PROBABILITY THEORY - PART 3
MARTINGALES

MANJUNATH KRISHNAPUR

CONTENTS

1. CONDITIONAL EXPECTATION

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space. Let $\mathcal{G} \subseteq \mathcal{F}$ be a sub sigma algebra of $\mathcal{F}$. Let $X : \Omega \to \mathbb{R}$ be a real-valued integrable random variable, i.e., $\mathbf{E}[|X|] < \infty$. A random variable $Y : \Omega \to \mathbb{R}$ is said to be a conditional expectation of X given $\mathcal{G}$ is (a) Y is $\mathcal{G}$-measurable, (b) $\mathbf{E}[|Y|] < \infty$, and (c) $\int\limits_A Y d\mathbf{P} = \int\limits_A X d\mathbf{P}$ for all $A \in \mathcal{G}$.

We shall say that any such Y is a version of $\mathbf{E}[X \mid \mathcal{G}]$. The notation is justified, since we shall show shortly that such a random variable always exists and is unique up to $\mathbf{P}$-null sets in $\mathcal{G}$.

Example 1. *Let $B, C \in \mathcal{F}$. Let $\mathcal{G} = \{\emptyset, B, B^c, \Omega\}$ and let $X = \mathbf{1}_C$. Since $\mathcal{G}$-measurable random variables must be constant on B and on B^c, we must take $Y = \alpha \mathbf{1}_B + \beta \mathbf{1}_{B^c}$. Writing the condition for equality of integrals of Y and X over B and over B^c, we get $\alpha \mathbf{P}(B) = \mathbf{P}(C \cap B)$, $\beta \mathbf{P}(B^c) = \mathbf{P}(C \cap B^c)$. It is easy to see that with then the equality also holds for integrals over $\emptyset$ and over Ω. Hence, the unique choice for conditional expectation of X given $\mathcal{G}$ is*

$$Y(\omega) = \begin{cases} \mathbf{P}(C \cap B)/\mathbf{P}(B) & \text{if } \omega \in B, \\ \mathbf{P}(C \cap B^c)/\mathbf{P}(B^c) & \text{if } \omega \in B^c. \end{cases}$$

This agrees with the notion that we learned in basic probability classes. If we get to know that B happened, we update our probability of C to $\mathbf{P}(C \cap B)/\mathbf{P}(B)$ and if we get to know that B^c happened, we update it to $\mathbf{P}(C \cap B^c)/\mathbf{P}(B^c)$.

Exercise 2. *Suppose $\Omega = \sqcup_{k=1}^n B_k$ is a partition of Ω where $B_k \in \mathcal{F}$ and $\mathbf{P}(B_k) > 0$ for each k. Then show that the unique conditional expectation of $\mathbf{1}_C$ given $\mathcal{G}$ is*

$$\sum_{k=1}^n \frac{\mathbf{P}(C \cap B_k)}{\mathbf{P}(B_k)} \mathbf{1}_{B_k}.$$

Example 3. *Suppose Z is $\mathbb{R}^d$-valued and (X, Z) has density $f(x, z)$ with respect to Lebesgue measure on $\mathbb{R} \times \mathbb{R}^d$. Let $\mathcal{G} = \sigma(Z)$. Then, show that a version of $\mathbf{E}[X \mid \mathcal{G}]$ is*

$$Y(\omega) = \begin{cases} \frac{\int\limits_{\mathbb{R}} x f(x, Z(\omega)) dx}{\int\limits_{\mathbb{R}} f(x, Z(\omega)) dx} & \text{if the denominator is positive}, \\ 0 & \text{otherwise}. \end{cases}$$

Recall that $\mathcal{G}$-measurable random variables are precisely those of the form $h(Z)$ where $h : \mathbb{R}^d \to \mathbb{R}$ is a Borel measurable function. Here, it is clear that the set of ω for which $\int f(x, Z(\omega)) dx$ is zero is a $\mathcal{G}$-measurable set. Hence, Y defined above is $\mathcal{G}$-measurable.

We leave it as an exercise to check that Y is a version of $\mathbf{E}[X \mid \mathcal{G}]$.

