mathcal{E}_n := X^{-1}(\mathcal{E}'_n)$ be the σ-algebra of events that are invariant under all permutations $\varrho \in S(n)$. Further, let

$$\mathcal{E}' := \bigcap_{n=1}^\infty \mathcal{E}'_n = \sigma\big(F : F : E^{\mathbb{N}} \to \mathbb{R} \text{ is measurable and symmetric}\big)$$

and let $\mathcal{E}_n := \bigcap_{n=1}^\infty \mathcal{E}_n = X^{-1}(\mathcal{E}')$ be the σ-algebra of exchangeable events for X, or briefly the *exchangeable σ-algebra*.

Remark 12.7 If $A \in \sigma(X_n, n \in \mathbb{N})$ is an event, then there is a measurable $B \subset E^{\mathbb{N}}$ with $A = \{X \in B\}$. If we denote $A^{\varrho} = \{X^{\varrho} \in B\}$ for $\varrho \in S(n)$, then $\mathcal{E}_n = \{A : A^{\varrho} = A \text{ for all } \varrho \in S(n)\}$. This justifies the name "exchangeable event". ◊

Remark 12.8 If we write $\Xi_n(\omega) := \xi_n(X(\omega)) = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i(\omega)}$ for the nth empirical distribution, then, by Exercise 12.1.1, $\mathcal{E}_n = \sigma(\Xi_n)$. ◊

Remark 12.9 Denote by $\mathcal{T} = \bigcap_{n\in\mathbb{N}} \sigma(X_{n+1}, X_{n+2}, \ldots)$ the tail σ-algebra. Then $\mathcal{T} \subset \mathcal{E}$, and strict inclusion is possible.

Indeed, evidently $\sigma(X_{n+1}, X_{n+2}, \ldots) \subset \mathcal{E}_n$ for $n \in \mathbb{N}$; hence $\mathcal{T} \subset \mathcal{E}$. Now let $E = \{0, 1\}$ and let $X_1, X_2, \ldots$ be independent random variables with $\mathbf{P}[X_n = 1] \in (0, 1)$ for all $n \in \mathbb{N}$. The random variable $S := \sum_{n=1}^{\infty} X_n$ is measurable with respect to $\mathcal{E}$ but not with respect to $\mathcal{T}$. ◊

Theorem 12.10 *Let $X = (X_n)_{n\in\mathbb{N}}$ be exchangeable. If $\varphi : E^{\mathbb{N}} \to \mathbb{R}$ is measurable and if $\mathbf{E}[|\varphi(X)|] < \infty$, then for all $n \in \mathbb{N}$ and all $\varrho \in S(n)$,*

$$\mathbf{E}\big[\varphi(X) \,\big|\, \mathcal{E}_n\big] = \mathbf{E}\big[\varphi\big(X^{\varrho}\big) \,\big|\, \mathcal{E}_n\big]. \tag{12.2}$$

In particular,

$$\mathbf{E}\big[\varphi(X) \,\big|\, \mathcal{E}_n\big] = A_n(\varphi) := \frac{1}{n!}\sum_{\varrho\in S(n)} \varphi\big(X^{\varrho}\big). \tag{12.3}$$

Proof Let $A \in \mathcal{E}_n$. Then there exists a $B \in \mathcal{E}'_n$ such that $A = X^{-1}(B)$. Let $F = \mathbb{1}_B$. Then $F \circ X = \mathbb{1}_A$. By the definition of $\mathcal{E}_n$, the map $F : E^{\mathbb{N}} \to \mathbb{R}$ is measurable, n-symmetric and bounded. Therefore,

$$\mathbf{E}\big[\varphi(X)F(X)\big] = \mathbf{E}\big[\varphi\big(X^{\varrho}\big)F\big(X^{\varrho}\big)\big] = \mathbf{E}\big[\varphi\big(X^{\varrho}\big)F(X)\big].$$

Here we used the exchangeability of X in the first equality and the symmetry of F in the second equality. From this (12.2) follows. However, $A_n(\varphi)$ is $\mathcal{E}_n$-measurable and hence

$$\mathbf{E}\big[\varphi(X) \,\big|\, \mathcal{E}_n\big] = \mathbf{E}\Bigg[\frac{1}{n!}\sum_{\varrho\in S(n)} \varphi\big(X^{\varrho}\big) \,\Bigg|\, \mathcal{E}_n\Bigg] = \frac{1}{n!}\sum_{\varrho\in S(n)} \varphi\big(X^{\varrho}\big).$$ □

