13

Conditional Expectations and Martingales

13.1 Conditional Expectations

For two events $A, B \in \mathcal{F}$ in a probability space $(\Omega, \mathcal{F}, \mathrm{P})$, we previously defined the conditional probability of A given B as

$$\mathrm{P}(A|B) = \frac{\mathrm{P}(A \bigcap B)}{\mathrm{P}(B)}.$$

Similarly, we can define the conditional expectation of a random variable f given B as

$$\mathrm{E}(f|B) = \frac{\int_B f(\omega) d\mathrm{P}(\omega)}{\mathrm{P}(B)},$$

provided that the integral on the right-hand side is finite and the denominator is different from zero.

We now introduce an important generalization of this notion by defining the conditional expectation of a random variable given a σ-subalgebra $\mathcal{G} \subseteq \mathcal{F}$.

Definition 13.1. *Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a probability space, $\mathcal{G}$ a σ-subalgebra of $\mathcal{F}$, and $f \in L^1(\Omega, \mathcal{F}, \mathrm{P})$. The conditional expectation of f given $\mathcal{G}$, denoted by $\mathrm{E}(f|\mathcal{G})$, is the random variable $g \in L^1(\Omega, \mathcal{G}, \mathrm{P})$ such that for any $A \in \mathcal{G}$*

$$\int_A f d\mathrm{P} = \int_A g d\mathrm{P}. \tag{13.1}$$

Note that for fixed f, the left-hand side of (13.1) is a σ-additive function defined on the σ-algebra $\mathcal{G}$. Therefore, the existence and uniqueness (up to a set of measure zero) of the function g are guaranteed by the Radon-Nikodym Theorem. Here are several simple examples.

If f is measurable with respect to $\mathcal{G}$, then clearly $\mathrm{E}(f|\mathcal{G}) = f$. If f is independent of the σ-algebra $\mathcal{G}$, then $\mathrm{E}(f|\mathcal{G}) = \mathrm{E}f$, since $\int_A f d\mathrm{P} = \mathrm{P}(A)\mathrm{E}f$ in this case. Thus the conditional expectation is reduced to ordinary expectation

L. Koralov and Y.G. Sinai, *Theory of Probability and Random Processes*, Universitext, DOI 10.1007/978-3-540-68829-7_13,

if f is independent of $\mathcal{G}$. This is the case, in particular, when $\mathcal{G}$ is the trivial σ-algebra, $\mathcal{G} = \{\emptyset, \Omega\}$.

If $\mathcal{G} = \{B, \Omega \backslash B, \emptyset, \Omega\}$, where $0 < \mathrm{P}(B) < 1$, then

$$\mathrm{E}(f|\mathcal{G}) = \mathrm{E}(f|B)\chi_B + \mathrm{E}(f|(\Omega\backslash B))\chi_{\Omega\backslash B}.$$

Thus, the conditional expectation of f with respect to the smallest σ-algebra containing B is equal to the constant $\mathrm{E}(f|B)$ on the set B.

Concerning the notations, we shall often write $\mathrm{E}(f|g)$ instead of $\mathrm{E}(f|\sigma(g))$, if f and g are random variables on $(\Omega, \mathcal{F}, \mathrm{P})$. Likewise, we shall often write $\mathrm{P}(A|\mathcal{G})$ instead of $\mathrm{E}(\chi_A|\mathcal{G})$ to denote the conditional expectation of the indicator function of a set $A \in \mathcal{F}$. The function $\mathrm{P}(A|\mathcal{G})$ will be referred to as the conditional probability of A given the σ-algebra $\mathcal{G}$.

13.2 Properties of Conditional Expectations

Let us list several important properties of conditional expectations. Note that since the conditional expectation is defined up to a set of measure zero, all the equalities and inequalities below hold almost surely.

1. If $f_1, f_2 \in L^1(\Omega, \mathcal{F}, \mathrm{P})$ and a, b are constants, then
$$\mathrm{E}(af_1 + bf_2|\mathcal{G}) = a\mathrm{E}(f_1|\mathcal{G}) + b\mathrm{E}(f_2|\mathcal{G}).$$
2. If $f \in L^1(\Omega, \mathcal{F}, \mathrm{P})$, and $\mathcal{G}_1$ and $\mathcal{G}_2$ are σ-subalgebras of $\mathcal{F}$ such that $\mathcal{G}_2 \subseteq \mathcal{G}_1 \subseteq \mathcal{F}$, then
$$\mathrm{E}(f|\mathcal{G}_2) = \mathrm{E}(\mathrm{E}(f|\mathcal{G}_1)|\mathcal{G}_2).$$
3. If $f_1, f_2 \in L^1(\Omega, \mathcal{F}, \mathrm{P})$ and $f_1 \le f_2$, then $\mathrm{E}(f_1|\mathcal{G}) \le \mathrm{E}(f_2|\mathcal{G})$.
4. $\mathrm{E}(\mathrm{E}(f|\mathcal{G})) = \mathrm{E}f$.
5. (Conditional Dominated Convergence Theorem) If a sequence of measurable functions f_n converges to a measurable function f almost surely, and
$$|f_n| \le \varphi,$$
where φ is integrable on Ω, then $\lim_{n\to\infty} \mathrm{E}(f_n|\mathcal{G}) = \mathrm{E}(f|\mathcal{G})$ almost surely.
6. If $g, fg \in L^1(\Omega, \mathcal{F}, \mathrm{P})$, and f is measurable with respect to $\mathcal{G}$, then $\mathrm{E}(fg|\mathcal{G}) = f\mathrm{E}(g|\mathcal{G})$.

Properties 1–3 are clear. To prove property 4, it suffices to take $A = \Omega$ in the equality $\int_A f d\mathrm{P} = \int_A \mathrm{E}(f|\mathcal{G}) d\mathrm{P}$ defining the conditional expectation.

To prove the Conditional Dominated Convergence Theorem, let us first assume that f_n is a monotonic sequence. Without loss of generality we may assume that f_n is monotonically non-decreasing (the case of a non-increasing sequence is treated similarly). Thus the sequence of functions $\mathrm{E}(f_n|\mathcal{G})$ satisfies the assumptions of the Levi Convergence Theorem (see Sect. 3.5).

Let $g = \lim_{n\to\infty} \mathrm{E}(f_n|\mathcal{G})$. Then g is $\mathcal{G}$-measurable and $\int_A g d\mathrm{P} = \int_A f d\mathrm{P}$ for any $A \in \mathcal{G}$, again by the Levi Theorem.

If the sequence f_n is not necessarily monotonic, we can consider the auxiliary sequences $\overline{f}_n = \inf_{m\geq n} f_m$ and $\overline{\overline{f}}_n = \sup_{m\geq n} f_m$. These sequences are already monotonic and satisfy the assumptions placed on the sequence f_n. Therefore,

$$\lim_{n\to\infty} \mathrm{E}(\overline{f}_n|\mathcal{G}) = \lim_{n\to\infty} \mathrm{E}(\overline{\overline{f}}_n|\mathcal{G}) = \mathrm{E}(f|\mathcal{G}).$$

Since $\overline{f}_n \leq f_n \leq \overline{\overline{f}}_n$, the Dominated Convergence Theorem follows from the monotonicity of the conditional expectation (property 3).

To prove the last property, first we consider the case when f is the indicator function of a set $B \in \mathcal{G}$. Then for any $A \in \mathcal{G}$

$$\int_A \chi_B \mathrm{E}(g|\mathcal{G}) d\mathrm{P} = \int_{A\bigcap B} \mathrm{E}(g|\mathcal{G}) d\mathrm{P} = \int_{A\bigcap B} g d\mathrm{P} = \int_A \chi_B g d\mathrm{P},$$

which proves the statement for $f = \chi_B$. By linearity, the statement is also true for simple functions taking a finite number of values. Next, without loss of generality, we may assume that $f, g \geq 0$. Then we can find a non-decreasing sequence of simple functions f_n, each taking a finite number of values such that $\lim_{n\to\infty} f_n = f$ almost surely. We have $f_n g \to fg$ almost surely, and the Dominated Convergence Theorem for conditional expectations can be applied to the sequence $f_n g$ to conclude that

$$\mathrm{E}(fg|\mathcal{G}) = \lim_{n\to\infty} \mathrm{E}(f_n g|\mathcal{G}) = \lim_{n\to\infty} f_n \mathrm{E}(g|\mathcal{G}) = f\mathrm{E}(g|\mathcal{G}).$$

We now state Jensen's Inequality and the Conditional Jensen's Inequality, essential to our discussion of conditional expectations and martingales. The proofs of these statements can be found in many other textbooks, and we shall not provide them here (see "Real Analysis and Probability" by R. M. Dudley).

