# Chapter VII:
## Conditioning and Martingales

Up to this point we 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 us to make explicit use of the concept of *conditioning*, although, as we will see shortly, this concept has been lurking silently in the background.

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 $\Omega$ 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 $X = a$. Obviously, unless one is very lucky and $\mathbb{P}(X = a) > 0$, dividing by $\mathbb{P}(X = a)$ is not going to do the job. Hence, 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 in this chapter we will present Kolmogorov's elegant solution to this problem.

### § 7.1 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 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 $\Omega$ into elements of $\mathcal{F}$ of 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 we attempt 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

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

Obviously, $f$ is not only $\mathcal{F}$-measurable, it is measurable with respect to the $\sigma$-algebra $\sigma(\mathcal{P})$ over $\Omega$ 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 $A \in \sigma(\mathcal{P})$ and $n$, either $A_n \subseteq A$ or $A \cap A_n = \emptyset$, we have that

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

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

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

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 would 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.

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

$$\mathbb{E}^{\mathbb{P}}\big[X_\Sigma,\,A\big] = \mathbb{E}^{\mathbb{P}}\big[X,\,A\big] \quad \text{for every } A \in \Sigma. \tag{7.1.1}$$

Obviously, having made this definition, our first order of business is to show that such an $X_\Sigma$ always exists and to discover in what sense it is uniquely determined. The latter problem is dealt with by the following lemma.

Lemma 7.1.2. *Let $\Sigma$ be a sub-$\sigma$-algebra of $\mathcal{F}$, and suppose that $X_\Sigma$ and $Y_\Sigma$ are a pair of $(-\infty,\infty]$-valued $\Sigma$-measurable random variables for which $X_\Sigma^-$ and $Y_\Sigma^-$ are both $\mathbb{P}$-integrable. Then*

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

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

Proof: Without loss in generality, we may and will assume that $\Sigma=\mathcal{F}$ and will therefore drop the subscript $\Sigma$; and, since the "if" implication is completely trivial, we will only discuss 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 7.1.3. *Let $\Sigma$ be a sub-$\sigma$-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 $X_\Sigma$ 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 $Y_\Sigma$ which is a conditional expectation value of $Y$ given $\Sigma$. In particular, if $X=Y$* (a.s., $\mathbb{P}$)*, then $\big\{Y_\Sigma\ne X_\Sigma\big\}$ is a $\Sigma$-measurable, $\mathbb{P}$-null set.**

Proof: In view of Lemma 7.1.2, it suffices for us 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 $X_\Sigma$ exists, and let $\mathcal{G}^+$ denote the nonnegative elements of $\mathcal{G}$. If $\{X_n\}_1^\infty\subseteq\mathcal{G}^+$ is nondecreasing and, for each $n\in\mathbb{Z}^+$, $\big(X_n\big)_\Sigma$ denotes a conditional expectation of $X_n$ given $\Sigma$, 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 $X_\Sigma$ are the pointwise limits of the $X_n$'s and $\big(X_n\big)_\Sigma$'s, respectively, then

* 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.

the Monotone Convergence Theorem guarantees that $X_\Sigma$ is a conditional expectation of $X$ given $\Sigma$. Hence, we now know that $\mathcal{G}^+$ is closed under nondecreasing, pointwise limits, and therefore we will know that $\mathcal{G}^+$ contains all nonnegative 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 $L^2(P;\mathbb{R})$) and $\mathbf{L}_\Sigma = L^2(\Omega,\Sigma,\mathbb{P};\mathbb{R})$ is the subspace of $L^2(P;\mathbb{R})$ consisting of its $\Sigma$-measurable elements, then the orthogonal projection $X_\Sigma$ of $X$ onto $\mathbf{L}_\Sigma$ is a $\Sigma$-measurable random variable which is square $\mathbb{P}$-integrable and satisfies (7.1.1).

So far we have proved that $\mathcal{G}^+$ contains all nonnegative, $\mathcal{F}$-measurable $X$'s. Furthermore, if $X$ is nonnegative, then (by Lemma 7.1.2) $X_\Sigma \geq 0$ (a.s., $\mathbb{P}$) and so $X_\Sigma$ is $\mathbb{P}$-integrable precisely when $X$ itself is. In particular, we can arrange to make $X_\Sigma$ take its values in $[0,\infty)$ when $X$ is nonnegative 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 $\Sigma$. $\square$

**Convention.** Because it is determined only up to a $\Sigma$-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 $\Sigma$-measurable $X_\Sigma$'s which satisfy (7.1.1), and we will adopt the notation $\mathbb{E}^{\mathbb{P}}\big[X\big|\Sigma\big]$ 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 $\mathbb{E}^{\mathbb{P}}\big[X\big|\Sigma\big]$ and the members of that equivalence class. Thus (just as one would when dealing with the Lebesgue spaces) we will abuse notation by using $\mathbb{E}^{\mathbb{P}}\big[X\big|\Sigma\big]$ to denote a generic element of the equivalence class $\mathbb{E}^{\mathbb{P}}\big[X\big|\Sigma\big]$, and will be more precise only when $\mathbb{E}^{\mathbb{P}}\big[X\big|\Sigma\big]$ contains some particularly distinguished member. For example, recall the random variables $\mathbf{T}_n$ entering the definition of the simple Poisson process $\big\{N(t) : t \in (0,\infty)\big\}$ in § 4.2.1, then, it is clear that we can take

$$\mathbb{E}^{\mathbb{P}}\Big[\mathbf{1}_{\{n\}}\big(\mathbf{N}(t)\big)\,\Big|\,\sigma\big(\mathbf{T}_1,\dots,\mathbf{T}_n\big)\Big] = \mathbf{1}_{[0,t]}\big(\mathbf{T}_n\big)e^{-(t-\mathbf{T}_n)},$$

and we would be foolish to take any other representative. More generally, we will always take nonnegative representatives of $\mathbb{E}^{\mathbb{P}}\big[X\big|\Sigma\big]$ when $X$ itself is nonnegative 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 $\mathbb{E}^{\mathbb{P}}\big[\mathbf{1}_B\big|\Sigma\big]$ the **conditional probability of** $B$ **given** $\Sigma$ and to write $\mathbb{P}(B|\Sigma)$ instead of $\mathbb{E}^{\mathbb{P}}\big[\mathbf{1}_B\big|\Sigma\big]$. 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 7.1.4. *Let $\Sigma$ be a sub-$\sigma$-algebra of $\mathcal{F}$. If $X$ is a $\mathbb{P}$-integrable random variable and $\mathcal{C} \subseteq \Sigma$ is a $\pi$-system (cf. Exercise 1.1.10) which generates $\Sigma$, then*

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

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

$$\big|\mathbb{E}^{\mathbb{P}}\big[X\big|\Sigma\big]\big| \le \mathbb{E}^{\mathbb{P}}\big[|X|\big|\Sigma\big]; \tag{7.1.5}$$

$$\mathbb{E}^{\mathbb{P}}\big[X\big|\mathcal{T}\big] = \mathbb{E}^{\mathbb{P}}\Big[\mathbb{E}^{\mathbb{P}}\big[X\big|\Sigma\big]\,\Big|\mathcal{T}\Big] \tag{7.1.6}$$

*when $\mathcal{T}$ is a sub-$\sigma$-algebra of $\Sigma$; 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 $\mathbb{E}^{\mathbb{P}}\big[Y^-\big] < \infty$. Then, $\mathbb{P}$-almost surely:*

$$\mathbb{E}^{\mathbb{P}}\big[\alpha X + \beta Y\big|\Sigma\big] = \alpha\mathbb{E}^{\mathbb{P}}\big[X\big|\Sigma\big] + \beta\mathbb{E}^{\mathbb{P}}\big[Y\big|\Sigma\big]$$

*for each $\alpha$, $\beta \in [0, \infty)$, and*

$$\mathbb{E}^{\mathbb{P}}\big[YX\big|\Sigma\big] = Y\,\mathbb{E}^{\mathbb{P}}\big[X\big|\Sigma\big] \tag{7.1.7}$$

*if $Y$ is $\Sigma$-measurable and $(XY)^-$ is $\mathbb{P}$-integrable. Finally, suppose that $\big\{X_n\big\}_1^\infty$ is a sequence of $(-\infty, \infty]$-valued random variables. Then, $\mathbb{P}$-almost surely:*

$$\mathbb{E}^{\mathbb{P}}\big[X_n\big|\Sigma\big] \nearrow \mathbb{E}^{\mathbb{P}}\big[X\big|\Sigma\big] \tag{7.1.8}$$

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

$$\mathbb{E}^{\mathbb{P}}\left[\varliminf_{n\to\infty} X_n\middle|\,\Sigma\right] \le \varliminf_{n\to\infty} \mathbb{E}^{\mathbb{P}}\big[X_n\big|\Sigma\big]. \tag{7.1.9}$$

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

Proof: To prove the 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 1.1.10) a $\lambda$-system which contains $\mathcal{C}$ and therefore $\Sigma$. Next, clearly (7.1.5) is just an application of Lemma 7.1.2, while (7.1.6) 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 $\Sigma$, use linearity to extend it to simple $\Sigma$-measurable functions, and complete the job by taking monotone limits. Finally, (7.1.8) is an immediate application of the Monotone Convergence Theorem; whereas (7.1.9) comes from the conjunction of

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

with (7.1.8). $\square$

§ **7.1.2. Regular Conditional Probability Distributions.** It probably will have occurred to most readers that the properties discussed in Theorem 7.1.4 give strong evidence that, for fixed $\omega \in \Omega$, $X \longmapsto \mathbb{E}^{\mathbb{P}}\big[X\big|\Sigma\big](\omega)$ 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 $X \longmapsto \mathbb{E}^{\mathbb{P}}\big[X\big|\Sigma\big](\omega)$ 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 $X \longmapsto \mathbb{E}^{\mathbb{P}}\big[X\big|\Sigma\big](\omega)$ for fixed $\omega \in \Omega$. One answer to this question is contained in the following theorem.

Theorem 7.1.10. *Suppose that $\Omega$ is a Polish space and that $\mathcal{F} = \mathcal{B}_\Omega$. Then, for every sub-$\sigma$-algebra $\Sigma$ of $\mathcal{F}$, there is a $\mathbb{P}$-almost surely unique $\Sigma$-measurable map $\omega \in \Omega \longmapsto P^\Sigma_\omega \in \mathbf{M}_1(\Omega)$ with the property that*

$$\mathbb{P}\big(A \cap B\big) = \int_A P^\Sigma_\omega(B)\,\mathbb{P}(d\omega) \quad \text{for all } A \in \Sigma \text{ and } B \in \mathcal{F}.$$

*In particular, for each $(-\infty,\infty]$-valued random variable $X$ which is bounded below,*

$$\omega \in \Omega \longmapsto \mathbb{E}^{\mathbb{P}^\Sigma_\omega}\big[X\big]$$

*is a conditional expectation value of $X$ given $\Sigma$. Finally, if $\Sigma$ is countably generated, then there is a $\mathbb{P}$-null set $\mathcal{N} \in \Sigma$ with the property that*

$$\mathbb{P}^\Sigma_\omega(A) = \mathbf{1}_A(\omega) \quad \text{for all } \omega \notin \mathcal{N} \text{ and } A \in \Sigma.$$

Proof: To prove the uniqueness, suppose $\omega \in \Omega \longmapsto Q^\Sigma_\omega \in \mathbf{M}_1(\Omega)$ were a second such mapping. We would then know that, for each $B \in \mathcal{F}$, $\mathbb{Q}^\Sigma_\omega(B) =$

$P_\omega^\Sigma(B)$ for $\mathbb{P}$-almost every $\omega \in \Omega$. Hence, since $\mathcal{F}$ (as the Borel field over a second countable topological space) is countably generated, we could find one $\Sigma$-measurable $\mathbb{P}$-null set off of which $\mathbb{Q}_\omega^\Sigma = P_\omega^\Sigma$. Similarly, to prove the final assertion when $\Sigma$ is countably generated, note (cf. (7.1.7)) that, for each $A \in \Sigma$, $\mathbb{P}_\omega^\Sigma(A) = \mathbf{1}_A(\omega) = \delta_\omega(A)$ for $\mathbb{P}$-almost every $\omega \in \Omega$. Thus, once again countability allows us to choose one $\Sigma$-measurable $\mathbb{P}$-null set $\mathcal{N}$ such that $\mathbb{P}_\omega^\Sigma \upharpoonright \Sigma = \delta_\omega \upharpoonright \Sigma$ if $\omega \notin \mathcal{N}$.

We next turn to the question of existence. For this purpose, first choose (cf. (**ii**) of Lemma 6.1.3) $\rho$ to be a totally bounded metric for $\Omega$, and let $\mathcal{U} = U_{\mathrm{b}}^\rho(\Omega;\mathbb{R})$ be the space of bounded, $\rho$-uniformly continuous, $\mathbb{R}$-valued functions on $\Omega$. Then (cf. (**iii**) of Lemma 6.1.3) $\mathcal{U}$ is a separable Banach space with respect to the uniform norm. In particular, we can choose a sequence $\{f_n\}_0^\infty \subseteq \mathcal{U}$ so that $f_0 = \mathbf{1}$, the functions $f_0, \dots, f_n$ are linearly independent for each $n \in \mathbb{Z}^+$, and the linear span $\mathcal{S}$ of $\{f_n : n \in \mathbb{N}\}$ is dense in $\mathcal{U}$. Set $g_0 = \mathbf{1}$, and, for each $n \in \mathbb{Z}^+$, let $g_n$ be some fixed representative of $\mathbb{E}^{\mathbb{P}}\left[f_n \middle| \Sigma\right]$. Next, set

$$\mathfrak{R} = \left\{\boldsymbol{\alpha} \in \mathbb{R}^{\mathbb{N}} : \exists m \in \mathbb{N}\ \alpha_n = 0 \text{ for all } n \geq m\right\}$$

and define

$$f_{\boldsymbol{\alpha}} = \sum_{n=0}^\infty \alpha_n f_n \quad \text{and} \quad g_{\boldsymbol{\alpha}} = \sum_{n=0}^\infty \alpha_n g_n$$

for $\boldsymbol{\alpha} \in \mathfrak{R}$. Because of the linear independence of the $f_n$'s, we know that $f_{\boldsymbol{\alpha}} = f_{\boldsymbol{\beta}}$ if and only if $\boldsymbol{\alpha} = \boldsymbol{\beta}$. Hence, for each $\omega \in \Omega$, we can define the (not necessarily continuous) linear functional $\Lambda_\omega : \mathcal{S} \longrightarrow \mathbb{R}$ so that

$$\Lambda_\omega\bigl(f_{\boldsymbol{\alpha}}\bigr) = g_{\boldsymbol{\alpha}}(\omega), \qquad \boldsymbol{\alpha} \in \mathfrak{R}.$$

Clearly, $\Lambda_\omega(\mathbf{1}) = \mathbf{1}$ for all $\omega \in \Omega$. On the other hand, we cannot say that $\Lambda_\omega$ is always nonnegative as a linear functional on $\mathcal{S}$. In fact, the best we can do is extract a $\Sigma$-measurable $\mathbb{P}$-null set $\mathcal{N}$ so that $\Lambda_\omega$ is a nonnegative linear functional on $\mathcal{S}$ whenever $\omega \notin \mathcal{N}$. To this end, let $\mathbb{Q}$ denote the rational reals and set

$$\mathfrak{Q}^+ = \left\{\boldsymbol{\alpha} \in \mathfrak{R} \cap \mathbb{Q}^{\mathbb{N}} : f_{\boldsymbol{\alpha}} \geq 0\right\}.$$

Since $g_{\boldsymbol{\alpha}} \geq 0$ (a.s., $\mathbb{P}$) for every $\boldsymbol{\alpha} \in \mathfrak{Q}^+$ and $\mathfrak{Q}^+$ is countable,

$$\mathcal{N} \equiv \left\{\omega \in \Omega : \exists \boldsymbol{\alpha} \in \mathfrak{Q}^+ \quad g_{\boldsymbol{\alpha}}(\omega) < 0\right\}$$

is a $\Sigma$-measurable, $\mathbb{P}$-null set. In addition, it is obvious that, for every $\omega \notin \mathcal{N}$, $\Lambda_\omega(f) \geq 0$ whenever $f$ is a nonnegative element of $\mathcal{S}$. In particular, for $\omega \notin \mathcal{N}$,

$$\|f\|_{\mathrm{u}} \pm \Lambda_\omega(f) = \Lambda_\omega\bigl(\|f\|_{\mathrm{u}}\,\mathbf{1} \pm f\bigr) \geq 0, \quad f \in \mathcal{S},$$

and therefore $\Lambda_\omega$ admits a unique extension as a nonnegative, linear functional on $\mathcal{U}$ which takes $\mathbf{1}$ to 1. Furthermore, it is an easy matter to check that, for every $f \in \mathcal{U}$, the function

$$g(\omega) = \begin{cases} \Lambda_\omega(f) & \text{for} \quad \omega \notin \mathcal{N} \\ \mathbb{E}^{\mathbb{P}}[f] & \text{for} \quad \omega \in \mathcal{N} \end{cases}$$

is a conditional expectation value of $f$ given $\Sigma$.

