Chapter V
Conditioning and Martingales

Up to this point I have been dealing with random variables which are either themselves mutually independent or are built out of other random variables which are. For this reason, it has not been necessary for me to make explicit use of the concept of *conditioning*, although, as we will see shortly, this concept has been lurking silently in the background. In this chapter I will first give the modern formulation of conditional expectations and then provide an example of the way in which conditional expectations can be used.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and suppose that $A \in \mathcal{F}$ is a set having positive $\mathbb{P}$-measure. For reasons which are most easily understand when Ω is finite and $\mathbb{P}$ is uniform, the ratio

$$\mathbb{P}(B|A) \equiv \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}, \quad B \in \mathcal{F},$$

is called the **conditional probability of** B **given** A. As one learns in an elementary course, the introduction of conditional probabilities makes many calculations much simpler; in particular, conditional probabilities help to clarify dependence relations between the events represented by A and B. For example, B is independent of A precisely when $\mathbb{P}(B|A) = P(B)$ or, in words, *when the condition that* A *occurs does not change the probability that* B *occurs*. Thus, it is unfortunate that the naïve definition of conditioning as described above does not cover many important situations. For example, suppose that X and Y are random variables and that one wants to talk about the conditional probability that $Y \leq b$ given that $X = a$. Unless one is very lucky and $\mathbb{P}(X = a) > 0$, dividing by $\mathbb{P}(X = a)$ is not going to do the job. As this example illustrates, it is of great importance to generalize the concept of conditional probability to include situations when the event on which one is conditioning has $\mathbb{P}$-measure 0, and so the next section is devoted to Kolmogorov's elegant solution to the problem of doing so.

§ 5.1 Conditioning

In order to appreciate the idea behind Kolmogorov's solution, imagine someone told you the conditional probability that the event B occurs given that the event A occurs. Obviously, since you have no way of saying anything about the probability of B when A does not occur, she has provided you with incomplete information about B. Thus, before you are satisfied, you should demand to know also what is the conditional probability of B given that A does not occur. Of course, this second piece of information is relevant only if A *is not certain*, in which case $\mathbb{P}(A) < 1$ and therefore $\mathbb{P}\big(B\big|A\complement\big)$ is well defined. More generally, suppose that $\mathcal{P} = \big\{A_1, \dots, A_N\big\}$ (N here may be either finite or countably infinite) is a partition of Ω into elements of $\mathcal{F}$ having positive $\mathbb{P}$-measure. Then, in order to have complete information about the probability of $B \in \mathcal{F}$ relative to $\mathcal{P}$, one has to know the entire list of the numbers $\mathbb{P}\big(B\big|A_n\big)$, $1 \le n \le N$. Next, suppose that one attempts to describe this list in a way which does not depend explicitly on the positivity of the numbers $\mathbb{P}(A_n)$. For this purpose, consider the function

$$\omega \in \Omega \longmapsto f(\omega) \equiv \sum_{n=1}^{N} \mathbb{P}\big(B\big|A_n\big)\, \mathbf{1}_{A_n}(\omega).$$

Clearly, f is not only $\mathcal{F}$-measurable, it is measurable with respect to the σ-algebra $\sigma(\mathcal{P})$ over Ω generated by $\mathcal{P}$. In particular (because the only $\sigma(\mathcal{P})$-measurable set of $\mathbb{P}$-measure 0 is empty), f is uniquely determined by its $\mathbb{P}$-integrals $\mathbb{E}^{\mathbb{P}}[f, A]$ over sets $A \in \sigma(\mathcal{P})$. Moreover, because, for each $B \in \sigma(\mathcal{P})$ and n, either $A_n \subseteq B$ or $B \cap A_n = \emptyset$, we have that

$$\mathbb{E}^{\mathbb{P}}\big[f, A\big] = \sum_{n=1}^{N} \mathbb{P}\big(B \cap A_n\big) = \sum_{\{n: A_n \subseteq B\}} \mathbb{P}\big(A_n \cap B\big) = \mathbb{P}\big(A \cap B\big).$$

Hence, the function f is uniquely determined by the property that

$$\mathbb{E}^{\mathbb{P}}\big[f, A\big] = \mathbb{P}\big(A \cap B\big) \quad \text{for every} \quad A \in \sigma(\mathcal{P}).$$

The beauty of this description is that it makes perfectly good sense even if some of the A_n's have $\mathbb{P}$-measure 0, only in that case the description does not determine f pointwise but merely up to a $\sigma(\mathcal{P})$-measurable **$\mathbb{P}$-null set** (i.e., a set of $\mathbb{P}$-measure 0), which is the very least one should expect to pay for *dividing by* 0.

§ **5.1.1. Kolmogorov's Definition.** With the preceding discussion in mind, one ought to find the following formulation reasonable. Namely, given a sub-σ-algebra $\Sigma \subseteq \mathcal{F}$ and a $(-\infty, \infty]$-valued random variable X whose negative part $X^-\big(\equiv -(X \wedge 0)\big)$ is $\mathbb{P}$-integrable, I will say that the random variable X_Σ is a **conditional expectation of** X **given** Σ if X_Σ is $(-\infty, \infty]$-valued and Σ-measurable, $\big(X_\Sigma\big)^-$ is $\mathbb{P}$-integrable, and

$$\mathbb{E}^{\mathbb{P}}\big[X_\Sigma, A\big] = \mathbb{E}^{\mathbb{P}}\big[X, A\big] \quad \text{for every } A \in \Sigma. \tag{5.1.1}$$

Obviously, having made this definition, my first duty is to show that such an X_Σ always exists and to discover in what sense it is uniquely determined. The latter problem is dealt with in the following lemma.

Lemma 5.1.2. *Let Σ be a sub-σ-algebra of $\mathcal{F}$, and suppose that X_Σ and Y_Σ are a pair of $(-\infty,\infty]$-valued Σ-measurable random variables for which X_Σ^- and Y_Σ^- are both $\mathbb{P}$-integrable. Then*

$$\mathbb{E}^{\mathbb{P}}\big[X_\Sigma,\, A\big] \le \mathbb{E}^{\mathbb{P}}\big[Y_\Sigma,\, A\big] \quad \text{for every } A \in \Sigma,$$

if and only if $X_\Sigma \le Y_\Sigma$ (a.s., $\mathbb{P}$).

Proof: Without loss in generality, I may and will assume that $\Sigma = \mathcal{F}$ and will therefore drop the subscript Σ; and, since the "if" implication is completely trivial, I will discuss only the minimally less trivial "only if" assertion. Thus, suppose that $\mathbb{P}$-integrals of Y dominate those of X and yet that $X > Y$ on a set of positive $\mathbb{P}$-measure. We could then choose an $M \in [1,\infty)$ so that $\mathbb{P}(A) \vee P(B) > 0$ where

$$A \equiv \big\{X \le M \text{ and } Y \le X - \tfrac{1}{M}\big\} \quad \text{and} \quad B \equiv \big\{X = \infty \text{ and } Y \le M\big\}.$$

But if $\mathbb{P}(A) > 0$, then

$$\mathbb{E}^{\mathbb{P}}\big[X,\, A\big] \le \mathbb{E}^{\mathbb{P}}\big[Y,\, A\big] \le \mathbb{E}^{\mathbb{P}}\big[X,\, A\big] - \tfrac{1}{M}P(A),$$

which, because $\mathbb{E}^{\mathbb{P}}\big[X,\, A\big]$ is a finite number, is impossible. At the same time, if $\mathbb{P}(B) > 0$, then

$$\infty = \mathbb{E}^{\mathbb{P}}\big[X,\, B\big] \le \mathbb{E}^{\mathbb{P}}\big[Y,\, B\big] \le M < \infty,$$

which is also impossible. □

Theorem 5.1.3. *Let Σ be a sub-σ-algebra of $\mathcal{F}$ and X a $(-\infty,\infty]$-valued random variable for which X^- is $\mathbb{P}$-integrable. Then there exists a conditional expectation value X_Σ of X. Moreover, if Y is a second $(-\infty,\infty]$-valued random variable and $Y \ge X$ (a.s., $\mathbb{P}$), then Y^- is $\mathbb{P}$-integrable and $Y_\Sigma \ge X_\Sigma$ (a.s., $\mathbb{P}$) for any Y_Σ which is a conditional expectation value of Y given Σ. In particular, if $X = Y$ (a.s., $\mathbb{P}$), then $\big\{Y_\Sigma \ne X_\Sigma\big\}$ is a Σ-measurable, $\mathbb{P}$-null set.**

Proof: In view of Lemma 5.1.2, it suffices for me to handle the initial existence statement. To this end, let $\mathcal{G}$ denote the class of X satisfying $\mathbb{E}^{\mathbb{P}}\big[X^-\big] < \infty$ for which an X_Σ exists, and let $\mathcal{G}^+$ denote the non-negative elements of $\mathcal{G}$. If $\{X_n : n \ge 1\} \subseteq \mathcal{G}^+$ is non-decreasing and, for each $n \in \mathbb{Z}^+$, $\big(X_n\big)_\Sigma$ denotes a conditional expectation of X_n given Σ, then $0 \le \big(X_n\big)_\Sigma \le \big(X_{n+1}\big)_\Sigma$ (a.s., $\mathbb{P}$), and therefore we can arrange that $0 \le \big(X_n\big)_\Sigma \le \big(X_{n+1}\big)_\Sigma$ everywhere. In particular, if X and X_Σ are the pointwise limits of the X_n's and $\big(X_n\big)_\Sigma$'s, respectively, then the Monotone Convergence Theorem guarantees that X_Σ is a

* Kolmogorov himself, and most authors ever since, have obtained the existence of conditional expectation values as a consequence of the Radon–Nikodym Theorem. Because I find projections more intuitively appealing, I prefer the approach given here.

conditional expectation of X given Σ. Hence, we now know that $\mathcal{G}^+$ is closed under non-decreasing, pointwise limits, and therefore we will know that $\mathcal{G}^+$ contains all non-negative random variables X as soon as we show that $\mathfrak{G}$ contains all bounded X's. But if X is bounded (and is therefore an element of $L^2(\mathbb{P};\mathbb{R})$) and $L_\Sigma = L^2(\Omega, \Sigma, \mathbb{P}; \mathbb{R})$ is the subspace of $L^2(\mathbb{P};\mathbb{R})$ consisting of its Σ-measurable elements, then the orthogonal projection X_Σ of X onto L_Σ is a Σ-measurable random variable which is square $\mathbb{P}$-integrable and satisfies (5.1.1).

So far I have proved that $\mathcal{G}^+$ contains all non-negative, $\mathcal{F}$-measurable X's. Furthermore, if X is non-negative, then (by Lemma 5.1.2) $X_\Sigma \geq 0$ (a.s., $\mathbb{P}$) and so X_Σ is $\mathbb{P}$-integrable precisely when X itself is. In particular, we can arrange to make X_Σ take its values in $[0,\infty)$ when X is non-negative and $\mathbb{P}$-integrable. Finally, to see that $X \in \mathcal{G}$ for every X with $\mathbb{E}^{\mathbb{P}}\big[X^-\big] < \infty$, simply consider X^+ and X^- separately, apply the preceding to show that $\big(X^\pm\big)_\Sigma \geq 0$ (a.s., $\mathbb{P}$) and that $\big(X^-\big)_\Sigma$ is $\mathbb{P}$-integrable, and check that the random variable

$$X_\Sigma \equiv \begin{cases} \big(X^+\big)_\Sigma - \big(X^-\big)_\Sigma & \text{when} \quad \big(X^\pm\big)_\Sigma \geq 0 \text{ and } \big(X^-\big)_\Sigma < \infty \\ 0 & \text{otherwise} \end{cases}$$

is a conditional expectation of X given Σ. □

Convention. Because it is determined only up to a Σ-measurable $\mathbb{P}$-null set, one cannot, in general, talk about *the* conditional expectation of X as a *function*. Instead, the best that one can do is say that **the conditional expectation of** X is the equivalence class of Σ-measurable X_Σ's which satisfy (5.1.1), and I will adopt the notation $\mathbb{E}^{\mathbb{P}}[X|\Sigma]$ to denote this equivalence class. On the other hand, because one is usually interested only in $\mathbb{P}$-integrals of conditional expectations, it has become common practice to ignore, for the most part, the distinction between the equivalence class $\mathbb{E}^{\mathbb{P}}[X|\Sigma]$ and the members of that equivalence class. Thus (just as one would when dealing with the Lebesgue spaces) I will abuse notation by using $\mathbb{E}^{\mathbb{P}}[X|\Sigma]$ to denote a generic element of the equivalence class $\mathbb{E}^{\mathbb{P}}[X|\Sigma]$ and will be more precise only when $\mathbb{E}^{\mathbb{P}}[X|\Sigma]$ contains some particularly distinguished member. For example, recall the random variables T_n entering the definition of the simple Poisson process $\{N(t) : t \in (0,\infty)\}$ in § 4.2.1. It is then clear that we can take

$$\mathbb{E}^{\mathbb{P}}\Big[\mathbf{1}_{\{n\}}\big(N(t)\big)\,\Big|\,\sigma\big(T_1,\dots,T_n\big)\Big] = \mathbf{1}_{[0,t]}\big(T_n\big)e^{-(t-T_n)},$$

and one would be foolish to take any other representative. More generally, I will always take non-negative representatives of $\mathbb{E}^{\mathbb{P}}[X|\Sigma]$ when X itself is non-negative and $\mathbb{R}$-valued representatives when X is $\mathbb{P}$-integrable. Finally, for historical reasons, it is usual to distinguish the case when X is the indicator function $\mathbf{1}_B$ of a set $B \in \mathcal{F}$ and to call $\mathbb{E}^{\mathbb{P}}[\mathbf{1}_B|\Sigma]$ the **conditional probability of** B **given** Σ and to write $\mathbb{P}(B|\Sigma)$ instead of $\mathbb{E}^{\mathbb{P}}[\mathbf{1}_B|\Sigma]$. Of course, representatives of $\mathbb{P}(B|\Sigma)$ will always be assumed to take their values in $[0,1]$.

Once one has established the existence and uniqueness of conditional expectations, there is a long list of more or less obvious properties which one can easily verify. The following theorem contains some of the more important items which ought to appear on such a list.

THEOREM 5.1.4. *Let Σ be a sub-σ-algebra of $\mathcal{F}$. If X is a $\mathbb{P}$-integrable random variable and $\mathcal{C} \subseteq \Sigma$ is a π-system (cf. Exercise 1.1.12) which generates Σ, then*

$$Y = \mathbb{E}^{\mathbb{P}}\big[X\big|\Sigma\big] \quad (\text{a.s.}, \mathbb{P}) \iff$$
$$Y \in L^1(\Omega, \Sigma, \mathbb{P}; \mathbb{R}) \text{ and } \mathbb{E}^{\mathbb{P}}\big[Y,\, A\big] = \mathbb{E}^{\mathbb{P}}\big[X,\, A\big] \text{ for } A \in \mathcal{C} \cup \{\Omega\}.$$

Moreover, if X is any $(-\infty, \infty]$-valued random variable which satisfies $\mathbb{E}^{\mathbb{P}}[X^-] < \infty$, then each of the following relations holds $\mathbb{P}$-almost surely:

$$\big|\mathbb{E}^{\mathbb{P}}\big[X\big|\Sigma\big]\big| \le \mathbb{E}^{\mathbb{P}}\big[|X|\big|\Sigma\big]; \tag{5.1.5}$$

$$\mathbb{E}^{\mathbb{P}}\big[X\big|\mathcal{T}\big] = \mathbb{E}^{\mathbb{P}}\Big[\mathbb{E}^{\mathbb{P}}\big[X\big|\Sigma\big]\,\Big|\mathcal{T}\Big] \tag{5.1.6}$$

when $\mathcal{T}$ is a sub-σ-algebra of Σ; and, when X is $\mathbb{R}$-valued and $\mathbb{P}$-integrable,

$$\mathbb{E}^{\mathbb{P}}\big[-X\big|\Sigma\big] = -\mathbb{E}^{\mathbb{P}}\big[X\big|\Sigma\big].$$

Next, let Y be a second $(-\infty, \infty]$-valued random variable with $\mathbb{E}^{\mathbb{P}}[Y^-] < \infty$. Then, $\mathbb{P}$-almost surely:

$$\mathbb{E}^{\mathbb{P}}\big[\alpha X + \beta Y\big|\Sigma\big] = \alpha\mathbb{E}^{\mathbb{P}}\big[X\big|\Sigma\big] + \beta\mathbb{E}^{\mathbb{P}}\big[Y\big|\Sigma\big] \quad \text{for each } \alpha,\, \beta \in [0, \infty),$$

and

$$\mathbb{E}^{\mathbb{P}}\big[YX\big|\Sigma\big] = Y\,\mathbb{E}^{\mathbb{P}}\big[X\big|\Sigma\big] \tag{5.1.7}$$

if Y is Σ-measurable and $(XY)^-$ is $\mathbb{P}$-integrable. Finally, suppose that $\{X_n : n \ge 1\}$ is a sequence of $(-\infty, \infty]$-valued random variables. Then, $\mathbb{P}$-almost surely:

$$\mathbb{E}^{\mathbb{P}}\big[X_n\big|\Sigma\big] \nearrow \mathbb{E}^{\mathbb{P}}\big[X\big|\Sigma\big] \tag{5.1.8}$$

if $\mathbb{E}^{\mathbb{P}}[X_1^-] < \infty$ and $X_n \nearrow X$ (a.s., $\mathbb{P}$)*; and, more generally,*

$$\mathbb{E}^{\mathbb{P}}\left[\varliminf_{n\to\infty} X_n\middle|\,\Sigma\right] \le \varliminf_{n\to\infty} \mathbb{E}^{\mathbb{P}}\big[X_n\big|\Sigma\big]. \tag{5.1.9}$$

if $X_n \ge 0$ (a.s., $\mathbb{P}$) for each $n \in \mathbb{Z}^+$.