Heuristic for the Structure of Exchangeable Families

Consider a finite exchangeable family $X_1, \ldots, X_N$ of E-valued random variables. For $n \le N$, what is the conditional distribution of $(X_1, \ldots, X_n)$ given Ξ_N? For any measurable $A \subset E$, $\{X_i \in A\}$ occurs for exactly $N\,\Xi_N(A)$ of the $i \in \{1, \ldots, N\}$, where the order does not change the probability. Hence we are in the situation of drawing colored balls *without* replacement. More precisely, let the pairwise distinct points $e_1, \ldots, e_k \in E$ be the atoms of Ξ_N and let $N_1, \ldots, N_k$ be the corresponding absolute frequencies. Hence $\Xi_N = \sum_{i=1}^{k} (N_i/N)\delta_{e_i}$. We thus deal with balls of k different colors and with N_i balls of the ith color. We draw n of these balls without replacement but respecting the order. Up to the order, the resulting distribution is thus the generalized hypergeometric distribution (see (1.19) on page 44). Hence, for pairwise disjoint, measurable sets $A_1, \ldots, A_k$ with $\biguplus_{l=1}^{k} A_l = E$, for $i_1, \ldots, i_n \in \{1, \ldots, k\}$, pairwise distinct $j_1, \ldots, j_n \in \{1, \ldots, N\}$ and with the convention $m_l := \#\{r \in \{1, \ldots, n\} : i_r = l\}$ for $l \in \{1, \ldots, k\}$, we have

$$\mathbf{P}[X_{j_r} \in A_{i_r} \text{ for all } r = 1, \ldots, n \mid \Xi_N] = \frac{1}{(N)_n} \prod_{l=1}^{k} \big(N\,\Xi_N(A_l)\big)^{m_l}. \tag{12.4}$$

Here we defined $(n)_l := n(n-1)\ldots(n-l+1)$.

What happens if we let $N \to \infty$? For simplicity, assume that for all $l = 1, \ldots, k$, the limit $\Xi_\infty(A_l) = \lim_{N\to\infty} \Xi_N(A_l)$ exists in a suitable sense. Then (12.4) formally becomes

$$\mathbf{P}[X_{j_r} \in A_{i_r} \text{ for all } r = 1, \ldots, n \mid \Xi_\infty] = \prod_{l=1}^{k} \Xi_\infty(A_l)^{m_l}. \tag{12.5}$$

Drawing without replacements thus asymptotically turns into drawing *with* replacements. Hence the random variables $X_1, X_2, \ldots$ are independent with distribution Ξ_∞ given Ξ_∞.

For a formal proof along the lines of this heuristic, see Section 13.4.

In order to formulate (and prove) this statement (de Finetti's theorem) rigorously in Section 12.3, we need some more technical tools (e.g., the notion of conditional independence). A further tool will be the convergence theorem for backwards martingales that will be formulated in Section 12.2. For further reading on exchangeable random variables, we refer to [4, 33, 98, 105].

Exercise 12.1.1 *Let $n \in \mathbb{N}$. Show that every symmetric function $f : E^n \to \mathbb{R}$ can be written in the form $f(x) = g(\frac{1}{n}\sum_{i=1}^{n} \delta_{x_i})$, where g has to be chosen appropriately (depending on f).*

Exercise 12.1.2 *Derive equation (12.4) formally.*

Exercise 12.1.3 *Let $X_1, \ldots, X_n$ be exchangeable, square integrable random variables. Show that*

$$\mathbf{Cov}[X_1, X_2] \geq -\frac{1}{n-1}\mathbf{Var}[X_1]. \tag{12.6}$$

For $n \geq 2$, give a nontrivial example for equality in (12.6).

Exercise 12.1.4 *Let $X_1, X_2, X_3 \ldots$ be exchangeable, square integrable random variables. Show that* $\mathbf{Cov}[X_1, X_2] \geq 0$.

Exercise 12.1.5 *Show that for all $n \in \mathbb{N} \setminus \{1\}$, there is an exchangeable family of random variables $X_1, \ldots, X_n$ that cannot be extended to an infinite exchangeable family $X_1, X_2, \ldots$.*

12.2 Backwards Martingales

The concepts of filtration and martingale do not require the index set I (interpreted as time) to be a subset of $[0, \infty)$. Hence we can consider the case $I = -\mathbb{N}_0$.