At this point, all that remains is to show that, for $\mathbb{P}$-almost every $\omega \notin \mathcal{N}$, $\Lambda_\omega$ is given by integration with respect to a $\mathbb{P}_\omega \in \mathbf{M}_1(\Omega)$. In particular, by the Riesz Representation Theorem, there is nothing more to do in the case when $\Omega$ is compact. To treat the case when $\Omega$ is not compact, we want to use Lemma 6.1.6. For this purpose, first choose (cf. the last part of Lemma 6.1.6) a nondecreasing sequence of sets $K_n \subset\subset \Omega$, $n \in \mathbb{Z}^+$, with the property that $\mathbb{P}\big(K_n\complement\big) \le \frac{1}{2^n}$. Next, define

$$\eta_{m,n}(\omega) = \frac{m\,\rho(\omega, K_n)}{1 + m\,\rho(\omega, K_n)} \quad \text{for } m,\, n \in \mathbb{Z}^+.$$

Clearly, $\eta_{m,n} \in \mathcal{U}$ for each pair $(m, n)$ and $0 \le \eta_{m,n} \nearrow \mathbf{1}_{K_n\complement}$ as $m \to \infty$ for each $n \in \mathbb{Z}^+$. Thus, by the Monotone Convergence Theorem, for each $n \in \mathbb{Z}^+$,

$$\begin{aligned}\int_{\mathcal{N}\complement} \sup_{m \in \mathbb{Z}^+} \Lambda_\omega\big(\eta_{m,n}\big)\,\mathbb{P}(d\omega) &= \lim_{m\to\infty} \int_{\mathcal{N}\complement} \Lambda_\omega\big(\eta_{m,n}\big)\,\mathbb{P}(d\omega) \\ &= \lim_{m\to\infty} \mathbb{E}^{\mathbb{P}}\big[\eta_{m,n}\big] \le \frac{1}{2^n};\end{aligned}$$

and so, by the Borel–Cantelli Lemma, we can find a $\Sigma$-measurable $\mathbb{P}$-null set $\mathcal{N}' \supseteq \mathcal{N}$ such that

$$M(\omega) \equiv \sup_{n \in \mathbb{Z}^+} n \left( \sup_{m \in \mathbb{Z}^+} \Lambda_\omega\big(\eta_{m,n}\big) \right) < \infty \quad \text{for every } \omega \notin \mathcal{N}'.$$

Hence, if $\omega \notin \mathcal{N}'$, then, for every $f \in \mathcal{U}$ and $n \in \mathbb{Z}^+$,

$$\begin{aligned}\big|\Lambda_\omega(f)\big| &\le \big|\Lambda_\omega\big((1 - \eta_{m,n})\,f\big)\big| + \big|\Lambda_\omega\big(\eta_{m,n}\,f\big)\big| \\ &\le \big\|(1 - \eta_{m,n})\,f\big\|_{\mathrm{u}} + \frac{M(\omega)}{n}\,\|f\|_{\mathrm{u}}\end{aligned}$$

for all $m \in \mathbb{Z}^+$. But $\big\|(1 - \eta_{m,n})\,f\big\|_{\mathrm{u}} \longrightarrow \|f\|_{\mathrm{u},K_n}$ as $m \to \infty$, and so we now see that the condition in (6.1.7) is satisfied by $\Lambda_\omega$ for every $\omega \notin \mathcal{N}'$. In other words, we have shown that, for each $\omega \notin \mathcal{N}'$, there is a unique $\mathbb{P}^\Sigma_\omega \in \mathbf{M}_1(\Omega)$ such that $\Lambda_\omega(f) = \mathbb{E}^{\mathbb{P}^\Sigma_\omega}\big[f\big]$ for all $f \in \mathcal{U}$. Finally, if we complete the definition of the map $\omega \in \Omega \longmapsto P^\Sigma_\omega$ by taking $\mathbb{P}^\Sigma_\omega = P$ for $\omega \in \mathcal{N}'$, then this map is $\Sigma$-measurable and

$$\mathbb{E}^{\mathbb{P}}\big[f,\, A\big] = \int_\Omega \mathbb{E}^{\mathbb{P}^\Sigma_\omega}\big[f\big]\,\mathbb{P}(d\omega), \quad A \in \Sigma,$$

first for all $f \in \mathcal{U}$ and thence for all bounded $\mathcal{F}$-measurable $f$'s. $\square$

Given a measurable space $(\Omega, \mathcal{F})$ and a sub-$\sigma$-algebra $\Sigma$, a **$\Sigma$-measurable transition probability** is a map $(\omega, B) \in \Omega \times \mathcal{F} \longmapsto \mathbb{P}(\omega, B) \in [0,1]$ with the properties that $B \in \mathcal{F} \longmapsto \mathbb{P}(\omega, B) \in [0,1]$ is a probability measure for each $\omega \in \Omega$ and $\omega \in \Omega \longmapsto \mathbb{P}(\omega, B) \in [0,1]$ is $\Sigma$-measurable for each $B \in \mathcal{F}$. In particular, if $\mathbb{P}$ is a probability measure on $(\Omega, \mathcal{F})$, then a **conditional probability distribution of $\mathbb{P}$ given $\Sigma$** is a $\Sigma$-measurable transition $(\omega, B) \longmapsto P^{\Sigma}_{\omega}(B)$ probability for which $\omega \rightsquigarrow \mathbb{P}^{\Sigma}(B)$ is a conditional probability of $B$ given $\Sigma$. If, in addition, for $\omega$ outside a $\Sigma$-measurable, $\mathbb{P}$-null set and all $A \in \Sigma$, $\mathbb{P}^{\Sigma}(A) = \mathbf{1}_A(\omega)$, then the conditional probability distribution is said to be **regular**. Notice that, although they may not always exist, conditional probability distributions are always unique up to a $\Sigma$-measurable $\mathbb{P}$-null set so long as $\mathcal{F}$ is countably generated. Moreover, Theorem 7.1.10 says that they will always exist if $\Omega$ is Polish and $\mathcal{F} = \mathcal{B}_{\Omega}$. Finally, whenever a conditional probability distribution of $\mathbb{P}$ given $\Sigma$ exists, the argument leading to the last part of Theorem 7.1.10 when $\Sigma$ is countably generated is completely general and shows that a regular version can be found.

**§ 7.1.3. Some Extensions.** For various applications, it is convenient to have two extensions of the basic theory developed above. Specifically, as we 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, $\mu$ will be an arbitrary (nonnegative) measure on $(\Omega, \mathcal{F})$ and $\big(E, \|\cdot\|_E\big)$ will be a separable Banach space; and we begin by reviewing a few elementary facts about $\mu$-integration for $E$-valued random variables.

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

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

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

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

(we use $E^*$ to denote the dual of $E$ and $\big\langle \mathbf{x}, \lambda \big\rangle$ to denote the action of $\lambda \in E^*$ on $\mathbf{x} \in E$), and therefore that the mapping taking $\mu$-simple $\mathbf{X}$ to $\mathbb{E}^{\mu}[\mathbf{X}]$ is linear. Next, because $E$ is separable and therefore there exists a sequence $\{\lambda_n\}_1^{\infty}$ from the unit ball in $E^*$ with the property that

$$\sup_{n \in \mathbb{Z}^+} \big\langle \mathbf{x}, \lambda_n \big\rangle = \|\mathbf{x}\|_E \quad \text{for all} \quad \mathbf{x} \in E, \tag{7.1.11}$$

we know that $\omega \in \Omega \longmapsto \|\mathbf{X}(\omega)\|_E \in \mathbb{R}$ is $\mathcal{F}$-measurable if $\mathbf{X} : \Omega \longrightarrow E$ is $\mathcal{F}$-measurable. In particular, for $\mathcal{F}$-measurable $\mathbf{X} : \Omega \longrightarrow E$, we can set

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

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

The definition of $\mu$-integration for $E$-valued $\mathbf{X}$ is completed in the following lemma.

LEMMA 7.1.12. *For each $\mu$-integrable $\mathbf{X} : \Omega \longrightarrow E$ there is a unique element $\mathbb{E}^\mu[\mathbf{X}] \in E$ for which $\big\langle \mathbb{E}^{\mathbb{P}}[\mathbf{X}], \lambda\big\rangle = \mathbb{E}^{\mathbb{P}}\big[\langle \mathbf{X}, \lambda\rangle\big]$ for all $\lambda \in E^*$. In particular, the mapping $\mathbf{X} \in L^1(\mu; E) \longmapsto \mathbb{E}^\mu[\mathbf{X}] \in E$ is linear and satisfies*

$$\big\|\mathbb{E}^\mu[\mathbf{X}]\big\|_E \le \mathbb{E}^\mu\Big[\|\mathbf{X}\|_E\Big]. \tag{7.1.13}$$

*Finally, if $\mathbf{X} \in L^p(\mu; E)$ where $p \in [1,\infty)$, then there exists a sequence $\{\mathbf{X}_n\}_1^\infty$ of $E$-valued, $\mu$-simple functions with the property that $\|\mathbf{X}_n - \mathbf{X}\|_{L^p(\mu;E)} \longrightarrow 0$.*

PROOF: Clearly uniqueness, linearity, and (7.1.13) all follow immediately from the given characterization of $\mathbb{E}^\mu[\mathbf{X}]$. 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 (7.1.13), $\mathbb{E}^\mu[\mathbf{X}]$ can be taken equal to $\lim_{n\to\infty} \mathbb{E}^\mu[\mathbf{X}_n]$ if $\|\mathbf{X} - \mathbf{X}_n\|_{L^1(\mu;E)} \longrightarrow 0$.

To prove the approximation assertion, we begin with the case when $\mu$ is finite and $M = \sup_{\omega\in\Omega} \|\mathbf{X}(\omega)\|_E < \infty$. Next, choose a dense sequence $\{\mathbf{x}_\ell\}_1^\infty$ in $E$, set $A_{0,n} = \emptyset$, and, for $\ell \in \mathbb{Z}^+$,

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

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 $\mathbf{X}_n : \Omega \longrightarrow E$ is defined so that

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

and $\mathbf{X}_n(\omega) = \mathbf{0}$ when $\omega \notin \bigcup_1^{L_n} A_{\ell,n}$, then $\mathbf{X}_n$ is $\mu$-simple and

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

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

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

where

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

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

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

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

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

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

all to denote the quantity $\mathbb{E}^\mu\big[\mathbf{1}_B\, \mathbf{X}\big]$. Also, when discussing the spaces $L^p(\mu; E)$, we will adopt the usual convention of blurring the distinction between a particular $\mathcal{F}$-measurable $\mathbf{X} : \Omega \longrightarrow E$ belonging to $L^p(\mu; E)$ and the equivalence class of those $\mathcal{F}$-measurable $\mathbf{Y}$'s which differ from $\mathbf{X}$ on a $\mu$-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 Banach space.

Theorem 7.1.14. *Let $(\Omega, \mathcal{F}, \mu)$ be a $\sigma$-finite measure space and $\mathbf{X} : \Omega \longrightarrow E$ a $\mu$-locally integrable function. Then*

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

*Next, assume that $\Sigma$ is a sub-$\sigma$-algebra for which $\mu \restriction \Sigma$ is $\sigma$-finite. Then for each $\mu$-locally integrable $\mathbf{X} : \Omega \longrightarrow E$ there is a $\mu$-almost everywhere unique $\mu$-locally integrable, $\Sigma$-measurable $\mathbf{X}_\Sigma : \Omega \longrightarrow E$ such that*

$$\mathbb{E}^\mu\big[\mathbf{X}_\Sigma,\, A\big] = \mathbb{E}^\mu\big[\mathbf{X},\, A\big] \quad \text{for every } A \in \Sigma \text{ with } \mu(A) < \infty. \tag{7.1.15}$$

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

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

*Finally,*

$$\big\|\mathbf{X}_\Sigma\big\|_E \le \big(\|\mathbf{X}\|_E\big)_\Sigma \quad \text{(a.e., } \mu\text{)}. \tag{7.1.16}$$

*Hence, not only does (7.1.15) continue to hold for any* $A \in \Sigma$ *with* $\mathbf{1}_A\mathbf{X} \in L^1(\mu; E)$*; but also, for each* $p \in [1,\infty]$*, the mapping* $\mathbf{X} \in L^p(\mu; E) \longmapsto \mathbf{X}_\Sigma \in L^p(\mu; E)$ *is a linear contraction.*

Proof: Clearly, it is only necessary to prove the "$\Longleftarrow$" part of the first assertion. Thus, suppose that $\mu(\mathbf{X} \neq \mathbf{0}) > 0$. Then there exists an $\epsilon > 0$ and a $\lambda \in E^*$ with the property that $\mu\big(\langle \mathbf{X}, \lambda\rangle \ge \epsilon\big) > 0$; from which it follows (by $\sigma$-finiteness) that there is an $A \in \mathcal{F}$ for which $\mu(A) < \infty$ and

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

We turn next to the uniqueness and other properties of $\mathbf{X}_\Sigma$. But it is obvious that uniqueness is an immediate consequence of the first assertion and that linearity follows from uniqueness. As for (7.1.16), notice that if $\lambda \in E^*$ and $\|\lambda\|_{E^*} \le 1$, then

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

for every $A \in \Sigma$ with $\mu(A) < \infty$. Hence, at least when $\mu$ is a probability measure, Theorem 7.1.3 implies that $\big\langle \mathbf{X}_\Sigma, \lambda\big\rangle \le \big(\|\mathbf{X}\|_E\big)_\Sigma$ (a.e., $\mu$) for each element $\lambda$ from the unit ball in $E^*$; and so, because $E$ is separable, (7.1.16) follows in this case from (7.1.11). To handle $\mu$'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 $\mu$ to make it a probability measure, or $\mu(\Omega) = \infty$, in which case we can use the $\sigma$-finiteness of $\mu \restriction \Sigma$ to reduce ourselves to the countable, disjoint union of the preceding cases.

Finally, to prove the existence of $\mathbf{X}_\Sigma$, we proceed as in the last part of the preceding paragraph to reduce ourselves to the case when $\mu$ is a probability measure $\mathbb{P}$. Next, suppose that $\mathbf{X}$ is $\mathbb{P}$-simple, let $R$ denote its range, and note that

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

has the required properties. In order to handle general $\mathbf{X} \in L^1(P; E)$, we use the approximation result in Lemma 7.1.12 to find a sequence $\{\mathbf{X}_n\}_1^\infty$ of $\mathbb{P}$-simple functions which tend to $\mathbf{X}$ in $L^1(P; E)$. Then, since

$$(\mathbf{X}_n)_\Sigma - (\mathbf{X}_m)_\Sigma = \big(\mathbf{X}_n - \mathbf{X}_m\big)_\Sigma \quad (\text{a.s.}, \mathbb{P})$$

and therefore, by (7.1.16),

$$\left\|(\mathbf{X}_n)_\Sigma - (\mathbf{X}_m)_\Sigma\right\|_{L^1(P;E)} \le \left\|\mathbf{X}_n - \mathbf{X}_m\right\|_{L^1(P;E)},$$

we know that there exists a $\Sigma$-measurable $\mathbf{X}_\Sigma \in L^1(P;E)$ to which the sequence $\left\{(\mathbf{X}_n)_\Sigma\right\}_1^\infty$ converges; and clearly $\mathbf{X}_\Sigma$ has the required properties. □

Referring to the setting in the second part of Theorem 7.1.14, we will extend the convention introduced following Theorem 7.1.3 and call the $\mu$-equivalence class of $\mathbf{X}_\Sigma$'s satisfying (7.1.15) the **$\mu$-conditional expectation of $\mathbf{X}$ given** $\Sigma$, will use $\mathbb{E}^\mu[\mathbf{X}|\Sigma]$ to denote this $\mu$-equivalence class, and will, in general, ignore the distinction between the equivalence class and a generic representative of that class. In addition, if $\mathbf{X} : \Omega \longrightarrow E$ is $\mu$-locally integrable, then, just as in Theorem 7.1.4, the following are essentially immediate consequences of uniqueness:

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

and

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

whenever $\mathcal{T}$ is a sub-$\sigma$-algebra of $\Sigma$ for which $\mu \restriction \mathcal{T}$ is $\sigma$-finite.