Example 4. *This is really a class of examples. Assume that $\mathbf{E}[X^2] < \infty$. Then, we can show that existence of $\mathbf{E}[X \mid \mathcal{G}]$ by an elementary Hilbert space argument. Recall that $H = L^2(\Omega, \mathcal{F}, \mathbf{P})$ and $W = L^2(\Omega, \mathcal{G}, \mathbf{P})$ (here we write $\mathbf{P}$ again to mean $\mathbf{P}$ restricted to $\mathcal{G}$) are Hilber spaces and W is a closed subspace of H.*

Since elements of H and W are equivalence classes of random variables, let us write $[X]$ for the equivalence class of a random variables X (strictly, we should write $[X]_H$ and $[X]_W$, but who has the time?). By elementary Hilbert space theory, there exists a unique projection operator $P_W : H \to W$ such that $P_W v \in W$ and $v - P_W v \in W^\perp$ for each $v \in H$. In fact, $P_W v$ is the closest point in W to u.

Now let Y be any $\mathcal{G}$-measurable random variable such that $[Y] = P_W[X]$. We claim that Y is a version of $\mathbf{E}[X \mid \mathcal{G}]$. Indeed, if Z is $\mathcal{G}$-measurable and square integrable, then $\mathbf{E}[(X - Y)Z] = 0$ because $[X] - [Y] \in W^\perp$ and $[Z] \in W$. In particular, taking $Z = \mathbf{1}_A$ for any $A \in \mathcal{G}$, we get $\mathbf{E}[X\mathbf{1}_A] = \mathbf{E}[Y\mathbf{1}_A]$. This is the defining property of conditional expectation.

For later purpose, we note that the projection operator that occurs above has a special property (which does not even make sense for a general orthogonal projection in a Hilbert space).

Exercise 5. *If $X \geq 0$ $a.s.$ and $\mathbf{E}[X^2] < \infty$, show that $P_W[X] \geq 0$ $a.s.$ [**Hint:** If $[Y] = P_W[X]$, then $\mathbf{E}[(X - Y_+)^2] \leq \mathbf{E}[(X - Y)^2]$ with equality if and only if $Y \geq 0$ $a.s.$]*

Uniqueness of conditional expectation: Suppose Y_1, Y_2 are two versions of $\mathbf{E}[X \mid \mathcal{G}]$. Then $\int_A Y_1 d\mathbf{P} = \int_A Y_2 d\mathbf{P}$ for all $A \in \mathcal{G}$, since both are equal to $\int_A X d\mathbf{P}$. Let $A = \{\omega : Y_1(\omega) > Y_2(\omega)\}$. Then the equality $\int_A (Y_1 - Y_2) d\mathbf{P} = 0$ can hold if and only if $\mathbf{P}(A) = 0$ (since the integrand is positive on A). Similarly $\mathbf{P}\{Y_2 - Y_1 > 0\} = 0$. This, $Y_1 = Y_2$ $a.s.$ (which means that $\{Y_1 \neq Y_2\}$ is $\mathcal{G}$-measurable and has zero probability under $\mathbf{P}$).

Thus, conditional expectation, if it exists, is unique up to almost sure equality.

Existence of conditional expectation: There are two approaches to this question.

First approach: Radon-Nikodym theorem: Let $X \geq 0$ and $\mathbf{E}[X] < \infty$. Then consider the measure $\nu : \mathcal{G} \to [0, 1]$ defined by $\nu(A) = \int_A X d\mathbf{P}$ (we assumed non-negativity so that $\nu(A) \geq 0$ for all $A \in \mathcal{G}$). Further, $\mathbf{P}$ is a probability measure when restricted to $\mathcal{G}$ (we continue to denote it by $\mathbf{P}$). It is clear that if $A \in \mathcal{G}$ and $\mathbf{P}(A) = 0$, then $\nu(A) = 0$. In other words, ν is absolutely continuous to $\mathbf{P}$ on $(\Omega, \mathcal{G})$. By the Radon-Nikodym theorem, there exists $Y \in L^1(\Omega, \mathcal{G}, \mathbf{P})$ such that $\nu(A) = \int_A Y d\mathbf{P}$ for all $A \in \mathcal{G}$. Thus, Y is $\mathcal{G}$-measurable and $\int_A Y d\mathbf{P} = \int_A X d\mathbf{P}$ (the right side is $\nu(A)$). Thus, Y is a version of $\mathbf{E}[X \mid \mathcal{G}]$.