Definition 12.11 (Backwards martingale) Let $\mathbb{F} = (\mathcal{F}_n)_{n \in -\mathbb{N}_0}$ be a filtration and let $X = (X_n)_{n \in -\mathbb{N}_0}$ be an $\mathbb{F}$-martingale. Then $X = (X_{-n})_{n \in \mathbb{N}_0}$ is called a *backwards martingale*.

Remark 12.12 A backwards martingale is always uniformly integrable. This follows from Corollary 8.22 and the fact that $X_{-n} = \mathbf{E}[X_0 \mid \mathcal{F}_{-n}]$ for any $n \in \mathbb{N}_0$. ◊

Example 12.13 Let $X_1, X_2, \ldots$ be exchangeable real random variables. For $n \in \mathbb{N}$, let $\mathcal{F}_{-n} = \mathcal{E}_n$ and

$$Y_{-n} = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

We show that $(Y_{-n})_{n \in \mathbb{N}}$ is an $\mathbb{F}$-backwards martingale. Clearly, Y is adapted. Furthermore, by Theorem 12.10 (with $k = n$ and $\varphi(X_1, \ldots, X_n) = \frac{1}{n-1}(X_1 + \ldots + X_{n-1})$),

$$\mathbf{E}[Y_{-n+1} \mid \mathcal{F}_{-n}] = \frac{1}{n!} \sum_{\varrho \in S(n)} \frac{1}{n-1} (X_{\varrho(1)} + \ldots + X_{\varrho(n-1)}) = Y_{-n}.$$

Now replace $\mathbb{F}$ by the smaller filtration $\mathbb{G} = (\mathcal{G}_n)_{n \in -\mathbb{N}}$ that is defined by $\mathcal{G}_{-n} = \sigma(Y_{-n}, X_{n+1}, X_{n+2}, \ldots) = \sigma(Y_{-n}, Y_{-n-1}, Y_{-n-2}, \ldots)$ for $n \in \mathbb{N}$. This is the filtration generated by Y; thus Y is also a $\mathbb{G}$-backwards martingale (see Remark 9.29). ◊

Let $a < b$ and $n \in \mathbb{N}$. Let $U_{-n}^{a,b}$ be the number of upcrossings of X over $[a, b]$ between times $-n$ and 0. Further, let $U^{a,b} = \lim_{n \to \infty} U_{-n}^{a,b}$. By the upcrossing inequality (Lemma 11.3), we have $\mathbf{E}[U_{-n}^{a,b}] \le \frac{1}{b-a} \mathbf{E}[(X_0 - a)^+]$; hence $\mathbf{P}[U^{a,b} < \infty] = 1$. As in the proof of the martingale convergence theorem (Theorem 11.4), we infer the following.

Theorem 12.14 (Convergence theorem for backwards martingales) *Let $(X_n)_{n \in -\mathbb{N}_0}$ be a martingale with respect to $\mathbb{F} = (\mathcal{F}_n)_{n \in -\mathbb{N}_0}$. Then there exists $X_{-\infty} = \lim_{n \to \infty} X_{-n}$ almost surely and in L^1. Furthermore, $X_{-\infty} = \mathbf{E}[X_0 \mid \mathcal{F}_{-\infty}]$, where $\mathcal{F}_{-\infty} = \bigcap_{n=1}^{\infty} \mathcal{F}_{-n}$.*

Example 12.15 Let $X_1, X_2, \dots$ be exchangeable, integrable random variables. Further, let $\mathcal{T} = \bigcap_{n=1}^{\infty} \sigma(X_m, m \ge n)$ be the tail σ-algebra of $X_1, X_2, \dots$ and let $\mathcal{E}$ be the exchangeable σ-algebra. Then $\mathbf{E}[X_1 \mid \mathcal{T}] = \mathbf{E}[X_1 \mid \mathcal{E}]$ a.s. and

$$\frac{1}{n}\sum_{i=1}^{n} X_i \stackrel{n\to\infty}{\longrightarrow} \mathbf{E}[X_1 \mid \mathcal{E}] \quad \text{a.s. and in } L^1.$$