## Exercises for § 7.1

EXERCISE 7.1.17. As the proof of existence in Theorem 7.1.4 makes clear, the operation $X \in L^2(P;\mathbb{R}) \longmapsto \mathbb{E}^{\mathbb{P}}\left[X\middle|\Sigma\right]$ is just the operation of orthogonal projection from $L^2(P;\mathbb{R})$ onto the space $L^2(\Omega,\Sigma,\mathbb{P};\mathbb{R})$ of $\Sigma$-measurable elements of $L^2(P;\mathbb{R})$. For this reason, one might be inclined to think that the concept of conditional expectation is basically a Hilbert space notion. However, as we will show in this exercise, 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 $\mathbf{L}$ be a closed linear subspace of $L^2(\mathbb{P};\mathbb{R})$ and let $\Sigma_{\mathbf{L}} = \sigma\big(X : X \in \mathbf{L}\big)$ be the $\sigma$-algebra over $\Omega$ generated by $X \in \mathbf{L}$. Show that $\mathbf{L} = L^2\big(\Omega,\Sigma_{\mathbf{L}},\mathbb{P};\mathbb{R}\big)$ if and only if $\mathbf{1} \in \mathbf{L}$ and $X^+ \in \mathbf{L}$ whenever $X \in \mathbf{L}$.

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

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

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

(**ii**) Let $\Pi$ be an orthogonal projection operator on $L^2(\mathbb{P};\mathbb{R})$, set $\mathbf{L} = \text{Range}(\Pi)$, and let $\Sigma = \Sigma_{\mathbf{L}}$, where $\Sigma_{\mathbf{L}}$ is defined as in part (**i**). Show that $\Pi X = \mathbb{E}^{\mathbb{P}}\left[X\middle|\Sigma\right]$ (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 $\big(\Pi X\big)^n \in \mathbf{L}$, $n \in \mathbb{Z}^+$, for every $X \in L^\infty(\mathbb{P};\mathbb{R})$. Next, using the preceding together with Weierstrass's Approximation Theorem, show that $(\Pi X)^+ \in \mathbf{L}$, 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 $\mathbf{L} = L^2\big(\Omega,\Sigma,\mathbb{P};\mathbb{R}\big)$.

(**iii**) Just in case the situation is not completely clarified by part (**ii**), consider once again a closed linear subspace $\mathbf{L}$ of $L^2(\mathbb{P};\mathbb{R})$ and let $\Pi_{\mathbf{L}}$ be the orthogonal projection mapping onto $\mathbf{L}$. Given $X \in L^2(\mathbb{P};\mathbb{R})$, recall that $\Pi_{\mathbf{L}}X$ is characterized as the unique element of $\mathbf{L}$ for which $X - \Pi_{\mathbf{L}}X \perp \mathbf{L}$, and show that $\mathbb{E}^{\mathbb{P}}\big[X\big|\Sigma_{\mathbf{L}}\big]$ is the unique element of $L^2(\Omega,\Sigma_{\mathbf{L}},\mathbb{P};\mathbb{R})$ 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 $Y_1,\dots,Y_n \in \mathbf{L}$. In particular, $\Pi_{\mathbf{L}}X = \mathbb{E}^{\mathbb{P}}\big[X\big|\Sigma_{\mathbf{L}}\big]$ if and only if $X - \Pi_{\mathbf{L}}X$ is perpendicular not only to all *linear* functions of the $Y$'s in $\mathbf{L}$ but even to all *nonlinear* ones.

Exercise 7.1.18. There is an important, non-trivial situation in which conditioning and projecting turn out to be the same thing. Namely, let $\mathfrak{G} \subseteq L^2(\mathbb{P};\mathbb{R})$ be a centered Gaussian family (cf. Lemma 5.1.1), and show that for any closed linear subspace $\mathbf{L}$ of $\mathfrak{G}$ and any $X \in \mathfrak{G}$, $\Pi_{\mathbf{L}}X = \mathbb{E}^{\mathbb{P}}\big[X\big|\Sigma_{\mathbf{L}}\big]$.

To give an example of the preceding, consider an $\mathbb{R}^N$-valued Brownian motion $\{\mathbf{B}(t) : t \ge 0\}$ on $(\Omega,\mathcal{F},\mathbb{P})$, and, for $T \in (0,\infty)$, set

$$\mathbf{B}_T(t,\omega) = \mathbf{B}(t,\omega) - L_T\big(t,\mathbf{B}(T,\omega)\big) \quad \text{where } L_T(\mathbf{y}) \equiv \frac{t\wedge T}{T}\mathbf{y},$$

and define $\mu_{T,\mathbf{y}} \in \mathbf{M}_1\big(C(\mathbb{R}^N)\big)$ to be the distribution of $\omega \rightsquigarrow \mathbf{B}_T(\,\cdot\,,\omega) + L_T(\,\cdot\,,\mathbf{y})$ under $\mathbb{P}$. Remembering that $\mathfrak{G}(\mathbf{B}) = \operatorname{span}\big(\{\mathbf{B}(t) : t \ge 0\}\big)$ is a Gaussian family, show that

$$\mathbb{P}\Big(\mathbf{B}(\,\cdot\,) \in \tau \,\Big|\, \sigma\big(\mathbf{B}(T)\big)\Big) = \mu_{T,\mathbf{B}(T)}(\Gamma), \quad \Gamma \in \mathcal{B}_{C(\mathbb{R}^N)}.$$

Equivalently, if $\mathcal{W}^{(N)} \in \mathbf{M}_1\big(C(\mathbb{R}^N)\big)$ is the distribuiton of $\omega \rightsquigarrow \mathbf{B}(\,\cdot\,,\omega)$ under $\mathbb{P}$, then $\omega \rightsquigarrow \mu_{T,\mathbf{B}(T,\omega)}$ is a regular conditional probabilility of $\mathcal{W}^{(N)}$ given $\sigma\big(\mathbf{B}(T)\big)$. In particular, $\mu_{T,\mathbf{0}}$ is the distribution of Brownian motion conditioned to be back at $\mathbf{0}$ at time $T$, and for this reason the stochastic process $\{\mathbf{B}_T(t) : t \ge 0\}$ is said to be a **pinned Brownian motion**

EXERCISE 7.1.19. 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's measure to $[0,1)$. Next, for $n \in \mathbb{N}$, take $\mathcal{F}_n$ to be the $\sigma$-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 $L^2([0,1];\mathbb{C})$. (See part (**iv**) of Exercise (?) and Exercise (?) for a continuation of this exercise.)

EXERCISE 7.1.20. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $\Sigma$ a sub-$\sigma$-algebra of $\mathcal{F}$ with the property that $\mu \restriction \Sigma$ is $\sigma$-finite. Next, let $E$ be a separable Hilbert space, $p \in [1,\infty]$, $\mathbf{X} \in L^p(\mu; E)$, and $\mathbf{Y}$ a $\Sigma$-measurable element of $L^{p'}(\mu; E)$ ($p'$ is the Hölder conjugate of $p$). Show that

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

**Hint**: First observe that it suffices to check that

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

Next, choose an orthonormal basis $\{\mathbf{e}_n\}_1^{\infty}$ for $E$ and justify the steps in

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

## § 7.2: Discrete Parameter Martingales

In this section we 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 we will be doing is developing an abstract version of differentiation theory (cf. Corollary 7.2.18 and Theorem 7.2.19).

Although we will want to make some extensions later (cf. § 7.3), we start with the following setting. $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and $\{\mathcal{F}_n : n \in \mathbb{N}\}$ is a nondecreasing sequence of sub-$\sigma$-algebra's of $\mathcal{F}$. Given a measurable space $(E, \mathcal{B})$, we 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 and

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

for each $n \in \mathbb{N}$; and 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

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

for each $n \in \mathbb{N}$. In the future, we will abbreviate these statements by saying that the triple $\big(X_n, \mathcal{F}_n, \mathbb{P}\big)$ is a submartingale or martingale.

**Examples.** The most trivial example of a submartingale is provided by a nondecreasing sequence $\{a_n\}_1^\infty$. That is, if $X_n \equiv a_n$, $n \in \mathbb{Z}^+$, then $\big(X_n, \mathcal{F}_n, \mathbb{P}\big)$ is a submartingale on any probability space $\big(\Omega, \mathcal{F}, \mathbb{P}\big)$ relative to any nondecreasing $\{\mathcal{F}_n : n \in \mathbb{N}\}$. More interesting examples are those given below.*

(**i**) Let $\{Y_n\}_1^\infty$ be a sequence of mutually independent $(-\infty, \infty]$-valued random variables with $\mathbb{E}^{\mathbb{P}}\big[Y_n^-\big] < \infty$, $n \in \mathbb{N}$, set $\mathcal{F}_0 = \{\emptyset, \Omega\}$, $\mathcal{F}_n = \sigma\big(Y_1, \dots, Y_n\big)$ for $n \in \mathbb{Z}^+$, and define $X_n = \sum_{1 \le m \le n}^n Y_m$, where summation over the empty set is taken to be 0. Then, because $\mathbb{E}^{\mathbb{P}}\big[Y_{n+1}\big|\mathcal{F}_n\big] = \mathbb{E}^{\mathbb{P}}\big[Y_{n+1}\big]$ (a.s., $\mathbb{P}$) and therefore

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

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 $\mathbb{E}^{\mathbb{P}}\big[Y_n\big] \ge 0$ 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 $\mathbb{E}^{\mathbb{P}}\big[Y_n\big] = 0$ for all $n \in \mathbb{Z}^+$. Finally, if $\{Y_n\}_1^\infty \subseteq L^2(\mathbb{P})$ and $\mathbb{E}^{\mathbb{P}}[Y_n] = 0$ for each $n \in \mathbb{Z}^+$, then

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

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

---

* 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.

(**ii**) If $X$ is an $\mathbb{R}$-valued, $\mathbb{P}$-integrable random variable and $\{\mathcal{F}_n : n \in \mathbb{N}\}$ is any nondecreasing sequence of sub-$\sigma$-algebras of $\mathcal{F}$, then, by (7.1.6),

$$\left(\mathbb{E}^{\mathbb{P}}\left[X \middle| \mathcal{F}_n\right], \mathcal{F}_n, \mathbb{P}\right)$$

is a martingale.

(**iii**) If $(X_n, \mathcal{F}_n, \mathbb{P})$ is a martingale, then, by (7.1.5), $(|X_n|, \mathcal{F}_n, \mathbb{P})$ is a submartingale.

**§ 7.2:.1. Doob's Inequality and Marcinkewitz's Theorem.** In view of (**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 7.2.1 **Doob's Inequality.** *Assume that $(X_n, \mathcal{F}_n, \mathbb{P})$ is a submartingale. Then, for every $N \in \mathbb{Z}^+$ and $\alpha \in (0, \infty)$:*

$$\mathbb{P}\left(\max_{0 \le n \le N} X_n \ge \alpha\right) \le \frac{1}{\alpha} \mathbb{E}^{\mathbb{P}}\left[X_N,\ \max_{0 \le n \le N} X_n \ge \alpha\right]. \tag{7.2.2}$$

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

$$\mathbb{E}^{\mathbb{P}}\left[\sup_{n \in \mathbb{N}} X_n^p\right]^{\frac{1}{p}} \le \frac{p}{p-1} \sup_{n \in \mathbb{N}} \mathbb{E}^{\mathbb{P}}\left[X_n^p\right]^{\frac{1}{p}}. \tag{7.2.3}$$

PROOF: To prove (7.2.2), set $A_0 = \{X_0 \ge \alpha\}$ 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,

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

Now assume that the $X_n$'s are nonnegative. Given (7.2.2), (7.2.3) becomes an easy application of Exercise 1.4.17. $\square$

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. (See Theorem (?) for more about the relationship between weak and strong type inequalities.) In order to demonstrate how powerful such a result can be, we will now apply Doob's Inequality to prove a theorem of Marcinkewitz. Because it is an argument to which we will return several times, 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 7.2.4. *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 nondecreasing sequence $\{\mathcal{F}_n : n \in \mathbb{N}\}$ of sub-$\sigma$-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 $\mathbb{E}^{\mathbb{P}}\big[X\big|\mathcal{F}_n\big] \longrightarrow X$ (a.s., $\mathbb{P}$) and in $L^p(\mathbb{P};\mathbb{R})$.*

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

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

$$\mathbb{P}\left(\sup_{n\in\mathbb{N}} |X_n| \geq \alpha\right) \leq \frac{1}{\alpha}\mathbb{E}^{\mathbb{P}}\left[|X|, \sup_{n\in\mathbb{N}} |X_n| \geq \alpha\right], \quad \alpha \in (0,\infty); \tag{7.2.5}$$

and, since, by (7.1.5), $|X_n| \leq \mathbb{E}^{\mathbb{P}}\big[|X|\,\big|\mathcal{F}_n\big]$ (a.s., $\mathbb{P}$), while proving (7.2.5) we may and will assume that $X$ and all the $X_n$'s are nonnegative. But then, by (7.2.2),

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

for all $N \in \mathbb{Z}^+$; and therefore (7.2.5) follows when $N \to \infty$ and one takes right limits in $\alpha$.

As our first application of (7.2.5), we note that $\{X_n\}_0^\infty$ is uniformly $\mathbb{P}$-integrable. Indeed, because $|X_n| \le \mathbb{E}^{\mathbb{P}}\big[|X|\,\big|\mathcal{F}_n\big]$, we have from (7.2.5) that

$$\sup_{n\in\mathbb{N}} \mathbb{E}^{\mathbb{P}}\Big[|X_n|,\,|X_n| \ge \alpha\Big] \le \sup_{n\in\mathbb{N}} \mathbb{E}^{\mathbb{P}}\Big[|X|,\,|X_n| \ge \alpha\Big] \le \mathbb{E}^{\mathbb{P}}\left[|X|,\,\sup_{n\in\mathbb{N}}|X_n| \ge \alpha\right] \longrightarrow 0$$

as $\alpha \to \infty$. Thus, we will know that the asserted convergence in 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 (7.2.3) and Exercise 1.4.17, 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 $\big\{X^{(k)}\big\}_1^\infty \subseteq \mathcal{G}$ and $X^{(k)} \longrightarrow X$ in $L^1(\mathbb{P};\mathbb{R})$, then, by (7.2.5),

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

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

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

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

Before moving on to more sophisticated convergence results, we will spend a little time showing that Corollary 7.2.4 is already interesting. In order to introduce our main application, recall our preliminary discussion of conditioning when we were attempting to explain Kolmogorov's idea at the beginning of this chapter. As we said there, the most easily understood situation occurs when one conditions with respect to a sub-$\sigma$-algebra $\Sigma$ which is generated by a countable partition $\mathcal{P}$. Indeed, in that case one can easily verify that

$$\mathbb{E}^{\mathbb{P}}\big[X\big|\Sigma\big] = \sum_{A\in\mathcal{P}} \frac{\mathbb{E}^{\mathbb{P}}\big[X,\, A\big]}{\mathbb{P}(A)}\,\mathbf{1}_A, \tag{7.2.6}$$

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, $\Sigma$ need not be (cf. Exercise 1.1.16). Furthermore, just because $\Sigma$ is countably generated, it will be seldom true that its generators can be chosen to form a countable partition. (For example, as soon as $\Sigma$ contains an uncountable number of atoms, such a partition cannot exist.) Nonetheless, if $\Sigma$ is any countably generated $\sigma$-algebra, then we can find a sequence $\{\mathcal{P}_n\}_0^\infty$ 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 $\{A_n\}_0^\infty$ for $\Sigma$ 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 $0 \le m \le n$.

Theorem 7.2.7. *Let $\Sigma$ be a countably generated sub-$\sigma$-algebra of $\mathcal{F}$ and choose $\{\mathcal{P}_n\}_0^\infty$ 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 (7.2.6) with $\mathcal{P} = \mathcal{P}_n$. Then $X_n \longrightarrow \mathbb{E}^{\mathbb{P}}\big[X\big|\Sigma\big]$ both $\mathbb{P}$-almost surely and in $L^p(\mathbb{P};\mathbb{R})$. In particular, even if $\Sigma$ is not countably generated, for each separable, closed subspace $\mathbf{L}$ of $L^p(\mathbb{P};\mathbb{R})$ there exists a sequence of finite partitions $\mathcal{P}_n$, $n \in \mathbb{N}$, such that*

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

*for every $X \in \mathbf{L}$.*

Proof: To prove the first part, simply set $\mathcal{F}_n = \sigma\big(\mathcal{P}_n\big)$, then identify $X_n$ as $\mathbb{E}^{\mathbb{P}}\big[X\big|\mathcal{F}_n\big]$, and finally apply Corollary 7.2.4. As for the second part, let $\Sigma(\mathbf{L})$ be the $\sigma$-algebra generated by

$$\Big\{\mathbb{E}^{\mathbb{P}}\big[X\big|\Sigma\big] : X \in \mathbf{L}\Big\},$$

note that $\Sigma(\mathbf{L})$ is countably generated and that

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

and apply the first part with $\Sigma$ replaced by $\Sigma(\mathbf{L})$. $\square$

Theorem 7.2.7 makes it easy to transfer the usual Jensen's Inequality to conditional expectations.