Proof: To prove the first assertion, note that the set of $A \in \Sigma$ for which $\mathbb{E}^{\mathbb{P}}[X, A] = \mathbb{E}^{\mathbb{P}}[Y, A]$ is (cf. Exercise 1.1.12) a λ-system which contains $\mathcal{C}$ and therefore Σ. Next, clearly (5.1.5) is just an application of Lemma 5.1.2, while (5.1.6) and the two equations which follow it are all expressions of uniqueness. As for the next equation, one can first reduce to the case with both X and Y are non-negative. Then one can use uniqueness to check it when Y is the indicator function of a element of Σ, use linearity to extend it to simple Σ-measurable functions, and complete the job by taking monotone limits. Finally, (5.1.8) is an immediate application of the Monotone Convergence Theorem; whereas (5.1.9) comes from the conjunction of

$$\mathbb{E}^{\mathbb{P}}\left[\inf_{n\geq m} X_n \,\middle|\, \Sigma\right] \leq \inf_{n\geq m} \mathbb{E}^{\mathbb{P}}\big[X_n \,\big|\, \Sigma\big] \quad (\text{a.s.}, \mathbb{P}), \quad m \in \mathbb{Z}^+,$$

with (5.1.8). $\square$

It probably will have occurred to most readers that the properties discussed in Theorem 5.1.4 give strong evidence that, for fixed $\omega \in \Omega$, $X \longmapsto \mathbb{E}^{\mathbb{P}}[X|\Sigma](\omega)$ behaves like an integral (in the sense of Daniell) and therefore ought to be expressible in terms of integration with respect to a probability measure $\mathbb{P}_\omega$. Indeed, if one could actually talk about $X \longmapsto \mathbb{E}^{\mathbb{P}}[X|\Sigma](\omega)$ for a fixed (as opposed to $\mathbb{P}$-almost every) $\omega \in \Omega$, then there is no doubt that such a $\mathbb{P}_\omega$ would have to exist. Thus, it is reasonable to ask whether there are circumstances in which one can gain sufficient control over all the $\mathbb{P}$-null sets involved to really make sense out of $X \longmapsto \mathbb{E}^{\mathbb{P}}[X|\Sigma](\omega)$ for fixed $\omega \in \Omega$. Of course, when Σ is generated by a countable partition $\mathcal{P}$, we already know what to do. Namely, when $\omega \in A \in \mathcal{P}$, we can take

$$\mathbb{E}^{\mathbb{P}}[X|\Sigma](\omega) = \begin{cases} 0 & \text{if } \mathbb{P}(A) = 0 \\ \frac{\mathbb{E}^{\mathbb{P}}[X,\, A]}{\mathbb{P}(A)} & \text{if } \mathbb{P}(A) > 0. \end{cases}$$

Even when Σ does not arise in this way, one can often find a satisfactory representation of conditional expectations as expectations. A quite general statement of this sort is the content of Theorem 9.2.1 in Chapter IX.

§ **5.1.2. Some Extensions.** For various applications it is convenient to have two extensions of the basic theory developed in § 5.1.1. Specifically, as I will now show, the theory is not restricted to probability (or even finite) measures and can be applied to random variables which take their values in a separable Banach space. Thus, from now on, μ will be an arbitrary (non-negative) measure on $(\Omega, \mathcal{F})$ and $\big(E, \|\cdot\|_E\big)$ will be a separable Banach space; and I begin by reviewing a few elementary facts about μ-integration for E-valued random variables.*

* The integration which I outline below is what functional analysts call the Bochner integral for Banach space valued functions. There is a more subtle and intricate theory due to Pettis, but Bochner's theory seems adequate for most probabilistic considerations.

A function $X : \Omega \longrightarrow E$ is said to be **μ-simple** if X is $\mathcal{F}$-measurable, X takes only finitely many values, and $\mu\big(X \neq 0\big) < \infty$, in which case its integral with respect to μ is the element of E given by:

$$\mathbb{E}^\mu[X] = \int_\Omega X(\omega)\,\mu(d\omega) \equiv \sum_{x \in E\setminus\{0\}} x\,\mu(X = x).$$

Notice that another description of $\mathbb{E}^\mu[X]$ is as the unique element of E with the property that

$$\big\langle \mathbb{E}^\mu[X], x^*\big\rangle = \mathbb{E}^\mu\big[\langle X, x^*\rangle\big] \quad \text{for all } x^* \in E^*$$

(I use E^* to denote the dual of E and $\langle x, x^*\rangle$ to denote the action of $x^* \in E^*$ on $x \in E$), and therefore that the mapping taking μ-simple X to $\mathbb{E}^\mu[X]$ is linear. Next, observe that $\omega \in \Omega \longmapsto \|X(\omega)\|_E \in \mathbb{R}$ is $\mathcal{F}$-measurable if $X : \Omega \longrightarrow E$ is $\mathcal{F}$-measurable. In particular, for $\mathcal{F}$-measurable $X : \Omega \longrightarrow E$, we can set

$$\|X\|_{L^p(\mu;E)} = \begin{cases} \mathbb{E}^\mu\big[\|X\|_E^p\big]^{\frac{1}{p}} & \text{if} \quad p \in [1,\infty) \\ \inf\big\{M :\, \mu\big(\|X\|_E > M\big) = 0\big\} & \text{if} \quad p = \infty \end{cases}$$

and will write $X \in L^p(\mu; E)$ when $\|X\|_{L^p(\mu;E)} < \infty$. Also, I will say the $X : \Omega \longrightarrow E$ is **μ-integrable** if $X \in L^1(\mu; E)$; and I will say that X is **locally μ-integrable** if $\mathbf{1}_A\, X$ is μ-integrable for every $A \in \mathcal{F}$ with $\mu(A) < \infty$.

The definition of μ-integration for E-valued X is completed in the following lemma.

Lemma 5.1.10. *For each μ-integrable $X : \Omega \longrightarrow E$ there is a unique element $\mathbb{E}^\mu[X] \in E$ for which $\big\langle \mathbb{E}^\mathbb{P}[X], x^*\big\rangle = \mathbb{E}^\mathbb{P}[\langle X, x^*\rangle]$ for all $x^* \in E^*$. In particular, the mapping $X \in L^1(\mu; E) \longmapsto \mathbb{E}^\mu[X] \in E$ is linear and satisfies*

$$\big\|\mathbb{E}^\mu[X]\big\|_E \leq \mathbb{E}^\mu\big[\|X\|_E\big]. \tag{5.1.11}$$

Finally, if $X \in L^p(\mu; E)$ where $p \in [1, \infty)$, then there is a sequence $\{X_n : n \geq 1\}$ of E-valued, μ-simple functions with the property that $\|X_n - X\|_{L^p(\mu;E)} \longrightarrow 0$.

Proof: Clearly uniqueness, linearity, and (5.1.11) all follow immediately from the given characterization of $\mathbb{E}^\mu[X]$. Thus, all that remains is to prove existence and the final approximation assertion. In fact, once the approximation assertion is proved, then existence will follow immediately from the observation that, by (5.1.11), $\mathbb{E}^\mu[X]$ can be taken equal to $\lim_{n\to\infty} \mathbb{E}^\mu[X_n]$ if $\|X - X_n\|_{L^1(\mu;E)} \longrightarrow 0$.

To prove the approximation assertion, I begin with the case when μ is finite and $M = \sup_{\omega\in\Omega} \|X(\omega)\|_E < \infty$. Next, choose a dense sequence $\{x_\ell : \ell \geq 1\}$ in E, set $A_{0,n} = \emptyset$, and

$$A_{\ell,n} = \Big\{\omega :\, \|X(\omega) - x_\ell\|_E < \tfrac{1}{n}\Big\} \quad \text{for } (\ell, n) \in \mathbb{Z}^+ \times \mathbb{Z}^+.$$

Then, for each $n \in \mathbb{Z}^+$ there exists an $L_n \in \mathbb{Z}^+$ with the property that

$$\mu\left(\Omega \setminus \bigcup_{\ell=1}^{L_n} A_{\ell,n}\right) < \frac{1}{n^p}.$$

Hence, if $X_n : \Omega \longrightarrow E$ is defined so that

$$X_n(\omega) = x_\ell \quad \text{when } 1 \le \ell \le L_n \text{ and } \omega \in A_{\ell,n} \setminus \bigcup_{k=0}^{\ell-1} A_{k,n}$$

and $X_n(\omega) = 0$ when $\omega \notin \bigcup_1^{L_n} A_{\ell,n}$, then X_n is μ-simple and

$$\|X - X_n\|_{L^p(\mu;E)} \le \frac{M + \mu(E)}{n}.$$

In order to handle the general case, let $X \in L^p(\mu;E)$ and $n \in \mathbb{Z}^+$ be given. We can then find an $r_n \in (0,1]$ with the property that

$$\int\limits_{\Omega(r_n)\complement} \|X(\omega)\|_E^p \, \mu(d\omega) \le \frac{1}{(2n)^p},$$

where

$$\Omega(r) \equiv \left\{\omega : \, r \le \|X(\omega)\|_E \le \tfrac{1}{r}\right\} \quad \text{for } r \in (0,1].$$

Since, for any $r \in (0,1]$, $r^p \mu\big(\Omega(r)\big) \le \|X\|^p_{L^p(\mu;E)}$, we can apply the preceding to the restrictions of μ and X to $\Omega(r_n)$ and thereby find a μ-simple $X_n : \Omega(r_n) \longrightarrow E$ with the property

$$\left(\int_{\Omega(r_n)} \|X(\omega) - X_n(\omega)\|_E^p \, \mu(d\omega)\right)^{\frac{1}{p}} \le \frac{1}{2n}.$$

Hence, after extending X_n to Ω by taking it to be 0 off of $\Omega(r_n)$, we arrive at a μ-simple X_n for which $\|X - X_n\|_{L^p(\mu;E)} \le \frac{1}{n}$. $\square$

Given an $\mathcal{F}$-measurable $X : \Omega \longrightarrow E$ and a $B \in \mathcal{F}$ for which $\mathbf{1}_B X \in L^1(\mu;E)$, I will use the notation

$$\mathbb{E}^\mu\big[X, \, B\big] \quad \text{or} \quad \int_B X \, d\mu \quad \text{or} \quad \int_B X(\omega) \, \mu(d\omega)$$

all to denote the quantity $\mathbb{E}^\mu[\mathbf{1}_B X]$. Also, when discussing the spaces $L^p(\mu;E)$, I will adopt the usual convention of blurring the distinction between a particular $\mathcal{F}$-measurable $X : \Omega \longrightarrow E$ belonging to $L^p(\mu;E)$ and the equivalence class of those $\mathcal{F}$-measurable Y's which differ from X on a μ-null set. Thus, with this convention, $\|\cdot\|_{L^p(\mu;E)}$ becomes a bona fide norm (not just a seminorm) on $L^p(\mu;E)$ with respect to which $L^p(\mu;E)$ becomes a normed vector space. Finally, by the same procedure with which one proves the $L^p(\mu;\mathbb{R})$ spaces are complete, one can prove that the spaces $L^p(\mu;E)$ are complete for any separable Banach space.

THEOREM 5.1.12. *Let $(\Omega,\mathcal{F},\mu)$ be a σ-finite measure space and $X:\Omega\longrightarrow E$ a locally μ-integrable function. Then*

$$\mu\big(X\neq 0\big)=0 \iff \mathbb{E}^{\mu}\big[X,\,A\big]=0 \text{ for } A\in\mathcal{F} \text{ with } \mu(A)<\infty.$$

Next, assume that Σ is a sub-σ-algebra for which $\mu\restriction\Sigma$ is σ-finite. Then for each locally μ-integrable $X:\Omega\longrightarrow E$ there is a μ-almost everywhere unique locally μ-integrable, Σ-measurable $X_\Sigma:\Omega\longrightarrow E$ such that

$$\mathbb{E}^{\mu}\big[X_\Sigma,\,A\big]=\mathbb{E}^{\mu}\big[X,\,A\big] \quad \text{for every } A\in\Sigma \text{ with } \mu(A)<\infty. \tag{5.1.13}$$

In particular, if $Y:\Omega\longrightarrow E$ is a second locally μ-integrable function, then, for all $\alpha,\,\beta\in\mathbb{R}$,

$$\big(\alpha X+\beta Y\big)_\Sigma=\alpha X_\Sigma+\beta Y_\Sigma \quad \text{(a.e., } \mu).$$

Finally,

$$\big\|X_\Sigma\big\|_E\le\big(\|X\|_E\big)_\Sigma \quad \text{(a.e., } \mu). \tag{5.1.14}$$

Hence, not only does (5.1.13) continue to hold for any $A\in\Sigma$ with $\mathbf{1}_A X\in L^1(\mu;E)$; but also, for each $p\in[1,\infty]$, the mapping $X\in L^p(\mu;E)\longmapsto X_\Sigma\in L^p(\mu;E)$ is a linear contraction.

PROOF: Clearly, it is only necessary to prove the "$\Longleftarrow$" part of the first assertion. Thus, suppose that $\mu(X\neq 0)>0$. Then, because E is separable and therefore (cf. Exercise 5.1.19) E^* with the weak* topology is also separable, there exists an $\epsilon>0$ and a $x^*\in E^*$ with the property that $\mu\big(\langle X,x^*\rangle\ge\epsilon\big)>0$; from which it follows (by σ-finiteness) that there is an $A\in\mathcal{F}$ for which $\mu(A)<\infty$ and

$$\Big\langle\mathbb{E}^{\mu}\big[X,\,A\big],x^*\Big\rangle=\mathbb{E}^{\mu}\Big[\big\langle X,x^*\big\rangle,\,A\Big]\neq 0.$$

I turn next to the uniqueness and other properties of X_Σ. But it is obvious that uniqueness is an immediate consequence of the first assertion and that linearity follows from uniqueness. As for (5.1.14), notice that if $x^*\in E^*$ and $\|x^*\|_{E^*}\le 1$, then

$$\mathbb{E}^{\mu}\big[\big\langle X_\Sigma,x^*\big\rangle,\,A\big]=\mathbb{E}^{\mu}\big[\big\langle X,x^*\big\rangle,\,A\big]\le\mathbb{E}^{\mu}\big[\|X\|_E,\,A\big]=\mathbb{E}^{\mu}\big[\big(\|X\|_E\big)_\Sigma,\,A\big]$$

for every $A\in\Sigma$ with $\mu(A)<\infty$. Hence, at least when μ is a probability measure, Theorem 5.1.3 implies that $\big\langle X_\Sigma,x^*\big\rangle\le\big(\|X\|_E\big)_\Sigma$ (a.e., μ) for each element x^* from the unit ball in E^*; and so, because E^* with the weak* topology is separable, (5.1.14) follows in this case. To handle μ's which are not probability measures, note that either $\mu(\Omega)=0$, in which case everything is trivial, or $\mu(\Omega)\in(0,\infty)$, in which case we can renormalize μ to make it a probability measure, or $\mu(\Omega) = \infty$, in which case we can use the σ-finiteness of $\mu \upharpoonright \Sigma$ to reduce ourselves to the countable, disjoint union of the preceding cases.

Finally, to prove the existence of X_Σ, I proceed as in the last part of the preceding paragraph to reduce myself to the case when μ is a probability measure $\mathbb{P}$. Next, suppose that X is simple, let R denote its range, and note that

$$X_\Sigma \equiv \sum_{x \in R} x\, \mathbb{P}\big(X = x \,\big|\, \Sigma\big)$$

has the required properties. In order to handle general $X \in L^1(\mathbb{P}; E)$, I use the approximation result in Lemma 5.1.10 to find a sequence $\{X_n : n \geq 1\}$ of simple functions which tend to X in $L^1(\mathbb{P}; E)$. Then, since

$$(X_n)_\Sigma - (X_m)_\Sigma = \big(X_n - X_m\big)_\Sigma \quad \text{(a.s., } \mathbb{P})$$

and therefore, by (5.1.14),

$$\big\|(X_n)_\Sigma - (X_m)_\Sigma\big\|_{L^1(\mathbb{P};E)} \leq \|X_n - X_m\|_{L^1(\mathbb{P};E)},$$

we know that there exists a Σ-measurable $X_\Sigma \in L^1(\mathbb{P}; E)$ to which the sequence $\big\{(X_n)_\Sigma : n \geq 1\big\}$ converges; and clearly X_Σ has the required properties. $\square$

Referring to the setting in the second part of Theorem 5.1.12, I will extend the convention introduced following Theorem 5.1.3 and call the μ-equivalence class of X_Σ's satisfying (5.1.13) the **μ-conditional expectation of X given** Σ, will use $\mathbb{E}^\mu[X|\Sigma]$ to denote this μ-equivalence class, and will, in general, ignore the distinction between the equivalence class and a generic representative of that class. In addition, if $X : \Omega \longrightarrow E$ is locally μ-integrable, then, just as in Theorem 5.1.4, the following are essentially immediate consequences of uniqueness:

$$\mathbb{E}^\mu\big[YX\big|\Sigma\big] = Y\,\mathbb{E}^\mu\big[X\big|\Sigma\big] \quad \text{(a.e., } \mu) \quad \text{for } Y \in L^\infty(\Omega, \Sigma, \mu; \mathbb{R}),$$

and

$$\mathbb{E}^\mu\big[X\big|\mathcal{T}\big] = \mathbb{E}^\mu\Big[\mathbb{E}^\mu\big[X\big|\Sigma\big]\Big|\mathcal{T}\Big] \quad \text{(a.e., } \mu)$$

whenever $\mathcal{T}$ is a sub-σ-algebra of Σ for which $\mu \upharpoonright \mathcal{T}$ is σ-finite.

Exercises for § 5.1

EXERCISE 5.1.15. As the proof of existence in Theorem 5.1.4 makes clear, the operation $X \in L^2(\mathbb{P}; \mathbb{R}) \longmapsto \mathbb{E}^{\mathbb{P}}[X|\Sigma]$ is just the operation of orthogonal projection from $L^2(\mathbb{P}; \mathbb{R})$ onto the space $L^2(\Omega, \Sigma, \mathbb{P}; \mathbb{R})$ of Σ-measurable elements of $L^2(\mathbb{P}; \mathbb{R})$. For this reason, one might be inclined to think that the concept of conditional expectation is basically a Hilbert space notion. However, as this exercise shows, that inclination should be resisted. The point is that, although conditional expectation is definitely an orthogonal projection, not every orthogonal projection is a conditional expectation!

(**i**) Let L be a closed linear subspace of $L^2(\mathbb{P};\mathbb{R})$, and let $\Sigma_L = \sigma\big(\{X : X \in L\}\big)$ be the σ-algebra over Ω generated by $X \in L$. Show that $L = L^2\big(\Omega, \Sigma_L, \mathbb{P}; \mathbb{R}\big)$ if and only if $\mathbf{1} \in L$ and $X^+ \in L$ whenever $X \in L$.

Hint: To prove the "if" assertion, let $X \in L$ be given, and show that

$$X_n \equiv \Big[n\big(X - \alpha\mathbf{1}\big)^+ \wedge \mathbf{1}\Big] \in L \quad \text{for every } \alpha \in \mathbb{R} \text{ and } n \in \mathbb{Z}^+.$$

Conclude that $X_n \nearrow \mathbf{1}_{(\alpha,\infty)} \circ X$ must be an element of L.