Indeed, if we let $Y_{-n} := \frac{1}{n}\sum_{i=1}^{n} X_i$, then (by Example 12.13) $(Y_{-n})_{n\in\mathbb{N}}$ is a backwards martingale with respect to $(\mathcal{F}_n)_{n\in-\mathbb{N}} = (\mathcal{E}_{-n})_{n\in-\mathbb{N}}$ and thus

$$Y_{-n} \stackrel{n\to\infty}{\longrightarrow} Y_{-\infty} = \mathbf{E}[X_1 \mid \mathcal{E}] \quad \text{a.s. and in } L^1.$$

Now, by Example 2.36(ii), $Y_{-\infty}$ is $\mathcal{T}$-measurable; hence (since $\mathcal{T} \subset \mathcal{E}$ and by virtue of the tower property of conditional expectation) $Y_{-\infty} = \mathbf{E}[X_1 \mid \mathcal{T}]$. ◊

Example 12.16 (Strong law of large numbers) If $Z_1, Z_2, \dots$ are real and i.i.d. with $\mathbf{E}[|Z_1|] < \infty$, then

$$\frac{1}{n}\sum_{i=1}^{n} Z_i \stackrel{n\to\infty}{\longrightarrow} \mathbf{E}[Z_1] \quad \text{almost surely.}$$

By Kolmogorov's 0–1 law (Theorem 2.37), the tail σ-algebra $\mathcal{T}$ is trivial; hence we have

$$\mathbf{E}[Z_1 \mid \mathcal{T}] = \mathbf{E}[Z_1] \quad \text{almost surely.}$$

In Corollary 12.19, we will see that in the case of independent random variables, $\mathcal{E}$ is also $\mathbf{P}$-trivial. This implies $\mathbf{E}[Z_1 \mid \mathcal{E}] = \mathbf{E}[Z_1]$. ◊

We close this section with a generalization of Example 12.15 to mean values of functions of $k \in \mathbb{N}$ variables. This conclusion from the convergence theorem for backwards martingales will be used in an essential way in the next section.

Theorem 12.17 *Let $X = (X_n)_{n\in\mathbb{N}}$ be an exchangeable family of random variables with values in E. Assume that $k \in \mathbb{N}$ and let $\varphi : E^k \to \mathbb{R}$ be measurable with $\mathbf{E}[|\varphi(X_1, \dots, X_k)|] < \infty$. Denote $\varphi(X) = \varphi(X_1, \dots, X_k)$ and let $A_n(\varphi) := \frac{1}{n!}\sum_{\varrho\in S(n)} \varphi(X^{\varrho})$. Then*

$$\mathbf{E}\big[\varphi(X) \,\big|\, \mathcal{E}\big] = \mathbf{E}\big[\varphi(X) \,\big|\, \mathcal{T}\big] = \lim_{n\to\infty} A_n(\varphi) \quad \textit{a.s. and in } L^1. \tag{12.7}$$

Proof By Theorem 12.10, $A_n(\varphi) = \mathbf{E}[\varphi(X) \mid \mathcal{E}_n]$. Hence $(A_{-n}(\varphi))_{n\ge k}$ is a backwards martingale with respect to $(\mathcal{E}_{-n})_{n\in-\mathbb{N}}$. Hence, by Theorem 12.14,

$$A_n(\varphi) \stackrel{n\to\infty}{\longrightarrow} \mathbf{E}\big[\varphi(X) \,\big|\, \mathcal{E}\big] \quad \text{a.s. and in } L^1. \tag{12.8}$$

As for the arithmetic mean (Example 12.16), we can argue that $\lim_{n\to\infty} A_n(\varphi)$ is $\mathcal{T}$-measurable. Indeed,

$$\limsup_{n\to\infty} \frac{\#\{\varrho \in S(n) : \varrho^{-1}(i) \le l \text{ for some } i \in \{1, \ldots, k\}\}}{n!} = 0 \quad \text{for all } l \in \mathbb{N}.$$

Thus, for large n, the dependence of $A_n(\varphi)$ on the first l coordinates is negligible. Together with (12.8), we get (12.7). □

Corollary 12.18 *Let $X = (X_n)_{n\in\mathbb{N}}$ be exchangeable. Then, for any $A \in \mathcal{E}$ there exists a $B \in \mathcal{T}$ with $\mathbf{P}[A \triangle B] = 0$.*

Note that $\mathcal{T} \subset \mathcal{E}$; hence the statement is trivially true if the roles of $\mathcal{E}$ and $\mathcal{T}$ are interchanged.