We shall consider a random variable f with values in $\mathbb{R}^d$ defined on a probability space $(\Omega, \mathcal{F}, \mathrm{P})$. Recall that a function $g : \mathbb{R}^d \to \mathbb{R}$ is called convex if $g(cx + (1-c)y) \leq cg(x) + (1-c)g(y)$ for all $x, y \in \mathbb{R}^d$, $0 \leq c \leq 1$.

Theorem 13.2 (Jensen's Inequality). *Let g be a convex (and consequently continuous) function on $\mathbb{R}^d$ and f a random variable with values in $\mathbb{R}^d$ such that $\mathrm{E}|f| < \infty$. Then, either $\mathrm{E}g(f) = +\infty$, or*

$$g(\mathrm{E}f) \leq \mathrm{E}g(f) < \infty.$$

Theorem 13.3 (Conditional Jensen's Inequality). *Let g be a convex function on $\mathbb{R}^d$ and f a random variable with values in $\mathbb{R}^d$ such that*

$$\mathrm{E}|f|, \mathrm{E}|g(f)| < \infty.$$

Let $\mathcal{G}$ be a σ-subalgebra of $\mathcal{F}$. Then almost surely

$$g(\mathrm{E}(f|\mathcal{G})) \leq \mathrm{E}(g(f)|\mathcal{G}).$$

Let $\mathcal{G}$ be a σ-subalgebra of $\mathcal{F}$. Let $H = L^2(\Omega, \mathcal{G}, \mathrm{P})$ be the closed linear subspace of the Hilbert space $L^2(\Omega, \mathcal{F}, \mathrm{P})$. Let us illustrate the use of the Conditional Jensen's Inequality by proving that for a random variable $f \in L^2(\Omega, \mathcal{F}, \mathrm{P})$, taking the conditional expectation $\mathrm{E}(f|\mathcal{G})$ is the same as taking the projection on H.

Lemma 13.4. *Let $f \in L^2(\Omega, \mathcal{F}, \mathrm{P})$ and P_H be the projection operator on the space H. Then*

$$\mathrm{E}(f|\mathcal{G}) = P_H f.$$

Proof. The function $\mathrm{E}(f|\mathcal{G})$ is square-integrable by the Conditional Jensen's Inequality applied to $g(x) = x^2$. Thus, $\mathrm{E}(f|\mathcal{G}) \in H$. It remains to show that $f - \mathrm{E}(f|\mathcal{G})$ is orthogonal to any $h \in H$. Since h is $\mathcal{G}$-measurable,

$$\mathrm{E}((f - \mathrm{E}(f|\mathcal{G}))\overline{h}) = \mathrm{E}\mathrm{E}((f - \mathrm{E}(f|\mathcal{G}))\overline{h}|\mathcal{G}) = \mathrm{E}(\overline{h}\mathrm{E}((f - \mathrm{E}(f|\mathcal{G}))|\mathcal{G})) = 0.$$

□

13.3 Regular Conditional Probabilities

Let f and g be random variables on a probability space $(\Omega, \mathcal{F}, \mathrm{P})$. If g takes a finite or countable number of values $y_1, y_2, \ldots$, and the probabilities of the events $\{\omega : g(\omega) = y_i\}$ are positive, we can write, similarly to (4.1), the formula of full expectation

$$\mathrm{E}f = \sum_i \mathrm{E}(f|g = y_i)\mathrm{P}(g = y_i).$$

Let us derive an analogue to this formula, which will work when the number of values of g is not necessarily finite or countable. The sets $\Omega_y = \{\omega : g(\omega) = y\}$, where $y \in \mathbb{R}$, still form a partition of the probability space Ω, but the probability of each Ω_y may be equal to zero. Thus, we need to attribute meaning to the expression $\mathrm{E}(f|\Omega_y)$ (also denoted by $\mathrm{E}(f|g = y)$). One way to do this is with the help of the concept of a regular conditional probability, which we introduce below.

Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a probability space and $\mathcal{G} \subseteq \mathcal{F}$ a σ-subalgebra. Let h be a measurable function from $(\Omega, \mathcal{F})$ to a measurable space $(X, \mathcal{B})$. To motivate the formal definition of a regular conditional probability, let us first assume that $\mathcal{G}$ is generated by a finite or countable partition $A_1, A_2, \ldots$ such that $\mathrm{P}(A_i) > 0$ for all i. In this case, for a fixed $B \in \mathcal{B}$, the conditional probability $\mathrm{P}(h \in B|\mathcal{G})$ is constant on each A_i equal to $\mathrm{P}(h \in B|A_i)$, as follows from the definition of the conditional probability. As a function of B, this expression is a probability measure on $(X, \mathcal{B})$. The concept of a regular conditional probability allows us to view $\mathrm{P}(h \in B|\mathcal{G})(\omega)$, for fixed ω, as a probability measure, even without the assumption that $\mathcal{G}$ is generated by a finite or countable partition.

Definition 13.5. *A function $Q : \mathcal{B} \times \Omega \to [0,1]$ is called a regular conditional probability of h given $\mathcal{G}$ if:*

1. *For each $\omega \in \Omega$, the function $Q(\cdot, \omega) : \mathcal{B} \to [0,1]$ is a probability measure on $(X, \mathcal{B})$.*
2. *For each $B \in \mathcal{B}$, the function $Q(B, \cdot) : \Omega \to [0,1]$ is $\mathcal{G}$-measurable.*
3. *For each $B \in \mathcal{B}$, the equality $\mathrm{P}(h \in B|\mathcal{G})(\omega) = Q(B, \omega)$ holds almost surely.*

We have the following theorem, which guarantees the existence and uniqueness of a regular conditional probability when X is a complete separable metric space. (The proof of this theorem can be found in "Real Analysis and Probability" by R. M. Dudley.)

Theorem 13.6. *Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a probability space and $\mathcal{G} \subseteq \mathcal{F}$ a σ-subalgebra. Let X be a complete separable metric space and $\mathcal{B}$ the σ-algebra of Borel sets of X. Take a measurable function h from $(\Omega, \mathcal{F})$ to $(X, \mathcal{B})$. Then there exists a regular conditional probability of h given $\mathcal{G}$. It is unique in the sense that if Q and Q' are regular conditional probabilities, then the measures $Q(\cdot, \omega)$ and $Q'(\cdot, \omega)$ coincide for almost all ω.*

The next lemma states that when the regular conditional probability exists, the conditional expectation can be written as an integral with respect to the measure $Q(\cdot, \omega)$.

Lemma 13.7. *Let the assumptions of Theorem 13.6 hold, and $f : X \to \mathbb{R}$ be a measurable function such that $\mathrm{E}(f(h(\omega))$ is finite. Then, for almost all ω, the function f is integrable with respect to $Q(\cdot, \omega)$, and*

$$\mathrm{E}(f(h)|\mathcal{G})(\omega) = \int_X f(x) Q(dx, \omega) \quad \text{for almost all } \omega. \tag{13.2}$$

Proof. First, let f be an indicator function of a measurable set, that is $f = \chi_B$ for $B \in \mathcal{B}$. In this case, the statement of the lemma is reduced to

$$\mathrm{P}(h \in B|\mathcal{G})(\omega) = Q(B, \omega),$$

which follows from the definition of the regular conditional probability.