Corollary 7.2.8 **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 $\Sigma$ a sub-$\sigma$-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*

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

Proof: By the classical Jensen's Inequality, $Y \equiv g(\mathbf{X})$ is $\mathbb{P}$-integrable. Hence, by the second part of Theorem 7.2.7, we can find finite partitions $\mathcal{P}_n$, $n \in \mathbb{N}$, so that

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

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,

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

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

$$Y(\omega) \equiv \begin{cases} \lim_{n\to\infty} Y_n(\omega) & \text{if} \quad \omega \in \Lambda \\ 0 & \text{if} \quad \omega \notin \Lambda, \end{cases}$$

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

COROLLARY 7.2.9. *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 $f : I \longrightarrow \mathbb{R} \cup \{+\infty\}$,*

$$f\Big(\mathbb{E}^{\mathbb{P}}\big[X\big|\Sigma\big]\Big) \le \mathbb{E}^{\mathbb{P}}\big[f(X)\big|\Sigma\big] \quad (\text{a.s.}, \mathbb{P}) \tag{7.2.10}$$

*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 $p \in [1, \infty)$,*

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

*Finally, if $\big(X_n, \mathcal{F}_n, \mathbb{P}\big)$ is an $I$-valued martingale and either $f$ is as above or if $\big(X_n, \mathcal{F}_n, P\big)$ is an $I$-valued submartingale and $f$ is bounded below and nondecreasing (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 7.2.8, we know that an $I$-valued representative of $\mathbb{E}^{\mathbb{P}}\big[X\big|\Sigma\big]$ exists when $X$ is $\mathbb{P}$-integrable, and the general case follows after a trivial truncation procedure. In order to prove (7.2.10), first assume that $f$ is bounded above by some $M < \infty$ and that $X \in L^1(\mathbb{P})$. Then (7.2.10) is an immediate consequence of the last part of Corollary 7.2.8 with $g = M - f$. To handle the case when $f$ is bounded below and to the left, first observe that either $f$ is nonincreasing everywhere, or there is an $a \in I \cap \mathbb{R}$ with the property that $f$ is nonincreasing to the left of $a$ and nondecreasing 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

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

Moreover, in the case when $f$ is nonincreasing, $\big\{f(X_n) : n \ge m\big\}$ is bounded and nonincreasing; and, in the other case, $\big\{f(X_n) : n \ge m \vee a\big\}$ is bounded below and nondecreasing. Hence, in both cases, (7.2.10) follows from the preceding after an application of the version of the Monotone Convergence Theorem in (7.1.8).

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

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

$\mathbb{P}$-almost surely. $\square$

**§ 7.2:.2. Doob's Stopping Time Theorem.** Perhaps the most subtle 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 facilitate the proofs of these results, we begin with an easy but seminal observation of Doob's.

LEMMA 7.2.11 **Doob's Decomposition.** *Let $(X_n, \mathcal{F}_n, \mathbb{P})$ be a $\mathbb{P}$-integrable submartingale. Then there is a $\mathbb{P}$-almost surely unique sequence $\{I_n : n \in \mathbb{N}\}$ of random variables with the properties that $I_0 \equiv 0$,*

$$I_{n-1} \le I_n \in L^1\big(\Omega, \mathcal{F}_{n-1}, \mathbb{P}; [0,\infty)\big) \quad \text{for each } n \in \mathbb{Z}^+,$$

*and $(X_n - I_n, \mathcal{F}_n, \mathbb{P})$ is a martingale.*

PROOF: To prove the existence, set $I_0 \equiv 0$ and

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

To prove the uniqueness, suppose that $\{J_n : n \in \mathbb{N}\}$ is a second such sequence and set $\Delta_n = I_n - J_n$, $n \in \mathbb{N}$. Then $\Delta_0 \equiv 0$, $\Delta_n$ is $\mathcal{F}_{n-1}$-measurable for each $n \in \mathbb{Z}^+$, and $(\Delta_n, \mathcal{F}_n, \mathbb{P})$ is a martingale. Hence

$$\Delta_n = \mathbb{E}^{\mathbb{P}}\big[\Delta_n\big|\mathcal{F}_{n-1}\big] = \Delta_{n-1} \quad (\text{a.s.}, \mathbb{P}) \quad \text{for each} \quad n \in \mathbb{Z}^+,$$

and so, by induction, $\Delta_n = 0$ (a.s., $\mathbb{P}$) for all $n \in \mathbb{N}$. $\square$

In order to state the next result, we need to introduce the notion of a stopping time in this setting. Namely, we will say that the function $\tau : \Omega \longrightarrow \mathbb{N} \cup \{\infty\}$ is an $\{\mathcal{F}_n : n \in \mathbb{N}\}$**-stopping time** if $\{\omega : \tau(\omega) = n\} \in \mathcal{F}_n$ for each $n \in \mathbb{N}$. In addition, given an $\{\mathcal{F}_n : n \in \mathbb{N}\}$-stopping time $\tau$, we use $\mathcal{F}_\tau$ to denotethe $\sigma$-algebra of $A \in \mathcal{F}$ such that $A \cap \{\tau = n\} \in \mathcal{F}_n$, $n \in \mathbb{Z}^+$. Notice that $\mathcal{F}_\sigma \subseteq \mathcal{F}_\tau$ if $\sigma \le \tau$. In addition, if $\{X_n : n \in \mathbb{N}\}$ is $\{\mathcal{F}_n : n \in \mathbb{N}\}$-progressively measurable, check that the random variable $X_\tau$ given by $X_\tau(\omega) = X_{\tau(\omega)}(\omega)$ is $\mathcal{F}_\tau$-measurable on $\{\tau < \infty\}$.

THEOREM 7.2.12 (**Hunt**). *Let $(X_n, \mathcal{F}_n, \mathbb{P})$ be a $\mathbb{P}$-integrable submartingale. Given bounded $\{\mathcal{F}_n : n \in \mathbb{N}\}$-stopping times $\sigma$ and $\tau$ satisfying $\sigma \le \tau$,*

$$X_\sigma \le \mathbb{E}^{\mathbb{P}}\big[X_\tau\big|\mathcal{F}_\sigma\big] \quad (\text{a.s.}, \mathbb{P}), \tag{7.2.13}$$

*and the inequality can be replaced by equality when $(X_n, \mathcal{F}_n, \mathbb{P})$ is a martingale. (Cf. Exercise 7.2.23) for unbounded stopping times.)*

PROOF: Choose $\{I_n : n \in \mathbb{N}\}$ as in Lemma 7.2.11, and set $Y_n = X_n - I_n$, $n \in \mathbb{N}$. Then, because $I_\sigma \le I_\tau$ and $I_\sigma$ is $\mathcal{F}_\sigma$-measurable:

$$\mathbb{E}^{\mathbb{P}}\big[X_\tau \big| \mathcal{F}_\sigma\big] \ge \mathbb{E}^{\mathbb{P}}\big[Y_\tau + I_\sigma \big| \mathcal{F}_\sigma\big] = \mathbb{E}^{\mathbb{P}}\big[Y_\tau \big| \mathcal{F}_\sigma\big] + I_\sigma,$$

and so it suffices to prove that equality holds in (7.2.13) when $(X_n, \mathcal{F}_n, \mathbb{P})$ is a martingale. To this end, choose $N \in \mathbb{Z}^+$ to be an upper bound for $\tau$, let $A \in \mathcal{F}_\sigma$ be given, and note that

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

and similarly (since $A \in \mathcal{F}_\sigma \subseteq \mathcal{F}_\tau$), $\mathbb{E}^{\mathbb{P}}\big[X_N,\, A\big] = \mathbb{E}^{\mathbb{P}}\big[X_\tau,\, A\big]$. □

The following easy consequence of Hunt's Theorem becomes more interesting when it is interpreted from the point of view of a gambler who is trying to *beat the system*. Namely, let $(X_n, \mathcal{F}_n, \mathbb{P})$ be a martingale with mean-value 0, and think of $X_n$ as the gambler's fortune after $n$ plays of a *fair game*. With this model in mind, it is natural to interpret an $\{\mathcal{F}_n : n \in \mathbb{N}\}$-stopping time $\tau$ as a *strategy*. That is, $\tau$ can be considered as a feasible (i.e., one which does not require the power of prophesy) scheme which the gambler can use to determine when he should stop playing. When couched in these terms, the next result predicts that *there is no strategy with which the gambler can alter his expected take!*

COROLLARY 7.2.14 (**Doob's Stopping Time Theorem**). *For any $\mathbb{P}$-integrable submartingale (martingale) $(X_n, \mathcal{F}_n, \mathbb{P})$ and any $\{\mathcal{F}_n : n \in \mathbb{N}\}$-stopping time $\tau$, $(X_{n\wedge\tau}, \mathcal{F}_n, P)$ is again a $\mathbb{P}$-integrable submartingale (martingale).*

PROOF: Let $A \in \mathcal{F}_{n-1}$. Then, since $A \cap \{\tau > n-1\} \in \mathcal{F}_{(n-1)\wedge\tau}$, (7.2.13) implies that

$$\begin{aligned}&\mathbb{E}^{\mathbb{P}}\big[X_{n\wedge\tau},\, A\big] \\ &\quad= \mathbb{E}^{\mathbb{P}}\big[X_{n\wedge\tau},\, A \cap \{\tau \le n-1\}\big] + \mathbb{E}^{\mathbb{P}}\big[X_{n\wedge\tau},\, A \cap \{\tau > n-1\}\big] \\ &\quad\ge \mathbb{E}^{\mathbb{P}}\big[X_\tau,\, A \cap \{\tau \le n-1\}\big] + \mathbb{E}^{\mathbb{P}}\big[X_{(n-1)\wedge\tau},\, A \cap \{\tau > n-1\}\big] \\ &\quad= \mathbb{E}^{\mathbb{P}}\big[X_{(n-1)\wedge\tau},\, A\big];\end{aligned}$$

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

**§ 7.2:.3. Martingale Convergence Theorem.** Our next goal is to show that, even when they are not given in the form covered by Corollary 7.2.4, *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 7.2.4. We will follow Doob and base our 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 nondecreasing process. Although the first of these observations will take a little time to develop, the second one is nearly trivial and is covered by the following simple lemma.

Given a sequence $\{x_n\}_0^\infty \subseteq \mathbb{R}$ and $-\infty < a < b < \infty$, we say that $\{x_n\}_0^\infty$ **upcrosses the interval** $[a,b]$ **at least** $N$ **times** if there exist integers $0 \le m_1 < n_1 < \cdots < m_N < n_N$ such that $x_{m_i} \le a$ and $x_{n_i} \ge b$ for each $1 \le i \le N$ 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 $\varliminf_{n\to\infty} x_n < \varlimsup_{n\to\infty} x_n$ if and only if there exist rational numbers $a < b$ such that $\{x_n\}$ upcrosses $[a,b]$ at least $N$ times for every $N \in \mathbb{Z}^+$.

Theorem 7.2.15 (**Doob's Martingale Convergence Theorem**). * *Let $\big(X_n, \mathcal{F}_n, \mathbb{P}\big)$ be a $\mathbb{P}$-integrable submartingale, and, for $-\infty < a < b < \infty$, let $U_{[a,b]}(\omega)$ denote the precise number of times that $\big\{X_n(\omega)\big\}_0^\infty$ upcrosses $[a,b]$. Then*

$$\mathbb{E}^{\mathbb{P}}\big[U_{[a,b]}\big] \le \sup_{n\in\mathbb{N}} \frac{\mathbb{E}^{\mathbb{P}}\big[(X_n - a)^+\big]}{b-a}. \tag{7.2.16}$$

*In particular, if*

$$\sup_{n\in\mathbb{N}} \mathbb{E}^{\mathbb{P}}\big[X_n^+\big] < \infty, \tag{7.2.17}$$

*then there exists a $\mathbb{P}$-integrable random variable $X$ to which $\{X_n\}_1^\infty$ converges $\mathbb{P}$-almost surely. (See Exercises (?) and (?) for other derivations.)*

Proof: Set $Y_n = \frac{(X_n-a)^+}{b-a}$ and note that (by Corollary 7.2.9) $\big(Y_n, \mathcal{F}_n, \mathbb{P}\big)$ is a $\mathbb{P}$-integrable submartingale. Next, set $\tau_0 = 0$, and, for $k \in \mathbb{Z}^+$, define

$$\sigma_k = \inf\big\{n \ge \tau_{k-1} : X_n \le a\big\} \quad \text{and} \quad \tau_k = \inf\big\{n \ge \sigma_k : X_n \ge b\big\}.$$

Proceeding by induction, it is an easy matter to check that all the $\sigma_k$'s and $\tau_k$'s are $\big\{\mathcal{F}_n : n \in \mathbb{N}\big\}$-stopping times. Moreover, if $N \in \mathbb{Z}^+$ and $U_{[a,b]}^{(N)}(\omega)$ is the

* 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.

precise number of times $\{X_{n\wedge N}(\omega)\}_0^\infty$ upcrosses $[a,b]$, then

$$\begin{aligned} U_{[a,b]}^{(N)} &\le \sum_{k=1}^N \bigl(Y_{N\wedge\tau_k} - Y_{N\wedge\sigma_k}\bigr) = Y_N - Y_0 - \sum_{k=1}^N \bigl(Y_{N\wedge\sigma_k} - Y_{N\wedge\tau_{k-1}}\bigr) \\ &\le Y_N - \sum_{k=1}^N \bigl(Y_{N\wedge\sigma_k} - Y_{N\wedge\tau_{k-1}}\bigr). \end{aligned}$$

Hence, since $\tau_{k-1} \le \sigma_k$ and therefore, by (7.2.13),

$$\mathbb{E}^{\mathbb{P}}\bigl[Y_{N\wedge\sigma_k} - Y_{N\wedge\tau_{k-1}}\bigr] \ge 0 \quad \text{for all} \quad k \in \mathbb{Z}^+,$$

we see that $\mathbb{E}^{\mathbb{P}}\bigl[U_{[a,b]}^{(N)}\bigr] \le \mathbb{E}^{\mathbb{P}}\bigl[Y_N\bigr]$; and, clearly (7.2.16) follows from this after one lets $N \to \infty$.

Given (7.2.16), the convergence result is easy. Namely, if (7.2.17) is satisfied, then (7.2.16) implies that there is a set $\Lambda$ 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 $\omega \in \Lambda$, $\{X_n(\omega)\}_0^\infty$ converges to some $X(\omega) \in [-\infty,\infty]$. Hence, we will be done as soon as we show that $\mathbb{E}^{\mathbb{P}}\bigl[|X|,\,\Lambda\bigr] < \infty$. But

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

and therefore Fatou's Lemma plus (7.2.17) shows that $X$ is $\mathbb{P}$-integrable on $\Lambda$. $\square$

The inequality in (7.2.16) is quite famous and is known as **Doob's upcrossing inequality**.

Corollary 7.2.18. *Let $\bigl(X_n, \mathcal{F}_n, \mathbb{P}\bigr)$ be a martingale. Then there exists an $X \in L^1(\mathbb{P};\mathbb{R})$ such that $X_n = \mathbb{E}^{\mathbb{P}}\bigl[X\big|\mathcal{F}_n\bigr]$ (a.s., $\mathbb{P}$) for each $n \in \mathbb{N}$ if and only if the sequence $\{X_n\}_0^\infty$ 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 $X_n = \mathbb{E}^{\mathbb{P}}\bigl[X\big|\mathcal{F}_n\bigr]$ (a.s., $\mathbb{P}$) for each $n \in \mathbb{N}$ if and only if $\{X_n\}_0^\infty$ is a bounded subset of $L^p(\mathbb{P};\mathbb{R})$.*

Proof: Because of Corollary 7.2.4 and (7.2.3), we need only check the "if" statement in the first assertion. But, if $\{X_n\}_0^\infty$ is uniformly $\mathbb{P}$-integrable, then (7.2.17) 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 (7.1.5), for each $m \in \mathbb{N}$,

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

Just as Corollary 7.2.4 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 7.2.19 (**Jessen**). *Let $\mathbb{P}$ and $\mathbb{Q}$ be a pair of probability measures on the measurable space $(\Omega, \mathcal{F})$ and $\{\mathcal{F}_n : n \in \mathbb{N}\}$ a nondecreasing sequence of sub-$\sigma$-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 $\mathbb{Q}_\mathrm{a}$ be the absolutely continuous part 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 $\{X_n\}_0^\infty$ 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 $G \equiv \{\lim_{n\to\infty} X_n \in (0,\infty)\}$, then $\mathbb{Q}_\mathrm{a}(A) = Q(A \cap G)$ 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, we 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.