Proof Since $\mathcal{E} \subset \sigma(X_1, X_2, \ldots)$, by the approximation theorem for measures, there exists a sequence of measurable sets $(A_k)_{k\in\mathbb{N}}$ with $A_k \in \sigma(X_1, \ldots, X_k)$ and such that $\mathbf{P}[A \triangle A_k] \stackrel{k\to\infty}{\longrightarrow} 0$. Let $C_k \in E^k$ be measurable with

$$A_k = \big\{(X_1, \ldots, X_k) \in C_k\big\}$$

for all $k \in \mathbb{N}$. Letting $\varphi_k := \mathbb{1}_{C_k}$, Theorem 12.17 implies that

$$\begin{aligned}\mathbb{1}_A = \mathbf{E}[\mathbb{1}_A \mid \mathcal{E}] &= \mathbf{E}\Big[\lim_{k\to\infty} \varphi_k(X) \,\Big|\, \mathcal{E}\Big] = \lim_{k\to\infty} \mathbf{E}\big[\varphi_k(X) \,\big|\, \mathcal{E}\big] \\ &= \lim_{k\to\infty} \mathbf{E}\big[\varphi_k(X) \,\big|\, \mathcal{T}\big] =: \psi \quad \text{almost surely.}\end{aligned}$$

Hence there is a $\mathcal{T}$-measurable function ψ with $\psi = \mathbb{1}_A$ almost surely. We can assume that $\psi = \mathbb{1}_B$ for some $B \in \mathcal{T}$. □

As a further application, we get the 0–1 law of Hewitt and Savage [72].

Corollary 12.19 (0–1 law of Hewitt–Savage) *Let $X_1, X_2, \ldots$ be i.i.d. random variables. Then the exchangeable σ-algebra is $\mathbf{P}$-trivial; that is, $\mathbf{P}[A] \in \{0, 1\}$ for all $A \in \mathcal{E}$.*

Proof By Kolmogorov's 0–1 law (Theorem 2.37), $\mathcal{T}$ is trivial. Hence the claim follows immediately from Corollary 12.18. □

12.3 De Finetti's Theorem

In this section, we show the structural theorem for countably infinite exchangeable families that was heuristically motivated at the end of Section 12.1. Hence we shall show that a countably infinite exchangeable family of random variables is an i.i.d.

family given the exchangeable σ-algebra $\mathcal{E}$. Furthermore, we compute the conditional distribution of the individual random variables. As a first step, we define conditional independence formally (see [25, Chapter 7.3]).

Definition 12.20 (Conditional independence) Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space, let $\mathcal{A} \subset \mathcal{F}$ be a sub-σ-algebra and let $(\mathcal{A}_i)_{i \in I}$ be an arbitrary family of sub-σ-algebras of $\mathcal{F}$. Assume that for any finite $J \subset I$, any choice of $A_j \in \mathcal{A}_j$ and for all $j \in J$,

$$\mathbf{P}\Bigg[\bigcap_{j \in J} A_j \,\Bigg|\, \mathcal{A}\Bigg] = \prod_{j \in J} \mathbf{P}[A_j \mid \mathcal{A}] \quad \text{almost surely.} \tag{12.9}$$

Then the family $(\mathcal{A}_i)_{i \in I}$ is called *independent given* $\mathcal{A}$.

A family $(X_i)_{i \in I}$ of random variables on $(\Omega, \mathcal{F}, \mathbf{P})$ is called independent (and identically distributed) given $\mathcal{A}$ if the generated σ-algebras $(\sigma(X_i))_{i \in I}$ are independent given $\mathcal{A}$ (and the conditional distributions $\mathbf{P}[X_i \in \cdot \mid \mathcal{A}]$ are equal).

Example 12.21 Any family $(\mathcal{A}_i)_{i \in I}$ of sub-σ-algebras of $\mathcal{F}$ is independent given $\mathcal{F}$. Indeed, letting $A = \bigcap_{j \in J} A_j$,

$$\mathbf{P}[A \mid \mathcal{F}] = \mathbb{1}_A = \prod_{j \in J} \mathbb{1}_{A_j} = \prod_{j \in J} \mathbf{P}[A_j \mid \mathcal{F}] \quad \text{almost surely.}$$

◇

Example 12.22 If $(\mathcal{A}_i)_{i \in I}$ is an independent family of σ-algebras and if $\mathcal{A}$ is trivial, then $(\mathcal{A}_i)_{i \in I}$ is independent given $\mathcal{A}$. ◇

Example 12.23 There is no "monotonicity" for conditional independence in the following sense: If $\mathcal{F}_1$, $\mathcal{F}_2$ and $\mathcal{F}_3$ are σ-algebras with $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}_3$ and such that $(\mathcal{A}_i)_{i \in I}$ is independent given $\mathcal{F}_1$ as well as given $\mathcal{F}_3$, then this does not imply independence given $\mathcal{F}_2$.