For a general integrable random variable X, let $X = X_+ - X_-$ and let Y_+ and Y_- be versions of $\mathbf{E}[X_+ \mid \mathcal{G}]$ and $\mathbf{E}[X_- \mid \mathcal{G}]$, respectively. Then $Y = X_+ - X_-$ is a version of $\mathbf{E}[X \mid \mathcal{G}]$.

Remark 6. *Where did we use the integrability of X in all this? When $X \geq 0$, we did not! In other words, for a non-negative random variable X (even if not integrable), there exists a Y taking values in $\mathbb{R}_+ \cup \{+\infty\}$*

such that Y is $\mathcal{G}$-measurable and $\int_A Y d\mathbf{P} = \int_A X d\mathbf{P}$. However, it is worth noting that if X is integrable, so is Y.

In the more general case, if Y_+ and Y_- are both infinite on a set of positive measure and then $Y_+ - Y_-$ is ill-defined on that set. Therefore, it is best to assume that $\mathbf{E}[|X|] < \infty$ so that Y_+ and Y_- are finite $a.s.$

Second approach: Approximation by square integrable random variables: Let $X \geq 0$ be an integrable random variable. Let $X_n = X \wedge n$ so that X_n are square integrable (in fact bounded) and $X_n \uparrow X$. Let Y_n be versions of $\mathbf{E}[X_n \mid \mathcal{G}]$, defined by the projections $P_W[X_n]$ as discussed earlier.

Now, $X_{n+1} - X_n \geq 0$, hence by the exercise above $P_W[X_{n+1} - X_n] \geq 0$ $a.s.$, hence by the linearity of projection, $P_W[X_n] \leq P_W[X_{n+1}]$ $a.s.$ In other words, $Y_n(\omega) \leq Y_{n+1}(\omega)$ for all $\omega \in \Omega_n$ where $\Omega_n \in \mathcal{G}$ is such that $\mathbf{P}(\Omega_n) = 1$. Then, $\Omega' := \cap_n \Omega_n$ is in $\mathcal{G}$ and has probability 1, and for $\omega \in \Omega'$, the sequence $Y_n(\omega)$ is non-decreasing.

Define $Y(\omega) = \lim_n Y_n(\omega)$ if $\omega \in \Omega'$ and $Y(\omega) = 0$ for $\omega \notin \Omega'$. Then Y is $\mathcal{G}$-measurable. Further, for any $A \in \mathcal{G}$, by MCT we see that $\int_A Y_n d\mathbf{P} \uparrow \int_A Y d\mathbf{P}$ and $\int_A X_n d\mathbf{P} \uparrow X d\mathbf{P}$. If $A \in \mathcal{G}$, then $\int_A Y_n d\mathbf{P} = \int_A X_n d\mathbf{P}$. Thus, $\int_A Y d\mathbf{P} \int_A X d\mathbf{P}$. This proves that Y is a conditional expectation of X given $\mathcal{G}$.

2. Conditional probability

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space and let $\mathcal{G}$ be a sub sigma algebra of $\mathcal{F}$. *Regular conditional probability* of $\mathbf{P}$ given $\mathcal{G}$ is any function $Q : \Omega \times \mathcal{F} \to [0, 1]$ such that

(1) For $\mathbf{P}$-$a.e.$ $\omega \in \Omega$, the map $A \to Q(\omega, A)$ is a probability measure on $\mathcal{F}$.

(2) For each $A \in \mathcal{G}$, then map $\omega \to Q(\omega, A)$ is a version of $\mathbf{E}[\mathbf{1}_A \mid \mathcal{G}]$.

The second condition of course means that for any $A \in \mathcal{F}$, the random variable $Q(\cdot, A)$ is $\mathcal{G}$-measurable and $\int_B Q(\omega, A) d\mathbf{P}(\omega) = \mathbf{P}(A \cap B)$ for all $B \in \mathcal{G}$.