Since both sides of (13.2) are linear in f, the lemma also holds when f is a simple function with a finite number of values. Now, let f be a non-negative measurable function such that $\mathrm{E}(f(h(\omega))$ is finite. One can find a sequence of non-negative simple functions f_n, each taking a finite number of values, such that $f_n \to f$ monotonically from below. Thus, $\mathrm{E}(f_n(h)|\mathcal{G})(\omega) \to \mathrm{E}(f(h)|\mathcal{G})(\omega)$ almost surely by the Conditional Dominated Convergence Theorem. Therefore, the sequence $\int_X f_n(x) Q(dx, \omega)$ is bounded almost surely, and $\int_X f_n(x) Q(dx, \omega) \to \int_X f(x) Q(dx, \omega)$ for almost all ω by the Levi Monotonic Convergence Theorem. This justifies (13.2) for non-negative f.

Finally, if f is not necessarily non-negative, it can be represented as a difference of two non-negative functions. $\square$

Example. Assume that Ω is a complete separable metric space, $\mathcal{F}$ is the σ-algebra of its Borel sets, and $(X, \mathcal{B}) = (\Omega, \mathcal{F})$. Let P be a probability measure on $(\Omega, \mathcal{F})$, and f and g be random variables on $(\Omega, \mathcal{F}, \mathrm{P})$. Let h be the identity mapping from Ω to itself, and let $\mathcal{G} = \sigma(g)$. In this case, (13.2) takes the form

$$\mathrm{E}(f|g)(\omega) = \int_{\Omega} f(\widetilde{\omega})Q(d\widetilde{\omega}, \omega) \quad \text{for almost all } \omega. \tag{13.3}$$

Let P_g be the measure on $\mathbb{R}$ induced by the mapping $g : \Omega \to \mathbb{R}$. For any $B \in \mathcal{B}$, the function $Q(B, \cdot)$ is constant on each level set of g, since it is measurable with respect to $\sigma(g)$. Therefore, for almost all y (with respect to the measure P_g), we can define measures $\mathrm{Q}_y(\cdot)$ on $(\Omega, \mathcal{F})$ by putting $\mathrm{Q}_{g(\omega)}(B) = Q(B, \omega)$.

The function $\mathrm{E}(f|g)$ is constant on each level set of g. Therefore, we can define $\mathrm{E}(f|g = y) = \mathrm{E}(f|g)(\omega)$, where ω is such that $g(\omega) = y$. This function is defined up to a set of measure zero (with respect to the measure P_g). In order to calculate the expectation of f, we can write

$$\mathrm{E}f = \mathrm{E}(\mathrm{E}(f|g)) = \int_{\mathbb{R}} \mathrm{E}(f|g = y)d\mathrm{P}_g(y) = \int_{\mathbb{R}} (\int_{\Omega} f(\widetilde{\omega})d\mathrm{Q}_y(\widetilde{\omega}))d\mathrm{P}_g(y),$$

where the second equality follows from the change of variable formula in the Lebesgue integral. It is possible to show that the measure Q_y is supported on the event $\Omega_y = \{\omega : g(\omega) = y\}$ for P_g-almost all y (we do not prove this statement here). Therefore, we can write the expectation as a double integral

$$\mathrm{E}f = \int_{\mathbb{R}} (\int_{\Omega_y} f(\widetilde{\omega})d\mathrm{Q}_y(\widetilde{\omega}))d\mathrm{P}_g(y).$$

This is the formula of the full mathematical expectation.

Example. Let h be a random variable with values in $\mathbb{R}$, f the identity mapping on $\mathbb{R}$, and $\mathcal{G} = \sigma(g)$. Then Lemma 13.7 states that

$$\mathrm{E}(h|g)(\omega) = \int_{\mathbb{R}} xQ(dx, \omega) \quad \text{for almost all } \omega,$$

where Q is the regular conditional probability of h given $\sigma(g)$. Assume that h and g have a joint probability density $p(x, y)$, which is a continuous function satisfying $0 < \int_{\mathbb{R}} p(x, y)dx < \infty$ for all y. It is easy to check that

$$Q(B, \omega) = \int_B p(x, g(\omega))dx(\int_{\mathbb{R}} p(x, g(\omega))dx)^{-1}$$

has the properties required of the regular conditional probability. Therefore,

$$\mathrm{E}(h|g)(\omega) = \int_{\mathbb{R}} xp(x, g(\omega))dx(\int_{\mathbb{R}} p(x, g(\omega))dx)^{-1} \quad \text{for almost all } \omega,$$

and

$$\mathrm{E}(h|g = y) = \int_{\mathbb{R}} xp(x, y)dx(\int_{\mathbb{R}} p(x, y)dx)^{-1} \quad \text{for } \mathrm{P}_g\text{−almost all } y.$$

13.4 Filtrations, Stopping Times, and Martingales

Let $(\Omega, \mathcal{F})$ be a measurable space and T a subset of $\mathbb{R}$ or $\mathbb{Z}$.

Definition 13.8. *A collection of σ-subalgebras $\mathcal{F}_t \subseteq \mathcal{F}$, $t \in T$, is called a filtration if $\mathcal{F}_s \subseteq \mathcal{F}_t$ for all $s \leq t$.*

Definition 13.9. *A random variable τ with values in the parameter set T is a stopping time of the filtration $\mathcal{F}_t$ if $\{\tau \leq t\} \in \mathcal{F}_t$ for each $t \in T$.*

Remark 13.10. Sometimes it will be convenient to allow τ to take values in $T \cup \{\infty\}$. In this case, τ is still called a stopping time if $\{\tau \leq t\} \in \mathcal{F}_t$ for each $t \in T$.

Example. Let $T = \mathbb{N}$ and Ω be the space of all functions $\omega : \mathbb{N} \to \{-1, 1\}$. (In other words, Ω is the space of infinite sequences made of -1's and 1's.) Let $\mathcal{F}_n$ be the smallest σ-algebra which contains all the sets of the form

$$\{\omega : \omega(1) = a_1, \ldots, \omega(n) = a_n\},$$

where $a_1, \ldots, a_n \in \{-1, 1\}$. Let $\mathcal{F}$ be the smallest σ-algebra containing all $\mathcal{F}_n$, $n \geq 1$. The space $(\Omega, \mathcal{F})$ can be used to model an infinite sequence of games, where the outcome of each game is either a loss or a gain of one dollar. Let

$$\tau(\omega) = \min\{n : \sum_{i=1}^{n} \omega(i) = 3\}.$$

Thus, τ is the first time when a gambler playing the game accumulates three dollars in winnings. (Note that $\tau(\omega) = \infty$ for some ω.) It is easy to demonstrate that τ is a stopping time. Let

$$\sigma(\omega) = \min\{n : \omega(n+1) = -1\}.$$

Thus, a gambler stops at time σ if the next game will result in a loss. Following such a strategy involves looking at the outcome of a future game before deciding whether to play it. Indeed, it is easy to check that σ does not satisfy the definition of a stopping time.

Remark 13.11. Recall the following notation: if x and y are real numbers, then $x \wedge y = \min(x, y)$ and $x \vee y = \max(x, y)$.

Lemma 13.12. *If σ and τ are stopping times of a filtration $\mathcal{F}_t$, then $\sigma \wedge \tau$ is also a stopping time.*

Proof. We need to show that $\{\sigma \wedge \tau \leq t\} \in \mathcal{F}_t$ for any $t \in T$, which immediately follows from

$$\{\sigma \wedge \tau \leq t\} = \{\sigma \leq t\} \bigcup \{\tau \leq t\} \in \mathcal{F}_t.$$

$\square$

In fact, if σ and τ are stopping times, then $\sigma \vee \tau$ is also a stopping time. If, in addition, $\sigma, \tau \geq 0$, then $\sigma + \tau$ is also a stopping time (see Problem 7).