We 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, $(-X_n, \mathcal{F}_n, \mathbb{P})$ is a nonpositive submartingale. Moreover, in the case when $\mathbb{Q}_n \ll P_n$, $n \in \mathbb{N}$, the same argument shows that $(X_n, \mathcal{F}_n, \mathbb{P})$ is a nonnegative martingale. Thus, in any case, there is a nonnegative, $\mathbb{P}$-integrable random variable $X$ with the property that $X_n \longrightarrow X$ (a.s., $\mathbb{P}$). In order to identify $X$ as $Y$, we 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 $\mathbb{E}^{\mathbb{P}}\big[X, A\big] \leq \mathbb{Q}(A)$ for every $A \in \mathcal{F}$. In particular, by choosing $B \in \mathcal{F}$ so that $\mathbb{Q}_\mathrm{a}(A) = \mathbb{Q}(A \cap B)$, $A \in \mathcal{F}$, and $\mathbb{P}(B\complement) = 0$, we arrive at

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

which means that $X \leq Y$ (a.s., $\mathbb{P}$). On the other hand, if $Y_n = \mathbb{E}^{\mathbb{P}}\big[Y \big| \mathcal{F}_n\big]$ 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 $Y_n \le X_n$ (a.s., $\mathbb{P}$) for each $n \in \mathbb{N}$. But, since $Y_n \longrightarrow Y$ and $X_n \longrightarrow X$ $\mathbb{P}$-almost surely, this means that $Y \le X$ (a.s., $\mathbb{P}$).

Next, assume that $\mathbb{Q}_n \ll P_n$ for each $n \in \mathbb{N}$ and therefore that $\bigl(X_n, \mathcal{F}_n, \mathbb{P}\bigr)$ is a nonnegative martingale. If $\{X_n\}_0^\infty$ 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, $\mathbb{Q} \ll P$ when $\{X_n\}_0^\infty$ is uniformly $\mathbb{P}$-integrable. Conversely, if $\mathbb{Q} \ll P$, then it is easy to see that $X_n = \mathbb{E}^{\mathbb{P}}\bigl[Y \big| \mathcal{F}_n\bigr]$ for each $n \in \mathbb{N}$, and therefore (by Corollary 7.2.4) that $\{X_n\}_0^\infty$ is uniformly $\mathbb{P}$-integrable.

Finally, assume that $\mathbb{Q}_n \sim P_n$ for each $n \in \mathbb{N}$. Then, the $X_n$'s can be chosen to take their values in $(0,\infty)$ and $Y_n \equiv \frac{1}{X_n} = \frac{dP_n}{d\mathbb{Q}_n}$. 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 $Y = \frac{dP_{\mathrm{a}}}{d\mathbb{Q}}$ and so $\mathbb{P}_{\mathrm{a}}(A) = \mathbb{E}^{\mathbb{Q}}\bigl[Y,\, A\bigr]$ for all $A \in \mathcal{F}$. Thus, when $C \in \mathcal{F}$ is chosen so that $\mathbb{P}_{\mathrm{s}}(C) = 0 = \mathbb{Q}(C\complement)$, then, since $Y = \frac{1}{X}$ (a.s., $\mathbb{Q}$) on $G$, it is becomes clear that

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

for all $A \in \mathcal{F}$. $\square$

**Exercises for § 7.2**

Exercise 7.2.20. In this exercise we will present 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. (7.2.2)), the result is nearly trivial for martingales having two bounded moments and, second, that everything can be reduced to that case.

(**i**) Let $\bigl(M_n, \mathcal{F}_n, \mathbb{P}\bigr)$ be a martingale for which

$$
\sup_{n\in\mathbb{N}} \mathbb{E}^{\mathbb{P}}\bigl[M_n^2\bigr] < \infty.
$$

Note that

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

and starting from these, 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, by applying (7.2.5) to the submartingale $\Bigl(\bigl|M_{n\vee m} - M_m\bigr|, \mathcal{F}_n, \mathbb{P}\Bigr)$, show that, for every $\epsilon > 0$,

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

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

(**ii**) Let $(X_n, \mathcal{F}_n, \mathbb{P})$ be a nonnegative submartingale with the property that $\sup_{n\in\mathbb{N}} \mathbb{E}^{\mathbb{P}}\big[X_n^2\big] < \infty$, define the sequence $\{I_n : n \in \mathbb{N}\}$ as in Lemma 7.2.11, and set $M_n = X_n - I_n$, $n \in \mathbb{N}$. Then $(M_n, \mathcal{F}_n, \mathbb{P})$ is a martingale, and clearly both $M_n$ and $I_n$ are square $\mathbb{P}$-integrable for each $n \in \mathbb{N}$. In fact, check that

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

and therefore that

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

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

(**iii**) Let $(X_n, \mathcal{F}_n, \mathbb{P})$ be a nonnegative martingale, set $Y_n = e^{-X_n}$, $n \in \mathbb{N}$, use Corollary 7.2.9 to see that $(Y_n, \mathcal{F}_n, \mathbb{P})$ is a uniformly bounded, nonnegative, submartingale, and apply part (**ii**) to conclude that $\{X_n\}_0^\infty$ converges $\mathbb{P}$-almost surely to a nonnegative $X \in L^1(\mathbb{P};\mathbb{R})$.

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

$$\sup_{n\in\mathbb{N}} \mathbb{E}^{\mathbb{P}}\big[|X_n|\big] < \infty. \tag{7.2.21}$$

For each $m \in \mathbb{N}$, define

$$Y_{n,m}^{\pm} = \mathbb{E}^{\mathbb{P}}\Big[X_{n\vee m}^{\pm}\Big|\mathcal{F}_m\Big] \vee 0, \quad n \in \mathbb{N}.$$

Show that $Y_{n+1,m}^{\pm} \ge Y_{n,m}^{\pm}$ (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 nonnegative martingales with

$$\mathbb{E}^{\mathbb{P}}\big[Y_0^+ + Y_0^-\big] \le \sup_{n\in\mathbb{N}} \mathbb{E}^{\mathbb{P}}\big[|X_n|\big],$$

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* (*7.2.21*) *admits a* **Hahn decomposition*** *as the difference of two nonnegative martingales whose sum has expectation value dominated by the left-hand side of* (7.2.21). Finally, use this observation together with (**iii**) to see that every such martingale converges $\mathbb{P}$-almost surely to some $X \in L^1(\mathbb{P};\mathbb{R})$.

* This useful observation was made by Klaus Krickeberg.

(**v**) By combining the final assertion in (**iv**) together with Doob's Decomposition in Lemma 7.2.11, give another proof of the convergence assertion in Theorem 7.2.15.

EXERCISE 7.2.22. In this exercise we will develop yet 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 $\{I_n : n \in \mathbb{N}\}$ for $(-Z_n, \mathcal{F}_n, \mathbb{P})$ as in Lemma 7.2.11, note that $I_n$'s can be chosen so that $0 \le I_n - I_{n-1} \le R$ for all $n \in \mathbb{Z}^+$, and set $M_n = Z_n + I_n$, $n \in \mathbb{N}$. Check that $(M_n, \mathcal{F}_n, \mathbb{P})$ is a nonnegative martingale with $M_n \le (n+1)R$ for each $n \in \mathbb{N}$. Next, show that

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

and conclude that

$$\mathbb{E}^{\mathbb{P}}\big[I_n^2\big] \le \mathbb{E}^{\mathbb{P}}\big[M_n^2\big] \le 3R\mathbb{E}^{\mathbb{P}}\big[Z_0\big], \quad n \in \mathbb{N}.$$

(**ii**) Let $(X_n, \mathcal{F}_n, \mathbb{P})$ be a nonnegative martingale. Show that, for each $R \in (0, \infty)$,

$$X_n = M_n^{(R)} - I_n^{(R)} + \Delta_n^{(R)}, \quad n \in \mathbb{N},$$

where $(M_n^{(R)}, \mathcal{F}_n, \mathbb{P})$ is a nonnegative martingale satisfying

$$\mathbb{E}^{\mathbb{P}}\Big[\big(M_n^{(R)}\big)^2\Big] \le 3R\,\mathbb{E}^{\mathbb{P}}\big[X_0\big], \quad n \in \mathbb{N},$$

$\{I_n^{(R)} : n \in \mathbb{N}\}$ is a nondecreasing sequence of random variables with the properties that $I_0^{(R)} \equiv 0$, $I_n^{(R)}$ is $\mathcal{F}_{n-1}$-measurable and

$$\mathbb{E}^{\mathbb{P}}\Big[\big(I_n^{(R)}\big)^2\Big] \le 3R\,\mathbb{E}^{\mathbb{P}}\big[X_0\big] \quad \text{for } n \in \mathbb{Z}^+,$$

and $\{\Delta_n^{(R)} : n \in \mathbb{N}\}$ is a $\{\mathcal{F}_n : n \in \mathbb{N}\}$-progressively measurable sequence with the property that

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

**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 $\Delta_n^{(R)} \ne 0$ for some $n \in \mathbb{N}$.

(**iii**) Let $\left(X_n, \mathcal{F}_n, \mathbb{P}\right)$ be any martingale. Show that, for each $R \in (0,\infty)$,

$$X_n = M_n^{(R)} + V_n^{(R)} + \Delta_n^{(R)}, \quad n \in \mathbb{N},$$

where $\left(M_n^{(R)}, \mathcal{F}_n, \mathbb{P}\right)$ is a martingale satisfying

$$\mathbb{E}^{\mathbb{P}}\left[\left(M_n^{(R)}\right)^2\right] \le 12\, R\mathbb{E}^{\mathbb{P}}\left[|X_n|\right],$$

$\left\{V_n^{(R)} : n \in \mathbb{N}\right\}$ is a sequence of random variables satisfying

$$V_0^{(R)} \equiv 0 \quad \text{and} \quad V_n^{(R)} \text{ is } \mathcal{F}_{n-1}\text{-measurable}$$

and

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

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

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

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

**Hint**: Use part (**iv**) of Exercise 7.2.20 to reduce the present case to the one just treated in (**ii**).

(**iv**) Let $\left(X_n, \mathcal{F}_n, \mathbb{P}\right)$ martingale which satisfies (7.2.21), and use part (**iii**) above together with part (**i**) of Exercise 7.2.20 to show that, for each $R \in (0,\infty)$, $\left\{X_n\right\}_0^\infty$ converges off of a set whose $\mathbb{P}$-measure is no more than $\frac{1}{R}$ times the supremum over $n \in \mathbb{N}$ of $\mathbb{E}^{\mathbb{P}}\left[|X_n|\right]$. In particular, when combined with Lemma 7.2.11, the preceding line of reasoning leads to yet another proof of the convergence result in Theorem 7.2.15.

Exercise 7.2.23. In this exercise we want to extend Hunt's Theorem (cf. Theorem 7.2.12) to allow for unbounded stopping times. To this end, let $\big(X_n, \mathcal{F}_n, P\big)$ be a uniformly $\mathbb{P}$-integrable submartingale on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and set $M_n = X_n - I_n,\ n \in \mathbb{N}$, where $\left\{I_n : n \in \mathbb{N}\right\}$ is the sequence discussed in Lemma 7.2.11. After checking that $\left(M_n, \mathcal{F}_n, \mathbb{P}\right)$ is a uniformly $\mathbb{P}$-integrable martingale, show that, for any $\left\{\mathcal{F}_n : n \in \mathbb{N}\right\}$-stopping time $\tau$:

$$X_\tau = \mathbb{E}^{\mathbb{P}}\left[M_\infty \middle| \mathcal{F}_\tau\right] + I_\tau \quad (\text{a.s.}, \mathbb{P}),$$

where, $X_\infty$, $M_\infty$, and $I_\infty$ are, respectively, the $\mathbb{P}$-almost sure limits of $\{X_n\}_0^\infty$, $\{M_n\}_0^\infty$, and $\{I_n\}_0^\infty$. In particular, if $\sigma$ and $\tau$ are a pair of $\{\mathcal{F}_n : n \in \mathbb{N}\}$-stopping times and $\sigma \le \tau$, conclude that

$$X_\sigma \le \mathbb{E}^{\mathbb{P}}\big[X_\tau \big| \mathcal{F}_\sigma\big] \quad (\text{a.s.}, \mathbb{P}).$$

Exercise 7.2.24. 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 $\rho(R) \ge R$ for all $R$ and $X_{n+1} \le \rho\big(X_n\big)$ (a.e., $\mathbb{P}$) for each $n \in \mathbb{N}$.

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

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

In particular, if $X_0$ is $\mathbb{P}$-integrable, show that $\big\{X_n(\omega)\big\}_0^\infty$ converges in $\mathbb{R}$ for $\mathbb{P}$-almost every $\omega$ for which $\big\{X_n(\omega)\big\}_0^\infty$ is bounded above.

**Hint**: After observing that

$$\sup_{n\in\mathbb{N}} \mathbb{E}^{\mathbb{P}}\big[X^+_{n\wedge\tau_R}\big] < \infty \quad \text{for every} \quad R \in (0,\infty),$$

conclude that, for each $R \in (0,\infty)$, $\big\{X_n\big\}_0^\infty$ converges $\mathbb{P}$-almost everywhere on $\{\tau_R = \infty\}$.

(**ii**) Let $\big\{Y_n\big\}_1^\infty$ be a sequence of mutually independent, $\mathbb{P}$-integrable random variables, assume that

$$\mathbb{E}^{\mathbb{P}}\big[Y_n\big] \ge 0 \text{ for } n \in \mathbb{N} \quad \text{and} \quad \sup_{n\in\mathbb{N}} \|Y_n^+\|_{L^\infty(\mathbb{P})} < \infty,$$

and set $S_n = \sum_1^n Y_m$. Show that $\big\{S_n(\omega)\big\}_1^\infty$ 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 nondecreasing sequence of sub-$\sigma$-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 Borel–Cantelli Lemma (cf. Lemma 1.1.3).

EXERCISE 7.2.25. For each $n \in \mathbb{N}$, let $(E_n, \mathcal{B}_n)$ be a measurable space and $\mu_n$ and $\nu_n$ 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 1.1.12) 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 $\pi_n$ is the natural projection from $\Omega$ 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

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

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 $\mathbb{Q}(B) \in \{0, 1\}$, where $B \equiv \left\{\varliminf_{n\to\infty} X_n = 0\right\}$, and combine this with the last part of Theorem 7.2.19 to conclude that $\mathbb{Q} \not\perp \mathbb{P} \implies \mathbb{Q} \ll P$. 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 7.2.26. Let $(\Omega, \mathcal{F})$ be a measurable space and $\Sigma$ a sub-$\sigma$-algebra of $\mathcal{F}$. Given a pair of probability measures $\mathbb{P}$ and $\mathbb{Q}$ on $(\Omega, \mathcal{F})$, let $X_\Sigma$ and $Y_\Sigma$ be nonnegative Radon–Nikodym derivatives of $\mathbb{P}_\Sigma \equiv P \restriction \Sigma$ and $\mathbb{Q}_\Sigma \equiv \mathbb{Q} \restriction \Sigma$, respectively, 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 $R$ is any $\sigma$-finite measure on $(\Omega, \Sigma)$ with the property that $\mathbb{P}_\Sigma \ll R$ and $\mathbb{Q}_\Sigma \ll R$, then the number $\left(\mathbb{P}, \mathbb{Q}\right)_\Sigma$ given above is equal to

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

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 $\{\mathcal{F}_n : n \in \mathbb{N}\}$ is a nondecreasing sequence of sub-$\sigma$-algebras of $\mathcal{F}$, and show that

$$\big(\mathbb{P},\mathbb{Q}\big)_{\mathcal{F}_n} \longrightarrow \big(\mathbb{P},\mathbb{Q}\big)_{\bigvee_0^\infty \mathcal{F}_n}.$$

(**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 nonnegative 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

$$\mathbb{E}^{\mathbb{P}}\Big[\sqrt{X_n}\Big] \longrightarrow 0 \quad \text{as} \quad n \to \infty.$$

(**iv**) Let $\{\sigma_n\}_0^\infty \subseteq (0,\infty)$, and, for each $n \in \mathbb{N}$, let $\mu_n$ and $\nu_n$ be Gaussian measures on $\mathbb{R}$ with variance $\sigma_n^2$. If $a_n$ and $b_n$, respectively, are the mean value of $\mu_n$ and $\nu_n$, 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 7.2.27. Again let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space, only this time suppose that $\{\mathcal{F}_n : n \in \mathbb{N}\}$ is a sequence of sub-$\sigma$-algebras which is *nonincreasing*. Given a sequence $\{X_n : n \in \mathbb{N}\}$ of $(-\infty,\infty]$-valued random variables, we say that the triple $\big(X_n,\mathcal{F}_n,\mathbb{P}\big)$ is a **reversed submartingale** or a **reversed martingale** according to whether $\big(X_{N-n\wedge N},\mathcal{F}_{N-n\wedge N},\mathbb{P}\big)$ is a submartingale or martingale for every $N \in \mathbb{N}$.

(**i**) Reproduce the result in Corollary 7.2.9 when $\big(X_n,\mathcal{F}_n,\mathbb{P}\big)$ is either a reversed martingale or a reversed submartingale. In particular, conclude that $\big(|X_n|,\mathcal{F}_n,\mathbb{P}\big)$ is a reversed submartingale if $\big(X_n,\mathcal{F}_n,\mathbb{P}\big)$ is a reversed martingale.