(**ii**) Let Π be an orthogonal projection operator on $L^2(\mathbb{P};\mathbb{R})$, set $L = \text{Range}(\Pi)$, and let $\Sigma = \Sigma_L$, where Σ_L is defined as in part (**i**). Show that $\Pi X = \mathbb{E}^{\mathbb{P}}[X|\Sigma]$ (a.s., $\mathbb{P}$) for all $X \in L^2(\mathbb{P};\mathbb{R})$ if and only if $\Pi\mathbf{1} = \mathbf{1}$ and

$$(*) \qquad \Pi\big(X\,\Pi Y\big) = (\Pi X)(\Pi Y) \quad \text{for all} \quad X,\, Y \in L^\infty(\mathbb{P};\mathbb{R}).$$

Hint: Assume that $\Pi\mathbf{1} = \mathbf{1}$ and that (*) holds. Given $X \in L^\infty(\mathbb{P};\mathbb{R})$, use induction to show that

$$\|\Pi X\|^n_{L^{2n}(\mathbb{P})} \le \|X\|^{n-1}_{L^\infty(\mathbb{P})}\|X\|_{L^2(\mathbb{P})} \quad \text{and} \quad \big(\Pi X\big)^n = \Pi\big(X(\Pi X)^{n-1}\big)$$

for all $n \in \mathbb{Z}^+$. Conclude that $\|\Pi X\|_{L^\infty(\mathbb{P})} \le \|X\|_{L^\infty(\mathbb{P})}$ and that $\big(\Pi X\big)^n \in L$, $n \in \mathbb{Z}^+$, for every $X \in L^\infty(\mathbb{P};\mathbb{R})$. Next, using the preceding together with Weierstrass's Approximation Theorem, show that $(\Pi X)^+ \in L$, first for $X \in L^\infty(\mathbb{P};\mathbb{R})$ and then for all $X \in L^2(\mathbb{P};\mathbb{R})$. Finally, apply (**i**) to arrive at $L = L^2\big(\Omega, \Sigma, \mathbb{P}; \mathbb{R}\big)$.

(**iii**) Just in case the situation is not completely clarified by part (**ii**), consider once again a closed linear subspace L of $L^2(\mathbb{P};\mathbb{R})$ and let Π_L be orthogonal projection onto L. Given $X \in L^2(\mathbb{P};\mathbb{R})$, recall that $\Pi_L X$ is characterized as the unique element of L for which $X - \Pi_L X \perp L$, and show that $\mathbb{E}^{\mathbb{P}}[X|\Sigma_L]$ is the unique element of $L^2(\Omega, \Sigma_L, \mathbb{P}; \mathbb{R})$ with the property that

$$X - \mathbb{E}^{\mathbb{P}}\big[X\big|\Sigma_L\big] \perp f\big(Y_1, \dots, Y_n\big)$$

for all $n \in \mathbb{Z}^+$, $f \in C_{\mathrm{b}}\big(\mathbb{R}^n;\mathbb{R}\big)$, and $Y_1, \dots, Y_n \in L$. In particular, $\Pi_L X = \mathbb{E}^{\mathbb{P}}[X|\Sigma_L]$ if and only if $X - \Pi_L X$ is perpendicular not only to all *linear* functions of the Y's in L but even to all *nonlinear* ones.

Exercise 5.1.16. In spite of the preceding, there is a situation in which orthogonal projection coincides with conditioning. Namely, suppose that $\mathfrak{G}$ is a closed Gaussian family in $L^2(\mathbb{P};\mathbb{R})$, and let L be a closed, linear subspace of $\mathfrak{G}$. As an application of Lemma 4.3.1, show that, for any $X \in \mathfrak{G}$, the orthogonal projection $\Pi_L X$ of X onto L is a conditional expectation value of X given the σ-algebra Σ_L generated by the elements of L.

EXERCISE 5.1.17. Because most projections are not conditional expectations, it is an unfortunate fact of life that, for the most part, partial sums of Fourier series cannot be interpreted as conditional expectations. Be that as it may, there are special cases in which such an interpretation is possible. To see this, take $\Omega = [0,1)$, $\mathcal{F} = \mathcal{B}_{[0,1)}$, and $\mathbb{P}$ to be the restriction of Lebesgue measure to $[0,1)$. Next, for $n \in \mathbb{N}$, take $\mathcal{F}_n$ to be the σ-algebra generated by those $f \in C([0,1);\mathbb{C})$ which are periodic with period 2^{-n}. Finally, set $\mathbf{e}_k(x) = \exp\bigl[\sqrt{-1}k2\pi x\bigr]$ for $k \in \mathbb{Z}$, and use elementary Fourier analysis to show that, for each $n \in \mathbb{N}$, $\bigl\{\mathbf{e}_{k2^n} : k \in \mathbb{Z}\bigr\}$ is an orthonormal basis for $L^2(\Omega, \mathcal{F}_n, \mathbb{P}; \mathbb{C})$. In particular, conclude that, for every $f \in L^2(\mathbb{P};\mathbb{C})$:

$$\mathbb{E}^{\mathbb{P}}\bigl[f \,\big|\, \mathcal{F}_n\bigr] = \mathbb{E}^{\mathbb{P}}[f] + \sum_{k\in\mathbb{Z}} \bigl(f, \mathbf{e}_{k2^n}\bigr)_{L^2([0,1);\mathbb{C})} \mathbf{e}_{k2^n},$$

where the convergence is in $L^2([0,1];\mathbb{C})$. (Also see Exercise 5.2.37.)

EXERCISE 5.1.18. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and Σ a sub-σ-algebra of $\mathcal{F}$ with the property that $\mu \restriction \Sigma$ is σ-finite. Next, let E be a separable Hilbert space, $p \in [1,\infty]$, $X \in L^p(\mu;E)$, and Y a Σ-measurable element of $L^{p'}(\mu;E)$ (p' is the Hölder conjugate of p). Show that

$$\mathbb{E}^{\mu}\Bigl[\bigl(Y,X\bigr)_E\Big|\Sigma\Bigr] = \Bigl(Y, \mathbb{E}^{\mu}\bigl[X\big|\Sigma\bigr]\Bigr)_E \quad \mu\text{-almost surely.}$$

Hint: First observe that it suffices to check that

$$\mathbb{E}^{\mu}\Bigl[\bigl(Y,X\bigr)_E\Bigr] = \mathbb{E}^{\mu}\Bigl[\Bigl(Y, \mathbb{E}^{\mu}\bigl[X\big|\Sigma\bigr]\Bigr)_E\Bigr].$$

Next, choose an orthonormal basis $\{\mathbf{e}_n : n \ge 0\}$ for E, and justify the steps in

$$\begin{aligned}\mathbb{E}^{\mu}\bigl[\bigl(Y,X\bigr)_E\bigr] &= \sum_1^{\infty} \mathbb{E}^{\mu}\bigl[\bigl(Y,\mathbf{e}_n\bigr)_E\bigl(\mathbf{e}_n,X\bigr)_E\bigr] \\ &= \sum_1^{\infty} \mathbb{E}^{\mu}\Bigl[\bigl(Y,\mathbf{e}_n\bigr)_E\, \mathbb{E}^{\mu}\bigl[\bigl(\mathbf{e}_n,X\bigr)_E\big|\Sigma\bigr]\Bigr] = \mathbb{E}^{\mu}\bigl[\bigl(Y,\mathbb{E}^{\mu}[X|\Sigma]\bigr)_E\bigr].\end{aligned}$$

EXERCISE 5.1.19. Let E be a separable Banach space, and show that, for each $R > 0$, the closed ball $\overline{B_{E^*}(0,R)}$ with the weak* topology is a compact metric space. Conclude from this that the weak* topology on E^* is second countable and therefore separable.

Hint: Choose a countable, dense subset $\{x_n : n \ge 1\}$ in the unit ball $B_E(0,1)$, and define

$$\rho(x^*, y^*) = \sum_{n=1}^{\infty} 2^{-n}\bigl|\langle x_n, x^* - y^*\rangle\bigr| \quad \text{for } x^*, y^* \in \overline{B_{E^*}(0,R)}.$$

Show that ρ is a metric for the weak* topology on $\overline{B_{E^*}(0,R)}$. Next, choose $\{x_{n_m} : m \geq 1\}$ so that $x_{n_1} = x_1$ and $x_{n_{m+1}} = x_n$ if n is the first $n > n_m$ such that x_n is linearly independent of $\{x_1, \dots, x_{n-1}\}$. Given a sequence $\{x_\ell^* : \ell \geq 1\}$ in $\overline{B_{E^*}(0,R)}$, use a diagnalization argument to find a subsequence $\{x_{\ell_k}^* : k \geq 1\}$ such that $a_m = \lim_{k\to\infty}\langle x_{n_m}, x_{\ell_k}^*\rangle$ exists for each $m \geq 1$. Now define f on the span S of $\{x_{n_m} : m \geq 1\}$ so that $f(x) = \sum_{m=1}^M \alpha_m a_m$ if $x = \sum_{m=1}^M \alpha_m x_{n_m}$, note that $f(x) = \lim_{k\to\infty}\langle x, x_{\ell_k}^*\rangle$ for $x \in S$, and conclude that f is linear on S and satisfies the estimate $|f(x)| \leq R\|x\|_E$ there. Since S is dense in E, there is a unique extension of f as a bounded linear functional on E satisfying the same estimate, and so there exists an $x^* \in \overline{B_{E^*}(0,R)}$ such that $\langle x, x^*\rangle = \lim_{k\to\infty}\langle x, x_{\ell_k}^*\rangle$ for all $x \in S$. Finally, check that this convergence continues to hold for all $x \in E$, and conclude that $x_{\ell_k}^* \longrightarrow x^*$ in the weak* topology.

§ 5.2 Discrete Parameter Martingales

In this section I will introduce an interesting and useful class of stochastic processes which unifies and simplifies several branches of probability theory as well as other branches of analysis. From the analytic point of view, what I will be doing is developing an abstract version of differentiation theory (cf. Theorem 6.1.8).

Although I will want to make some extensions in § 5.3, I start in the following setting. $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and $\{\mathcal{F}_n : n \in \mathbb{N}\}$ is a non-decreasing sequence of sub-σ-algebra's of $\mathcal{F}$. Given a measurable space $(E, \mathcal{B})$, say that the family $\{X_n : n \in \mathbb{N}\}$ of E-valued random variables is $\{\mathcal{F}_n : n \in \mathbb{N}\}$**-progressively measurable** if X_n is $\mathcal{F}_n$-measurable for each $n \in \mathbb{N}$. Next, a family $\{X_n : n \in \mathbb{N}\}$ of $(-\infty, \infty]$-valued random variables is said to be a $\mathbb{P}$**-submartingale with respect to** $\{\mathcal{F}_n : n \in \mathbb{N}\}$ if it is $\{\mathcal{F}_n : n \in \mathbb{N}\}$-progressively measurable, $\mathbb{E}^{\mathbb{P}}[X_n^-] < \infty$, and, for each $n \in \mathbb{N}$, $X_n \leq \mathbb{E}^{\mathbb{P}}[X_{n+1}|\mathcal{F}_n]$ (a.s., $\mathbb{P}$). It is said to be a $\mathbb{P}$**-martingale with respect to** $\{\mathcal{F}_n : n \in \mathbb{N}\}$ if $\{X_n : n \in \mathbb{N}\}$ is an $\{\mathcal{F}_n : n \in \mathbb{N}\}$-progressively measurable family of $\mathbb{R}$-valued, $\mathbb{P}$-integrable random variables satisfying $X_n = \mathbb{E}^{\mathbb{P}}[X_{n+1}|\mathcal{F}_n]$ (a.s., P) for each $n \in \mathbb{N}$. In the future, I will abbreviate these statements by saying that the triple $(X_n, \mathcal{F}_n, \mathbb{P})$ is a submartingale or martingale.

Examples. The most trivial example of a submartingale is provided by a non-decreasing sequence $\{a_n : n \geq 0\}$. That is, if $X_n \equiv a_n$, $n \in \mathbb{N}$, then $(X_n, \mathcal{F}_n, \mathbb{P})$ is a submartingale on any probability space $(\Omega, \mathcal{F}, \mathbb{P})$ relative to any non-decreasing $\{\mathcal{F}_n : n \in \mathbb{N}\}$. More interesting examples are those given below.*

(**i**) Let $\{Y_n : n \geq 1\}$ be a sequence of mutually independent $(-\infty, \infty]$-valued random variables with $\mathbb{E}^{\mathbb{P}}[Y_n^-] < \infty$, $n \in \mathbb{N}$, set $\mathcal{F}_0 = \{\emptyset, \Omega\}$, $\mathcal{F}_n = \sigma(\{Y_1, \dots, Y_n\})$

* For a much more interesting and complete list of examples, the reader might want to consult J. Neveu's *Discrete-parameter Martingales*, publ. in 1975 by North–Holland.

for $n \in \mathbb{Z}^+$, and define $X_n = \sum_{1 \le m \le n} Y_m$, where summation over the empty set is taken to be 0. Then, because $\mathbb{E}^{\mathbb{P}}[Y_{n+1}|\mathcal{F}_n] = \mathbb{E}^{\mathbb{P}}[Y_{n+1}]$ (a.s., $\mathbb{P}$) and therefore

$$\mathbb{E}^{\mathbb{P}}\big[X_{n+1}\big|\mathcal{F}_n\big] = X_n + \mathbb{E}^{\mathbb{P}}\big[Y_{n+1}\big] \quad \text{(a.s.,} \mathbb{P}\text{)}$$

for every $n \in \mathbb{N}$, we see that $\big(X_n, \mathcal{F}_n, \mathbb{P}\big)$ is a submartingale if and only if $\mathbb{E}^{\mathbb{P}}[Y_n] \ge 0$ for all $n \in \mathbb{Z}^+$. In fact, if the Y_n's are $\mathbb{R}$-valued and $\mathbb{P}$-integrable, then the same line of reasoning shows that $\big(X_n, \mathcal{F}_n, \mathbb{P}\big)$ is a martingale if and only if $\mathbb{E}^{\mathbb{P}}[Y_n] = 0$ for all $n \in \mathbb{Z}^+$. Finally, if $\{Y_n : n \ge 0\} \subseteq L^2(\mathbb{P};\mathbb{R})$ and $\mathbb{E}^{\mathbb{P}}[Y_n] = 0$ for each $n \in \mathbb{Z}^+$, then

$$\mathbb{E}^{\mathbb{P}}\big[X_{n+1}^2 \,\big|\, \mathcal{F}_n\big] = X_n^2 + \mathbb{E}^{\mathbb{P}}\big[Y_{n+1}^2 \,\big|\, \mathcal{F}_n\big] \ge X_n^2 \quad \text{(a.s.,} \mathbb{P}\text{)},$$

and so $\big(X_n^2, \mathcal{F}_n, \mathbb{P}\big)$ is a submartingale.

(**ii**) If X is an $\mathbb{R}$-valued, $\mathbb{P}$-integrable random variable and $\big\{\mathcal{F}_n : n \in \mathbb{N}\big\}$ is a non-decreasing sequence of sub-σ-algebras of $\mathcal{F}$, then, by (5.1.6), $\big(\mathbb{E}^{\mathbb{P}}[X|\mathcal{F}_n], \mathcal{F}_n, \mathbb{P}\big)$ is a martingale.

(**iii**) If $\big(X_n, \mathcal{F}_n, \mathbb{P}\big)$ is a martingale, then, by (5.1.5), $\big(|X_n|, \mathcal{F}_n, \mathbb{P}\big)$ is a submartingale.

§ **5.2.1. Doob's Inequality and Marcinkewitz's Theorem.** In view of Example (**i**) above, we see that partial sums of independent random variables with mean-value 0 are a source of martingales and that their squares are a source of submartingales. Hence, it is reasonable to ask whether some of the important facts about such partial sums will continue to be true for all martingales; and perhaps the single most important indication that the answer may be "yes" is contained in the following generalization of Kolmogorov's Inequality (cf. Theorem 1.4.5). Like most of the foundational results in martingale theory, this one is due to J.L. Doob.

THEOREM 5.2.1 (**Doob's Inequality**). *Assume that $\big(X_n, \mathcal{F}_n, \mathbb{P}\big)$ is a submartingale. Then, for every $N \in \mathbb{Z}^+$ and $\alpha \in (0,\infty)$:*

$$\mathbb{P}\left(\max_{0 \le n \le N} X_n \ge \alpha\right) \le \frac{1}{\alpha}\mathbb{E}^{\mathbb{P}}\left[X_N,\ \max_{0 \le n \le N} X_n \ge \alpha\right]. \tag{5.2.2}$$

In particular, if the X_n's are non-negative, then, for each $p \in (1,\infty)$,

$$\mathbb{E}^{\mathbb{P}}\left[\sup_{n \in \mathbb{N}} X_n^p\right]^{\frac{1}{p}} \le \frac{p}{p-1} \sup_{n \in \mathbb{N}} \mathbb{E}^{\mathbb{P}}\big[X_n^p\big]^{\frac{1}{p}}. \tag{5.2.3}$$

PROOF: To prove (5.2.2), set $A_0 = \big\{X_0 \ge \alpha\big\}$ and

$$A_n = \left\{X_n \ge \alpha \text{ but } \max_{0 \le m < n} X_m < \alpha\right\} \quad \text{for} \quad n \in \mathbb{Z}^+.$$

Then the A_n's are mutually disjoint and $A_n \in \mathcal{F}_n$ for each $n \in \mathbb{N}$. Thus,

$$P\left(\max_{0\le n\le N} X_n \ge \alpha\right) = \sum_{n=0}^N P\big(A_n\big) \le \sum_{n=0}^N \frac{\mathbb{E}^{\mathbb{P}}\big[X_n,\, A_n\big]}{\alpha}$$
$$\le \sum_{n=0}^N \frac{\mathbb{E}^{\mathbb{P}}\big[X_N,\, A_n\big]}{\alpha} = \frac{1}{\alpha}\mathbb{E}^{\mathbb{P}}\left[X_N,\, \max_{0\le n\le N} X_n \ge \alpha\right].$$

Now assume that the X_n's are non-negative. Given (5.2.2), (5.2.3) becomes an easy application of Exercise 1.4.18. □

Doob's inequality is an example of what analysts call a **weak-type inequality**. To be more precise, it is a *weak-type* 1–1 inequality. The terminology derives from the fact that such an inequality follows immediately from an L^1-norm, or *strong-type* 1–1, inequality between the objects under consideration; but, in general, it is strictly weaker. In order to demonstrate how powerful such a result can be, I will now apply Doob's Inequality to prove a theorem of Marcinkewitz. Because it is an argument to which we will return again, the reader would do well to become comfortable with the line of reasoning which allows one to pass from a *weak-type inequality*, like Doob's, to almost sure convergence results.