Unlike conditional expectation, conditional probability does not necessarily exist[1]. Why is that? Given $(\Omega, \mathcal{F}, \mathbf{P})$ and $\mathcal{G} \subseteq \mathcal{F}$, consider the conditional expectations $Q(\omega, B) := \mathbf{E}[\mathbf{1}_B \mid \mathcal{G}]$ for $\omega \in \Omega$ and $B \in \mathcal{F}$. Can we not simply prove that Q is a conditional probability? The second property is satisfied by definition. We require $B \to Q(\omega, B)$ to be a probability measure for $a.e.$ ω. In fact, if $B_n \uparrow B$, then the conditional MCT says that $\mathbf{E}[\mathbf{1}_{B_n} \mid \mathcal{G}] \uparrow \mathbf{E}[\mathbf{1}_B \mid \mathcal{G}]$ $a.s.$ Should this not give countable additivity of $Q(\omega, \cdot)$? The issue is in the choice of versions and the fact that there are uncountably many such sequences. For each $B \in \mathcal{F}$, if we choose (and fix) a version of $\mathbf{E}[\mathbf{1}_B \mid \mathcal{G}]$, then for each ω, it might well be possible to find a sequence $B_n \uparrow B$ for which $\mathbf{E}[\mathbf{1}_{B_n} \mid \mathcal{G}](\omega)$ does not increase to $\mathbf{E}[\mathbf{1}_B \mid \mathcal{G}]$. This is why, the existence of conditional probability is not trivial.

But it does exist in all cases of interest.

[1]So I have heard. If I ever saw a counterexample, I have forgotten it.

Theorem 7. *Let M be a complete and separable metric space and let $\mathcal{B}_M$ be its Borel sigma algebra. Then, for any Borel probability measure $\mathbf{P}$ on $(M, \mathcal{B}_M)$ and any sub sigma algebra $\mathcal{G} \subseteq \mathcal{B}_M$, a regular conditional probability Q exists. It is unique in the sense that if Q' is another regular conditional probability, then $Q(\omega, \cdot) = Q'(\omega, \cdot)$ for $\mathbf{P}$-$a.e.$ $\omega \in M$.*

We shall prove this for the special case when $\Omega = \mathbb{R}$. The same proof can be easily written for $\Omega = \mathbb{R}^d$, with only minor notational complication. The above general fact can be deduced from the following fact that we state without proof[2].

Theorem 8. *Let (M, d) be a complete and separable metric space. Then, there exists a bijection $\varphi : M \to [0, 1]$ such that φ and φ^{-1} are both Borel measurable.*

With this fact, any question of measures on a complete, separable metric space can be transferred to the case of $[0, 1]$. In particular, the existence of regular conditional probabilities can be deduced. Note that the above theorem applies to (say) $M = (0, 1)$ although it is not complete in the usual metric. Indeed, one can put a complete metric on $(0, 1)$ (how?) without changing the topology (and hence the Borel sigma algebra) and then apply the above theorem.

A topological space whose topology can be induced by a metric that makes it complete and separable is called a *Polish space* (named after Polish mathematicians who studied it, perhaps Ulam and others). In this language, Theorem 8 says that any Polish space is Borel isomorphic to $(\mathbb{R}, \mathcal{B}_\mathbb{R})$ and Theorem 7 says that regular conditional probabilities exist for any Borel probability measure on $\mathcal{B}$ with respect to an arbitrary sub sigma algebra thereof.

Proof of Theorem 7 when $M = \mathbb{R}$. We start with a Borel probability measure $\mathbf{P}$ on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$ and $\mathcal{G} \subseteq \mathcal{B}_\mathbb{R}$. For each $t \in \mathbb{Q}$, let Y_t be a version of $\mathbf{E}[\mathbf{1}_{(-\infty,t]} \mid \mathcal{G}]$. For any rational $t < t'$, we know that $Y_t(\omega) \le Y_{t'}(\omega)$ for all $\omega \notin N_{t,t'}$ where $N_{t,t'}$ is a Borel set with $\mathbf{P}(N_{t,t'}) = 0$. Further, by the conditional MCT, there exists a Borel set N_* with $\mathbf{P}(N) = 0$ such that for $\omega \notin N_*$, we have $\lim_{t\to\infty} Y_t(\omega) = 1$ and $\lim_{t\to-\infty} Y_t = 0$ where the limits are taken through rationals only.