(**ii**) Given a reversed submartingale $\big(X_n,\mathcal{F}_n,\mathbb{P}\big)$, show that

$$\mathbb{P}\left(\sup_{n\in\mathbb{N}} X_n \ge \alpha\right) \le \frac{1}{\alpha}\mathbb{E}^{\mathbb{P}}\left[X_0,\, \sup_{n\in\mathbb{N}} X_n \ge \alpha\right], \quad \alpha \in (0,\infty).$$

In particular, if $\big(X_n,\mathcal{F}_n,\mathbb{P}\big)$ is a nonnegative reversed submartingale and $X_0 \in L^p(\mathbb{P};\mathbb{R})$ for some $p \in [1,\infty]$, conclude from the preceding that $\{X_n : n \in \mathbb{N}\}$ is uniformly $\mathbb{P}$-integrable and that

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

(**iii**) Given a reversed submartingale $\big(X_n,\mathcal{F}_n,\mathbb{P}\big)$ with $X_0 \in L^1(\mathbb{P};\mathbb{R})$, show that there is a $\mathbb{P}$-almost surely unique, $\mathcal{F}_\infty \equiv \bigcap_0^\infty \mathcal{F}_n$-measurable $X : \Omega \longrightarrow [-\infty,\infty)$ such that $X_n \longrightarrow X$ (a.s., $\mathbb{P}$). Further, show that $X \in L^1(\mathbb{P};\mathbb{R})$ if $\sup_{n\in\mathbb{N}} \mathbb{E}^{\mathbb{P}}\big[X_n^-\big] < \infty$. Finally, if $X_0 \in L^p(\mathbb{P};\mathbb{R})$ for some $p \in [1,\infty)$ and $\big(X_n,\mathcal{F}_n,\mathbb{P}\big)$ is either nonnegative or a reversed martingale, show that $X_n \longrightarrow X$ in $L^p(\mathbb{P};\mathbb{R})$.

**Hint**: Given $a < b$, let $D_{[a,b]}$ be the precise number of times that $\{X_n : n \in \mathbb{N}\}$ **downcrosses** $[a, b]$ (i.e., the precise number of times that $\{-X_n : n \in \mathbb{N}\}$ upcrosses $[-b, -a]$), and prove the **downcrossing inequality**

$$\mathbb{E}^{\mathbb{P}}\big[D_{[a,b]}\big] \leq \frac{\mathbb{E}^{\mathbb{P}}\big[(X_0 - a)^+\big]}{b - a}.$$

EXERCISE 7.2.28. An important application of reversed martingales is provided by De Finetti's theory of **exchangeability** . To describe this theory, let $\Sigma$ denote the group of all *finite permutations* of $\mathbb{Z}^+$. That is, the elements $\sigma$ of $\Sigma$ are isomorphisms of $\mathbb{Z}^+$ which fix all but a finite number of elements. Alternatively, $\Sigma = \bigcup_{m=1}^{\infty} \Sigma_m$, where $\Sigma_m$ is the group of isomorphisms $\sigma$ of $\mathbb{Z}^+$ with the property that $n = \sigma(n)$ for all $n > m$. Next, let $(E, \mathcal{B})$ be a measurable space, and, for each $\sigma \in \Sigma$, define $S_\sigma : E^{\mathbb{Z}^+} \longrightarrow E^{\mathbb{Z}^+}$ so that

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

Obviously, each $S_\sigma$ is a $\mathcal{B}^{\mathbb{Z}^+}$-measurable isomorphism of $E^{\mathbb{Z}^+}$. Also, if for $m \in \mathbb{Z}^+$,

$$\mathcal{I}_m \equiv \big\{B \in \mathcal{B}^{\mathbb{Z}^+} : B = S_\sigma B \text{ for all } \sigma \in \Sigma_m\big\},$$

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

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

Now suppose that $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and that

$$\omega \in \Omega \longmapsto \mathbf{X}(\omega) \equiv \big(X_1(\omega), \dots, X_n(\omega), \dots\big) \in E^{\mathbb{Z}^+}$$

is a measurable map which is $\mathbb{P}$-*exchangeable* in the sense that the distribution of $\omega \in \Omega \longmapsto S_\sigma \circ \mathbf{X}(\omega) \in E^{\mathbb{Z}^+}$ under $\mathbb{P}$ is the same for all $\sigma \in \Sigma$.

(**i**) Given a measurable $g : E \longrightarrow \mathbb{R}$ satisfying $g \circ X_1 \in L^1(\mathbb{P}; \mathbb{R})$, show that $\mathbb{E}^{\mathbb{P}}[g \circ X_\ell \,|\, \mathbf{X}^{-1}(\mathcal{I}_m)] = \mathbb{E}^{\mathbb{P}}[g \circ X_1 \,|\, \mathbf{X}^{-1}(\mathcal{I}_m)]$ for all $1 \leq \ell \leq m$, and conclude first that

$$\mathbb{E}^{\mathbb{P}}\big[g \circ X_1 \,\big|\, \mathbf{X}^{-1}(\mathcal{I}_m)\big] = \frac{1}{m} \sum_{\ell=1}^{m} g \circ X_\ell,$$

and then that

$$\mathbb{E}^{\mathbb{P}}\big[g \circ X_1 \,\big|\, \mathbf{X}^{-1}(\mathcal{I}_\infty)\big] = \lim_{m \to \infty} \frac{1}{m} \sum_{1}^{m} g \circ X_\ell$$

both $\mathbb{P}$-almost surely and in $L^1(\mathbb{P}; \mathbb{R})$.

(**ii**) Use $\mathcal{T}$ to denote the tail field $\bigcap_{m=1}^{\infty}\sigma\big(\{X_n : n \geq m\}\big)$ determined by $\{X_n : n \geq 1\}$, and observe that $\mathcal{T} \subseteq \mathbf{X}^{-1}(\mathcal{I}_\infty)$. As a consequence of (**i**), conclude that

$$\mathbb{E}^{\mathbb{P}}\big[g \circ X_1 \,\big|\, \mathbf{X}^{-1}(\mathcal{I}_\infty)\big] = \lim_{m\to\infty} \frac{1}{m}\sum_{1}^{m} g \circ X_\ell = \mathbb{E}^{\mathbb{P}}\big[g \circ X_1 \,\big|\, \mathcal{T}\big]$$

$\mathbb{P}$-almost surely. In particular, if $\mathcal{T}$ is $\mathbb{P}$-trivial (i.e., all its elements have $\mathbb{P}$-probability 0 or 1), then one gets

$$\lim_{m\to\infty} \frac{1}{m}\sum_{1}^{m} g \circ X_\ell = \mathbb{E}\big[g \circ X_1\big] \quad P\text{-almost surely.}$$

Note that a particular case is the one when the $X_n$'s are mutually independent and identically distributed. Because of Kolmogorov's 0–1 Law (cf. Theorem 1.1.2), $\mathcal{T}$ is $\mathbb{P}$-trivial in this case, and so we have yet another derivation of the Strong Law of Large Numbers (cf. Theorem 1.4.11).*

Exercise 7.2.29. We continue in the setting of Exercise 7.2.28. The goal of this exercise is to prove that for every $F \in L^1(\mathbb{P};\mathbb{R})$,

$$\mathbb{E}^{\mathbb{P}}\big[F \,\big|\, \mathbf{X}(\mathcal{I}_\infty)\big] = \mathbb{E}^{\mathbb{P}}\big[F \,\big|\, \mathcal{T}\big] \quad P\text{-almost surely.} \tag{7.2.30}$$

Equivalently, since we already know that $\mathcal{T} \subseteq \mathcal{I}_\infty$, what we want to show is that for every $I \in \mathcal{I}_\infty$ there is a $T \in \mathcal{T}$ such that the symmetric difference $I\Delta T = (I \setminus T) \cup (T \setminus I)$ is a $\mathbb{P}$-null set. Thus, by Kolmogorov's 0–1, when the $X_n$'s are mutually independent, $\mathbb{P}(I) \in \{0,1\}$ for every $I \in \mathbf{X}^{-1}(\mathcal{I}_\infty)$, a statement which is called the **Hewitt–Savage** 0–1 **Law**. Equivalently, independence implies that every $\mathbf{X}^{-1}(\mathcal{I}_\infty)$-measurable $F : \Omega \longrightarrow \mathbb{R}$ is $\mathbb{P}$-almost surely constant.

(**i**) Let $N \in \mathbb{Z}^+$ be given, set

$$\Sigma^{(N)} = \big\{\sigma \in \Sigma : \sigma(mN + \ell) = \sigma(mN) + \ell \text{ for all } m \in \mathbb{N} \text{ and } 1 \leq \ell \leq N\big\},$$

and define

$$\mathcal{I}_\infty^{(N)} = \big\{B \in \mathcal{B}^{\mathbb{Z}^+} : B = S_\sigma B \text{ for all } \sigma \in \Sigma^{(N)}\big\}.$$

Next, suppose that $F \in L^1(\mathbb{P};\mathbb{R})$ is $\sigma\big(\{X_n : 1 \leq n \leq N\}\big)$-measurable, and observe that $F = g \circ (X_1, \dots, X_N)$ for some measurable $g : E^N \longrightarrow \mathbb{R}$. After replacing $E$ by $E^N$ in Exercise 7.2.28, show that

$$\mathbb{E}^{\mathbb{P}}\big[F \big| \mathbf{X}^{-1}(\mathcal{I}_\infty^{(N)})\big] = \mathbb{E}^{\mathbb{P}}\big[F \,\big|\, \mathcal{T}\big] \quad P\text{-almost surely.}$$

Hence, since $\mathcal{T} \subseteq \mathbf{X}^{-1}(\mathcal{I}_\infty^{(N)})$, (7.2.30) holds for every $\sigma\big(\{X_n : 1 \leq n \leq N\}\big)$-measurable $F \in L^1(\mathbb{P};\mathbb{R})$.

* It turns out that exchangeable random variables whose tail field is trivial are necessarily independent. Thus, the present line of reasoning does not really extend the Strong Law. On the other hand, the present derivation extends without alteration to the Banach space setting (cf. Exercise (?)).

(**ii**) To complete the proof of (7.2.30), first observe that, since $\mathcal{T} \subseteq \mathbf{X}^{-1}(\mathcal{I}_\infty) \subseteq \sigma\big(\{X_n : n \in \mathbb{Z}^+\}\big)$, it suffices to handle $F \in L^1(\mathbb{P};\mathbb{R})$ which are $\sigma\big(\{X_n : n \in \mathbb{Z}^+\}\big)$-measurable. Second, note that the class of $F \in L^1(\mathbb{P};\mathbb{R})$ for which (7.2.30) holds is closed. Finally, use the fact that

$$\mathbb{E}^{\mathbb{P}}\big[F \,\big|\, \sigma\big(\{X_n : 1 \le n \le N\}\big)\big] \stackrel{L^1(\mathbb{P};\mathbb{R})}{\longrightarrow} \mathbb{E}^{\mathbb{P}}\big[F \,\big|\, \sigma\big(\{X_n : n \in \mathbb{Z}^+\}\big)\big].$$

EXERCISE 7.2.31. 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 1.4.26 we showed that $\varliminf_{n\to\infty} |S_n| < \infty$ $\mathbb{P}$-almost surely. Here we will show that

$$\varliminf_{n\to\infty} |S_n| = 0 \quad P\text{-almost surely.} \tag{7.2.32}$$

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 $L \in [0,\infty)$ such that

$$L = \varliminf_{n\to\infty} |S_n| \quad (\text{a.s.}, \mathbb{P}).$$

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

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

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

**Hint**: Set $\rho = \frac{L}{3}$, and suppose that $\mathbb{P}(|S_m - x| < \rho) > 0$ for some $m \in \mathbb{Z}^+$. Observe that $\{S_{m+n} - S_m : n \in \mathbb{Z}^+\}$ has the same $\mathbb{P}$-distribution as $\{X_n : n \in \mathbb{Z}^+\}$, and therefore that $\mathbb{P}(|S_{m+n} - S_m| < 3\rho \text{ i.o.}) = 0$. Thus, since $|S_{m+n} - x| \ge |S_{m+n} - S_m| - |S_m - x|$, $\mathbb{P}(|S_n - x| < \rho \text{ i.o.}) \le \mathbb{P}(|S_m - x| \ge \rho) < 1$. But, by the Hewitt–Savage 0–1 Law, this means that $\mathbb{P}(|S_n - x| < \rho \text{ i.o.}) = 0$.

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

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

which, in view of (**i**), is a contradiction.

(**iii**) Knowing (7.2.32), conclude that, for each $x \in \mathbb{R}$ and $\epsilon > 0$, one has the dichotomy

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

Exercise 7.2.33. Here is a rather silly 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 7.1.19. 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 7.1.19,

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

After noting that $\{\mathcal{F}_n : n \in \mathbb{N}\}$ is nonincreasing, use the result obtained in part (**iii**) here 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 $L^2([0,1);\mathbb{C})$.*

## § 7.3: Some Extensions

It turns out that many of the results obtained in § 7.2 admit easy extensions to both infinite measures and Banach space valued random variables. Furthermore, in many applications, these extensions play a useful, and occasionally essential, role. Throughout the discussion which follows, $(\Omega, \mathcal{F}, \mu)$ will be a measure space and $\{\mathcal{F}_n : n \in \mathbb{N}\}$ will be a nondecreasing sequence of sub-$\sigma$-algebras with the property that $\mu \restriction \mathcal{F}_0$ is $\sigma$-finite. In particular, this means that the conditional expectation of a $\mu$-locally integrable random variable given $\mathcal{F}_n$ is well-defined (cf. Theorem 7.1.14 Theorem) even if the random variable takes values in a separable Banach space $E$. Thus, we will say that the sequence $\{\mathbf{X}_n;\, n \in \mathbb{N}\}$ of

* When $f$ is a function with the property that $(f, e_\ell)_{L^2([0,1);\mathbb{C})} = 0$ for all $\ell \in \mathbb{Z}\setminus\{2^m : m \in \mathbb{N}\}$, 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.

$E$-valued random variables is a **$\mu$-martingale with respect to** $\{\mathcal{F}_n : n \in \mathbb{N}\}$, or, more briefly, that the triple $(\mathbf{X}_n, \mathcal{F}_n, \mu)$ is a **martingale** if $\{\mathbf{X}_n : n \in \mathbb{N}\}$ is $\{\mathcal{F}_n : n \in \mathbb{N}\}$-progressively measurable, each $\mathbf{X}_n$ is $\mu$-locally integrable, and

$$\mathbf{X}_{n-1} = \mathbb{E}^{\mu}\big[\mathbf{X}_n \big| \mathcal{F}_{n-1}\big] \quad \text{(a.e., } \mu) \quad \text{for each } n \in \mathbb{Z}^+.$$

Furthermore, when $E = \mathbb{R}$, we will say that $\{X_n : n \in \mathbb{N}\}$ is a **$\mu$-submartingale with respect to** $\{\mathcal{F}_n : n \in \mathbb{N}\}$ (equivalently, the triple $(X_n, \mathcal{F}_n, \mu)$ is a **submartingale**) if $\{X_n : n \in \mathbb{N}\}$ is $\{\mathcal{F}_n : n \in \mathbb{N}\}$-progressively measurable, each $X_n$ is $\mu$-locally integrable, and

$$X_{n-1} \le \mathbb{E}^{\mu}\big[X_n \big| \mathcal{F}_{n-1}\big] \quad \text{(a.e., } \mu) \quad \text{for each } n \in \mathbb{Z}^+.$$

**§ 7.3:.1. Martingale Theory for a $\sigma$-Finite Measure Space.** Without any real effort, we can now prove the following variants of each of the basic results in § 7.2.