Definition 13.13. *Let τ be a stopping time of the filtration $\mathcal{F}_t$. The σ-algebra of events determined prior to the stopping time τ, denoted by $\mathcal{F}_\tau$, is the collection of events $A \in \mathcal{F}$ for which $A \bigcap \{\tau \leq t\} \in \mathcal{F}_t$ for each $t \in T$.*

Clearly, $\mathcal{F}_\tau$ is a σ-algebra. Moreover, τ is $\mathcal{F}_\tau$-measurable since

$$\{\tau \leq c\} \bigcap \{\tau \leq t\} = \{\tau \leq c \wedge t\} \in \mathcal{F}_t,$$

and therefore $\{\tau \leq c\} \in \mathcal{F}_\tau$ for each c. If σ and τ are two stopping times such that $\sigma \leq \tau$, then $\mathcal{F}_\sigma \subseteq \mathcal{F}_\tau$. Indeed, if $A \in \mathcal{F}_\sigma$, then

$$A \bigcap \{\tau \leq t\} = (A \bigcap \{\sigma \leq t\}) \bigcap \{\tau \leq t\} \in \mathcal{F}_t.$$

Now let us consider a process X_t together with a filtration $\mathcal{F}_t$ defined on a common probability space.

Definition 13.14. *A random process X_t is called adapted to a filtration $\mathcal{F}_t$ if X_t is $\mathcal{F}_t$-measurable for each $t \in T$.*

An example of a stopping time is provided by the first time when a continuous process hits a closed set.

Lemma 13.15. *Let X_t be a continuous $\mathbb{R}^d$-valued process adapted to a filtration $\mathcal{F}_t$, where $t \in \mathbb{R}^+$. Let K be a closed set in $\mathbb{R}^d$ and $s \geq 0$. Let*

$$\tau^s(\omega) = \inf\{t \geq s, X_t(\omega) \in K\}$$

be the first time, following s, when the process hits K. Then τ^s is a stopping time.

Proof. For an open set U, define

$$\tau_U^s(\omega) = \inf\{t \geq s, X_t(\omega) \in U\},$$

where the infimum of the empty set is $+\infty$. First, we show that the set $\{\omega : \tau_U^s(\omega) < t\}$ belongs to $\mathcal{F}_t$ for any $t \in \mathbb{R}^+$. Indeed, from the continuity of the process it easily follows that

$$\{\tau_U^s < t\} = \bigcup_{u \in \mathbb{Q}, s < u < t} \{X_u \in U\},$$

and the right-hand side of this equality belongs to $\mathcal{F}_t$. Now, for the set K, we define the open sets $U_n = \{x \in \mathbb{R}^d : \mathrm{dist}(x, K) < 1/n\}$. We claim that for $t > s$,

$$\{\tau^s \leq t\} = \bigcap_{n=1}^{\infty} \{\tau_{U_n}^s < t\}. \qquad (13.4)$$

Indeed, if $\tau^s(\omega) \le t$, then for each n the trajectory $X_u(\omega)$ enters the open set U_n for some u, $s < u < t$, due to the continuity of the process. Thus ω belongs to the event on the right-hand side of (13.4).

Conversely, if ω belongs to the event on the right-hand side of (13.4), then there is a non-decreasing sequence of times u_n such that $s < u_n < t$ and $X_{u_n}(\omega) \in U_n$. Taking $u = \lim_{n\to\infty} u_n$, we see that $u \le t$ and $X_u(\omega) \in K$, again due to the continuity of the process. This means that $\tau^s(\omega) \le t$, which justifies (13.4).

Since the event on the right-hand side of (13.4) belongs to $\mathcal{F}_t$, we see that $\{\tau^s \le t\}$ belongs to $\mathcal{F}_t$ for $t > s$. Furthermore, $\{\tau^s \le s\} = \{X_s \in K\} \in \mathcal{F}_s$. We have thus proved that τ^s is a stopping time. $\square$

For a given random process, a simple example of a filtration is that generated by the process itself:

$$\mathcal{F}_t^X = \sigma(X_s, s \le t).$$

Clearly, X_t is adapted to the filtration $\mathcal{F}_t^X$.

Definition 13.16. *A family $(X_t, \mathcal{F}_t)_{t\in T}$ is called a martingale if the process X_t is adapted to the filtration $\mathcal{F}_t$, $X_t \in L^1(\Omega, \mathcal{F}, \mathrm{P})$ for all t, and*

$$X_s = \mathrm{E}(X_t|\mathcal{F}_s) \qquad \text{for} \quad s \le t.$$

If the equal sign is replaced by $\le$ or $\ge$, then $(X_t, \mathcal{F}_t)_{t\in T}$ is called a submartingale or supermartingale respectively.

We shall often say that X_t is a martingale, without specifying a filtration, if it is clear from the context what the parameter set and the filtration are.

If one thinks of X_t as the fortune of a gambler at time t, then a martingale is a model of a fair game (any information available by time s does not affect the fact that the expected increment in the fortune over the time period from s to t is equal to zero). More precisely, $\mathrm{E}(X_t - X_s|\mathcal{F}_s) = 0$.

If $(X_t, \mathcal{F}_t)_{t\in T}$ is a martingale and f is a convex function such that $f(X_t)$ is integrable for all t, then $(f(X_t), \mathcal{F}_t)_{t\in T}$ is a submartingale. Indeed, by the Conditional Jensen's Inequality,

$$f(X_s) = f(\mathrm{E}(X_t|\mathcal{F}_s)) \le \mathrm{E}(f(X_t)|\mathcal{F}_s).$$

For example, if $(X_t, \mathcal{F}_t)_{t\in T}$ is a martingale, then $(|X_t|, \mathcal{F}_t)_{t\in T}$ is a submartingale, If, in addition, X_t is square-integrable, then $(X_t^2, \mathcal{F}_t)_{t\in T}$ is a submartingale.

13.5 Martingales with Discrete Time

In this section we study martingales with discrete time ($T = \mathbb{N}$). In the next section we shall state the corresponding results for continuous time martingales, which will lead us to the notion of an integral of a random process with respect to a continuous martingale.

Our first theorem states that any submartingale can be decomposed, in a unique way, into a sum of a martingale and a non-decreasing process adapted to the filtration $(\mathcal{F}_{n-1})_{n\geq 2}$.

Theorem 13.17 (Doob Decomposition). *If $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ is a submartingale, then there exist two random processes, M_n and A_n, with the following properties:*

1. $X_n = M_n + A_n$ *for* $n \geq 1$.
2. $(M_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ *is a martingale.*
3. $A_1 = 0$, A_n *is* $\mathcal{F}_{n-1}$*-measurable for* $n \geq 2$.
4. A_n *is non-decreasing, that is*

$$A_n(\omega) \leq A_{n+1}(\omega)$$

almost surely for all $n \geq 1$.

If another pair of processes $\overline{M}_n, \overline{A}_n$ *has the same properties, then* $M_n = \overline{M}_n$, $A_n = \overline{A}_n$ *almost surely.*

Proof. Assuming that the processes M_n and A_n with the required properties exist, we can write for $n \geq 2$

$$X_{n-1} = M_{n-1} + A_{n-1},$$

$$X_n = M_n + A_n.$$

Taking the difference and then the conditional expectation with respect to $\mathcal{F}_{n-1}$, we obtain

$$\mathrm{E}(X_n|\mathcal{F}_{n-1}) - X_{n-1} = A_n - A_{n-1}.$$

This shows that A_n is uniquely defined by the process X_n and the random variable A_{n-1}. The random variable M_n is also uniquely defined, since $M_n = X_n - A_n$. Since $M_1 = X_1$ and $A_1 = 0$, we see, by induction on n, that the pair of processes M_n, A_n with the required properties is unique.