Corollary 5.2.4. *Let X be an $\mathbb{R}$-valued random variable and $p \in [1,\infty)$. If $X \in L^p(\mathbb{P};\mathbb{R})$, then for any non-decreasing sequence $\{\mathcal{F}_n : n \in \mathbb{N}\}$ of sub-σ-algebras of $\mathcal{F}$:*

$$\mathbb{E}^{\mathbb{P}}\big[X\big|\mathcal{F}_n\big] \longrightarrow \mathbb{E}^{\mathbb{P}}\left[X\middle|\bigvee_0^\infty \mathcal{F}_n\right] \quad (\text{a.s.}, \mathbb{P}) \text{ and in } L^p(\mathbb{P};\mathbb{R})$$

as $n \to \infty$. In particular, if X is $\bigvee_0^\infty \mathcal{F}_n$-measurable, then $\mathbb{E}^{\mathbb{P}}[X|\mathcal{F}_n] \longrightarrow X$ (a.s., $\mathbb{P}$) and in $L^p(\mathbb{P};\mathbb{R})$.

Proof: Without loss in generality, assume that $\mathcal{F} = \bigvee_0^\infty \mathcal{F}_n$.

Given $X \in L^1(\mathbb{P};\mathbb{R})$, set $X_n = \mathbb{E}^{\mathbb{P}}[X|\mathcal{F}_n]$ for $n \in \mathbb{N}$. The key to my proof will be the inequality

$$\text{(5.2.5)} \qquad \mathbb{P}\left(\sup_{n\in\mathbb{N}} |X_n| \ge \alpha\right) \le \frac{1}{\alpha}\mathbb{E}^{\mathbb{P}}\left[|X|, \sup_{n\in\mathbb{N}} |X_n| \ge \alpha\right], \quad \alpha \in (0,\infty);$$

and, since, by (5.1.5), $|X_n| \le \mathbb{E}^{\mathbb{P}}[|X|\,|\mathcal{F}_n]$ (a.s., $\mathbb{P}$), while proving (5.2.5) I may and will assume that X and all the X_n's are non-negative. But then, by (5.2.2),

$$P\left(\sup_{0\le n\le N} X_n > \alpha\right) \le \frac{1}{\alpha}\mathbb{E}^{\mathbb{P}}\left[X_N,\, \sup_{0\le n\le N} X_n > \alpha\right]$$
$$= \frac{1}{\alpha}\mathbb{E}^{\mathbb{P}}\left[X,\, \sup_{0\le n\le N} X_n > \alpha\right]$$

for all $N \in \mathbb{Z}^+$; and therefore (5.2.5) follows when $N \to \infty$ and one takes right limits in α.

As my first application of (5.2.5), note that $\{X_n : n \geq 0\}$ is uniformly $\mathbb{P}$-integrable. Indeed, because $|X_n| \leq \mathbb{E}^{\mathbb{P}}[|X|\,|\mathcal{F}_n]$, we have from (5.2.5) that

$$\begin{aligned}\sup_{n\in\mathbb{N}} \mathbb{E}^{\mathbb{P}}\Big[|X_n|,\,|X_n| \geq \alpha\Big] &\leq \sup_{n\in\mathbb{N}} \mathbb{E}^{\mathbb{P}}\Big[|X|,\,|X_n| \geq \alpha\Big] \\ &\leq \mathbb{E}^{\mathbb{P}}\left[|X|,\,\sup_{n\in\mathbb{N}}|X_n| \geq \alpha\right] \longrightarrow 0\end{aligned}$$

as $\alpha \to \infty$. Thus, we will know that the asserted convergence takes place in $L^1(\mathbb{P};\mathbb{R})$ as soon as we show that it happens $\mathbb{P}$-almost surely. In addition, if $X \in L^p(\mathbb{P};\mathbb{R})$ for some $p \in (1,\infty)$, then, by (5.2.5) and Exercise 1.4.18, we see that $\big\{|X_n|^p : n \in \mathbb{N}\big\}$ is uniformly $\mathbb{P}$-integrable and, therefore, that $X_n \longrightarrow X$ in $L^p(\mu;\mathbb{R})$ as soon as it does (a.s., $\mathbb{P}$). In other words, everything comes down to checking the $\mathbb{P}$-almost sure convergence.

To prove the $\mathbb{P}$-almost sure convergence, let $\mathcal{G}$ be the set of $X \in L^1(\mathbb{P};\mathbb{R})$ for which $X_n \longrightarrow X$ (a.s., $\mathbb{P}$). Clearly, $X \in \mathcal{G}$ if $X \in L^1(\mathbb{P};\mathbb{R})$ is $\mathcal{F}_n$-measurable for some $n \in \mathbb{N}$; and, therefore, $\mathcal{G}$ is dense in $L^1(\mathbb{P};\mathbb{R})$. Thus, all that remains is to prove that $\mathcal{G}$ is closed in $L^1(\mathbb{P};\mathbb{R})$. But if $\{X^{(k)} : k \geq 1\} \subseteq \mathcal{G}$ and $X^{(k)} \longrightarrow X$ in $L^1(\mathbb{P};\mathbb{R})$, then, by (5.2.5),

$$\begin{aligned}&\mathbb{P}\left(\sup_{n\geq N}\big|X_n - X\big| \geq 3\alpha\right) \\ &\leq \mathbb{P}\left(\sup_{n\geq N}\big|X_n - X_n^{(k)}\big| \geq \alpha\right) + \mathbb{P}\left(\sup_{n\geq N}\big|X_n^{(k)} - X^{(k)}\big| \geq \alpha\right) \\ &\qquad + \mathbb{P}\Big(\big|X^{(k)} - X\big| \geq \alpha\Big) \\ &\leq \frac{2}{\alpha}\big\|X - X^{(k)}\big\|_{L^1(\mathbb{P})} + \mathbb{P}\left(\sup_{n\geq N}\big|X_n^{(k)} - X^{(k)}\big| \geq \alpha\right)\end{aligned}$$

for every $N \in \mathbb{Z}^+$, $\alpha \in (0,\infty)$, and $k \in \mathbb{Z}^+$. Hence, by first letting $N \to \infty$ and then $k \to \infty$, we see that

$$\lim_{N\to\infty} \mathbb{P}\left(\sup_{n\geq N}\big|X_n - X\big| \geq 3\alpha\right) = 0 \quad \text{for every } \alpha \in (0,\infty);$$

and this proves that $X \in \mathcal{G}$. □

Before moving on to more sophisticated convergence results, I will spend a little time showing that Corollary 5.2.4 is already interesting. In order to introduce my main application, recall my preliminary discussion of conditioning when I was attempting to explain Kolmogorov's idea at the beginning of this chapter. As I said there, the most easily understood situation occurs when one conditions with respect to a sub-σ-algebra Σ which is generated by a countable partition $\mathcal{P}$. Indeed, in that case one can easily verify that

$$\mathbb{E}^{\mathbb{P}}\big[X\big|\Sigma\big] = \sum_{A\in\mathcal{P}} \frac{\mathbb{E}^{\mathbb{P}}\big[X,\,A\big]}{\mathbb{P}(A)}\,\mathbf{1}_A, \tag{5.2.6}$$

where it is understood that

$$\frac{\mathbb{E}^{\mathbb{P}}\big[X,\,A\big]}{\mathbb{P}(A)} \equiv 0 \quad \text{when} \quad \mathbb{P}(A) = 0.$$

Unfortunately, even when $\mathcal{F}$ is countably generated, Σ need not be (cf. Exercise 1.1.18). Furthermore, just because Σ is countably generated, it will be seldom true that its generators can be chosen to form a countable partition. (For example, as soon as Σ contains an uncountable number of atoms, such a partition cannot exist.) Nonetheless, if Σ is any countably generated σ-algebra, then we can find a sequence $\{\mathcal{P}_n : n \geq 0\}$ of finite partitions with the properties that

$$\Sigma = \sigma\left(\bigcup_0^\infty \mathcal{P}_n\right) \quad \text{and} \quad \sigma\big(\mathcal{P}_{n-1}\big) \subseteq \sigma\big(\mathcal{P}_n\big), \quad n \in \mathbb{Z}^+.$$

In fact, simply choose a countable generating sequence $\{A_n : n \geq 0\}$ for Σ and take $\mathcal{P}_n$ to be the collection of distinct sets of the form $B_0 \cap \cdots \cap B_n$, where $B_m \in \big\{A_m, A_m\complement\big\}$ for each $0 \leq m \leq n$.

Theorem 5.2.7. *Let Σ be a countably generated sub-σ-algebra of $\mathcal{F}$, and choose $\{\mathcal{P}_n : n \geq 0\}$ to be a sequence of finite partitions as above. Next, given $p \in [1,\infty)$ and a random variable $X \in L^p(\mathbb{P};\mathbb{R})$, define X_n for $n \in \mathbb{N}$ by the right-hand side of (5.2.6) with $\mathcal{P} = \mathcal{P}_n$. Then $X_n \longrightarrow \mathbb{E}^{\mathbb{P}}[X|\Sigma]$ both $\mathbb{P}$-almost surely and in $L^p(\mathbb{P};\mathbb{R})$. Moreover, even if Σ is not countably generated, for each separable, closed subspace L of $L^p(\mathbb{P};\mathbb{R})$ there exists a sequence of finite partitions $\mathcal{P}_n$, $n \in \mathbb{N}$, such that*

$$\sum_{A\in\mathcal{P}_n} \frac{\mathbb{E}^{\mathbb{P}}\big[X,\,A\big]}{\mathbb{P}(A)}\,\mathbf{1}_A \longrightarrow \mathbb{E}^{\mathbb{P}}\big[X\big|\Sigma\big] \quad (\text{a.s.},\mathbb{P}) \text{ and in } L^p(\mathbb{P};\mathbb{R})$$

for every $X \in L$.

Proof: To prove the first part, simply set $\mathcal{F}_n = \sigma\big(\mathcal{P}_n\big)$, then identify X_n as $\mathbb{E}^{\mathbb{P}}[X|\mathcal{F}_n]$, and finally apply Corollary 5.2.4. As for the second part, let $\Sigma(L)$ be the σ-algebra generated by $\big\{\mathbb{E}^{\mathbb{P}}[X|\Sigma] : X \in L\big\}$, note that $\Sigma(L)$ is countably generated and that

$$\mathbb{E}^{\mathbb{P}}\big[X\big|\Sigma\big] = \mathbb{E}^{\mathbb{P}}\big[X\big|\Sigma(L)\big] \quad (\text{a.s.},\mathbb{P}) \quad \text{for each } X \in L,$$

and apply the first part with Σ replaced by $\Sigma(L)$. □

Theorem 5.2.7 makes it easy to transfer the usual Jensen's Inequality to conditional expectations.

Corollary 5.2.8 (**Jensen's Inequality**). *Let C be a closed, convex subset of $\mathbb{R}^N$, $\mathbf{X}$ a C-valued, $\mathbb{P}$-integrable random variable, and Σ a sub-σ-algebra of $\mathcal{F}$. Then there is a C-valued representative $\mathbf{X}_\Sigma$ of*

$$\mathbb{E}^{\mathbb{P}}\big[\mathbf{X}\big|\Sigma\big] \equiv \begin{bmatrix} \mathbb{E}^{\mathbb{P}}\big[X_1\big|\Sigma\big] \\ \vdots \\ \mathbb{E}^{\mathbb{P}}\big[X_N\big|\Sigma\big] \end{bmatrix}.$$

In addition, if $g : C \longrightarrow [0,\infty)$ is continuous and concave, then

$$\mathbb{E}^{\mathbb{P}}\big[g(\mathbf{X})\big|\Sigma\big] \le g\big(\mathbf{X}_\Sigma\big) \quad (\text{a.s.}, \mathbb{P}).$$

Proof: By the classical Jensen's Inequality, $Y \equiv g(\mathbf{X})$ is $\mathbb{P}$-integrable. Hence, by the second part of Theorem 5.2.7, we can find finite partitions $\mathcal{P}_n$, $n \in \mathbb{N}$, so that

$$\mathbf{X}_n \equiv \sum_{A\in\mathcal{P}_n} \frac{\mathbb{E}^{\mathbb{P}}[\mathbf{X},\, A]}{\mathbb{P}(A)}\mathbf{1}_A \longrightarrow \mathbb{E}^{\mathbb{P}}[\mathbf{X}|\Sigma]$$

and

$$Y_n \equiv \sum_{A\in\mathcal{P}_n} \frac{\mathbb{E}^{\mathbb{P}}\big[g(\mathbf{X}),\, A\big]}{\mathbb{P}(A)}\mathbf{1}_A \longrightarrow \mathbb{E}^{\mathbb{P}}\big[g(\mathbf{X})\big|\Sigma\big]$$

$\mathbb{P}$-almost surely. Furthermore, again by the classical Jensen's Inequality,

$$\frac{\mathbb{E}^{\mathbb{P}}[\mathbf{X},\, A]}{\mathbb{P}(A)} \in C \quad \text{and} \quad \frac{\mathbb{E}^{\mathbb{P}}\big[g(\mathbf{X}),\, A\big]}{\mathbb{P}(A)} \le g\left(\frac{\mathbb{E}^{\mathbb{P}}[\mathbf{X},\, A]}{\mathbb{P}(A)}\right)$$

for all $A \in \mathcal{F}$ with $\mathbb{P}(A) > 0$. Hence, if $\Lambda \in \Sigma$ denotes the set of ω for which

$$\lim_{n\to\infty} \begin{bmatrix} \mathbf{X}_n(\omega) \\ Y_n(\omega) \end{bmatrix} \in \mathbb{R}^{N+1}$$

exists, $\mathbf{v}$ is a fixed element of C,

$$\mathbf{X}_\Sigma(\omega) \equiv \begin{cases} \lim_{n\to\infty} \mathbf{X}_n(\omega) & \text{if} \quad \omega \in \Lambda \\ \mathbf{v} & \text{if} \quad \omega \notin \Lambda, \end{cases}$$

and

$$Y(\omega) \equiv \begin{cases} \lim_{n\to\infty} Y_n(\omega) & \text{if} \quad \omega \in \Lambda \\ \mathbf{v} & \text{if} \quad \omega \notin \Lambda, \end{cases}$$

then $\mathbf{X}_\Sigma$ is a C-valued representative of $\mathbb{E}^{\mathbb{P}}[\mathbf{X}|\Sigma]$, Y is a representative of $\mathbb{E}^{\mathbb{P}}[g(\mathbf{X})|\Sigma]$, and $Y(\omega) \le g\big(\mathbf{X}_\Sigma(\omega)\big)$ for every $\omega \in \Omega$. $\square$

Corollary 5.2.9. *Let I be a non-trivial, closed interval in $\mathbb{R} \cup \{+\infty\}$ (i.e., either $I \subset \mathbb{R}$ is bounded on the right or $I \cap \mathbb{R}$ is unbounded on the right and I includes the point $+\infty$). Then every I-valued random variable X with $\mathbb{P}$-integrable negative part admits an I-valued representative of $\mathbb{E}^{\mathbb{P}}[X|\Sigma]$. Furthermore, given a continuous, convex function $f : I \longrightarrow \mathbb{R} \cup \{+\infty\}$,*

$$f\Big(\mathbb{E}^{\mathbb{P}}\big[X\big|\Sigma\big]\Big) \le \mathbb{E}^{\mathbb{P}}\big[f(X)\big|\Sigma\big] \quad (\text{a.s.}, \mathbb{P}) \tag{5.2.10}$$

if either f is bounded above and X is $\mathbb{P}$-integrable or f is bounded below and to the left (i.e., f is bounded on each interval of the form $I \cap (-\infty, a]$ with $a \in I \cap \mathbb{R}$). In particular, for each $p \in [1,\infty)$,

$$\Big\|\mathbb{E}^{\mathbb{P}}\big[X \,\big|\, \Sigma\big]\Big\|_{L^p(\mathbb{P};\mathbb{R})} \le \|X\|_{L^p(\mathbb{P};\mathbb{R})}.$$

Finally, if either $\big(X_n, \mathcal{F}_n, \mathbb{P}\big)$ is an I-valued martingale and f is as above or if $\big(X_n, \mathcal{F}_n, P\big)$ is an I-valued submartingale and f is bounded below and non-decreasing (as well as continuous and convex), then $\big(f(X_n), \mathcal{F}_n, \mathbb{P}\big)$ is a submartingale.

Proof: In view of Corollary 5.2.8, we know that an I-valued representative of $\mathbb{E}^{\mathbb{P}}[X|\Sigma]$ exists when X is $\mathbb{P}$-integrable, and the general case follows after a trivial truncation procedure. In order to prove (5.2.10), first assume that f is bounded above by some $M < \infty$ and that $X \in L^1(\mathbb{P};\mathbb{R})$. Then (5.2.10) is an immediate consequence of the last part of Corollary 5.2.8 with $g = M - f$. To handle the case when f is bounded below and to the left, first observe that either f is non-increasing everywhere, or there is an $a \in I \cap \mathbb{R}$ with the property that f is non-increasing to the left of a and non-decreasing to the right of a. Next, let an I-valued X with $X^- \in L^1(\mathbb{P})$ be given, and set $X_n = X \wedge n$. Then there exists an $m \in \mathbb{Z}^+$ such that X_n is I-valued for all $n \ge m$; and clearly, by the preceding, we know that

$$f\Big(\mathbb{E}^{\mathbb{P}}\big[X_n\big|\Sigma\big]\Big) \le \mathbb{E}^{\mathbb{P}}\big[f(X_n)\big|\Sigma\big] \quad (\text{a.s.}, \mathbb{P}) \quad \text{for all } n \ge m. \tag{*}$$

Moreover, in the case when f is non-increasing, $\big\{f(X_n) : n \ge m\big\}$ is bounded below and non-increasing; and, in the other case, $\big\{f(X_n) : n \ge m \vee a\big\}$ is bounded below and non-decreasing. Hence, in both cases, (5.2.10) follows (*) after an application of the version of the Monotone Convergence Theorem in (5.1.8).

To complete the proof, simply note that in either of the two cases given, the results just proved justify:

$$\mathbb{E}^{\mathbb{P}}\big[f(X_n)\big|\mathcal{F}_{n-1}\big] \ge f\Big(\mathbb{E}^{\mathbb{P}}\big[X_n\big|\mathcal{F}_{n-1}\big]\Big) \ge f\big(X_{n-1}\big) \quad \mathbb{P}\text{-almost surely.} \quad \square$$

§ **5.2.2. Doob's Stopping Time Theorem.** Perhaps the most far-reaching contribution that Doob made to martingale theory is his observation that one can "stop" a martingale without destroying the martingale property. Later, L. Snell showed that the analogous result is true for submartingales.