In order to illustrate this, assume that X and Y are nontrivial independent real random variables. Let $\mathcal{F}_1 = \{\emptyset, \Omega\}$, $\mathcal{F}_2 = \sigma(X + Y)$ and $\mathcal{F}_3 = \sigma(X, Y)$. Then $\sigma(X)$ and $\sigma(Y)$ are independent given $\mathcal{F}_1$ as well as given $\mathcal{F}_3$ but not given $\mathcal{F}_2$. ◇

Let $X = (X_n)_{n \in \mathbb{N}}$ be a stochastic process on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ with values in a Polish space E. Let $\mathcal{E}$ be the exchangeable σ-algebra and let $\mathcal{T}$ be the tail σ-algebra.

Theorem 12.24 (de Finetti) *The family $X = (X_n)_{n \in \mathbb{N}}$ is exchangeable if and only if there exists a σ-algebra $\mathcal{A} \subset \mathcal{F}$ such that $(X_n)_{n \in \mathbb{N}}$ is i.i.d. given $\mathcal{A}$. In this case, $\mathcal{A}$ can be chosen to equal the exchangeable σ-algebra $\mathcal{E}$ or the tail-σ-algebra $\mathcal{T}$.*

Proof "$\Longrightarrow$" Let X be exchangeable and let $\mathcal{A} = \mathcal{E}$ or $\mathcal{A} = \mathcal{T}$. For any $n \in \mathbb{N}$, let $f_n : E \to \mathbb{R}$ be a bounded measurable map. Let

$$\varphi_k(x_1, \ldots, x_k) = \prod_{i=1}^{k} f_i(x_i) \quad \text{for any } k \in \mathbb{N}.$$

Let $A_n(\varphi)$ be the symmetrized average from Theorem 12.17. Then

$$\begin{aligned} A_n(\varphi_{k-1})A_n(f_k) &= \frac{1}{n!} \sum_{\varrho \in S(n)} \varphi_{k-1}\big(X^{\varrho}\big) \frac{1}{n} \sum_{i=1}^{n} f_k(X_i) \\ &= \frac{1}{n!} \sum_{\varrho \in S(n)} \varphi_k\big(X^{\varrho}\big) + R_{n,k} = A_n(\varphi_k) + R_{n,k}, \end{aligned}$$

where

$$\begin{aligned} |R_{n,k}| &\le 2\|\varphi_{k-1}\|_\infty \cdot \|f_k\|_\infty \cdot \frac{1}{n!} \frac{1}{n} \sum_{\varrho \in S(n)} \sum_{i=1}^{n} \mathbb{1}_{\{i \in \{\varrho(1), \ldots, \varrho(k-1)\}\}} \\ &= 2\|\varphi_{k-1}\|_\infty \cdot \|f_k\|_\infty \cdot \frac{k-1}{n} \stackrel{n\to\infty}{\longrightarrow} 0. \end{aligned}$$

Together with Theorem 12.17, we conclude that

$$A_n(\varphi_{k-1})A_n(f_k) \stackrel{n\to\infty}{\longrightarrow} \mathbf{E}\big[\varphi_k(X_1, \ldots, X_k) \,\big|\, \mathcal{A}\big] \quad \text{a.s. and in } L^1.$$

On the other hand, again by Theorem 12.17,

$$A_n(\varphi_{k-1}) \stackrel{n\to\infty}{\longrightarrow} \mathbf{E}\big[\varphi_{k-1}(X_1, \ldots, X_{k-1}) \,\big|\, \mathcal{A}\big]$$

and

$$A_n(f_k) \stackrel{n\to\infty}{\longrightarrow} \mathbf{E}\big[f_k(X_1) \,\big|\, \mathcal{A}\big].$$

Hence

$$\mathbf{E}\big[\varphi_k(X_1, \ldots, X_k) \,\big|\, \mathcal{A}\big] = \mathbf{E}\big[\varphi_{k-1}(X_1, \ldots, X_{k-1}) \,\big|\, \mathcal{A}\big]\mathbf{E}\big[f_k(X_1) \,\big|\, \mathcal{A}\big].$$

Thus we get inductively

$$\mathbf{E}\left[\prod_{i=1}^{k} f_i(X_i) \,\middle|\, \mathcal{A}\right] = \prod_{i=1}^{k} \mathbf{E}\big[f_i(X_1) \,\big|\, \mathcal{A}\big].$$

Therefore, X is i.i.d. given $\mathcal{A}$.