Furthermore, given a submartingale X_n, we can use the relations

$$M_1 = X_1, \quad A_1 = 0,$$

$$A_n = \mathrm{E}(X_n|\mathcal{F}_{n-1}) - X_{n-1} + A_{n-1}, \quad M_n = X_n - A_n, \quad n \geq 2,$$

to define inductively the processes M_n and A_n. Clearly, they have properties 1, 3 and 4. In order to verify property 2, we write

$$\mathrm{E}(M_n|\mathcal{F}_{n-1}) = \mathrm{E}(X_n - A_n|\mathcal{F}_{n-1}) = \mathrm{E}(X_n|\mathcal{F}_{n-1}) - A_n$$

$$= X_{n-1} - A_{n-1} = M_{n-1}, \quad n \geq 2,$$

which proves that $(M_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ is a martingale. □

If $(X_n, \mathcal{F}_n)$ is an adapted process and τ is a stopping time, then $X_{\tau(\omega)}(\omega)$ is a random variable measurable with respect to the σ-algebra $\mathcal{F}_\tau$. Indeed, one needs to check that $\{X_\tau \in B\} \bigcap \{\tau \leq n\} \in \mathcal{F}_n$ for any Borel set B of the real line and each n. This is true since τ takes only integer values and $\{X_m \in B\} \in \mathcal{F}_n$ for each $m \leq n$.

In order to develop an intuitive understanding of the next theorem, one can again think of a martingale as a model of a fair game. In a fair game, a gambler cannot increase or decrease the expectation of his fortune by entering the game at a point of time $\sigma(\omega)$, and then quitting the game at $\tau(\omega)$, provided that he decides to enter and leave the game based only on the information available by the time of the decision (that is, without looking into the future).

Theorem 13.18 (Optional Sampling Theorem). *If $(X_n, \mathcal{F}_n)_{n \in \mathbb{N}}$ is a submartingale and σ and τ are two stopping times such that $\sigma \leq \tau \leq k$ for some $k \in \mathbb{N}$, then*

$$X_\sigma \leq \mathrm{E}(X_\tau | \mathcal{F}_\sigma).$$

If $(X_n, \mathcal{F}_n)_{n \in \mathbb{N}}$ is a martingale or a supermartingale, then the same statement holds with the $\leq$ sign replaced by $=$ or $\geq$ respectively.

Proof. The case of $(X_n, \mathcal{F}_n)_{n \in \mathbb{N}}$ being a supermartingale is equivalent to considering the submartingale $(-X_n, \mathcal{F}_n)_{n \in \mathbb{N}}$. Thus, without loss of generality, we may assume that $(X_n, \mathcal{F}_n)_{n \in \mathbb{N}}$ is a submartingale.

Let $A \in \mathcal{F}_\sigma$. For $1 \leq m \leq n$ we define

$$A_m = A \bigcap \{\sigma = m\}, \quad A_{m,n} = A_m \bigcap \{\tau = n\},$$

$$B_{m,n} = A_m \bigcap \{\tau > n\}, \quad C_{m,n} = A_m \bigcap \{\tau \geq n\}.$$

Note that $B_{m,n} \in \mathcal{F}_n$, since $\{\tau > n\} = \Omega \backslash \{\tau \leq n\} \in \mathcal{F}_n$. Therefore, by definition of a submartingale,

$$\int_{B_{m,n}} X_n d\mathrm{P} \leq \int_{B_{m,n}} X_{n+1} d\mathrm{P}.$$

Since $C_{m,n} = A_{m,n} \bigcup B_{m,n}$,

$$\int_{C_{m,n}} X_n d\mathrm{P} \leq \int_{A_{m,n}} X_n d\mathrm{P} + \int_{B_{m,n}} X_{n+1} d\mathrm{P},$$

and thus, since $B_{m,n} = C_{m,n+1}$,

$$\int_{C_{m,n}} X_n d\mathrm{P} - \int_{C_{m,n+1}} X_{n+1} d\mathrm{P} \leq \int_{A_{m,n}} X_n d\mathrm{P}.$$

By taking the sum from $n = m$ to k, and noting that we have a telescopic sum on the left-hand side, we obtain

$$\int_{A_m} X_m d\mathrm{P} \le \int_{A_m} X_\tau d\mathrm{P},$$

were we used that $A_m = C_{m,m}$. By taking the sum from $m = 1$ to k, we obtain

$$\int_A X_\sigma d\mathrm{P} \le \int_A X_\tau d\mathrm{P}.$$

Since $A \in \mathcal{F}_\sigma$ was arbitrary, this completes the proof of the theorem. $\square$

Definition 13.19. *A set of random variables $\{f_s\}_{s \in S}$ is said to be uniformly integrable if*

$$\lim_{\lambda \to \infty} \sup_{s \in S} \int_{\{|f_s| > \lambda\}} |f_s| d\mathrm{P} = 0.$$

Remark 13.20. The Optional Sampling Theorem is, in general, not true for unbounded stopping times σ and τ. If, however, we assume that the random variables $X_n, n \in \mathbb{N}$, are uniformly integrable, then the theorem remains valid even for unbounded σ and τ.

Remark 13.21. There is an equivalent way to define uniform integrability (see Problem 9). Namely, a set of random variables $\{f_s\}_{s \in S}$ is uniformly integrable if

(1) There is a constant K such that $\int_\Omega |f_s| dP \le K$ for all $s \in S$, and
(2) For any $\varepsilon > 0$ one can find $\delta > 0$ such that $\int_A |f_s(\omega)| d\mathrm{P}(\omega) \le \varepsilon$ for all $s \in S$, provided that $\mathrm{P}(A) \le \delta$.

For a random process X_n and a constant $\lambda > 0$, we define the event $A(\lambda, n) = \{\omega : \max_{1 \le i \le n} X_i(\omega) \ge \lambda\}$. From the Chebyshev Inequality it follows that $\lambda \mathrm{P}(\{X_n \ge \lambda\}) \le \mathrm{E} \max(X_n, 0)$. If $(X_n, \mathcal{F}_n)$ is a submartingale, we can make a stronger statement. Namely, we shall now use the Optional Sampling Theorem to show that the event $\{X_n \ge \lambda\}$ on the left-hand side can be replaced by $A(\lambda, n)$.

Theorem 13.22 (Doob Inequality). *If $(X_n, \mathcal{F}_n)$ is a submartingale, then for any $n \in \mathbb{N}$ and any $\lambda > 0$,*

$$\lambda \mathrm{P}(A(\lambda, n)) \le \int_{A(\lambda, n)} X_n d\mathrm{P} \le \mathrm{E} \max(X_n, 0).$$

Proof. We define the stopping time σ to be the first moment when $X_i \ge \lambda$ if $\max_{i \le n} X_i \ge \lambda$ and put $\sigma = n$ if $\max_{i \le n} X_i < \lambda$. The stopping time τ is defined simply as $\tau = n$. Since $\sigma \le \tau$, the Optional Sampling Theorem can be applied to the pair of stopping times σ and τ. Note that $A(\lambda, n) \in \mathcal{F}_\sigma$ since

$$A(\lambda, n) \bigcap \{\sigma \le m\} = \{\max_{i \le m} X_i \ge \lambda\} \in \mathcal{F}_m.$$

Therefore, since $X_\sigma \geq \lambda$ on $A(\lambda, n)$,

$$\lambda \mathrm{P}(A(\lambda, n)) \leq \int_{A(\lambda,n)} X_\sigma d\mathrm{P} \leq \int_{A(\lambda,n)} X_n d\mathrm{P} \leq \mathrm{E}\max(X_n, 0),$$

where the second inequality follows from the Optional Sampling Theorem. □

Remark 13.23. Suppose that ξ_1, ξ_2,... is a sequence of independent random variables with finite mathematical expectations and variances, $m_i = \mathrm{E}\xi_i$, $V_i = \mathrm{Var}\xi_i$. One can obtain the Kolmogorov Inequality of Sect. 7.1 by applying Doob's Inequality to the submartingale $\zeta_n = (\xi_1 + \ldots + \xi_n - m_1 - \ldots - m_n)^2$.

13.6 Martingales with Continuous Time

In this section we shall formulate the statements of the Doob Decomposition, the Optional Sampling Theorem, and the Doob Inequality for continuous time martingales. The proofs of these results rely primarily on the corresponding statements for the case of martingales with discrete time. We shall not provide additional technical details, but interested readers may refer to "Brownian Motion and Stochastic Calculus" by I. Karatzas and S. Shreve for the complete proofs.

Before formulating the results, we introduce some new notations and definitions.