Let $N = \bigcup_{t,t'} N_{t,t'} \cup N_*$ so that $\mathbf{P}(N) = 0$ by countable additivity. For $\omega \notin N$, the function $t \to Y_t(\omega)$ from $\mathbb{Q}$ to $[0, 1]$ is non-decreasing and has limits 1 and 0 at $+\infty$ and $-\infty$, respectively. Now define $F : \Omega \times \mathbb{R} \to [0, 1]$ by

$$F(\omega, t) = \begin{cases} \inf\{Y_s(\omega) : s \ge t, s \in \mathbb{Q}\} & \text{if } \omega \notin N, \\ 0 & \text{if } \omega \in N. \end{cases}$$

By exercise 9 below, for any $\omega \notin N$, we see that $F(\omega, \cdot)$ is the CDF of some probability measure μ_ω on $\mathbb{R}$, provided $\omega \notin N$. Define $Q : \Omega \times \mathcal{B}_\mathbb{R} \to [0, 1]$ by $Q(\omega, A) = \mu_\omega(A)$. We claim that Q is a conditional probability of $\mathbf{P}$ given $\mathcal{G}$.

The first condition, that $Q(\omega, \cdot)$ be a probability measure on $\mathcal{B}_\mathbb{R}$ is satisfied by construction. We only need to prove that $Q(\cdot, A)$ is a version of $\mathbf{E}[\mathbf{1}_A \mid \mathcal{G}]$. Observe that the collection of $A \in \mathcal{B}_\mathbb{R}$

[2] For a proof, see Chapter 13 of Dudley's book *Real analysis and probability*.

for which this is true, forms a sigma-algebra. Hence, it suffices to show that this statement is true for $A = (-\infty, t]$ for any $t \in \mathbb{R}$. For fixed t, by definition $Q(\omega, (-\infty, t])$ is the decreasing limit of $Y_s(\omega) = \mathbf{E}[\mathbf{1}_{(-\infty,s]} \,|\, \mathcal{G}](\omega)$ as $s \downarrow t$, whenever $\omega \notin N$. By the conditional MCT it follows that $Q(\cdot, (-\infty, t]) = \mathbf{E}[\mathbf{1}_{(-\infty,t]} \,|\, \mathcal{G}]$. This completes the proof. ■

The following exercise was used in the proof.

Exercise 9. *Let $f : \mathbb{Q} \to [0, 1]$ be a non-decreasing function such that $f(t)$ converges to 1 or 0 according as $t \to +\infty$ or $t \to -\infty$, respectively. Then define $F : \mathbb{R} \to [0, 1]$ by $F(t) = \inf\{f(q) : t \le q \in \mathbb{Q}\}$. Show that F is a CDF of a probability measure.*

Proof of Theorem 7 for general M, assuming Theorem 8. Let $\varphi : M \to \mathbb{R}$ be a Borel isomorphism. That is φ is bijective and φ, φ^{-1} are both Borel measurable. We are given a probability measure $\mathbf{P}$ on $(M, \mathcal{B}_M)$ and a sigma algebra $\mathcal{G} \subseteq \mathcal{B}_M$. Let $\mathbf{P}' = \mathbf{P} \circ \varphi^{-1}$ be its pushforward probability measure on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$. Let $\mathcal{G}' = \{\varphi(A) : A \in \mathcal{G}\}$, clearly a sub sigma algebra of $\mathcal{B}_\mathbb{R}$.

From the already proved case, we get $Q' : \mathbb{R} \times \mathcal{B}_M \to [0, 1]$, a conditional probability of $\mathbf{P}'$ given $\mathcal{G}'$. Define $Q : M \times \mathcal{B}_M \to [0, 1]$ by $Q(\omega, A) = Q'(\varphi(\omega), \varphi(A))$. Check that Q' is a conditional probability of $\mathbf{P}$ given $\mathcal{G}$. ■

3. Relationship between conditional probability and conditional expectation

Let M be a complete and separable metric space (or in terms introduced earlier, a Polish space). Let $\mathbf{P}$ be a probability measure on $\mathcal{B}_M$ and let $\mathcal{G} \subseteq \mathcal{B}_M$ be a sub sigma algebra. Let Q be a regular conditional probability for $\mathbf{P}$ given $\mathcal{G}$ which exists, as discussed in the previous section. Let $X : M \to \mathbb{R}$ be a Borel measurable, integrable random variable. We defined the conditional expectation $\mathbf{E}[X \,|\, \mathcal{G}]$ in the first section. We now claim that the conditional expectation is actually the expectation with respect to the conditional probability measure. In other words, we claim that

$$\mathbf{E}[X \,|\, \mathcal{G}](\omega) = \int_M X(\omega') dQ_\omega(\omega') \tag{1}$$

where $Q_\omega(\cdot)$ is a convenient notation probability measure $Q(\omega, \cdot)$ and $dQ_\omega(\omega')$ means that we use Lebesgue integral with respect to the probability measure Q_ω (thus ω' is a dummy variable which is integrated out).