THEOREM 7.3.1. *Let $(X_n, \mathcal{F}_n, \mu)$ be a $\mu$-submartingale. Then, for each $N \in \mathbb{N}$ and $A \in \mathcal{F}_0$ on which $X_N$ is $\mu$-integrable:*

$$\mu\left(\left\{\max_{0 \le n \le N} X_n \ge \alpha\right\} \cap A\right) \le \frac{1}{\alpha}\,\mathbb{E}^{\mu}\left[X_N,\, \left\{\max_{0 \le n \le N} X_n \ge \alpha\right\} \cap A\right] \tag{7.3.2}$$

*for all $\alpha \in (0, \infty)$. In particular, if the $X_n$'s are all nonnegative, then, for every $p \in (1, \infty)$ and $A \in \mathcal{F}_0$:*

$$\mathbb{E}^{\mu}\left[\sup_{n \in \mathbb{N}} |X_n|^p,\, A\right]^{\frac{1}{p}} \le \frac{p}{p-1} \sup_{n \in \mathbb{N}} \mathbb{E}^{\mu}\big[|X_n|^p,\, A\big]^{\frac{1}{p}}.$$

*Furthermore, for any bounded $\{\mathcal{F}_n : n \in \mathbb{N}\}$-stopping times $\sigma \le \tau$,*

$$X_\sigma \le \mathbb{E}^{\mu}\big[X_\tau \big| \mathcal{F}_\sigma\big] \quad \text{(a.e., } \mu),$$

*and the inequality is an equality in the martingale case. In particular, for each stopping time $\tau$, $(X_{n\wedge\tau}, \mathcal{F}_n, \mu)$ is a submartingale; and, when $(X_n, \mathcal{F}_n, \mu)$ is itself a martingale, then $(X_{n\wedge\tau}, \mathcal{F}_n, \mu)$ is again a martingale. Next, given $a < b$ and $A \in \mathcal{F}_0$,*

$$\mathbb{E}^{\mu}\big[U_{[a,b]},\, A\big] \le \sup_{n \in \mathbb{N}} \frac{\mathbb{E}^{\mu}\big[(X_n - a)^+,\, A\big]}{b-a},$$

*where $U_{[a,b]}(\omega)$ denotes the precise number of times that $\{X_n(\omega)\}_0^\infty$ upcrosses $[a, b]$ (cf. the discussion preceding Theorem 7.2.15). Finally,*

$$\sup_{n \in \mathbb{N}} \mathbb{E}^{\mu}\big[X_n^+,\, A\big] < \infty \text{ for every } A \in \mathcal{F}_0 \text{ with } \mu(A) < \infty$$
$$\Longrightarrow\; X_n \longrightarrow X \quad \text{(a.e., } \mu),$$

*where $X$ is $\bigvee_0^\infty \mathcal{F}_n$-measurable and $\mu$-locally integrable; and, for each $p \in (1,\infty)$, the convergence is in $L^p(\mu;\mathbb{R})$ if and only if $\{X_n : n \geq 0\}$ is bounded in $L^p(\mu;\mathbb{R})$. In fact, in the case of martingales, there is a $\bigvee_0^\infty \mathcal{F}_n$-measurable, $\mu$-locally integrable $X$ such that*

$$X_n = \mathbb{E}^\mu\big[X\big|\mathcal{F}_n\big] \quad \text{(a.e., } \mu) \quad \textit{for all } n \in \mathbb{N}$$

*if and only if $\{X_n : n \geq 0\}$ is uniformly $\mu$-integrable on each $A \in \mathcal{F}_0$ with $\mu(A) < \infty$, in which case $X$ is $\mu$-integrable if and only if $X_n \longrightarrow X$ in $L^1(\mu;\mathbb{R})$. On the other hand, if $p \in (1,\infty)$, then $X \in L^p(\mu : \mathbb{R};\mathbb{R})$ if and only if $\{X_n : n \geq 0\}$ is bounded in $L^p(\mu;\mathbb{R})$, in which case, $X_n \longrightarrow X$ in $L^p(\mu;\mathbb{R})$.*

Proof: Obviously, there is no problem unless $\mu(\Omega) = \infty$. However, even then, each of these results follows immediately from its counterpart in § 7.2 once one makes the following trivial observation. Namely, given $\Omega' \in \mathcal{F}_0$ with $\mu(\Omega') \in (0,\infty)$, set

$$\mathcal{F}' = \mathcal{F}[\Omega'], \quad \mathcal{F}'_n = \mathcal{F}_n[\Omega'], \quad X'_n = X_n \restriction \Omega', \quad \text{and } \mathbb{P} = \frac{\mu \restriction \mathcal{F}'}{\mu\Omega')}.$$

Then $\big(X'_n, \mathcal{F}'_n, \mathbb{P}'\big)$ is a submartingale or martingale depending on whether the original $\big(X_n, \mathcal{F}_n, \mu\big)$ was a submartingale or martingale. Hence, when $\mu(\Omega) = \infty$, we simply choose a sequence $\{\Omega_k\}_1^\infty$ of mutually disjoint, $\mu$-finite elements of $\mathcal{F}_0$ so that $\Omega = \bigcup_1^\infty \Omega_k$, work on each $\Omega_k$ separately, and, at the end, sum the results. $\square$

We will now spend a little time seeing how Theorem 7.3.1 can be applied to give simple proof of the **Hardy–Littlewood maximal inequality**. To state their result, we define the maximal function $\mathbf{M}f$ for $f \in L^1(\mathbb{R}^N;\mathbb{R})$ by

$$\mathbf{M}f(\mathbf{x}) = \sup_{Q \ni \mathbf{x}} \frac{1}{|Q|} \int_Q |f(\mathbf{y})|\, d\mathbf{y}, \quad \mathbf{x} \in \mathbb{R}^N,$$

where $Q$ is used to denote a generic **cube**

$$Q = \prod_{j=1}^N [a_j, a_j + r) \quad \text{with } \mathbf{a} \in \mathbb{R}^N \text{ and } r > 0. \tag{7.3.3}$$

As is easily checked, $\mathbf{M}f : \mathbb{R}^N \longrightarrow [0,\infty]$ is lower semicontinuous and therefore certainly Borel measurable. Furthermore, if we restrict our attention to *nicely meshed* families of cubes, then it is easy to relate $\mathbf{M}f$ to martingales. More precisely, for each $n \in \mathbb{Z}$, the **$n$th standard diadic partition of** $\mathbb{R}^N$ is the partition $\mathcal{P}_n$ of $\mathbb{R}^N$ into the cubes

$$C_n(\mathbf{k}) \equiv \prod_{i=1}^N \left[\frac{k_i}{2^n}, \frac{k_i+1}{2^n}\right), \quad \mathbf{k} \in \mathbb{Z}^N. \tag{7.3.4}$$

These partitions are nicely meshed in the sense that the $(n+1)$st is a refinement of the $n$th. Equivalently, if $\mathcal{F}_n$ denotes the $\sigma$-algebra over $\mathbb{R}^N$ generated by the partition $\mathcal{P}_n$, then $\mathcal{F}_n \subseteq \mathcal{F}_{n+1}$. Moreover, if $f \in L^1(\mathbb{R}^N;\mathbb{R})$ and

$$X_n^f(\mathbf{x}) \equiv 2^{nN} \int_{C_n(\mathbf{k})} \left|f(\mathbf{y})\right| d\mathbf{y} \quad \text{for } \mathbf{x} \in C_n(\mathbf{k}) \text{ and } \mathbf{k} \in \mathbb{Z}^N,$$

then

$$X_n^f = \mathbb{E}^{\text{Leb}}\left[|f|\middle|\mathcal{F}_n\right] \quad \text{(a.e., Leb)}$$

for each $n \in \mathbb{Z}$. In particular, for each $m \in \mathbb{Z}$,

$$\left(X_{m+n}^f, \mathcal{F}_{m+n}, \text{Leb}\right), \quad n \in \mathbb{N},$$

is a nonnegative martingale; and so, by applying (7.3.2) for each $m \in \mathbb{Z}$ and then letting $m \searrow -\infty$, we see that

$$\left|\left\{\mathbf{x} : \mathbf{M}^{(\mathbf{0})} f(\mathbf{x}) \geq \alpha\right\}\right| \leq \frac{1}{\alpha} \int\limits_{\{\mathbf{M}^{(\mathbf{0})} f \geq \alpha\}} f(\mathbf{y})\, dy, \quad \alpha \in (0,\infty), \tag{7.3.5}$$

where

$$\mathbf{M}^{(\mathbf{0})} f(\mathbf{x}) = \sup\left\{\frac{1}{|Q|}\int_Q f(\mathbf{y})\, d\mathbf{y} : \mathbf{x} \in Q \in \bigcup_{n\in\mathbb{Z}} \mathcal{P}_n\right\}$$

and we have used $|\Gamma|$ to denote the Lebesgue measure of $\Gamma$.

At first sight, one might hope that it should be possible to pass directly from (7.3.5) to analogous estimates on the level sets of $\mathbf{M}f$. However, the passage from (7.3.5) to control on $\mathbf{M}f$ is not so easy as it might at first appear: the "sup" in the definition of $\mathbf{M}f$ involves many more cubes than the one in the definition of $\mathbf{M}^{(\mathbf{0})}f$; and it is for this reason we will have to introduce additional families of meshed partitions. Namely, for each $\boldsymbol{\eta} \in \{0,1\}^N$ set

$$\mathcal{P}_n(\boldsymbol{\eta}) = \left\{\frac{(-1)^n \boldsymbol{\eta}}{3\,2^n} + C_n(\mathbf{k}) : \mathbf{k} \in \mathbb{Z}^N\right\},$$

where $C_n(\mathbf{k})$ is the cube described in above. It is then an easy matter to check that, for each $\boldsymbol{\eta} \in \{0,1\}^N$, $\left\{\mathcal{P}_n(\boldsymbol{\eta}) : n \in \mathbb{Z}\right\}$ is a family of meshed partitions of $\mathbb{R}^N$. Furthermore, if

$$\left[\mathbf{M}^{(\boldsymbol{\eta})} f\right](\mathbf{x}) = \sup\left\{\frac{1}{|Q|}\int_Q \left|f(\mathbf{y})\right| d\mathbf{y} : \mathbf{x} \in Q \in \bigcup_{n\in\mathbb{Z}} \mathcal{P}_n(\boldsymbol{\eta})\right\}, \quad \mathbf{x} \in \mathbb{R}^N,$$

then exactly the same argument which (when $\boldsymbol{\eta} = \mathbf{0}$) led us to (7.3.5) can now be used to get

$$\left|\left\{\mathbf{x} \in \mathbb{R}^N : \left[\mathbf{M}^{(\boldsymbol{\eta})} f\right](\mathbf{x}) \geq \alpha\right\}\right| \leq \frac{1}{\alpha} \int\limits_{\{\mathbf{M}^{(\boldsymbol{\eta})} f \geq \alpha\}} \left|f(\mathbf{y})\right| d\mathbf{y},$$

for each $\boldsymbol{\eta} \in \{0,1\}^N$ and $\alpha \in (0,\infty)$. Finally, if $Q$ is given by (7.3.3) and $r \leq \frac{1}{3\,2^n}$, then it is possible to find an $\boldsymbol{\eta} \in \{0,1\}^N$ and a $C \in \mathcal{P}_n(\boldsymbol{\eta})$ for which $Q \subseteq C$. (To see this, first reduce to the case when $N = 1$.) Hence,

$$\max_{\boldsymbol{\eta} \in \{0,1\}^N} \mathbf{M}^{(\boldsymbol{\eta})} f \leq \mathbf{M} f \leq 6^N \max_{\boldsymbol{\eta} \in \{0,1\}^N} \mathbf{M}^{(\boldsymbol{\eta})} f.$$

After combining this with the estimate on $\mathbf{M}^{(\boldsymbol{\eta})} f$, we arrive at the following version of the **Hardy–Littlewood inequality**

$$\left|\left\{\mathbf{x} \in \mathbb{R}^N : \mathbf{M} f(\mathbf{x}) \geq \alpha\right\}\right| \leq \frac{(12)^N}{\alpha} \int_{\mathbb{R}^N} |f(\mathbf{y})|\, d\mathbf{y}. \tag{7.3.6}$$

At the same time, because (cf. Exercise 1.4.17) the estimate for $\mathbf{M}^{(\boldsymbol{\eta})} f$ also implies that

$$\max_{\boldsymbol{\eta} \in \{0,1\}^N} \left\|\mathbf{M}^{(\boldsymbol{\eta})} f\right\|_{L^p(\mathbb{R}^N;\mathbb{R})} \leq \frac{p}{p-1} \|f\|_{L^p(\mathbb{R}^N;\mathbb{R})}, \quad p \in (1,\infty],$$

we can also use the same argument to obtain

$$\left\|\mathbf{M} f\right\|_{L^p(\mathbb{R}^N);\mathbb{R}} \leq \frac{(12)^N p}{p-1} \|f\|_{L^p(\mathbb{R}^N)}, \quad p \in (1,\infty]. \tag{7.3.7}$$

In this connection, notice that there is no hope of getting this estimate when $p = 1$, since it is clear that

$$\varliminf_{|\mathbf{x}| \to \infty} |\mathbf{x}|^N \mathbf{M} f(\mathbf{x}) > 0$$

whenever $f$ does not vanish Leb-almost everywhere.

The inequality in (7.3.6) plays the same rôle in classical analysis as Doob's inequality plays in martingale theory. For example, by essentially the same argument as we used to pass from Doob's inequality to Corollary 7.2.4, we obtain the following famous **Lebesgue Differentiation Theorem**.

Theorem 7.3.8. *For each* $f \in L^1\big(\mathbb{R}^N;\mathbb{R}\big)$,

$$\lim_{B\searrow\{\mathbf{x}\}} \frac{1}{|B|}\int_B \big|f(\mathbf{y})-f(\mathbf{x})\big|\,d\mathbf{y}=0 \tag{7.3.9}$$
$$\text{for Leb-almost every } \mathbf{x}\in\mathbb{R}^N,$$

*where, for each* $\mathbf{x}\in\mathbb{R}^N$*, the limit is taken over balls* $B$ *which contain* $\mathbf{x}$ *and tend to* $\mathbf{x}$ *in the sense that their radii shrink to* 0*. In particular,*

$$f(\mathbf{x}) = \lim_{B\searrow\{\mathbf{x}\}} \frac{1}{|B|}\int_B f(\mathbf{y})\,d\mathbf{y} \quad \text{for Leb-almost every } \mathbf{x}\in\mathbb{R}^N.$$

Proof: We begin with the observation that, for each $f\in L^1(\mathbb{R}^N;\mathbb{R})$,

$$\tilde{\mathbf{M}}f(\mathbf{x}) \equiv \sup_{B\ni\mathbf{x}} \frac{1}{|B|}\int_B \big|f(\mathbf{y})\big|\,d\mathbf{y} \le \kappa_N \mathbf{M}f(\mathbf{x}), \quad \mathbf{x}\in\mathbb{R}^N$$

where $\kappa_n = \frac{2^N}{\mathbf{\Omega}_N}$ with $\mathbf{\Omega}_N = \big|B_{\mathbb{R}^N}(\mathbf{0},1)\big|$. Second, we remark that (7.3.9) for *every* $\mathbf{x}\in\mathbb{R}^N$ is trivial when $f\in C_{\mathrm{c}}(\mathbb{R}^N;\mathbb{R})$. Hence, all that remains is to check that if $f_n \longrightarrow f$ in $L^1(\mathbb{R}^N;\mathbb{R})$ and if (7.3.9) holds for each $f_n$, then it holds for $f$. To this end, let $\epsilon>0$ be given and observe that, because of the preceding and (7.3.6),

$$\begin{aligned}
&\left|\left\{\mathbf{x}:\ \varlimsup_{B\searrow\{\mathbf{x}\}} \frac{1}{|B|}\int_B \big|f(\mathbf{y})-f(\mathbf{x})\big|\,d\mathbf{y}\ge\epsilon\right\}\right| \\
&\quad\le \left|\left\{\mathbf{x}:\ \tilde{\mathbf{M}}(f-f_n)(\mathbf{x})\ge\frac{\epsilon}{3}\right\}\right| \\
&\qquad + \left|\left\{\mathbf{x}:\ \varlimsup_{B\searrow\{\mathbf{x}\}} \frac{1}{|B|}\int_B \big|f_n(\mathbf{y})-f_n(\mathbf{x})\big|\,d\mathbf{y}\ge\frac{\epsilon}{3}\right\}\right| \\
&\qquad\quad + \left|\left\{\mathbf{x}:\ \big|f_n(\mathbf{x})-f(\mathbf{x})\big|\ge\tfrac{\epsilon}{3}\right\}\right| \\
&\quad\le \frac{3}{\epsilon}\big(1+(12)^N\kappa_N\big)\|f-f_n\|_{L^1(\mathbb{R}^N)}
\end{aligned}$$

for every $n\in\mathbb{Z}^+$. Hence, after letting $n\to\infty$, we see that (7.3.9) also holds for $f$. □

Although applications like Lebesgue's Differentiation Theorem might make one to think that (7.3.6) is most interesting because of what it says about averages over small cubes, its implications for large cubes are also significant. In fact, as we will see in (?), it allows us to prove Birkhoff's Individual Ergodic Theorem (cf. Theorem (?)), which may be viewed as *differentiation at infinity.* The link between ergodic theory and the Hardy–Littlewood Inequality is provided by the following deterministic version of the Maximal Ergodic Lemma (cf. Lemma (?)). Namely, let $\{a_{\mathbf{k}} : \mathbf{k} \in \mathbb{Z}^N\}$ be a summable subset of $[0,\infty)$ and set

$$\overline{S}_n(\mathbf{k}) = \frac{1}{(2n)^N} \sum_{\mathbf{j} \in Q_n} a_{\mathbf{j}+\mathbf{k}}, \quad n \in \mathbb{N} \text{ and } \mathbf{k} \in \mathbb{Z}^N,$$

where $Q_n = \{\mathbf{j} \in \mathbb{Z}^N : -n \le j_i < n \text{ for } 1 \le i \le N\}$. By applying (7.3.6) and (7.3.7) to the function $f$ given by (cf. (7.3.4))

$$f(\mathbf{x}) = a_{\mathbf{k}} \quad \text{when } \mathbf{x} \in C_0(\mathbf{k}),$$

we see that

$$\text{card}\left\{\mathbf{k} \in \mathbb{Z}^N : \sup_{n \in \mathbb{Z}^+} \overline{S}_n(\mathbf{k}) \ge \alpha\right\} \le \frac{(12)^N}{\alpha} \sum_{\mathbf{k} \in \mathbb{Z}^N} a_{\mathbf{k}}, \quad \alpha \in (0,\infty) \tag{7.3.10}$$

and

$$\left(\sum_{\mathbf{k} \in \mathbb{Z}^N} \sup_{n \in \mathbb{Z}^+} |\overline{S}_n(\mathbf{k})|^p\right)^{\frac{1}{p}} \le \frac{(12)^N p}{p-1} \left(\sum_{\mathbf{k} \in \mathbb{Z}^N} |a_{\mathbf{k}}|^p\right)^{\frac{1}{p}} \tag{7.3.11}$$

for each $p \in (1,\infty]$. The inequality in (7.3.10) is called **Hardy's Inequality**. Actually, Hardy was drawn to this line of research by his passion for the game of cricket. What Hardy wanted to find is the optimal order in which to arrange batters to maximize the average score per inning. Thus, he worked with a nonnegative sequence $\{a_k\}_0^\infty$ in which $a_k$ represented the expected number of runs scored by player $k$, and what he showed is that, for each $\alpha \in (0,\infty)$,

$$\left|\left\{k \in \mathbb{N} : \sup_{n \in \mathbb{Z}^+} \overline{S}_n(k) \ge \alpha\right\}\right|$$

is maximized when $\{a_n\}_0^\infty$ is nonincreasing; from which it is an easy application of Markov's inequality to prove that

$$\left|\left\{k \in \mathbb{N} : \sup_{n \in \mathbb{Z}^+} \overline{S}_n(k) \ge \alpha\right\}\right| \le \frac{1}{\alpha} \sum_0^\infty a_k, \quad \alpha \in (0,\infty).$$

Although this sharpened result can also be obtained as a corollary the *Sunrise Lemma*,* Hardy's approach remains the most appealing.