Given a filtration $(\mathcal{F}_t)_{t\in\mathbb{R}^+}$ on a probability space $(\Omega, \mathcal{F}, \mathrm{P})$, we define the filtration $(\mathcal{F}_{t+})_{t\in\mathbb{R}^+}$ as follows: $A \in \mathcal{F}_{t+}$ if and only if $A \in \mathcal{F}_{t+\delta}$ for any $\delta > 0$. We shall say that $(\mathcal{F}_t)_{t\in\mathbb{R}^+}$ is right-continuous if $\mathcal{F}_t = \mathcal{F}_{t+}$ for all $t \in \mathbb{R}^+$.

Recall that a set $A \subseteq \Omega$ is said to be P-negligible if there is an event $B \in \mathcal{F}$ such that $A \subseteq B$ and $\mathrm{P}(B) = 0$.

We shall often impose the following technical assumption on our filtration.

Definition 13.24. *A filtration $(\mathcal{F}_t)_{t\in\mathbb{R}^+}$ is said to satisfy the usual conditions if it is right-continuous and all the* P*-negligible events from $\mathcal{F}$ belong to $\mathcal{F}_0$.*

We shall primarily be interested in processes whose every realization is right-continuous (right-continuous processes), or every realization is continuous (continuous processes). It will be clear that in the results stated below the assumption that a process is right-continuous (continuous) can be replaced by the assumption that the process is indistinguishable from a right-continuous (continuous) process.

Later we shall need the following lemma, which we state now without a proof. (A proof can be found in "Brownian Motion and Stochastic Calculus" by I. Karatzas and S. Shreve.)

Lemma 13.25. *Let $(X_t, \mathcal{F}_t)_{t\in\mathbb{R}^+}$ be a submartingale with filtration which satisfies the usual conditions. If the function $f : t \to \mathrm{E}X_t$ from $\mathbb{R}^+$ to $\mathbb{R}$ is right-continuous, then there exists a right-continuous modification of the process X_t*

which is also adapted to the filtration $\mathcal{F}_t$ (and therefore is also a submartingale).

We formulate the theorem on the decomposition of continuous submartingales.

Theorem 13.26 (Doob-Meyer Decomposition). *Let $(X_t, \mathcal{F}_t)_{t\in\mathbb{R}^+}$ be a continuous submartingale with filtration which satisfies the usual conditions. Let S_a be the set of all stopping times bounded by a. Assume that for every $a > 0$ the set of random variables $\{X_\tau\}_{\tau\in S_a}$ is uniformly integrable. Then there exist two continuous random processes M_t and A_t such that:*

1. *$X_t = M_t + A_t$ for all $t \geq 0$ almost surely.*
2. *$(M_t, \mathcal{F}_t)_{t\in\mathbb{R}^+}$ is a martingale.*
3. *$A_0 = 0$, A_t is adapted to the filtration $\mathcal{F}_t$.*
4. *A_t is non-decreasing, that is $A_s(\omega) \leq A_t(\omega)$ if $s \leq t$ for every ω.*

If another pair of processes $\overline{M}_t, \overline{A}_t$ has the same properties, then M_t is indistinguishable from $\overline{M}_t$ and A_t is indistinguishable from $\overline{A}_t$.

We can also formulate the Optional Sampling Theorem for continuous time submartingales. If τ is a stopping time of a filtration $\mathcal{F}_t$, and the process X_t is adapted to the filtration $\mathcal{F}_t$ and right-continuous, then it is not difficult to show that X_τ is $\mathcal{F}_\tau$-measurable (see Problems 1 and 2 in Chap. 19).

Theorem 13.27 (Optional Sampling Theorem). *If $(X_t, \mathcal{F}_t)_{t\in\mathbb{R}^+}$ is a right-continuous submartingale, and σ and τ are two stopping times such that $\sigma \leq \tau \leq r$ for some $r \in \mathbb{R}^+$, then*

$$X_\sigma \leq \mathrm{E}(X_\tau | \mathcal{F}_\sigma).$$

If $(X_t, \mathcal{F}_t)_{t\in\mathbb{R}^+}$ is a either martingale or a supermartingale, then the same statement holds with the $\leq$ sign replaced by $=$ or $\geq$ respectively.

Remark 13.28. As in the case of discrete time, the Optional Sampling Theorem remains valid even for unbounded σ and τ if the random variables $X_t, t \in \mathbb{R}^+$, are uniformly integrable.

The proof of the following lemma relies on a simple application of the Optional Sampling Theorem.

Lemma 13.29. *If $(X_t, \mathcal{F}_t)_{t\in\mathbb{R}^+}$ is a right-continuous (continuous) martingale, τ is a stopping time of the filtration $\mathcal{F}_t$, and $Y_t = X_{t\wedge\tau}$, then $(Y_t, \mathcal{F}_t)_{t\in\mathbb{R}^+}$ is also a right-continuous (continuous) martingale.*

Proof. Let us show that $\mathrm{E}(Y_t - Y_s|\mathcal{F}_s) = 0$ for $s \leq t$. We have

$$\mathrm{E}(Y_t - Y_s|\mathcal{F}_s) = \mathrm{E}(X_{t\wedge\tau} - X_{s\wedge\tau}|\mathcal{F}_s) = \mathrm{E}((X_{(t\wedge\tau)\vee s} - X_s)|\mathcal{F}_s).$$

The expression on the right-hand side of this equality is equal to zero by the Optional Sampling Theorem. Since $t \wedge \tau$ is a continuous function of t, the right-continuity (continuity) of Y_t follows from the right-continuity (continuity) of X_t. □

Finally, we formulate the Doob Inequality for continuous time submartingales.

Theorem 13.30 (Doob Inequality). *If $(X_t, \mathcal{F}_t)$ is a right-continuous submartingale, then for any $t \in \mathbb{R}^+$ and any $\lambda > 0$*

$$\lambda \mathrm{P}(A(\lambda, t)) \le \int_{A(\lambda,t)} X_t d\mathrm{P} \le \mathrm{E}\max(X_t, 0),$$

where $A(\lambda, t) = \{\omega : \sup_{0\le s\le t} X_s(\omega) \ge \lambda\}$.

13.7 Convergence of Martingales

We first discuss convergence of martingales with discrete time.

Definition 13.31. *A martingale $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ is said to be right-closable if there is a random variable $X_\infty \in L^1(\Omega, \mathcal{F}, \mathrm{P})$ such that $\mathrm{E}(X_\infty|\mathcal{F}_n) = X_n$ for all $n \in \mathbb{N}$.*

The random variable X_∞ is sometimes referred to as the last element of the martingale.

We can define $\mathcal{F}_\infty$ as the minimal σ-algebra containing $\mathcal{F}_n$ for all n. For a right-closable martingale we can define $X'_\infty = \mathrm{E}(X_\infty|\mathcal{F}_\infty)$. Then X'_∞ also serves as the last element since

$$\mathrm{E}(X'_\infty|\mathcal{F}_n) = \mathrm{E}(\mathrm{E}(X_\infty|\mathcal{F}_\infty)|\mathcal{F}_n) = \mathrm{E}(X_\infty|\mathcal{F}_n) = X_n.$$

Therefore, without loss of generality, we shall assume from now on that, for a right-closable martingale, the last element X_∞ is $\mathcal{F}_\infty$-measurable.

Theorem 13.32. *A martingale is right-closable if and only if it is uniformly integrable (that is the sequence of random variables $X_n, n \in \mathbb{N}$, is uniformly integrable).*

We shall only prove that a right-closable martingale is uniformly integrable. The proof of the converse statement is slightly more complicated, and we omit it here. Interested readers may find it in "Real Analysis and Probability" by R. M. Dudley.