In order to state their results here, I need to introduce the notion of a stopping time in this setting. Namely, I will say that the function $\zeta : \Omega \longrightarrow \mathbb{N} \cup \{\infty\}$ is an **stopping time** relative to $\{\mathcal{F}_n : n \geq 0\}$ if $\big\{\omega : \zeta(\omega) = n\big\} \in \mathcal{F}_n$ for each $n \in \mathbb{N}$. In addition, given a stopping time ζ, I use $\mathcal{F}_\zeta$ to denote the σ-algebra of $A \in \mathcal{F}$ such that $A \cap \{\zeta = n\} \in \mathcal{F}_n$, $n \in \mathbb{Z}^+$. Notice that $\mathcal{F}_{\zeta_1} \subseteq \mathcal{F}_{\zeta_2}$ if $\zeta_1 \leq \zeta_2$. In addition, if $\big\{X_n : n \in \mathbb{N}\big\}$ is $\big\{\mathcal{F}_n : n \in \mathbb{N}\big\}$-progressively measurable, check that the random variable X_ζ given by $X_\zeta(\omega) = X_{\zeta(\omega)}(\omega)$ is $\mathcal{F}_\zeta$-measurable on $\{\zeta < \infty\}$.

Doob used stopping times to give a mathematically rigorous formulation of the W.C. Field's assertion that "you can't cheat an honest man." That is, consider a gambler who is trying to *beat the system*. Assuming that he is playing a fair game, it is reasonable to say his gain X_n after n plays will evolve as a martingale. More precisely, if $\mathcal{F}_n$ contains the history of the game up to and including the nth play, then $\big(X_n, \mathcal{F}_n, \mathbb{P}\big)$ will be a martingale. In the context of this model, a stopping time can be thought of as a feasible (i.e., one which does not require the gift of prophesy) strategy that the gambler can use to determine when he should stop playing in order to maximize his gains. When couched in these terms, the next result predicts that *there is no strategy with which the gambler can alter his expected gain.*

Theorem 5.2.11 (**Doob's Stopping Time Theorem**). *For any submartingale (martingale) $\big(X_n, \mathcal{F}_n, \mathbb{P}\big)$ which is $\mathbb{P}$-integrable and any stopping time ζ, $\big(X_{n\wedge\zeta}, \mathcal{F}_n, P\big)$ is again a $\mathbb{P}$-integrable submartingale (martingale).*

Proof: Let $A \in \mathcal{F}_{n-1}$. Then, since $A \cap \{\zeta > n-1\} \in \mathcal{F}_{n-1}$,

$$\begin{aligned}\mathbb{E}^{\mathbb{P}}\big[X_{n\wedge\zeta},\, A\big] &= \mathbb{E}^{\mathbb{P}}\big[X_\zeta,\, A \cap \{\zeta \leq n-1\}\big] + \mathbb{E}^{\mathbb{P}}\big[X_n,\, A \cap \{\zeta > n-1\}\big] \\ &\geq \mathbb{E}^{\mathbb{P}}\big[X_\zeta,\, A \cap \{\zeta \leq n-1\}\big] + \mathbb{E}^{\mathbb{P}}\big[X_{n-1},\, A \cap \{\zeta > n-1\}\big] = \mathbb{E}^{\mathbb{P}}\big[X_{(n-1)\wedge\zeta},\, A\big];\end{aligned}$$

and, in the case of martingales, the inequality in the preceding can be replaced by an equality. □

Closely related to Doob's Stopping Time Theorem is an important variant due to G. Hunt. In order to facilitate the proof of Hunt's result, I begin with an easy but seminal observation of Doob's.

Lemma 5.2.12 (**Doob's Decomposition**). *For each $n \in \mathbb{N}$ let X_n be an $\mathcal{F}_n$-measurable, $\mathbb{P}$-integrable random variable. Then, up to a $\mathbb{P}$-null set, there is at most one sequence $\{A_n : n \geq 0\} \subseteq L^1(\mathbb{P};\mathbb{R})$ such that $A_0 = 0$, A_n is $\mathcal{F}_{n-1}$-measurable for each $n \in \mathbb{Z}^+$, and $\big(X_n - A_n, \mathcal{F}_n, \mathbb{P}\big)$ is a martingale. Moreover, if $(X_n, \mathcal{F}_n, \mathbb{P})$ is an integrable submartingale, then such a sequence $\{A_n : n \geq 0\}$ exists, and $A_{n-1} \leq A_n$ $\mathbb{P}$-almost surely for all $n \in \mathbb{Z}^+$.*

PROOF: To prove the uniqueness assertion, suppose that $\{A_n : n \geq 0\}$ and $\{B_n : n \geq 0\}$ are two such sequences, and set $\Delta_n = B_n - A_n$. Then $\Delta_0 = 0$, Δ_n is $\mathcal{F}_{n-1}$-measurable for each $n \in \mathbb{Z}^+$, and $(\Delta_n, \mathcal{F}_n, \mathbb{P})$ is a martingale. But this means that $\Delta_n = \mathbb{E}^{\mathbb{P}}[\Delta_n \,|\, \mathcal{F}_{n-1}] = \Delta_{n-1}$ for all $n \in \mathbb{Z}^+$, and so $\Delta_n = 0$ for all $n \in \mathbb{N}$.

Now suppose that $(X_n, \mathcal{F}_n, \mathbb{P})$ is an integrable submartingale. To prove the asserted existence result, set $A_0 \equiv 0$ and

$$A_n = A_{n-1} + \mathbb{E}^{\mathbb{P}}\big[X_n - X_{n-1}\big|\mathcal{F}_{n-1}\big] \vee 0 \quad \text{for} \quad n \in \mathbb{Z}^+. \quad \square$$

THEOREM 5.2.13 (**Hunt**). *Let $\big(X_n, \mathcal{F}_n, \mathbb{P}\big)$ be a $\mathbb{P}$-integrable submartingale. Given bounded stopping times ζ and ζ' satisfying $\zeta \leq \zeta'$,*

$$X_\zeta \leq \mathbb{E}^{\mathbb{P}}\big[X_{\zeta'}\big|\mathcal{F}_\zeta\big] \quad (\text{a.s.}, \mathbb{P}), \tag{5.2.14}$$

and the inequality can be replaced by equality when $\big(X_n, \mathcal{F}_n, \mathbb{P}\big)$ is a martingale. (Cf. Exercise 5.2.31 for unbounded stopping times.)

PROOF: Choose $\big\{A_n : n \in \mathbb{N}\big\}$ as in Lemma 5.2.12, and set $Y_n = X_n - A_n$, $n \in \mathbb{N}$. Then, because $A_\zeta \leq A_{\zeta'}$ and A_ζ is $\mathcal{F}_\zeta$-measurable,

$$\mathbb{E}^{\mathbb{P}}\big[X_{\zeta'}\big|\mathcal{F}_\zeta\big] \geq \mathbb{E}^{\mathbb{P}}\big[Y_{\zeta'} + A_\zeta\big|\mathcal{F}_\zeta\big] = \mathbb{E}^{\mathbb{P}}\big[Y_{\zeta'}\big|\mathcal{F}_\zeta\big] + A_\zeta,$$

it suffices to prove that equality holds in (5.2.14) when $\big(X_n, \mathcal{F}_n, \mathbb{P}\big)$ is a martingale. To this end, choose $N \in \mathbb{Z}^+$ to be an upper bound for ζ', let $\Gamma \in \mathcal{F}_\zeta$ be given, and note that

$$\begin{aligned} \mathbb{E}^{\mathbb{P}}\big[X_N,\, \Gamma\big] &= \sum_{n=0}^{N} \mathbb{E}^{\mathbb{P}}\big[X_N,\, \Gamma \cap \{\zeta = n\}\big] \\ &= \sum_{n=0}^{N} \mathbb{E}^{\mathbb{P}}\big[X_n,\, \Gamma \cap \{\zeta = n\}\big] = \mathbb{E}^{\mathbb{P}}\big[X_\zeta,\, \Gamma\big]; \end{aligned}$$

and similarly (since $\Gamma \in \mathcal{F}_\zeta \subseteq \mathcal{F}_{\zeta'}$), $\mathbb{E}^{\mathbb{P}}[X_N,\, \Gamma] = \mathbb{E}^{\mathbb{P}}[X_{\zeta'},\, \Gamma]$. $\square$

§ **5.2.3. Martingale Convergence Theorem.** My next goal is to show that, even when they are not given in the form covered by Corollary 5.2.4, *martingales want to converge*. If for no other reason, such a result has got to be more difficult because one does not know ahead of time what, if it exists, the limit ought to be. Thus, the reasoning will have to be more subtle than that used in the proof of Corollary 5.2.4. I will follow Doob and base my argument on the idea that, in some sense, a martingale has got to be *nearly constant* and that a submartingale is the sum of a martingale and a non-decreasing process. In order to make mathematics out of this idea, I need to introduce a somewhat novel criterion for convergence of real numbers. Namely, given a sequence $\{x_n : n \geq 0\} \subseteq \mathbb{R}$ and $-\infty < a < b < \infty$, say that $\{x_n : n \geq 0\}$ **upcrosses the interval** $[a,b]$ **at least** N **times** if there exist integers $0 \leq m_1 < n_1 < \cdots < m_N < n_N$ such that $x_{m_i} \leq a$ and $x_{n_i} \geq b$ for each $1 \leq i \leq N$ and that it **upcrosses** $[a,b]$ **precisely** N **times** if it upcrosses $[a,b]$ at least N but does not upcross $[a,b]$ at least $N+1$ times. Notice that $\underline{\lim}_{n\to\infty} x_n < \overline{\lim}_{n\to\infty} x_n$ if and only if there exist rational numbers $a < b$ such that $\{x_n : n \geq 0\}$ upcrosses $[a,b]$ at least N times for every $N \in \mathbb{Z}^+$. Hence, $\{x_n : n \geq 0\}$ converges in $[-\infty, \infty]$ if and only if it upcrosses $[a,b]$ at most finitely often for each pair of rational numbers $a < b$.

THEOREM 5.2.15 (**Doob's Martingale Convergence Theorem**). * *Let $(X_n, \mathcal{F}_n, \mathbb{P})$ be a $\mathbb{P}$-integrable submartingale, and, for $-\infty < a < b < \infty$, let $U_{[a,b]}(\omega)$ denote the precise number of times that $\{X_n(\omega) : n \geq 0\}$ upcrosses $[a,b]$. Then*

$$\mathbb{E}^{\mathbb{P}}\big[U_{[a,b]}\big] \leq \sup_{n\in\mathbb{N}} \frac{\mathbb{E}^{\mathbb{P}}\big[(X_n - a)^+\big]}{b-a}. \tag{5.2.16}$$

In particular, if

$$\sup_{n\in\mathbb{N}} \mathbb{E}^{\mathbb{P}}\big[X_n^+\big] < \infty, \tag{5.2.17}$$

then there exists a $\mathbb{P}$-integrable random variable X to which $\{X_n : n \geq 0\}$ converges $\mathbb{P}$-almost surely. (See Exercises 5.2.28 and 5.2.30 for other derivations.)

PROOF: Set $Y_n = \frac{(X_n - a)^+}{b-a}$, and note that (by Corollary 5.2.9) $(Y_n, \mathcal{F}_n, \mathbb{P})$ is a $\mathbb{P}$-integrable submartingale. Next, let $N \in \mathbb{Z}^+$ be given, and set $\zeta_0' = 0$, and, for $k \in \mathbb{Z}^+$, define

$$\zeta_k = \inf\big\{n \geq \zeta_{k-1}' : X_n \leq a\big\} \wedge N \quad \text{and} \quad \zeta_k' = \inf\big\{n \geq \zeta_k : X_n \geq b\big\} \wedge N.$$

Proceeding by induction, it is an easy matter to check that all the ζ_k's and ζ_k''s are stopping times. Moreover, if $U^{(N)}_{[a,b]}(\omega)$ is the precise number of times $\big\{X_{n\wedge N}(\omega) : n \geq 0\big\}$ upcrosses $[a,b]$, then

$$\begin{aligned} U^{(N)}_{[a,b]} &\leq \sum_{k=1}^{N}\big(Y_{\zeta_k'} - Y_{\zeta_k}\big) = Y_N - Y_0 - \sum_{k=1}^{N}\big(Y_{\zeta_k} - Y_{\zeta_{k-1}'}\big) \\ &\leq Y_N - \sum_{k=1}^{N}\big(Y_{\zeta_k} - Y_{\zeta_{k-1}'}\big). \end{aligned}$$

* In the notes to Chapter VII of his *Stochastic Processes*, publ. by J. Wiley in 1953, Doob gives a thorough account of the relationship between his convergence result and earlier attempts in the same direction. In particular, he points out that, in 1946, S. Anderson and B. Jessen formulated and proved a closely related convergence theorem.

Hence, since $\zeta'_{k-1} \le \zeta_k$ and therefore, by (5.2.14), $\mathbb{E}^{\mathbb{P}}\big[Y_{\zeta_k} - Y_{\zeta'_{k-1}}\big] \ge 0$ for all $k \in \mathbb{Z}^+$, we see that $\mathbb{E}^{\mathbb{P}}[U^{(N)}_{[a,b]}] \le \mathbb{E}^{\mathbb{P}}[Y_N]$; and, clearly (5.2.16) follows from this after one lets $N \to \infty$.

Given (5.2.16), the convergence result is easy. Namely, if (5.2.17) is satisfied, then (5.2.16) implies that there is a set Λ of full $\mathbb{P}$-measure such that $U_{[a,b]}(\omega) < \infty$ for all rational $a < b$ and $\omega \in \Lambda$; and so, by the remark preceding the statement of this theorem, for each $\omega \in \Lambda$, $\{X_n(\omega) : n \ge 0\}$ converges to some $X(\omega) \in [-\infty, \infty]$. Hence, we will be done as soon as we know that $\mathbb{E}^{\mathbb{P}}[|X|, \Lambda] < \infty$. But

$$\mathbb{E}^{\mathbb{P}}\big[|X_n|\big] = 2\mathbb{E}^{\mathbb{P}}\big[X_n^+\big] - \mathbb{E}^{\mathbb{P}}\big[X_n\big] \le 2\mathbb{E}^{\mathbb{P}}\big[X_n^+\big] - \mathbb{E}^{\mathbb{P}}\big[X_0\big], \quad n \in \mathbb{N},$$

and therefore Fatou's Lemma plus (5.2.17) shows that X is $\mathbb{P}$-integrable. □

The inequality in (5.2.16) is quite famous and is known as **Doob's upcrossing inequality**.

Remark 5.2.18. The argument in the proof of Theorem 5.2.15 is so smooth that it is easy to miss the point which makes it work. Namely, the whole proof turns on the inequality $\mathbb{E}^{\mathbb{P}}[Y_{\zeta_k} - Y_{\zeta'_{k-1}}] \ge 0$. At first sight, this inequality seems to be wrong, since one is inclined to think that $Y_{\zeta_k} < Y_{\zeta'_{k-1}}$. However, Y_{ζ_k} need be less than $Y_{\zeta'_{k-1}}$ only if $\zeta_k < N$, which is precisely what, with high probability, the submartingale property is preventing from happening.

Corollary 5.2.19. *Let $\big(X_n, \mathcal{F}_n, \mathbb{P}\big)$ be a martingale. Then there exists an $X \in L^1(\mathbb{P};\mathbb{R})$ such that $X_n = \mathbb{E}^{\mathbb{P}}[X|\mathcal{F}_n]$ (a.s., $\mathbb{P}$) for each $n \in \mathbb{N}$ if and only if the sequence $\{X_n : n \ge 0\}$ is uniformly $\mathbb{P}$-integrable. In addition, if $p \in (1, \infty]$, then there is an $X \in L^p(\mathbb{P};\mathbb{R})$ such that $X_n = \mathbb{E}^{\mathbb{P}}[X|\mathcal{F}_n]$ (a.s., $\mathbb{P}$) for each $n \in \mathbb{N}$ if and only if $\{X_n : n \ge 0\}$ is a bounded subset of $L^p(\mathbb{P};\mathbb{R})$.*

Proof: Because of Corollary 5.2.4 and (5.2.3), we need only check the "if" statement in the first assertion. But, if $\{X_n : n \ge 0\}$ is uniformly $\mathbb{P}$-integrable, then (5.2.17) holds and therefore $X_n \longrightarrow X$ (a.s., $\mathbb{P}$) for some $\mathbb{P}$-integrable X. Moreover, uniform integrability together with almost sure convergence implies convergence in $L^1(\mathbb{P};\mathbb{R})$, and therefore, by (5.1.5), for each $m \in \mathbb{N}$,

$$X_m = \lim_{n\to\infty} \mathbb{E}^{\mathbb{P}}\big[X_n \big| \mathcal{F}_m\big] = \mathbb{E}^{\mathbb{P}}\big[X \big| \mathcal{F}_m\big] \quad \text{(a.s., } \mathbb{P}\text{)}. \quad \square$$

Just as Corollary 5.2.4 led us to an intuitively appealing way to construct conditional expectations, so Doob's Theorem gives us an appealing approximation procedure for Radon–Nikodym derivatives.