"$\Longleftarrow$" Now let X be i.i.d. given $\mathcal{A}$ for a suitable σ-algebra $\mathcal{A} \subset \mathcal{F}$. For any bounded measurable function $\varphi : E^n \to \mathbb{R}$ and for any $\varrho \in S(n)$, we have $\mathbf{E}[\varphi(X) \mid \mathcal{A}] = \mathbf{E}[\varphi(X^\varrho) \mid \mathcal{A}]$. Hence

$$\mathbf{E}\big[\varphi(X)\big] = \mathbf{E}\big[\mathbf{E}\big[\varphi(X) \bigm| \mathcal{A}\big]\big] = \mathbf{E}\big[\mathbf{E}\big[\varphi\big(X^\varrho\big) \bigm| \mathcal{A}\big]\big] = \mathbf{E}\big[\varphi\big(X^\varrho\big)\big],$$

whence X is exchangeable. □

Denote by $\mathcal{M}_1(E)$ the set of probability measures on E equipped with the topology of weak convergence (see Definition 13.12 and Remark 13.14). That is, a sequence $(\mu_n)_{n\in\mathbb{N}}$ in $\mathcal{M}_1(E)$ converges weakly to a $\mu \in \mathcal{M}_1(E)$ if and only if $\int f\,d\mu_n \overset{n\to\infty}{\longrightarrow} \int f\,d\mu$ for any bounded continuous function $f : E \to \mathbb{R}$. We will study weak convergence in Chapter 13 in greater detail. At this point, we use the topology only to make $\mathcal{M}_1(E)$ a measurable space, namely with the Borel σ-algebra $\mathcal{B}(\mathcal{M}_1(E))$. Now we can study random variables with values in $\mathcal{M}_1(E)$, so-called random measures (compare also Section 24.1). For $x \in E^{\mathbb{N}}$, let $\xi_n(x) = \frac{1}{n}\sum_{i=1}^n \delta_{x_i} \in \mathcal{M}_1(E)$.

Definition 12.25 The random measure

$$\Xi_n := \xi_n(X) := \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$$

is called the *empirical distribution* of $X_1, \ldots, X_n$.

Assume the conditions of Theorem 12.24 are in force.

Theorem 12.26 (de Finetti representation theorem) *The family $X = (X_n)_{n\in\mathbb{N}}$ is exchangeable if and only if there is a σ-algebra $\mathcal{A} \subset \mathcal{F}$ and an $\mathcal{A}$-measurable random variable $\Xi_\infty : \Omega \to \mathcal{M}_1(E)$ with the property that given Ξ_∞, $(X_n)_{n\in\mathbb{N}}$ is i.i.d. with $\mathcal{L}[X_1 \mid \Xi_\infty] = \Xi_\infty$. In this case, we can choose $\mathcal{A} = \mathcal{E}$ or $\mathcal{A} = \mathcal{T}$.*

Proof "$\Longleftarrow$" This follows as in the proof of Theorem 12.24.

"$\Longrightarrow$" Let X be exchangeable. Then, by Theorem 12.24, there exists a σ-algebra $\mathcal{A} \subset \mathcal{F}$ such that $(X_n)_{n\in\mathbb{N}}$ is i.i.d. given $\mathcal{A}$. As E is Polish, there exists a regular conditional distribution (see Theorem 8.37) $\Xi_\infty := \mathcal{L}[X_1 \mid \mathcal{A}]$. For measurable $A_1, \ldots, A_n \subset E$, we have $\mathbf{P}[X_i \in A_i \mid \mathcal{A}] = \Xi_\infty(A_i)$ for all $i = 1, \ldots, n$; hence

$$\mathbf{P}\left[\bigcap_{i=1}^n \{X_i \in A_i\} \;\middle|\; \Xi_\infty\right] = \mathbf{E}\left[\mathbf{P}\left[\bigcap_{i=1}^n \{X_i \in A_i\} \;\middle|\; \mathcal{A}\right] \;\middle|\; \Xi_\infty\right]$$

$$= \mathbf{E}\left[\prod_{i=1}^n \Xi_\infty(A_i) \;\middle|\; \Xi_\infty\right] = \prod_{i=1}^n \Xi_\infty(A_i).$$