Proof. We need to show that

$$\lim_{\lambda\to\infty} \sup_{n\in\mathbb{N}} \int_{\{|X_n|>\lambda\}} |X_n| d\mathrm{P} = 0.$$

Since $|\cdot|$ is a convex function,

$$|X_n| = |\mathrm{E}(X_\infty|\mathcal{F}_n)| \le \mathrm{E}(|X_\infty||\mathcal{F}_n)$$

by the Conditional Jensen's Inequality. Therefore,

$$\int_{\{|X_n|>\lambda\}} |X_n| d\mathrm{P} \le \int_{\{|X_n|>\lambda\}} |X_\infty| d\mathrm{P}.$$

Since $|X_\infty|$ is integrable and the integral is absolutely continuous with respect to the measure P, it is sufficient to prove that

$$\lim_{\lambda\to\infty} \sup_{n\in\mathbb{N}} \mathrm{P}\{|X_n| > \lambda\} = 0.$$

By the Chebyshev Inequality,

$$\lim_{\lambda\to\infty} \sup_{n\in\mathbb{N}} \mathrm{P}\{|X_n| > \lambda\} \le \lim_{\lambda\to\infty} \sup_{n\in\mathbb{N}} \mathrm{E}|X_n|/\lambda \le \lim_{\lambda\to\infty} \mathrm{E}|X_\infty|/\lambda = 0,$$

which proves that a right-closable martingale is uniformly integrable. □

The fact that a martingale is right-closable is sufficient to establish convergence in probability and in L^1.

Theorem 13.33 (Doob). *Let $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ be a right-closable martingale. Then*

$$\lim_{n\to\infty} X_n = X_\infty$$

almost surely and in $L^1(\Omega, \mathcal{F}, \mathrm{P})$.

Proof. (Due to C.W. Lamb.) Let $\mathcal{K} = \bigcup_{n\in\mathbb{N}} \mathcal{F}_n$. Let $\mathcal{G}$ be the collection of sets which can be approximated by sets from $\mathcal{K}$. Namely, $A \in \mathcal{G}$ if for any $\varepsilon > 0$ there is $B \in \mathcal{K}$ such that $\mathrm{P}(A\Delta B) < \varepsilon$. It is clear that $\mathcal{K}$ is a π-system, and that $\mathcal{G}$ is a Dynkin system. Therefore, $\mathcal{F}_\infty = \sigma(\mathcal{K}) \subseteq \mathcal{G}$ by Lemma 4.13.

Let F be the set of functions which are in $L^1(\Omega, \mathcal{F}, \mathrm{P})$ and are measurable with respect to $\mathcal{F}_n$ for some $n < \infty$. We claim that F is dense in $L^1(\Omega, \mathcal{F}_\infty, \mathrm{P})$. Indeed, any indicator function of a set from $\mathcal{F}_\infty$ can be approximated by elements of F, as we just demonstrated. Therefore, the same is true for finite linear combinations of indicator functions which, in turn, are dense in $L^1(\Omega, \mathcal{F}_\infty, \mathrm{P})$.

Since X_∞ is $\mathcal{F}_\infty$-measurable, for any $\varepsilon > 0$ we can find $Y_\infty \in F$ such that $\mathrm{E}|X_\infty - Y_\infty| \le \varepsilon^2$. Let $Y_n = \mathrm{E}(Y_\infty|\mathcal{F}_n)$. Then $(X_n - Y_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ is a martingale. Therefore, $(|X_n - Y_n|, \mathcal{F}_n)_{n\in\mathbb{N}}$ is a submartingale, as shown in Sect. 13.4, and $\mathrm{E}|X_n - Y_n| \le \mathrm{E}|X_\infty - Y_\infty|$. By Doob's Inequality (Theorem 13.22),

$$\mathrm{P}(\sup_{n\in\mathbb{N}} |X_n - Y_n| > \varepsilon) \le \sup_{n\in\mathbb{N}} \mathrm{E}|X_n - Y_n|/\varepsilon \le \mathrm{E}|X_\infty - Y_\infty|/\varepsilon \le \varepsilon.$$

Note that $Y_n = Y_\infty$ for large enough n, since Y_∞ is $\mathcal{F}_n$ measurable for some finite n. Therefore,

$$\mathrm{P}(\limsup_{n\to\infty} X_n - Y_\infty > \varepsilon) \le \varepsilon \quad \text{and} \quad \mathrm{P}(\liminf_{n\to\infty} X_n - Y_\infty < -\varepsilon) \le \varepsilon.$$

Also, by the Chebyshev Inequality, $\mathrm{P}(|X_\infty - Y_\infty| > \varepsilon) \leq \varepsilon$. Therefore,

$$\mathrm{P}(\limsup_{n\to\infty} X_n - X_\infty > 2\varepsilon) \leq 2\varepsilon \quad \text{and} \quad \mathrm{P}(\liminf_{n\to\infty} X_n - X_\infty < -2\varepsilon) \leq 2\varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, this implies that $\lim_{n\to\infty} X_n = X_\infty$ almost surely.

As shown above, for each $\varepsilon > 0$ we have the inequalities $\mathrm{E}|X_\infty - Y_\infty| \leq \varepsilon^2$, $\mathrm{E}|X_n - Y_n| \leq \mathrm{E}|X_\infty - Y_\infty|$, while $Y_n = Y_\infty$ for all sufficiently large n. This implies the convergence of X_n to X_∞ in $L^1(\Omega, \mathcal{F}, \mathrm{P})$. □

Example (Polya Urn Scheme). Consider an urn containing one black and one white ball. At time step n we take a ball randomly out of the urn and replace it with two balls of the same color.

More precisely, consider two processes A_n (number of black balls) and B_n (number of white balls). Then $A_0 = B_0 = 1$, and A_n, B_n, $n \geq 1$, are defined inductively as follows: $A_n = A_{n-1} + \xi_n$, $B_n = B_{n-1} + (1 - \xi_n)$, where ξ_n is a random variable such that

$$\mathrm{P}(\xi_n = 1|\mathcal{F}_{n-1}) = \frac{A_{n-1}}{A_{n-1} + B_{n-1}}, \quad \text{and} \quad \mathrm{P}(\xi_n = 0|\mathcal{F}_{n-1}) = \frac{B_{n-1}}{A_{n-1} + B_{n-1}},$$

and $\mathcal{F}_{n-1}$ is the σ-algebra generated by all A_k, B_k with $k \leq n-1$. Let $X_n = A_n/(A_n + B_n)$ be the proportion of black balls. Let us show that $(X_n, \mathcal{F}_n)_{n\geq 0}$ is a martingale. Indeed,

$$\mathrm{E}(X_n - X_{n-1}|\mathcal{F}_{n-1}) = \mathrm{E}(\frac{A_n}{A_n + B_n} - \frac{A_{n-1}}{A_{n-1} + B_{n-1}}|\mathcal{F}_{n-1}) =$$

$$\mathrm{E}(\frac{(A_{n-1} + B_{n-1})\xi_n - A_{n-1}}{(A_n + B_n)(A_{n-1} + B_{n-1})}|\mathcal{F}_{n-1}) =$$

$$\frac{1}{A_n + B_n}\mathrm{E}(\xi_n - \frac{A_{n-1}}{A_{n-1} + B_{n-1}})|\mathcal{F}_{n-1}) = 0,$$

as is required of a martingale. Here we used that $A_n + B_n = A_{n-1} + B_{n-1} + 1$, and is therefore $\mathcal{F}_{n-1}$-measurable. The martingale $(X_n, \mathcal{F}_n)_{n\geq 0}$ is uniformly integrable, simply because X_n are bounded by one. Therefore, by Theorem 13.33, there is a random variable X_∞ such that $\lim_{n\to\infty} X_n = X_\infty$ almost surely.

We can actually write the distribution of X_∞ explicitly. The variable A_n can take integer values between 1 and $n+1$. We claim that $\mathrm{P}(A_n = k) = 1/(n+1)$ for all $1 \leq k \leq n+1$. Indeed, the statement is obvious for $n = 0$. For $n \geq 1$, by induction,

$$\mathrm{P}(A_n = k) = \mathrm{P}(A_{n-1} = k-1; \xi_n = 1) + \mathrm{P}(A_{n-1} = k; \xi_n = 0) =$$

$$\frac{1}{n} \cdot \frac{k-1}{n+1} + \frac{1}{n} \cdot \frac{n-k+1}{n+1} = \frac{1}{n+1}.$$

This means that $\mathrm{P}(X_n = k/(n+2)) = 1/(n+1)$ for $1 \leq k \leq n+1$. Since the sequence X_n converges to X_∞ almost surely, it also converges in distribution. Therefore, the distribution of X_∞ is uniform on the interval $[0, 1]$.