Theorem 5.2.20 (**Jessen**). *Let $\mathbb{P}$ and $\mathbb{Q}$ be a pair of probability measures on the measurable space $(\Omega, \mathcal{F})$ and $\big\{\mathcal{F}_n : n \in \mathbb{N}\big\}$ a non-decreasing sequence of sub-σ-algebras whose union generates $\mathcal{F}$. For each $n \in \mathbb{N}$, let $\mathbb{Q}_{n,\mathrm{a}}$ and $\mathbb{Q}_{n,\mathrm{s}}$ denote,* *respectively, the absolutely continuous and singular parts of* $\mathbb{Q}_n \equiv \mathbb{Q} \restriction \mathcal{F}_n$ *with respect to* $\mathbb{P}_n \equiv \mathbb{P} \restriction \mathcal{F}_n$*, and set* $X_n = \frac{d\mathbb{Q}_{n,\mathrm{a}}}{d\mathbb{P}_n}$*. Also, let* $\mathbb{Q}_\mathrm{a}$ *and* $\mathbb{Q}_\mathrm{s}$ *be the absolutely and singular continuous parts of* $\mathbb{Q}$ *with respect to* $\mathbb{P}$*, and set* $Y = \frac{d\mathbb{Q}_\mathrm{a}}{d\mathbb{P}}$*. Then* $X_n \longrightarrow Y$ *(a.s.,* $\mathbb{P}$*). In particular,* $\mathbb{Q} \perp \mathbb{P}$ *if and only if* $X_n \longrightarrow 0$ *(a.s.,* $\mathbb{P}$*). Moreover, if* $\mathbb{Q}_n \ll \mathbb{P}_n$ *for each* $n \in \mathbb{N}$*, then* $\mathbb{Q} \ll \mathbb{P}$ *if and only if* $\{X_n : n \geq 0\}$ *is uniformly* $\mathbb{P}$*-integrable, in which case* $X_n \longrightarrow Y$ *in* $L^1(\mathbb{P};\mathbb{R})$ *as well as* $\mathbb{P}$*-almost surely. Finally, if* $\mathbb{Q}_n \sim \mathbb{P}_n$ *(i.e.,* $\mathbb{P}_n \ll \mathbb{Q}_n$ *as well as* $\mathbb{Q}_n \ll \mathbb{P}_n$*) for each* $n \in \mathbb{N}$ *and* $G \equiv \left\{\lim_{n\to\infty} X_n \in (0,\infty)\right\}$*, then* $\mathbb{Q}_\mathrm{a}(A) = \mathbb{Q}(A \cap G)$ *for all* $A \in \mathcal{F}$*, and therefore* $\mathbb{Q}(G) = 1 \iff \mathbb{Q} \ll \mathbb{P}$ *and* $\mathbb{Q}(G) = 0 \iff \mathbb{Q} \perp \mathbb{P}$*.*

Proof: Without loss in generality, I will assume throughout that all the X_n's as well as $Y \equiv \frac{dQ_\mathrm{a}}{dP}$ take values in $[0,\infty)$; and clearly, $\mathbb{E}^{\mathbb{P}}[X_n]$, $n \in \mathbb{N}$, and $\mathbb{E}^{\mathbb{P}}[Y]$ are all dominated by 1.

First note that

$$\mathbb{Q}_{n,\mathrm{s}}(A) = \sup\Big\{\mathbb{Q}(A \cap B) : \, B \in \mathcal{F}_n \text{ and } \mathbb{P}(B) = 0\Big\} \quad \text{for} \quad A \in \mathcal{F}_n.$$

Hence, $\mathbb{Q}_{n,\mathrm{s}} \restriction \mathcal{F}_{n-1} \geq \mathbb{Q}_{n-1,\mathrm{s}}$ for each $n \in \mathbb{Z}^+$, and so

$$\mathbb{E}^{\mathbb{P}}\big[X_n,\, A\big] = \mathbb{Q}_{n,\mathrm{a}}(A) \leq \mathbb{Q}_{n-1,\mathrm{a}}(A) = \mathbb{E}^{\mathbb{P}}\big[X_{n-1},\, A\big]$$

for all $n \in \mathbb{Z}^+$ and $A \in \mathcal{F}_{n-1}$. In other words, $\big(-X_n, \mathcal{F}_n, \mathbb{P}\big)$ is a non-positive submartingale. Moreover, in the case when $\mathbb{Q}_n \ll P_n$, $n \in \mathbb{N}$, the same argument shows that $\big(X_n, \mathcal{F}_n, \mathbb{P}\big)$ is a non-negative martingale. Thus, in any case, there is a non-negative, $\mathbb{P}$-integrable random variable X with the property that $X_n \longrightarrow X$ (a.s., $\mathbb{P}$). In order to identify X as Y, use Fatou's Lemma to see that, for any $m \in \mathbb{N}$ and $A \in \mathcal{F}_m$:

$$\mathbb{E}^{\mathbb{P}}\big[X,\, A\big] \leq \varliminf_{n\to\infty} \mathbb{E}^{\mathbb{P}}\big[X_n,\, A\big] = \varliminf_{n\to\infty} \mathbb{Q}_{n,\mathrm{a}}(A) \leq \mathbb{Q}(A);$$

and therefore $\mathbb{E}^{\mathbb{P}}[X,\, A] \leq \mathbb{Q}(A)$, first for $A \in \bigcup_0^\infty \mathcal{F}_m$ and then for every $A \in \mathcal{F}$. In particular, by choosing $B \in \mathcal{F}$ so that $\mathbb{Q}_\mathrm{s}(B) = 0 = \mathbb{P}\big(B\complement\big)$, we have that

$$\mathbb{E}^{\mathbb{P}}\big[X,\, A\big] = \mathbb{E}^{\mathbb{P}}\big[X,\, A \cap B\big] \leq \mathbb{Q}(A \cap B) = \mathbb{Q}_\mathrm{a}(A) = \mathbb{E}^{\mathbb{P}}\big[Y,\, A\big] \quad \text{for all } A \in \mathcal{F};$$

which means that $X \leq Y$ (a.s., $\mathbb{P}$). On the other hand, if $Y_n = \mathbb{E}^{\mathbb{P}}[Y|\mathcal{F}_n]$ for $n \in \mathbb{N}$, then

$$\mathbb{E}^{\mathbb{P}}\big[Y_n,\, A\big] = \mathbb{Q}_\mathrm{a}(A) \leq \mathbb{Q}_{n,\mathrm{a}}(A) = \mathbb{E}^{\mathbb{P}}\big[X_n,\, A\big] \quad \text{for all} \quad A \in \mathcal{F}_n,$$

and therefore $Y_n \leq X_n$ (a.s., $\mathbb{P}$) for each $n \in \mathbb{N}$. Thus, since $Y_n \longrightarrow Y$ and $X_n \longrightarrow X$ $\mathbb{P}$-almost surely, this means that $Y \leq X$ (a.s., $\mathbb{P}$).

Next, assume that $\mathbb{Q}_n \ll \mathbb{P}_n$ for each $n \in \mathbb{N}$ and therefore that $\big(X_n, \mathcal{F}_n, \mathbb{P}\big)$ is a non-negative martingale. If $\{X_n : n \geq 0\}$ is uniformly $\mathbb{P}$-integrable, then $X_n \longrightarrow Y$ in $L^1(\mathbb{P};\mathbb{R})$ and therefore $\mathbb{Q}_s(\Omega) = 1 - \mathbb{E}^{\mathbb{P}}[Y] = 0$. Hence, $\mathbb{Q} \ll \mathbb{P}$ when $\{X_n : n \geq 0\}$ is uniformly $\mathbb{P}$-integrable. Conversely, if $\mathbb{Q} \ll P$, then it is easy to see that $X_n = \mathbb{E}^{\mathbb{P}}[Y|\mathcal{F}_n]$ for each $n \in \mathbb{N}$, and therefore, by Corollary 5.2.4, that $\{X_n : n \geq 0\}$ is uniformly $\mathbb{P}$-integrable.

Finally, assume that $\mathbb{Q}_n \sim \mathbb{P}_n$ for each $n \in \mathbb{N}$. Then, the X_n's can be chosen to take their values in $(0,\infty)$ and $Y_n \equiv \frac{1}{X_n} = \frac{d\mathbb{P}_n}{d\mathbb{Q}_n}$. Hence, if $\mathbb{P}_{\mathrm{a}}$ and $\mathbb{P}_{\mathrm{s}}$ are the absolutely continuous and singular parts of $\mathbb{P}$ relative to $\mathbb{Q}$ and if $Y \equiv \underline{\lim}_{n\to\infty} Y_n$, then $Y = \frac{d\mathbb{P}_{\mathrm{a}}}{d\mathbb{Q}}$ and so $\mathbb{P}_{\mathrm{a}}(A) = \mathbb{E}^{\mathbb{Q}}[Y,\, A]$ for all $A \in \mathcal{F}$. Thus, when $B \in \mathcal{F}$ is chosen so that $\mathbb{P}_{\mathrm{s}}(B) = 0 = \mathbb{Q}(B\complement)$, then, since $Y = \frac{1}{X}$ on G and $\mathbb{E}^{\mathbb{P}}[X,\, C\cap G] = \mathbb{E}^{\mathbb{P}}[X,\, C]$ for all $C \in \mathcal{F}$, it is becomes clear that

$$\begin{aligned}\mathbb{Q}(A\cap G) &= \mathbb{E}^{\mathbb{Q}}\big[XY,\, A\cap G\big] = \mathbb{E}^{\mathbb{P}_{\mathrm{a}}}\big[X,\, A\cap G\big]\\ &= \mathbb{E}^{\mathbb{P}}\big[X,\, A\cap G\cap B\big] = \mathbb{E}^{\mathbb{P}}\big[X,\, A\cap B\big] = \mathbb{Q}_{\mathrm{a}}(A\cap B) = \mathbb{Q}_{\mathrm{a}}(A)\end{aligned}$$

for all $A \in \mathcal{F}$. □

§ **5.2.4. Reversed Martingales & Exchangeable Random Variables.** For some applications it is important to know what happens if one runs a submartingale or martingale backwards. Thus, again let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, only this time suppose that $\big\{\mathcal{F}_n : n \in \mathbb{N}\big\}$ is a sequence of sub-σ-algebras which is *non-increasing*. Given a sequence $\{X_n : n \geq 0\}$ of $(-\infty,\infty]$-valued random variables, I will say that the triple $\big(X_n, \mathcal{F}_n, \mathbb{P}\big)$ is either a **reversed submartingale** or a **reversed martingale** if, for each $n \in \mathbb{N}$, X_n is $\mathcal{F}_n$-measurable and either $X_n^- \in L^1(\mathbb{P};\mathbb{R})$ and $X_{n+1} \leq \mathbb{E}^{\mathbb{P}}[X_n \,|\, \mathcal{F}_{n+1}]$ or $X_n \in L^1(\mathbb{P};\mathbb{R})$ and $X_{n+1} = \mathbb{E}^{\mathbb{P}}[X_n \,|\, \mathcal{F}_{n+1}]$.

THEOREM 5.2.21. *If $(X_n, \mathcal{F}_n, \mathbb{P})$ is a reversed submartingale, then*

$$\mathbb{P}\left(\sup_{n\in\mathbb{N}} X_n \geq R\right) \leq \frac{1}{R}\mathbb{E}^{\mathbb{P}}\left[X_0,\, \sup_{n\in\mathbb{N}} X_n \geq R\right], \quad R \in (0,\infty). \tag{5.2.22}$$

Moreover, if $(X_n, \mathcal{F}_n, \mathbb{P})$ is a reversed martingale, then $\big(|X_n|, \mathcal{F}_n, \mathbb{P}\big)$ is a reversed submartingale. In particular, if $(X_n, \mathcal{F}_n, \mathbb{P})$ is a non-negative reversed submartingale and $X_0 \in L^1(\mathbb{P};\mathbb{R})$, then $\{X_n : n \geq 0\}$ is uniformly $\mathbb{P}$-integrable and

$$\left\|\sup_{n\in\mathbb{N}} X_n\right\|_{L^p(\mathbb{P};\mathbb{R})} \leq \frac{p}{p-1}\big\|X_0\big\|_{L^p(\mathbb{P};\mathbb{R})} \quad \text{when} \quad p \in (1,\infty). \tag{5.2.23}$$

Finally, if $(X_n, \mathcal{F}_n, \mathbb{P})$ is a reversed submartingale and $X_0 \in L^1(\mathbb{P};\mathbb{R})$, then there is a $\mathcal{F}_\infty \equiv \bigcap_{n=0}^\infty \mathcal{F}_n$-measurable $X : \Omega \longrightarrow [-\infty,\infty]$ to which X_n converges $\mathbb{P}$-almost surely. In fact, X will be $\mathbb{P}$-integrable if $\inf_{n\geq 0} \mathbb{E}^{\mathbb{P}}[|X_n|] < \infty$; and if $(X_n, \mathcal{F}_n, \mathbb{P})$ is either a non-negative reversed submartingale or a reversed martingale with $X_0 \in L^p(\mathbb{P};\mathbb{R})$ for some $p \in [1,\infty)$, then $X_n \longrightarrow X$ in $L^p(\mathbb{P};\mathbb{R})$.

PROOF: More or less everything here follows immediately from the observation that $(X_n, \mathcal{F}_n, \mathbb{P})$ is a reversed submartingale or a reversed martingale if and only if, for each $N \in \mathbb{Z}^+$, $(X_{N-n\wedge N}, \mathcal{F}_{N-n\wedge N}, \mathbb{P})$ is a submartingale or a martingale. Indeed, by this observation and (5.2.2) applied to $(X_{N-n\wedge N}, \mathcal{F}_{N-n\wedge N}, \mathbb{P})$,

$$\mathbb{P}\left(\max_{0\le n\le N} X_n > R\right) \le \frac{1}{R}\mathbb{E}^{\mathbb{P}}\left[X_0,\ \max_{0\le n\le N} X_n > R\right]$$

for every $N \ge 1$. When $N \to \infty$, the left hand side of the preceding tends to left hand side of $\mathbb{P}\left(\sup_{n\in\mathbb{N}} X_n > R\right)$ and

$$\begin{aligned}\mathbb{E}^{\mathbb{P}}\left[X_0,\ \max_{0\le n\le N} X_n > R\right] &= \mathbb{E}^{\mathbb{P}}\left[X_0^+,\ \max_{0\le n\le N} X_n > R\right] - \mathbb{E}^{\mathbb{P}}\left[X_0^-,\ \max_{0\le n\le N} X_n > R\right]\\ \longrightarrow \mathbb{E}^{\mathbb{P}}\left[X_0^+,\ \sup_{n\in\mathbb{N}} X_n > R\right] &- \mathbb{E}^{\mathbb{P}}\left[X_0^-,\ \sup_{n\in\mathbb{N}} X_n > R\right] = \mathbb{E}^{\mathbb{P}}\left[X_0,\ \sup_{n\in\mathbb{N}} X_n > R\right],\end{aligned}$$

since X_0^+ is non-negative, and therefore the Monotone Convergence Theorems applies, and X_0^- is integrable, and therefore Lebesgue's Dominated Convergence Theorem applies. Thus (5.2.22) follows after one takes right limits in R.

Another application of the same observation shows that $(|X_n|, \mathcal{F}_n, \mathbb{P})$ is a reversed submartingale when $(X_n, \mathcal{F}_n, \mathbb{P})$ is a reversed martingale, and once one knows this, (5.2.23) follows from (5.2.22) and Exercise 1.4.18. In addition, when $(X_n, \mathcal{F}_n, \mathbb{P})$ is either a non-negative reversed submartingale or a reversed martingale,

$$\sup_{n\in\mathbb{N}} \mathbb{E}^{\mathbb{P}}\big[|X_n|,\ |X_n| \ge R\big] \le \sup_{n\in\mathbb{N}} \mathbb{E}^{\mathbb{P}}\big[|X_0|,\ |X_n| \ge R\big] \le \mathbb{E}^{\mathbb{P}}\left[|X_0|,\ \sup_{n\in\mathbb{N}} |X_n| \ge R\right],$$

which, by the (5.2.22), tends to 0 as $R \to \infty$. Thus, $\{X_n : n \ge 0\}$ is uniformly $\mathbb{P}$-integrable.

It remains to prove the convergence assertions, and again the key is the same observation. Before seeing how it applies, first say that $\{x_n : n \ge 0\}$ downcrosses $[a, b]$ at least N times if there exist $0 \le m_1 < n_1 < \cdots < m_N < n_N$ such that $x_{m_i} \ge b$ and $x_{n_i} \le a$ for each $1 \le i \le N$. Clearly, the same argument which I used for upcrossings applies to downcrossings and shows that $\{x_n : n \ge 0\}$ converges in $[-\infty, \infty]$ if and only if it downcrosses $[a, b]$ finitely often for each rational pair $a < b$. In addition, $\{x_n : 0 \le n \le N\}$ downcrosses $[a, b]$ the same number of times as $\{x_{N-n} : 0 \le n \le N\}$ upcrosses it. Hence, if $D^{(N)}_{[a,b]}(\omega)$ is the number of times $\{X_{n\wedge N} : n \ge 0\}$ downcrosses $[a, b]$, then this observation together with the estimate in the proof of Theorem 5.2.15 for $\mathbb{E}^{\mathbb{P}}[U^{(N)}_{[a,b]}]$ show that

$$\mathbb{E}^{\mathbb{P}}\big[D^{(N)}_{[a,b]}\big] \le \frac{\mathbb{E}^{\mathbb{P}}\big[(X_0 - a)^+\big]}{b - a}.$$

Starting from here, the argument used to prove Theorem 5.2.15 shows that there exits a $\mathcal{F}_\infty$-measurable $X : \Omega \longrightarrow [-\infty, \infty]$ to which $\{X_n : n \geq 0\}$ converges $\mathbb{P}$-almost surely. Once one has this almost sure convergence result, the rest of the theorem is an easy application of standard measure theory and the uniform integrability estimates proved above. $\square$

An important application of reversed martingales is provided by De Finetti's theory of exchangeable random variables. To describe his theory, let Σ denote the group of all *finite permutations* of $\mathbb{Z}^+$. That is, an element π of Σ is an isomorphism of $\mathbb{Z}^+$ which moves only a finite number of integers. Alternatively, $\Sigma = \bigcup_{m=1}^\infty \Sigma_m$, where Σ_m is the group of isomorphisms π of $\mathbb{Z}^+$ with the property that $n = \pi(n)$ for all $n > m$. Next, let $(E, \mathcal{B})$ be a measurable space, and, for each $\pi \in \Sigma$, define $S_\pi : E^{\mathbb{Z}^+} \longrightarrow E^{\mathbb{Z}^+}$ so that

$$S_\pi \mathbf{x} = \big(x_{\pi(1)}, \dots, x_{\pi(n)}, \dots\big) \quad \text{if } \mathbf{x} = \big(x_1, \dots, x_n, \dots\big).$$

Obviously, each S_π is a $\mathcal{B}^{\mathbb{Z}^+}$-measurable isomorphism from $E^{\mathbb{Z}^+}$ onto itself. Also, if

$$\mathcal{A}_m \equiv \big\{B \in \mathcal{B}^{\mathbb{Z}^+} : B = S_\pi B \text{ for all } \pi \in \Sigma_m\big\} \quad \text{for } m \in \mathbb{Z}^+,$$

then the $\mathcal{A}_m$'s form a non-increasing sequence of sub π-algebras of $\mathcal{B}^{\mathbb{Z}^+}$, and

$$\bigcap_{m=1}^\infty \mathcal{A}_m = \mathcal{A}_\infty \equiv \big\{B \in \mathcal{B}^{\mathbb{Z}^+} : B = S_\pi B \text{ for all } \pi \in \Sigma\big\}.$$

Now suppose that $\{X_n : n \geq 1\}$ is a sequence of E-valued random variables on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and set $\mathbf{X}(\omega) = \big(X_1(\omega), \dots, X_n(\omega), \dots\big) \in E^{\mathbb{Z}^+}$. The X_n's are said to be **exchangeable random variables** if $\mathbf{X}$ has the same $\mathbb{P}$-distribution as S_π for every $\pi \in \Sigma$. The central result of De Finetti's theory is **De Finetti's Strong Law**, which states that for any $g : E \longrightarrow \mathbb{R}$ satisfying $g \circ X_1 \in L^1(\mathbb{P}; \mathbb{R})$,

$$\mathbb{E}^{\mathbb{P}}\big[g \circ X_1 \,\big|\, \mathbf{X}^{-1}(\mathcal{A}_\infty)\big] = \lim_{n\to\infty} \frac{1}{n} \sum_1^n g \circ X_m \tag{5.2.24}$$

where the convergence is $\mathbb{P}$-almost sure and in $L^1(\mathbb{P}; \mathbb{R})$.