**§ 7.3:.2. Banach Space Valued Martingales.** We turn next to Banach space valued martingales. Actually, everything except the easiest aspects of this topic becomes extremely complicated and technical very quickly, and, for this reason, we will restrict our attention to those results which do not involve any deep properties of the geometry of Banach spaces. In fact, the only general theory with which we will deal is contained in the following.

* See Lemma 3.4.5 in my *A Concise Introduction to the Theory of Integration*, 3rd edition, published by Birkhauser in 1998.

THEOREM 7.3.12. *Let $E$ be a separable Banach space and $\big(\mathbf{X}_n, \mathcal{F}_n, \mu\big)$ an $E$-valued martingale. Then $\big(\|\mathbf{X}_n\|_E, \mathcal{F}_n, \mu\big)$ is a nonnegative submartingale and therefore, for each $N \in \mathbb{Z}^+$ and all $\alpha \in (0, \infty)$,*

$$\mu\left(\sup_{0 \le n \le N} \|\mathbf{X}_n\|_E \ge \alpha\right) \le \frac{1}{\alpha}\mathbb{E}^\mu\left[\|\mathbf{X}_N\|_E,\ \sup_{0 \le n \le N} \|\mathbf{X}_n\|_E \ge \alpha\right]. \tag{7.3.13}$$

*In particular, for each $p \in (1, \infty]$,*

$$\left\|\sup_{n \in \mathbb{N}} \|\mathbf{X}_n\|_E\right\|_{L^p(\mu;E)} \le \frac{p}{p-1} \sup_{n \in \mathbb{N}} \|\mathbf{X}_n\|_{L^p(\mu;E)}. \tag{7.3.14}$$

*Finally, if $\mathbf{X} \in L^p(\mu; E)$ for some $p \in [1, \infty)$, then*

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

PROOF: The fact $\big(\|\mathbf{X}_n\|_E, \mathcal{F}_n, \mu\big)$ is a submartingale is an easy application of the inequality in (7.1.16); and, given this fact, the inequalities in (7.3.13) and (7.3.14) follow from the corresponding inequalities in Theorem 7.3.1.

While proving the convergence statements, we may and will assume that $\mathcal{F} = \bigvee_0^\infty \mathcal{F}_n$. Now let $\mathbf{X} \in L^p(\mu; E)$ be given, and set $\mathbf{X}_n = \mathbb{E}^\mu[\mathbf{X}|\mathcal{F}_n]$, $n \in \mathbb{N}$. Because of (7.3.13) and (7.3.14), we know (cf. the proofs of Corollary 7.2.4 and Theorem 7.3.8) that the set of $\mathbf{X}$ for which $\mathbf{X}_n \longrightarrow \mathbf{X}$ (a.e., $\mu$) is a closed subset of $L^p(\mu; E)$. Moreover, if $\mathbf{X}$ is $\mu$-simple, then the $\mu$-almost everywhere convergence of $\mathbf{X}_n$ to $\mathbf{X}$ follows easily from the $\mathbb{R}$-valued result. Hence, we now know that $\mathbf{X}_n \longrightarrow \mathbf{X}$ (a.s, $\mu$) for each $\mathbf{X} \in L^1(\mu; E)$. In addition, because of (7.3.14), when $p \in (1, \infty)$, the convergence in $L^p(\mu; E)$ follows by Lebesgue's Dominated Convergence Theorem. Finally, to prove the convergence in $L^1(\mu; E)$ when $\mathbf{X} \in L^1(\mu; E)$, note that, by Fatou's Lemma,

$$\|\mathbf{X}\|_{L^1(\mu;E)} \le \varliminf_{n \to \infty} \|\mathbf{X}_n\|_{L^1(\mu;E)};$$

whereas (7.1.16) guarantees that

$$\|\mathbf{X}\|_{L^1(\mu;E)} \ge \varlimsup_{n \to \infty} \|\mathbf{X}_n\|_{L^1(\mu;E)}.$$

Hence, because

$$\Big|\,\|\mathbf{X}_n\|_E - \|\mathbf{X}\|_E - \|\mathbf{X}_n - \mathbf{X}\|_E\,\Big| \le 2\|\mathbf{X}\|_E,$$

the convergence in $L^1(\mu; E)$ is again an application of Lebesgue's Dominated Convergence Theorem. □

Going beyond the convergence result in Theorem 7.3.12 to get an analogue of Doob's Martingale Convergence Theorem is hard. For one thing, a naïve analog is not even true for general separable Banach spaces, and a rather deep analysis of the geometry of Banach spaces is required in order to determine exactly when it is true.

**Exercises for § 7.3**

EXERCISE 7.3.15. In this exercise, we will develop Jensen's inequality in the Banach space setting. Thus, $(\Omega, \mathcal{F}, \mathbb{P})$ will be a probability space, $C$ will be a closed, convex subset of the separable Banach space $E$, and $\mathbf{X}$ will be a $C$-valued element of $L^1(\mathbb{P}; E)$.

(**i**) Show that there exists a sequence $\{\mathbf{X}_n\}_1^\infty$ of $C$-valued, $\mathbb{P}$-simple functions which tend to $\mathbf{X}$ both $\mathbb{P}$-almost surely and in $L^1(\mathbb{P}; E)$.

(**ii**) Show that $\mathbb{E}^{\mathbb{P}}[\mathbf{X}] \in C$ and that

$$\mathbb{E}^{\mathbb{P}}\big[g(\mathbf{X})\big] \le g\big(\mathbb{E}^{\mathbb{P}}[\mathbf{X}]\big)$$

for every continuous, concave $g : C \longrightarrow [0, \infty)$.

(**iii**) Given a sub-$\sigma$-algebra $\Sigma$ of $\mathcal{F}$, follow the argument in Corollary 7.2.8 to show that there exists a sequence $\{\mathcal{P}_n\}_0^\infty$ of finite, $\Sigma$-measurable partitions with the property that

$$\sum_{A \in \mathcal{P}_n} \frac{\mathbb{E}^{\mathbb{P}}[\mathbf{X},\, A]}{\mathbb{P}(A)} \mathbf{1}_A \longrightarrow \mathbb{E}^{\mathbb{P}}[\mathbf{X}|\Sigma] \quad \text{both } \mathbb{P}\text{-almost surely and in } L^1(\mathbb{P}; E).$$

In particular, conclude that there is a representative $\mathbf{X}_\Sigma$ of $\mathbb{E}^{\mathbb{P}}[\mathbf{X}|\Sigma]$ which is $C$-valued and that

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

for each continuous, convex $g : C \longrightarrow [0, \infty)$.

EXERCISE 7.3.16. Again let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $E$ a separable, real Banach space. Further, suppose that $\{\mathcal{F}_n\}_0^\infty$ is a *nonincreasing* sequence of sub-$\sigma$-algebras of $\mathcal{F}$, and set $\mathcal{F}_\infty = \bigcap_0^\infty \mathcal{F}_n$. Finally, let $\mathbf{X} \in L^1(\mathbb{P}; E)$.

(**i**) Show that

$$\mathbb{E}^{\mathbb{P}}\big[\mathbf{X}\big|\mathcal{F}_n\big] \longrightarrow \mathbb{E}^{\mathbb{P}}[\mathbf{X}|\mathcal{F}_\infty] \quad \text{both } \mathbb{P}\text{-almost surely and in } L^p(\mathbb{P}; E)$$

for any $p \in [1, \infty)$ with $\mathbf{X} \in L^p(\mathbb{P}; E)$.

**Hint**: Use (7.3.13) and the approximation result in Theorem 7.1.12 to reduce to the case when $\mathbf{X}$ is $\mathbb{P}$-simple. When $\mathbf{X}$ is $\mathbb{P}$-simple, get the result as an application of (**iii**) in Exercise 7.2.27.

(**ii**) Using part (**i**) and following the line of reasoning given in part (**ii**) of Exercise 7.2.28, give another proof of The Strong Law of Large Numbers for $E$-valued random variables. (See Exercise 5.1.11 for an entirely different approach.)

EXERCISE 7.3.17. As we saw in the proof of Theorem 7.3.8, the Hardy–Littlewood maximal function can be used to dominate other quantities of interest. As further confirmation of its importance, we will use it in this exercise to prove the analogue of Theorem 7.3.8 for a large class of approximate identities. That is, let $\psi \in L^1(\mathbb{R}^N;\mathbb{R})$ with $\int_{\mathbb{R}^N} \psi\, d\mathbf{x} = 1$ be given, and set

$$\psi_t(\mathbf{x}) = t^{-N}\psi\left(\tfrac{\mathbf{x}}{t}\right), \quad t \in (0,\infty) \text{ and } \mathbf{x} \in \mathbb{R}^N.$$

Then $\{\psi_t : t > 0\}$ forms an **approximate identity** in the sense that, as tempered distributions, $\psi_t \longrightarrow \delta_{\mathbf{0}}$ as $t \searrow 0$. In fact, because

$$\|\psi_t \star f\|_{L^p(\mathbb{R}^N;\mathbb{R})} \le \|\psi\|_{L^1(\mathbb{R}^N)}\, \|f\|_{L^p(\mathbb{R}^N;\mathbb{R})}, \quad t \in (0,\infty) \text{ and } p \in [1,\infty],$$

and

$$\psi_t \star f(\mathbf{x}) = \int_{\mathbb{R}^N} \psi(\mathbf{y})\, f(\mathbf{x} - t\mathbf{y})\, d\mathbf{y},$$

it is easy to see that, for each $p \in [1,\infty)$,

$$\lim_{t\searrow 0} \bigl\|\psi_t \star f - f\bigr\|_{L^p(\mathbb{R}^N;\mathbb{R})} = 0$$

first for $f \in C_{\mathrm{c}}(\mathbb{R}^N;\mathbb{R})$ and then for all $f \in L^p(\mathbb{R}^N;\mathbb{R})$.

The purpose of this exercise is to sharpen the preceding under the assumption that

$$\psi(\mathbf{x}) = \alpha\bigl(|\mathbf{x}|\bigr), \quad \mathbf{x} \in \mathbb{R}^N \setminus \{\mathbf{0}\} \quad \text{for some } \alpha \in C^1\bigl((0,\infty);\mathbb{R}\bigr) \text{ with}$$
$$A \equiv \int_{(0,\infty)} r^N |\alpha'(r)|\, dr < \infty.$$

Notice that when $\alpha$ is nonnegative and nonincreasing, integration by parts shows that $A = N$.

**(i)** Let $f \in C_{\mathrm{c}}(\mathbb{R}^N;\mathbb{R})$ be given, and set

$$\tilde{f}(r,\mathbf{x}) = \frac{1}{|B_{\mathbb{R}^N}(\mathbf{x},r)|} \int_{B_{\mathbb{R}^N}(\mathbf{x},r)} f(\mathbf{y})\, d\mathbf{y} \quad \text{for } r \in (0,\infty) \text{ and } \mathbf{x} \in \mathbb{R}^N.$$

Using integration by parts and the given hypotheses, show that

$$\psi_t \star f(\mathbf{x}) = -\tfrac{1}{N} \int_{(0,\infty)} r^N \alpha'(r)\, \tilde{f}(tr,\mathbf{x})\, dr,$$

and conclude that

$$\bigl|\psi_t \star f(\mathbf{x})\bigr| \le \tfrac{A}{N}\, \tilde{\mathbf{M}} f(\mathbf{x}),$$

where $\tilde{\mathbf{M}}f$ is the quantity introduced at the beginning of the proof of Theorem 7.3.8. In particular, conclude that there is a constant $K_N \in (0,\infty)$, depending only on $N \in \mathbb{Z}^+$, such that

$$\mathbf{M}_\psi f(\mathbf{x}) \equiv \sup_{t\in(0,\infty)} \big|\psi_t \star f(\mathbf{x})\big| \le K_N A\,\mathbf{M}f(\mathbf{x}), \quad \mathbf{x}\in\mathbb{R}^N.$$

(**ii**) Starting from the conclusion in (**i**), show that

$$\big|\{\mathbf{x}:\, M_\psi f(\mathbf{x}) \ge R\}\big| \le \frac{(12)^N K_N A\|f\|_{L^1(\mathbb{R}^N)}}{R}, \quad f\in L^1(\mathbb{R}^N;\mathbb{R}),$$

and that for $p\in(1,\infty]$,

$$\big\|M_\psi f\big\|_{L^p(\mathbb{R}^N;\mathbb{R})} \le \frac{(12)^N K_N A\,p}{p-1}\,\|f\|_{L^p(\mathbb{R}^N;\mathbb{R})}, \quad f\in L^p(\mathbb{R}^N;\mathbb{R}).$$

Finally, proceeding as in the proof of Theorem 7.3.8, use the first of these to prove that, for Leb-almost every $\mathbf{x}\in\mathbb{R}^N$ :

$$\begin{aligned}&\varlimsup_{t\searrow 0}\big|\psi_t\star f(\mathbf{x}) - f(\mathbf{x})\big| \\ &\qquad \le \varlimsup_{t\searrow 0}\int_{\mathbb{R}^N}\big|\psi_t(\mathbf{y})\big(f(\mathbf{x}-\mathbf{y}) - f(\mathbf{x})\big)\big|\,d\mathbf{y} = 0.\end{aligned}$$

Two of the most familiar examples to which the preceding applies are the Gauss kernel $g_t(\mathbf{x}) = (2\pi t)^{-\frac{N}{2}}\exp\big(-\frac{|\mathbf{x}|^2}{2}\big)$ and the Poisson kernel (cf. (3.2.35)) $\Pi_t^{\mathbb{R}^N}$. In both these cases, $A = N$.

Exercise 7.3.18. Let $E$ be a separable Hilbert space and $(\mathbf{X}_n,\mathcal{F},\mathbb{P})$ an $E$-valued martingale on some probability space $(\Omega,\mathcal{F},\mathbb{P})$ satisfying the condition

$$\sup_{n\in\mathbb{Z}^+}\mathbb{E}^{\mathbb{P}}\big[\|\mathbf{X}_n\|_E^2\big] < \infty.$$

Proceeding as in (**i**) of Exercise 7.2.20, first prove that there is a $\bigvee_1^\infty \mathcal{F}_n$-measurable $\mathbf{X}\in L^2(\mathbb{P};E)$ to which $\{\mathbf{X}_n\}_1^\infty$ converges in $L^2(\mathbb{P};E)$, next check that

$$\mathbf{X}_n = \mathbb{E}^{\mathbb{P}}\big[\mathbf{X}\big|\mathcal{F}_n\big] \quad (\text{a.s.},\mathbb{P}) \text{ for each } n\in\mathbb{Z}^+,$$

and finally apply the last part of Theorem 7.3.12 to see that $\mathbf{X}_n \longrightarrow \mathbf{X}$ $\mathbb{P}$-almost surely.