If $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ is bounded in $L^1(\Omega, \mathcal{F}, \mathrm{P})$ (that is $\mathrm{E}|X_n| \le c$ for some constant c and all n), we cannot claim that it is right-closable. Yet, the L^1-boundedness still guarantees almost sure convergence, although not necessarily to the last element of the martingale (which does not exist unless the martingale is uniformly integrable). We state the following theorem without a proof.

Theorem 13.34 (Doob). *Let $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ be a $L^1(\Omega, \mathcal{F}, \mathrm{P})$-bounded martingale. Then*

$$\lim_{n\to\infty} X_n = Y$$

almost surely, where Y is some random variable from $L^1(\Omega, \mathcal{F}, \mathrm{P})$.

Remark 13.35. Although the random variable Y belongs to $L^1(\Omega, \mathcal{F}, \mathrm{P})$, the sequence X_n need not converge to Y in $L^1(\Omega, \mathcal{F}, \mathrm{P})$.

Let us briefly examine the convergence of submartingales. Let $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ be an $L^1(\Omega, \mathcal{F}, \mathrm{P})$-bounded submartingale, and let $X_n = M_n + A_n$ be its Doob Decomposition. Then $\mathrm{E}A_n = \mathrm{E}(X_n - M_n) = \mathrm{E}(X_n - M_1)$. Thus, A_n is a monotonically non-decreasing sequence of random variables which is bounded in $L^1(\Omega, \mathcal{F}, \mathrm{P})$. By the Levi Monotonic Convergence Theorem, there exists the almost sure limit $A = \lim_{n\to\infty} A_n \in L^1(\Omega, \mathcal{F}, \mathrm{P})$.

Since A_n are bounded in $L^1(\Omega, \mathcal{F}, \mathrm{P})$, so too are M_n. Since A_n are non-negative random variables bounded from above by A, they are uniformly integrable. Therefore, if $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ is a uniformly integrable submartingale, then $(M_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ is a uniformly integrable martingale. Upon gathering the above arguments, and applying Theorems 13.33 and 13.34, we obtain the following lemma.

Lemma 13.36. *Let a submartingale $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ be bounded in $L^1(\Omega, \mathcal{F}, \mathrm{P})$. Then*

$$\lim_{n\to\infty} X_n = Y$$

almost surely, where Y is some random variable from $L^1(\Omega, \mathcal{F}, \mathrm{P})$. If X_n are uniformly integrable, then the convergence is also in $L^1(\Omega, \mathcal{F}, \mathrm{P})$.

Although our discussion of martingale convergence has been focused so far on martingales with discrete time, the same results carry over to the case of right-continuous martingales with continuous time. In Definition 13.31 and Theorems 13.33 and 13.34 we only need to replace the parameter $n \in \mathbb{N}$ by $t \in \mathbb{R}^+$. Since the proof of Lemma 13.36 in the continuous time case relies on the Doob-Meyer Decomposition, in order to make it valid in the continuous time case, we must additionally assume that the filtration satisfies the usual conditions and that the submartingale is continuous.

13.8 Problems

1. Let $g : \mathbb{R} \to \mathbb{R}$ be a measurable function which is not convex. Show that there is a random variable f on some probability space such that $\mathrm{E}|f| < \infty$ and $-\infty < \mathrm{E}g(f) < g(\mathrm{E}f) < \infty$.

2. Let ξ and η be two random variables with finite expectations such that $\mathrm{E}(\xi|\eta) \geq \eta$ and $\mathrm{E}(\eta|\xi) \geq \xi$. Prove that $\xi = \eta$ almost surely.

3. Let $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ be a square-integrable martingale with $\mathrm{E}X_1 = 0$. Show that for each $c > 0$

$$\mathrm{P}(\max_{1\leq i\leq n} X_i \geq c) \leq \frac{\mathrm{Var}(X_n)}{\mathrm{Var}(X_n) + c^2}.$$

4. Let $(\xi_1, \ldots, \xi_n)$ be a Gaussian vector with zero mean and covariance matrix B. Find the distribution of the random variable $\mathrm{E}(\xi_1|\xi_2, \ldots, \xi_n)$.

5. Let $A = \{(x, y) \in \mathbb{R}^2 : |x - y| < a,\ |x + y| < b\}$, where $a, b > 0$. Assume that the random vector (ξ_1, ξ_2) is uniformly distributed on A. Find the distribution of $\mathrm{E}(\xi_1|\xi_2)$.

6. Let ξ_1, ξ_2, ξ_3 be independent identically distributed bounded random variables with density $p(x)$. Find the distribution of

$$\mathrm{E}(\max(\xi_1, \xi_2, \xi_3)|\min(\xi_1, \xi_2, \xi_3))$$

in terms of the density p.

7. Prove that if σ and τ are stopping times of a filtration $\mathcal{F}_t$, then so is $\sigma \vee \tau$. If, in addition, $\sigma, \tau \geq 0$, then $\sigma + \tau$ is a stopping time.

8. Let $\xi_1, \xi_2, \ldots$ be independent $N(0, 1)$ distributed random variables. Let $S_n = \xi_1 + \ldots + \xi_n$ and $X_n = e^{S_n - n/2}$. Let $\mathcal{F}_n^X$ be the σ-algebra generated by $X_1, \ldots, X_n$. Prove that $(X_n, \mathcal{F}_n^X)_{n\in\mathbb{N}}$ is a martingale.

9. Prove that the definition of uniform integrability given in Remark 13.21 is equivalent to Definition 13.19.

10. A man tossing a coin wins one point for heads and five points for tails. The game stops when the man accumulates at least 1,000 points. Estimate with an accuracy ± 2 the expectation of the length of the game.

11. Let X_n be a process adapted to a filtration $\mathcal{F}_n$, $n \in \mathbb{N}$. Let $M > 0$ and $\tau(\omega) = \min(n : |X_n(\omega)| \geq M)$ (where $\tau(\omega) = \infty$ if $|X_n(\omega)| < M$ for all n). Prove that τ is a stopping time of the filtration $\mathcal{F}_n$.

12. Let a martingale $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ be uniformly integrable. Let the stopping time τ be defined as in the previous problem. Prove that $(X_{n\wedge\tau}, \mathcal{F}_{n\wedge\tau})_{n\in\mathbb{N}}$ is a uniformly integrable martingale.

13. Let N_n, $n \geq 1$, be the size of a population of bacteria at time step n. At each time step each bacteria produces a number of offspring and dies. The number of offspring is independent for each bacteria and is distributed according to the Poisson law with parameter $\lambda = 2$. Assuming that $N_1 = a > 0$, find the probability that the population will eventually die, that is find $\mathrm{P}\,(N_n = 0 \text{ for some } n \geq 1)$. (Hint: find c such that $\exp(-cN_n)$ is a martingale.)

14. Ann and Bob are gambling at a casino. In each game the probability of winning a dollar is 48%, and the probability of loosing a dollar is 52%. Ann decided to play 20 games, but will stop after 2 games if she wins them both. Bob decided to play 20 games, but will stop after 10 games if he wins at least 9 out of the first 10. What is larger: the amount of money Ann is expected to loose, or the amount of money Bob is expected to loose?

15. Let $(X_t, \mathcal{F}_t)_{t \in \mathbb{R}}$ be a martingale with continuous realizations. For $0 \le s \le t$, find $\mathrm{E}(\int_0^t X_u du | \mathcal{F}_s)$.

16. Consider an urn containing A_0 black balls and B_0 white balls. At time step n we take a ball randomly out of the urn and replace it with two balls of the same color. Let X_n denote the proportion of the black balls. Prove that X_n converges almost surely, and find the distribution of the limit.