To show this, it suffices to argue that the right hand side of (1) is $\mathcal{G}$-measurable, integrable and that its integral over $A \in \mathcal{G}$ is equal to $\int_A X d\mathbf{P}$.

Firstly, let $X = \mathbf{1}_B$ for some $B \in \mathcal{B}_M$. Then, the right hand side is equal to $Q_\omega(B) = Q(\omega, B)$. By definition, this is a version of $\mathbf{E}[\mathbf{1}_B \,|\, \mathcal{G}]$. By linearity, we see that (1) is valid whenever X is a simple random variable.

If X is a non-negative random variable, then we can find simple random variables $X_n \geq 0$ that increase to X. For each n

$$\mathbf{E}[X_n \,|\, \mathcal{G}](\omega) = \int\limits_M X_n(\omega')dQ_\omega(\omega') \;\text{ a.e.}\omega[\mathbf{P}].$$

The left side increases to $\mathbf{E}[X \,|\, \mathcal{G}]$ for $a.e..$ ω by the conditional MCT. For fixed $\omega \notin N$, the right side is an ordinary Lebesgue integral with respect to a probability measure Q_ω and hence the usual MCT shows that it increases to $\int\limits_M X(\omega')dQ_\omega(\omega')$. Thus, we get (1) for non-negative random variables.

For a general integrable random variable X, write it as $X = X_+ - X_-$ and use (1) individually for $X_\pm$ and deduce the same for X.

Remark 10. *Here we explain the reasons why we introduced conditional probability. In most books on martingales, only conditional expectation is introduced and is all that is needed. However, when conditional probability exists, conditional expectation becomes an actual expectation with respect to a probability measure. This makes it simpler to not have to prove many properties for conditional expectation as we shall see in the following section. Also, it is aesthetically pleasing to know that conditional probability exists in most circumstances of interest.*

A more important point is that, for discussing Markov processes (as we shall do when we discuss Brownian motion), conditional probability is the more natural language in which to speak.

4. Properties of conditional expectation

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space. We write $\mathcal{G}, \mathcal{G}_i$ for sub sigma algebras of $\mathcal{F}$ and X, X_i for integrable $\mathcal{F}$-measurable random variables on Ω.

(1) Linearity: For $\alpha, \beta \in \mathbb{R}$, we have $\mathbf{E}[\alpha X_1 + \beta X_2 \,|\, \mathcal{G}] = \alpha\mathbf{E}[X_1 \,|\, \mathcal{G}] + \beta\mathbf{E}[X_2 \,|\, \mathcal{G}]$ $a.s.$

(2) Positivity: If $X \geq 0$ $a.s.$, then $\mathbf{E}[X \,|\, \mathcal{G}] \geq 0$ $a.s.$ Consequently (consider $X_2 - X_1$), if $X_1 \leq X_2$, then $\mathbf{E}[X_1 \,|\, \mathcal{G}] \leq \mathbf{E}[X_2 \,|\, \mathcal{G}]$. Also $|\mathbf{E}[X \,|\, \mathcal{G}]| \leq \mathbf{E}[|X| \,|\, GG]$.

(3) Conditional MCT: If $0 \leq X_n \uparrow X$ $a.s.$, then $\mathbf{E}[X_n \,|\, \mathcal{G}] \uparrow \mathbf{E}[X \,|\, \mathcal{G}]$ $a.s.$ Here either assume that X is integrable or make sense of the conclusion using Remark 6.

(4) Conditional Fatou's: Let $0 \leq X_n$. Then, $\mathbf{E}[\liminf X_n \,|\, \mathcal{G}] \leq \liminf \mathbf{E}[X_n \,|\, \mathcal{G}]$ $a.s.$

(5) Conditional DCT: Let $X_n \stackrel{a.s.}{\to} X$ and assume that $|X_n| \leq Y$ for some Y with finite expectation, then $\mathbf{E}[X_n \,|\, \mathcal{G}] \stackrel{a.s.}{\to} \mathbf{E}[X \,|\, \mathcal{G}]$.