Therefore, $\mathcal{L}[X \mid \Xi_\infty] = \Xi_\infty^{\otimes\mathbb{N}}$. □

Remark 12.27

(i) In the case considered in the previous theorem, by the strong law of large numbers, for any bounded continuous function $f : E \to \mathbb{R}$,

$$\int f \, d\Xi_n \overset{n\to\infty}{\longrightarrow} \int f \, d\Xi_\infty \quad \text{almost surely.}$$

If in addition E is locally compact (e.g., $E = \mathbb{R}^d$), then one can even show that

$$\Xi_n \overset{n\to\infty}{\longrightarrow} \Xi_\infty \quad \text{almost surely.}$$

(ii) For finite families of random variables there is no perfect analog of de Finetti's theorem. See [33] for a detailed treatment of finite exchangeable families. ◊

Example 12.28 Let $(X_n)_{n\in\mathbb{N}}$ be exchangeable and assume $X_n \in \{0, 1\}$. Then there exists a random variable $Y : \Omega \to [0, 1]$ such that, for all finite $J \subset \mathbb{N}$,

$$\mathbf{P}[X_j = 1 \text{ for all } j \in J \mid Y] = Y^{\#J}.$$

In other words, $(X_n)_{n\in\mathbb{N}}$ is independent given Y and Ber_Y-distributed. Compare Example 12.3(iii). ◊

Example 12.29 (Pólya's urn model) (See Example 14.38, compare also [17, 58, 135].) Consider an urn with a total of N balls among which M are black and $M - N$ are white. At each step, a ball is drawn and is returned to the urn together with an *additional* ball of the same color. Let

$$X_n := \begin{cases} 1, & \text{if the } n\text{th ball is black,} \\ 0, & \text{else,} \end{cases}$$

and let $S_n = \sum_{i=1}^n X_i$. Then

$$\mathbf{P}[X_n = 1 \mid X_1, X_2, \ldots, X_{n-1}] = \frac{S_{n-1} + M}{N + n - 1}.$$

Inductively, for $x_1, \ldots, x_n \in \{0, 1\}$ and $s_k = \sum_{i=1}^k x_i$, we get

$$\begin{aligned} &\mathbf{P}[X_i = x_i \text{ for any } i = 1, \ldots, n] \\ &= \prod_{i \le n: x_i = 1} \frac{M + s_{i-1}}{N + i - 1} \prod_{i \le n: x_i = 0} \frac{N + i - 1 - M - s_{i-1}}{N + i - 1} \\ &= \frac{(N-1)!}{(N-1+n)!} \cdot \frac{(M + s_n - 1)!}{(M-1)!} \frac{(N - M - 1 + (n - s_n))!}{(N - M - 1)!}. \end{aligned}$$

The right-hand side depends on s_n only and not on the order of the $x_1, \ldots, x_n$. Hence $(X_n)_{n\in\mathbb{N}}$ is exchangeable. Let

$$Z = \lim_{n\to\infty} \frac{1}{n} S_n.$$

Then $(X_n)_{n\in\mathbb{N}}$ is i.i.d. Ber_Z-distributed given Z. Hence (see Example 12.28)

$$\begin{aligned}\mathbf{E}\big[Z^n\big] &= \mathbf{E}\big[\mathbf{P}[X_1 = \ldots = X_n = 1 \mid Z]\big] \\ &= \mathbf{P}[S_n = n] \\ &= \frac{(N-1)!}{(M-1)!}\frac{(M+n-1)!}{(N+n-1)!} \quad \text{for all } n \in \mathbb{N}.\end{aligned}$$

By Exercise 5.1.2, these are the moments of the Beta distribution $\beta_{M,N-M}$ on $[0, 1]$ with parameters $(M, N-M)$ (see Example 1.107(ii)). A distribution on $[0, 1]$ is uniquely characterized by its moments (see Theorem 15.4). Hence $Z \sim \beta_{M,N-M}$. $\Diamond$