To prove (5.2.24), observe that, for any $1 \leq m \leq n$, $\mathbb{E}^{\mathbb{P}}[g \circ X_m \,|\, \mathbf{X}^{-1}(\mathcal{A}_n)] = \mathbb{E}^{\mathbb{P}}[g \circ X_1 \,|\, \mathbf{X}^{-1}(\mathcal{A}_n)]$, which leads immediately to

$$\mathbb{E}^{\mathbb{P}}\big[g \circ X_1 \,\big|\, \mathbf{X}^{-1}(\mathcal{A}_n)\big] = \mathbb{E}^{\mathbb{P}}\left[\frac{1}{n}\sum_{m=1}^n g \circ X_m \,\middle|\, \mathbf{X}^{-1}(\mathcal{A}_n)\right] = \frac{1}{n}\sum_{m=1}^n g \circ X_m.$$

Hence, (5.2.24) follows as an application of Theorem 5.2.21.

De Finetti's Strong Law makes it important to get a handle on the σ-algebra $\mathbf{X}^{-1}(\mathcal{A}_\infty)$. In particular, one would like to know when $\mathbf{X}^{-1}(\mathcal{A}_\infty)$ is trivial in the sense that each of its elements has probability 0 or 1, in which case (5.2.24) self-improves to the statement that

$$\lim_{n\to\infty} \frac{1}{n}\sum_{1}^{n} g\circ X_m = \mathbb{E}^{\mathbb{P}}[g\circ X_1] \quad \mathbb{P}\text{-almost surely and in } L^1(\mathbb{P};\mathbb{R}). \tag{5.2.25}$$

The following lemma is the crucial step toward gaining an understanding of $\mathbf{X}^{-1}(\mathcal{A}_\infty)$.

Lemma 5.2.26. *Refer to the preceding, and let $\mathcal{T} = \bigcap_{m=1}^\infty \sigma\big(\{X_n : n \ge m\}\big)$ be the tail σ-algebra determined by $\{X_n : n \ge 1\}$. Then $\mathcal{T} \subseteq \mathbf{X}^{-1}(\mathcal{A}_\infty)$ and $\mathbf{X}^{-1}(\mathcal{A}_\infty)$ is contained in the completion of $\mathcal{T}$ with respect to $\mathbb{P}$. In particular, for each $F \in L^1(\mathbb{P};\mathbb{R})$,*

$$\mathbb{E}^{\mathbb{P}}\big[F \,\big|\, \mathbf{X}^{-1}(\mathcal{A}_\infty)\big] = \mathbb{E}^{\mathbb{P}}\big[F \,\big|\, \mathcal{T}\big] \quad (\text{a.s.}, \mathbb{P}). \tag{5.2.27}$$

Proof: The inclusion $\mathcal{T} \subseteq \mathbf{X}^{-1}(\mathcal{A}_\infty)$ is obvious. Thus, what remains to be proved is that, for any $F \in L^1(\mathbb{P};\mathbb{R})$, $\mathbb{E}^{\mathbb{P}}[F \,|\, \mathbf{X}^{-1}(\mathcal{A}_\infty)]$ is, up to a $\mathbb{P}$-null set, $\mathcal{T}$-measurable. To this end, begin by observing that it suffices to check this for F's which are $\sigma\big(\{X_n : 1 \le m \le N\}\big)$-measurable for some $N \in \mathbb{Z}^+$. Indeed, since $\mathbf{X}^{-1}(\mathcal{A}_\infty) \subseteq \sigma\big(\{X_n : n \ge 1\}\big)$, we know that

$$\begin{aligned}\mathbb{E}^{P}\big[F \,\big|\, \mathbf{X}^{-1}(\mathcal{A}_\infty)\big] &= \mathbb{E}^{\mathbb{P}}\Big[\mathbb{E}^{\mathbb{P}}\big[F \,\big|\, \sigma\big(\{X_n : n \ge 1\}\big)\big] \,\Big|\, \mathbf{X}^{-1}(\mathcal{A}_\infty)\Big] \\ &= \lim_{N\to\infty} \mathbb{E}^{\mathbb{P}}\Big[\mathbb{E}^{\mathbb{P}}\big[F \,\big|\, \sigma\big(\{X_m : 1 \le m \le N\}\big)\big] \,\Big|\, \mathbf{X}^{-1}(\mathcal{A}_\infty)\Big].\end{aligned}$$

Now suppose that F is $\sigma\big(\{X_m : 1 \le m \le N\}\big)$-measurable. Then there exists a $g : E^N \longrightarrow \mathbb{R}$ such that $F = g\big(X_1, \dots, X_N\big)$. If $N = 1$, then, because $\lim_{n\to\infty} \frac{1}{n}\sum_{m=1}^n g\circ X_m$ is $\mathcal{T}$-measurable, (5.2.24) says $E^{\mathbb{P}}[F \,|\, \mathbf{X}^{-1}(\mathcal{A}_\infty)]$ is $\mathcal{T}$-measurable. To get the same conclusion when $N \ge 2$, I want to apply the same reasoning, only now with E replaced by E^N. To be precise, define

$$\mathcal{A}_\infty^{(N)} = \big\{B \in \mathcal{B}^{\mathbb{Z}^+} : B = S_\sigma B \text{ for all } \pi \in \Sigma^{(N)}\big\} \text{ where}$$
$$\Sigma^{(N)} = \big\{\pi \in \Sigma : \pi(\ell N + m) = \pi(\ell N + 1) + m - 1 \text{ for all } \ell \in \mathbb{N} \text{ and } 1 \le m < N\big\}$$

is the group of finite permutations which transform $\mathbb{Z}^+$ in blocks of length N. By (5.2.24) applied with E^N replacing E, we find that $\mathbb{E}^{\mathbb{P}}\big[F \big| \mathbf{X}^{-1}(\mathcal{A}_\infty^{(N)})\big] = \mathbb{E}^{\mathbb{P}}\big[F \,\big|\, \mathcal{T}\big]$ $\mathbb{P}$-almost surely. Hence, since $\mathbf{X}^{-1}(\mathcal{I}_\infty) \subseteq \mathbf{X}^{-1}(\mathcal{I}_\infty^{(N)})$, (5.2.27) holds for every $\sigma\big(\{X_n : 1 \le n \le N\}\big)$-measurable $F \in L^1(\mathbb{P};\mathbb{R})$. □

The best known consequence of Lemma 5.2.26 is the **Hewitt–Savage 0–1 Law**, which says that $\mathbf{X}^{-1}(\mathcal{A}_\infty)$ is trivial if the X_n's are independent and identically distributed. Clearly, their result is a immediate consequence of Lemma 5.2.26 together with Kolmogorov's 0–1 law.

Seeing as the Strong Law of Large Numbers follows from (5.2.24) combined with the Hewitt–Savage 0–1 law, one might think that (5.2.24) represents an extension of the strong law. However, that is not really the case, since it can be shown that $\mathbf{X}^{-1}(\mathcal{A}_\infty)$ is trivial only if the X_n's are independent. On the other hand, the derivation of the Strong Law via (5.2.24) extends without alteration to the Banach space setting (cf. part (**ii**) of Exercise 6.1.16).

Exercises for § 5.2

EXERCISE 5.2.28. In this exercise I will outline a quite independent derivation of the convergence assertion in Doob's Martingale Convergence Theorem. The key observations here are first that, given Doob's Inequality (cf. (5.2.2)), the result is nearly trivial for martingales having two bounded moments and, second, that everything can be reduced to that case.

(**i**) Let $\big(M_n, \mathcal{F}_n, \mathbb{P}\big)$ be a martingale which is L^2-bounded in the sense that $\sup_{n\in\mathbb{N}} \mathbb{E}^{\mathbb{P}}[M_n^2] < \infty$. Note that

$$\mathbb{E}^{\mathbb{P}}\big[M_n^2\big] - \mathbb{E}^{\mathbb{P}}\big[M_{m-1}^2\big] = \mathbb{E}^{\mathbb{P}}\Big[\big(M_n - M_{m-1}\big)^2\Big] \quad \text{for} \quad 1 \le m \le n;$$

and starting from this, show that there is an $M \in L^2(\mathbb{P};\mathbb{R})$ such that $M_n \longrightarrow M$ in $L^2(\mathbb{P};\mathbb{R})$. Next apply (5.2.5) to the submartingale $\Big(\big|M_{n\vee m} - M_m\big|, \mathcal{F}_n, \mathbb{P}\Big)$ to show that, for every $\epsilon > 0$,

$$\mathbb{P}\left(\sup_{n\ge m} \big|M_n - M_m\big| \ge \epsilon\right) \le \frac{1}{\epsilon}\mathbb{E}^{\mathbb{P}}\Big[\big|M - M_m\big|\Big] \longrightarrow 0 \quad \text{as} \quad m \to \infty,$$

and conclude that $M_n \longrightarrow M$ (a.s., $\mathbb{P}$).

(**ii**) Let $\big(X_n, \mathcal{F}_n, \mathbb{P}\big)$ be a non-negative submartingale with the property that $\sup_{n\in\mathbb{N}} \mathbb{E}^{\mathbb{P}}[X_n^2] < \infty$, define the sequence $\big\{A_n : n \in \mathbb{N}\big\}$ as in Lemma 5.2.12, and set $M_n = X_n - A_n$, $n \in \mathbb{N}$. Then $\big(M_n, \mathcal{F}_n, \mathbb{P}\big)$ is a martingale, and clearly both M_n and A_n are square $\mathbb{P}$-integrable for each $n \in \mathbb{N}$. In fact, check that

$$\begin{aligned}\mathbb{E}^{\mathbb{P}}\big[M_n^2 - M_{n-1}^2\big] &= \mathbb{E}^{\mathbb{P}}\big[\big(M_n - M_{n-1}\big)\big(X_n + X_{n-1}\big)\big]\\ &= \mathbb{E}^{\mathbb{P}}\big[X_n^2 - X_{n-1}^2\big] - \mathbb{E}^{\mathbb{P}}\big[\big(A_n - A_{n-1}\big)\big(X_n + X_{n-1}\big)\big] \le \mathbb{E}^{\mathbb{P}}\big[X_n^2 - X_{n-1}^2\big],\end{aligned}$$

and therefore that

$$\mathbb{E}^{\mathbb{P}}\big[M_n^2\big] \le \mathbb{E}^{\mathbb{P}}\big[X_n^2\big] \quad \text{and} \quad \mathbb{E}^{\mathbb{P}}\big[A_n^2\big] \le 4\mathbb{E}^{\mathbb{P}}\big[X_n^2\big] \quad \text{for every } n \in \mathbb{N}.$$

Finally, show that there exist $M \in L^2(\mathbb{P};\mathbb{R})$ and $A \in L^2\big(\mathbb{P};[0,\infty)\big)$ such that $M_n \longrightarrow M$, $A_n \nearrow A$, and, therefore, $X_n \longrightarrow X \equiv M + A$ both $\mathbb{P}$-almost surely and in $L^2(\mathbb{P};\mathbb{R})$.

(**iii**) Let $(X_n, \mathcal{F}_n, \mathbb{P})$ be a non-negative martingale, set $Y_n = e^{-X_n}$, $n \in \mathbb{N}$, use Corollary 5.2.9 to see that $(Y_n, \mathcal{F}_n, \mathbb{P})$ is a uniformly bounded, non-negative, submartingale, and apply part (**ii**) to conclude that $\{X_n : n \geq 0\}$ converges $\mathbb{P}$-almost surely to a non-negative $X \in L^1(\mathbb{P}; \mathbb{R})$.

(**iv**) Let $(X_n, \mathcal{F}_n, \mathbb{P})$ be a martingale for which

$$\sup_{n \in \mathbb{N}} \mathbb{E}^{\mathbb{P}}\big[|X_n|\big] < \infty. \tag{5.2.29}$$

For each $m \in \mathbb{N}$, define $Y_{n,m}^{\pm} = \mathbb{E}^{\mathbb{P}}\big[X_{n \vee m}^{\pm} \big| \mathcal{F}_m\big] \vee 0$ for $n \in \mathbb{N}$. Show that $Y_{n+1,m}^{\pm} \geq Y_{n,m}^{\pm}$ (a.s., $\mathbb{P}$), define $Y_m^{\pm} = \varliminf_{n \to \infty} Y_{n,m}^{\pm}$, check that both $(Y_m^+, \mathcal{F}_m, \mathbb{P})$ and $(Y_m^-, \mathcal{F}_m, \mathbb{P})$ are non-negative martingales with $\mathbb{E}^{\mathbb{P}}\big[Y_0^+ + Y_0^-\big] \leq \sup_{n \in \mathbb{N}} \mathbb{E}^{\mathbb{P}}\big[|X_n|\big]$, and note that $X_m = Y_m^+ - Y_m^-$ (a.s., $\mathbb{P}$) for each $m \in \mathbb{N}$. In other words, *every martingale* $(X_n, \mathcal{F}_n, \mathbb{P})$ *satisfying (5.2.29) admits a* **Hahn decomposition*** *as the difference of two non-negative martingales whose sum has expectation value dominated by the left-hand side of* (5.2.29). Finally, use this observation together with (**iii**) to see that every such martingale converges $\mathbb{P}$-almost surely to some $X \in L^1(\mathbb{P}; \mathbb{R})$.

(**v**) By combining the final assertion in (**iv**) together with Doob's Decomposition in Lemma 5.2.12, give another proof of the convergence assertion in Theorem 5.2.15.

Exercise 5.2.30. In this exercise we will develop another way to reduce Doob's Martingale Convergence Theorem to the case of L^2-bounded martingales. The technique here is due to R. Gundy and derives from the ideas introduced by Calderón and Zygmund in connection with their famous work on weak-type 1–1 estimates for singular integrals.

(**i**) Let $\{Z_n : n \in \mathbb{N}\}$ be a $\{\mathcal{F}_n : n \in \mathbb{N}\}$-progressively measurable, $[0, R]$-valued sequence with the property that $(-Z_n, \mathcal{F}_n, \mathbb{P})$ is a submartingale. Next, choose $\{A_n : n \in \mathbb{N}\}$ for $(-Z_n, \mathcal{F}_n, \mathbb{P})$ as in Lemma 5.2.12, note that A_n's can be chosen so that $0 \leq A_n - A_{n-1} \leq R$ for all $n \in \mathbb{Z}^+$, and set $M_n = Z_n + A_n$, $n \in \mathbb{N}$. Check that $(M_n, \mathcal{F}_n, \mathbb{P})$ is a non-negative martingale with $M_n \leq (n+1)R$ for each $n \in \mathbb{N}$. Next, show that

$$\begin{aligned}
\mathbb{E}^{\mathbb{P}}\big[M_n^2 - M_{n-1}^2\big] &= \mathbb{E}^{\mathbb{P}}\big[(M_n - M_{n-1})(Z_n + Z_{n-1})\big] \\
&= \mathbb{E}^{\mathbb{P}}\big[Z_n^2 - Z_{n-1}^2\big] + \mathbb{E}^{\mathbb{P}}\big[(A_n - A_{n-1})(Z_n + Z_{n-1})\big] \\
&\leq \mathbb{E}^{\mathbb{P}}\big[Z_n^2 - Z_{n-1}^2\big] + 2R\,\mathbb{E}^{\mathbb{P}}\big[A_n - A_{n-1}\big],
\end{aligned}$$

and conclude that $\mathbb{E}^{\mathbb{P}}[A_n^2] \leq \mathbb{E}^{\mathbb{P}}[M_n^2] \leq 3R\mathbb{E}^{\mathbb{P}}[Z_0]$ for all $n \in \mathbb{N}$.

---

* This useful observation was made by Klaus Krickeberg.

(**ii**) Let $(X_n, \mathcal{F}_n, \mathbb{P})$ be a non-negative martingale. Show that, for each $R \in (0,\infty)$, $X_n = M_n^{(R)} - A_n^{(R)} + \Delta_n^{(R)}$, $n \in \mathbb{N}$, where $(M_n^{(R)}, \mathcal{F}_n, \mathbb{P})$ is a non-negative martingale satisfying $\mathbb{E}^{\mathbb{P}}\big[\big(M_n^{(R)}\big)^2\big] \le 3R\,\mathbb{E}^{\mathbb{P}}\big[X_0\big]$, $n \in \mathbb{N}$, $\{A_n^{(R)} : n \in \mathbb{N}\}$ is a non-decreasing sequence of random variables with the properties that $A_0^{(R)} \equiv 0$, $A_n^{(R)}$ is $\mathcal{F}_{n-1}$-measurable and $\mathbb{E}^{\mathbb{P}}\big[\big(A_n^{(R)}\big)^2\big] \le 3R\,\mathbb{E}^{\mathbb{P}}\big[X_0\big]$, $n \in \mathbb{Z}^+$, and $\{\Delta_n^{(R)} : n \in \mathbb{N}\}$ is a $\{\mathcal{F}_n : n \in \mathbb{N}\}$-progressively measurable sequence with the property that

$$\mathbb{P}\big(\exists n \in \mathbb{N}\ \Delta_n^{(R)} \neq 0\big) \le \frac{1}{R}\,\mathbb{E}^{\mathbb{P}}\big[X_0\big].$$

Hint: Set $Z_n^{(R)} = X_n \wedge R$ and $\Delta_n^{(R)} = X_n - Z_n^{(R)}$ for $n \in \mathbb{N}$, apply part (**i**) to $\{Z_n^{(R)} : n \in \mathbb{N}\}$, and use Doob's Inequality to estimate the probability that $\Delta_n^{(R)} \neq 0$ for some $n \in \mathbb{N}$.