(6) Conditional Jensen's inequality: If $\varphi : \mathbb{R} \to \mathbb{R}$ is convex and X and $\varphi(X)$ are integrable, then $\mathbf{E}[\varphi(X) \,|\, \mathcal{G}] \geq \varphi(\mathbf{E}[X \,|\, \mathcal{G}])$. In particular, of $\mathbf{E}[|X|^p] < \infty$ for some $p \geq 1$, then $\mathbf{E}[|X|^p \,|\, \mathcal{G}] \geq (\mathbf{E}[|X| \,|\, \mathcal{G}])^p$.

(7) Conditional Cauchy-Schwarz: If $\mathbf{E}[X^2], \mathbf{E}[Y^2] < \infty$, then $(\mathbf{E}[XY \,|\, \mathcal{G}])^2 \leq \mathbf{E}[X^2 \,|\, \mathcal{G}]\mathbf{E}[Y^2 \,|\, \mathcal{G}]$.

(8) Tower property: If $\mathcal{G}_1 \subseteq \mathcal{G}_2$, then $\mathbf{E}[\mathbf{E}[X \mid \mathcal{G}_2] \mid \mathcal{G}_1] = \mathbf{E}[X \mid \mathcal{G}_1]$ $a.s.$ In particular (taking $\mathcal{G} = \{\emptyset, \Omega\}$), we get $\mathbf{E}[\mathbf{E}[X \mid \mathcal{G}]] = \mathbf{E}[X]$.

(9) $\mathcal{G}$-measurable random variables are like constants for conditional expectation: For any bounded $\mathcal{G}$-measurable random variable Z, we have $\mathbf{E}[XZ \mid \mathcal{G}] = Z\mathbf{E}[X \mid \mathcal{G}]$ $a.s.$

If we assume that Ω is a Polish space and $\mathcal{F}$ is its Borel sigma algebra, then no proofs are needed! Indeed, all except the last two properties (highlighted in color) have analogues for expectation (Lebesue integral). And we saw that when conditional probability exists, then conditional expectation is just expectation with respect to conditional probability measure. Thus, ω by ω, the properties above hold for conditional expectations[3].

But the assumption that conditional probability exists is not necessary for the above properties to hold. In most books you will see direct proofs (which is why I presented it in this slightly different manner). To understand what those proofs actually do, let us revisit the reason why conditional probability may not exist in general. The difficulty there was that $B \to \mathbf{E}[\mathbf{1}_B \mid \mathcal{G}](\omega)$ fails countable additivity for each fixed ω, for some sequence $B_n \uparrow B$. But if we restrict attention to countably many sequences, then we can find a common set of zero probability outside of which there is no problem. Essentially in proving each of the above properties, we use only a finite or countable number of such sequences and that is what those proofs do. If you prefer it that way, then please consult the referred books.

In view of all this discussion, only the last two of the above stated properties needs to be proved.

Property (9): First consider the last property. If $Z\mathbf{1}_B$ for some $B \in \mathcal{G}$, it is the very definition of conditional expectation. From there, deduce the property when Z is a simple random variable, a non-negative random variable, a general integrable random variable. We leave the details as an exercise.

Property (8): Now consider the tower property which is of enormous importance to us. But the proof is straightforward. Let $Y_1 = \mathbf{E}[X \mid \mathcal{G}_1]$ and $Y_2 = \mathbf{E}[X \mid \mathcal{G}_2]$. If $A \in \mathcal{G}_1$, then by definition, $\int_A Y_1 d\mathbf{P} = \int_A X d\mathbf{P}$. Further, $\int_A Y_2 d\mathbf{P} = \int_A X d\mathbf{P}$ since $A \in \mathcal{G}_1$ too. This shows that $\int_A Y_1 d\mathbf{P} = \int_A Y_2 d\mathbf{P}$ for all $A \in \mathcal{G}_1$. Further, Y_1 is $\mathcal{G}_1$-measurable. Hence, it follows that $Y_1 = \mathbf{E}[Y_2 \mid \mathcal{G}]$. This is what is claimed there.

5. Cautionary tales on conditional probability

Even when knows all the definitions in and out, it is easy to make mistakes with conditional probability. Extreme caution is advocated! Practising some explicit computations also helps.

[3]You may complain that conditional MCT was used to show existence of conditional probability, then is it not circular reasoning to use conditional probability to prove conditional MCT? Indeed, at least a limited form of conditional MCT was already used. But the derivation of other properties using conditional probability is not circular.