(**iii**) Let $(X_n, \mathcal{F}_n, \mathbb{P})$ be any martingale. Using (**ii**) above and part (**iv**) of Exercise 5.2.28, show that, for each $R \in (0,\infty)$, $X_n = M_n^{(R)} + V_n^{(R)} + \Delta_n^{(R)}$, $n \in \mathbb{N}$, where $(M_n^{(R)}, \mathcal{F}_n, \mathbb{P})$ is a martingale satisfying $\mathbb{E}^{\mathbb{P}}\big[\big(M_n^{(R)}\big)^2\big] \le 12\,R\mathbb{E}^{\mathbb{P}}\big[|X_n|\big]$, $\{V_n^{(R)} : n \in \mathbb{N}\}$ is a sequence of random variables satisfying $V_0^{(R)} \equiv 0$, $V_n^{(R)}$ is $\mathcal{F}_{n-1}$-measurable, and

$$\mathbb{E}^{\mathbb{P}}\left[\left(\sum_1^n \big|V_m^{(R)} - V_{m-1}^{(R)}\big|\right)^2\right] \le 12R\mathbb{E}^{\mathbb{P}}\big[|X_n|\big]$$

for $n \in \mathbb{Z}^+$, and $\{\Delta_n :\in \mathbb{N}\}$ is an $\{\mathcal{F}_n : n \in \mathbb{N}\}$-progressively measurable sequence satisfying

$$\mathbb{P}\left(\exists\, 0 \le m \le n\ \Delta_m^{(R)} \neq 0\right) \le \frac{2}{R}\,\mathbb{E}^{\mathbb{P}}\big[|X_n|\big].$$

The preceding representation is called the **Calderón–Zygmund decomposition of the martingale** $(X_n, \mathcal{F}_n, \mathbb{P})$.

(**iv**) Let $(X_n, \mathcal{F}_n, \mathbb{P})$ be a martingale which satisfies (5.2.29), and use part (**iii**) above together with part (**i**) of Exercise 5.2.28 to show that, for each $R \in (0,\infty)$, $\{X_n : n \ge 0\}$ converges off of a set whose $\mathbb{P}$-measure is no more than $\frac{2}{R}$ times the supremum over $n \in \mathbb{N}$ of $\mathbb{E}^{\mathbb{P}}[|X_n|]$. In particular, when combined with Lemma 5.2.12, the preceding line of reasoning leads to the advertised alternate proof of the convergence result in Theorem 5.2.15.

Exercise 5.2.31. In this exercise we will extend Hunt's Theorem (cf. Theorem 5.2.13) to allow unbounded stopping times. To this end, let $(X_n, \mathcal{F}_n, \mathbb{P})$ be a uniformly $\mathbb{P}$-integrable submartingale on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and set $M_n = X_n - A_n$, $n \in \mathbb{N}$, where $\{A_n : n \in \mathbb{N}\}$ is the sequence discussed in Lemma

5.2.12. After checking that $\big(M_n, \mathcal{F}_n, \mathbb{P}\big)$ is a uniformly $\mathbb{P}$-integrable martingale, show that, for any stopping time ζ: $X_\zeta = \mathbb{E}^{\mathbb{P}}[M_\infty|\mathcal{F}_\zeta] + A_\zeta \quad (\text{a.s.}, \mathbb{P})$, where, X_∞, M_∞, and A_∞ are, respectively, the $\mathbb{P}$-almost sure limits of $\{X_n : n \geq 0\}$, $\{M_n : n \geq 0\}$, and $\{A_n : n \geq 0\}$. In particular, if ζ and ζ' are a pair of stopping times and $\zeta \leq \zeta'$, conclude that $X_\zeta \leq \mathbb{E}^{\mathbb{P}}[X_{\zeta'}|\mathcal{F}_\zeta] \quad (\text{a.s.}, \mathbb{P})$.

Exercise 5.2.32. There are times when submartingales converge even though they are not bounded in $L^1(\mathbb{P};\mathbb{R})$. For example, suppose that $(X_n, \mathcal{F}_n, \mathbb{P})$ is a submartingale for which there exists a non-decreasing function $\rho : \mathbb{R} \longmapsto \mathbb{R}$ with the properties that $\rho(R) \geq R$ for all R and $X_{n+1} \leq \rho\big(X_n\big)$ (a.e., $\mathbb{P}$) for each $n \in \mathbb{N}$.

(i) Set $\zeta_R(\omega) = \inf\big\{n \in \mathbb{N} : X_n(\omega) \geq R\big\}$ for $R \in (0, \infty)$, and note that

$$\sup_{n\in\mathbb{N}} X_{n\wedge\zeta_R} \leq X_0 \vee \rho(R) \quad (\text{a.e.}, \mathbb{P}).$$

In particular, if X_0 is $\mathbb{P}$-integrable, show that $\{X_n(\omega) : n \geq 0\}$ converges in $\mathbb{R}$ for $\mathbb{P}$-almost every ω for which $\{X_n(\omega) : n \geq 0\}$ is bounded above.

Hint: After observing that $\sup_{n\in\mathbb{N}} \mathbb{E}^{\mathbb{P}}[X^+_{n\wedge\zeta_R}] < \infty$ for every $R \in (0, \infty)$, conclude that, for each $R \in (0, \infty)$, $\{X_n : n \geq 0\}$ converges $\mathbb{P}$-almost everywhere on $\{\zeta_R = \infty\}$.

(ii) Let $\{Y_n : n \geq 1\}$ be a sequence of mutually independent, $\mathbb{P}$-integrable random variables, assume that $\mathbb{E}^{\mathbb{P}}[Y_n] \geq 0$ for $n \in \mathbb{N}$ and $\sup_{n\in\mathbb{N}} \|Y_n^+\|_{L^\infty(\mathbb{P};\mathbb{R})} < \infty$, and set $S_n = \sum_1^n Y_m$. Show that $\{S_n : n \geq 0\}$ is either $\mathbb{P}$-almost surely unbounded above or $\mathbb{P}$-almost surely convergent in $\mathbb{R}$.

(iii) Let $\big\{\mathcal{F}_n : n \in \mathbb{N}\big\}$ be a non-decreasing sequence of sub-σ-algebras and A_n an element of $\mathcal{F}_n$ for each $n \in \mathbb{N}$. Show that the set of $\omega \in \Omega$ for which either

$$\sum_{n=0}^{\infty} \mathbf{1}_{A_n}(\omega) < \infty \text{ but } \sum_{n=1}^{\infty} P\big(A_n \,\big|\, \mathcal{F}_{n-1}\big)(\omega) = \infty$$

or

$$\sum_{n=0}^{\infty} \mathbf{1}_{A_n}(\omega) = \infty \text{ but } \sum_{n=1}^{\infty} P\big(A_n \,\big|\, \mathcal{F}_{n-1}\big)(\omega) < \infty$$

has $\mathbb{P}$-measure 0. In particular, note that this gives another derivation of the second part of the Borel–Cantelli Lemma (cf. Lemma 1.1.3).

Exercise 5.2.33. For each $n \in \mathbb{N}$, let $(E_n, \mathcal{B}_n)$ be a measurable space and μ_n and ν_n a pair of probability measures on $(E_n, \mathcal{B}_n)$ with the property that $\nu_n \ll \mu_n$. Prove **Kakutani's Theorem** which says that (cf. Exercise 1.1.14) either $\prod_{n\in\mathbb{N}} \nu_n \perp \prod_{n\in\mathbb{N}} \mu_n$ or $\prod_{n\in\mathbb{N}} \nu_n \ll \prod_{n\in\mathbb{N}} \mu_n$.

Hint: Set

$$\Omega = \prod_{n\in\mathbb{N}} E_n, \quad \mathcal{F} = \prod_{n\in\mathbb{N}} \mathcal{B}_n, \quad \mathbb{P} = \prod_{n\in\mathbb{N}} \mu_n, \quad \text{and } \mathbb{Q} = \prod_{n\in\mathbb{N}} \nu_n.$$

Next, take $\mathcal{F}_n = \pi_n^{-1}\left(\prod_0^n \mathcal{B}_m\right)$, where π_n is the natural projection from Ω onto $\prod_0^n E_m$, set $\mathbb{P}_n = \mathbb{P} \restriction \mathcal{F}_n$ and $\mathbb{Q}_n = \mathbb{Q} \restriction \mathcal{F}_n$, and note that

$$X_n(x) \equiv \frac{d\mathbb{Q}_n}{d\mathbb{P}_n}(x) = \prod_0^n f_m(x_m), \quad x \in \Omega,$$

where $f_n \equiv \frac{d\nu_n}{d\mu_n}$. In particular, when $\nu_n \sim \mu_n$ for each $n \in \mathbb{N}$, use Kolmogorov's 0–1 Law (cf. Theorem 1.1.2) to see that $\mathbb{Q}(G) \in \{0,1\}$, where $G \equiv \left\{\lim_{n\to\infty} X_n \in (0,\infty)\right\}$, and combine this with the last part of Theorem 5.2.20 to conclude that $\mathbb{Q} \not\perp \mathbb{P} \implies \mathbb{Q} \ll \mathbb{P}$. Finally, to remove the assumption that $\nu_n \sim \mu_n$ for all n's, define $\tilde{\nu}_n$ on $(E_n, \mathcal{B}_n)$ by $\tilde{\nu}_n = \left(1 - 2^{-n-1}\right)\nu_n + 2^{-n-1}\mu_n$, check that $\tilde{\nu}_n \sim \mu_n$ and $\mathbb{Q} \ll \tilde{\mathbb{Q}} \equiv \prod_{n\in\mathbb{N}} \tilde{\nu}_n$, and use the preceding to complete the proof.

Exercise 5.2.34. Let $(\Omega, \mathcal{F})$ be a measurable space and Σ a sub-σ-algebra of $\mathcal{F}$. Given a pair of probability measures $\mathbb{P}$ and $\mathbb{Q}$ on $(\Omega, \mathcal{F})$, let X_Σ and Y_Σ be non-negative Radon–Nikodym derivatives of, respectively, $\mathbb{P}_\Sigma \equiv \mathbb{P} \restriction \Sigma$ and $\mathbb{Q}_\Sigma \equiv \mathbb{Q} \restriction \Sigma$ with respect to $\left(\mathbb{P}_\Sigma + \mathbb{Q}_\Sigma\right)$, and define

$$\left(\mathbb{P}, \mathbb{Q}\right)_\Sigma = \int X_\Sigma^{\frac{1}{2}}\, Y_\Sigma^{\frac{1}{2}}\, d(\mathbb{P} + \mathbb{Q}).$$

(**i**) Show that if μ is any σ-finite measure on (Ω, Σ) with the property that $\mathbb{P}_\Sigma \ll \mu$ and $\mathbb{Q}_\Sigma \ll \mu$, then the number $\left(\mathbb{P}, \mathbb{Q}\right)_\Sigma$ given above is equal to

$$\int \left(\frac{d\mathbb{P}_\Sigma}{d\mu}\right)^{\frac{1}{2}} \left(\frac{d\mathbb{Q}_\Sigma}{d\mu}\right)^{\frac{1}{2}} d\mu.$$

Also, check that $\mathbb{P}_\Sigma \perp \mathbb{Q}_\Sigma$ if and only if $\left(\mathbb{P}, \mathbb{Q}\right)_\Sigma = 0$.

(**ii**) Suppose that $\left\{\mathcal{F}_n : n \in \mathbb{N}\right\}$ is a non-decreasing sequence of sub-σ-algebras of $\mathcal{F}$, and show that $(\mathbb{P}, \mathbb{Q})_{\mathcal{F}_n} \longrightarrow (\mathbb{P}, \mathbb{Q})_{\bigvee_0^\infty \mathcal{F}_n}$.

(**iii**) Referring to part (**ii**), assume that $\mathbb{Q} \restriction \mathcal{F}_n \ll P \restriction \mathcal{F}_n$ for each $n \in \mathbb{N}$, let X_n be a non-negative Radon–Nikodym derivative of $\mathbb{Q} \restriction \mathcal{F}_n$ with respect to $\mathbb{P} \restriction \mathcal{F}_n$, and show that $\mathbb{Q} \restriction \bigvee_0^\infty \mathcal{F}_n$ is singular to $\mathbb{P} \restriction \bigvee_0^\infty \mathcal{F}_n$ if and only if $\mathbb{E}^{\mathbb{P}}\left[\sqrt{X_n}\right] \longrightarrow 0$ as $n \to \infty$.

(**iv**) Let $\{\sigma_n\}_0^\infty \subseteq (0,\infty)$, and, for each $n \in \mathbb{N}$, let μ_n and ν_n be Gaussian measures on $\mathbb{R}$ with variance σ_n^2. If a_n and b_n are the mean value of, respectively, μ_n and ν_n, show that

$$\prod_{n\in\mathbb{N}} \nu_n \sim \prod_{n\in\mathbb{N}} \mu_n \quad \text{or} \quad \prod_{n\in\mathbb{N}} \nu_n \perp \prod_{n\in\mathbb{N}} \mu_n$$

depending on whether $\sum_0^\infty \sigma_n^{-2}(b_n - a_n)^2$ converges or diverges.

EXERCISE 5.2.35. Let $\{X_n : n \in \mathbb{Z}^+\}$ be a sequence of identically distributed, mutually independent, integrable, mean-value 0, $\mathbb{R}$-valued random variables on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and set $S_n = \sum_1^n X_m$ for $n \in \mathbb{Z}^+$. In Exercise 1.4.28 we showed that $\underline{\lim}_{n\to\infty} |S_n| < \infty$ $\mathbb{P}$-almost surely. Here we will show that

$$\underline{\lim}_{n\to\infty} |S_n| = 0 \quad P\text{-almost surely.} \tag{5.2.36}$$

As was mentioned before, this result was proved first by K.L. Chung and W.H. Fuchs. The basic observation behind the present proof is due to A. Perlin, who noticed that, by the Hewitt–Savage 0–1 Law, there is a constant $L \in [0,\infty)$ which equals $\underline{\lim}_{n\to\infty} |S_n| = L$ $\mathbb{P}$-almost surely. Thus, the problem is to show that $L = 0$, and we will do this by an simple argument invented by A. Yushkevich.

(**i**) Assuming that $L > 0$, use the Hewitt–Savage 0–1 to show that

$$\mathbb{P}\Big(|S_n - x| < \tfrac{L}{3} \text{ i.o.}\Big) = 0 \quad \text{for any } x \in \mathbb{R},$$

where "i.o." stands for "infinitely often" and means here "for infinitely many n's".

Hint: Set $\rho = \frac{L}{3}$. Begin by observing that, because $\{S_{m+n} - S_m : n \in \mathbb{Z}^+\}$ has the same $\mathbb{P}$-distribution as $\{S_n : n \in \mathbb{Z}^+\}$, $\mathbb{P}(|S_{m+n} - S_m| < 2\rho \text{ i.o.}) = 0$ for any $m \in \mathbb{Z}^+$. Thus, since $|S_{m+n} - x| \geq |S_{m+n} - S_m| - |S_m - x|$, $\mathbb{P}(|S_n - x| < \rho \text{ i.o.}) \leq \mathbb{P}(|S_m - x| \geq \rho)$ for any $m \in \mathbb{Z}^+$. Moreover, by the Hewitt-Savage 0–1 Law, $\mathbb{P}(|S_n - x| < \rho \text{ i.o.}) \in \{0,1\}$. Hence, either $\mathbb{P}(|S_n - x| < \rho \text{ i.o.}) = 0$, or one has the contradiction that $\mathbb{P}(|S_m - x| < \rho) = 0$ for all $m \in \mathbb{Z}^+$ and yet $\mathbb{P}(|S_n - x| < \rho \text{ i.o.}) = 1$.

(**ii**) Still assuming that $L > 0$, argue that

$$\mathbb{P}\big(|S_n - L| < \tfrac{L}{3} \text{ i.o.}\big) \vee \mathbb{P}\big(|S_n + L| < \tfrac{L}{3} \text{ i.o.}\big) = 1,$$

which, in view of (**i**), is a contradiction. Conclude that (5.2.36) holds.

(**iii**) Knowing (5.2.36) and the Hewitt–Savage 0–1 Law, show that, for each $x \in \mathbb{R}$ and $\epsilon > 0$, one has the dichotomy

$$P\big(|S_n - x| < \epsilon\big) = 0 \text{ for all } n \geq 1 \quad \text{or} \quad P\big(|S_n - x| < \epsilon \text{ i.o.}\big) = 1.$$

EXERCISE 5.2.37. Here is a rather frivolous application of reversed martingales. Let $(\Omega, \mathcal{F}, \mathbb{P})$, $\{\mathcal{F}_n : n \in \mathbb{N}\}$, and $\{\mathbf{e}_k : k \in \mathbb{Z}\}$ be as in part (**v**) of Exercise 5.1.17. Next, take $S_m = \{(2k+1)2^m : k \in \mathbb{Z}\}$ for each $m \in \mathbb{N}$, and, for $f \in L^2\big([0,1);\mathbb{C}\big)$, set

$$\Delta_m(f) = \sum_{\ell \in S_m} \big(f, \mathbf{e}_\ell\big)_{L^2([0,1);\mathbb{C})} \mathbf{e}_\ell,$$

where the convergence is in $L^2(([0,1];\mathbb{C})$. Note that, by Exercise 5.1.17,

$$f - \mathbb{E}^{\mathbb{P}}\big[f \,\big|\, \mathcal{F}_{n+1}\big] = \sum_{m=0}^{n} \Delta_m(f).$$

After noting that $\{\mathcal{F}_n : n \in \mathbb{N}\}$ is non-increasing, use the convergence result for reversed martingales in Theorem 5.2.21 to see that the expansion

$$f = \big(f, \mathbf{1}\big)_{L^2([0,1);\mathbb{C})} + \sum_{m=0}^{\infty} \Delta_m(f)$$

converges both almost everywhere as well as in $L^2([0,1);\mathbb{C})$.*

* When f is a function with the property that $(f, e_\ell)_{L^2([0,1);\mathbb{C})} = 0$ for all $\ell \in \mathbb{Z}\backslash\{2^m : m \in \mathbb{N}\}$, the preceding almost everywhere convergence result can be interpreted as saying that the Fourier series of f converges almost everywhere, a result which was discovered originally by Kolmogorov. The proof suggested here is based on fading memories of a conversation with N. Varopolous. Of course, ever since L. Carleson's definitive theorem on the almost every convergence of the Fourier series of an arbitrary square integrable function, the interest in this result of Kolmogorov is mostly historical.