Here is an example. Let (U,V) be uniform on $[0,1]^2$. Consider the line segment $L = \{(u,v) \in [0,1]^2 : u = 2v\}$. What is the distribution of (U,V) conditioned on the even that it lies on L? This question is ambiguous as we shall see.

If one want to condition on an event of positive probability, there is never an ambiguity, the answer is $\mathbf{P}(A \mid B) = P(A \cap B)/\mathbf{P}(B)$. If you take the more advanced viewpoint of conditioning on the sigma-algebra $\mathcal{G} = \{\emptyset, B, B^c, \Omega\}$, the answer is the same. Or more precisely, you will say that $\mathbf{E}[\mathbf{1}_A \mid \mathcal{G}] = \mathbf{P}(A \mid B)\mathbf{1}_B + \mathbf{P}(A \mid B^c)\mathbf{1}_{B^c}$ where $\mathbf{P}(A \mid B)$ and $\mathbf{P}(A \mid B^c)$ are given by the elementary conditioning formulas. But the meaning is the same.

If B has zero probability, then it does not matter how we define the conditional probability given $\mathcal{G}$ for $\omega \in B$, hence this is not the way to go. So what do we mean by conditioning on the event $B = \{(U,V) \in L\}$ which has zero probability?

One possible meaning is to set $Y = 2U - V$ and condition on $\sigma\{Y\}$. We get a conditional measure μ_y for y in the range of Y. Again, it is possible to change μ_y for finitely many y, but it would be unnatural, since $y \mapsto \mu_y$ can be checked to be continuous (in whatever reasonable sense you think of). Hence, there is well-defined μ_0 satisfying $\mu_0(L) = 1$. That is perhaps means by the conditional distribution of (U,V) given that it belongs to L?

But this is not the only possible way. We could set $Z = V/U$ and condition on $\sigma\{Z\}$ and get conditional measure ν_z for z in the range of Z. Again, $z \mapsto \nu_z$ is continuous, and a well-defined measure ν_2, satisfying $\nu_2(L) = 1$. It is a rival candidate for the answer to our question.

Which is the correct one? Both! The question was ambiguous to start with. To summarize, conditioning on a positive probability even is unambiguous. When conditioning on a zero probability event B, we must condition on a random variable Y such that $B = Y^{-1}\{y_0\}$ for some y_0. If there is some sort of continuity in the conditional distributions, we get a probability measure μ_{y_0} supported on B. But the choice of Y is not unique, hence the answer depends on what Y you choose.

A more naive, but valid way to think of this is to approximate B by positive probability events B_n such that $B_n \downarrow B$. If the probability measures $\mathbf{P}(\cdot \mid B_n)$ have a limit as $n \to \infty$, that can be candidate for our measure. But again, there are different ways to approximate B by B_ns, so there is no unambiguous answer.

6. Martingales

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space. Let $\mathcal{F}_\bullet = (\mathcal{F}_n)_{n \in \mathbb{N}}$ be a collection of sigma subalgebras of $\mathcal{F}$ indexed by natural numbers such that $\mathcal{F}_m \subseteq \mathcal{F}_n$ whenever $m < n$. Then we say that $\mathcal{F}_\bullet$ is a *filtration*. Instead of $\mathbb{N}$, a filtration may be indexed by other totally ordered sets like $\mathbb{R}_+$ or $\mathbb{Z}$ or $\{0, 1, \ldots, n\}$ etc. A sequence of random variables $X = (X_n)_{n \in \mathbb{N}}$ defined on $(\Omega, \mathcal{F}, \mathbf{P})$ is said to be *adapted* to the filtration $\mathcal{F}_\bullet$ if $X_n \in \mathcal{F}_n$ for each n.

Definition 11. In the above setting, let $X = (X_n)_{n \in \mathbb{N}}$ be adapted to $\mathcal{F}_\bullet$. We say that X is a *super-martingale* if $\mathbf{E}|X_n| < \infty$ for each $n \geq 0$ and $\mathbf{E}[X_n \mid \mathcal{F}_{n-1}] \leq X_{n-1}$ $a.s.$ for each $n \geq 1$.

We say that X is a *sub-martingale* if $-X$ is a super-martingale. If X is both a super-martingale and a sub-martingale, then we say that X is a *martingale*. When we want to explicitly mention the filtration, we write $\mathcal{F}_\bullet$-martingale or $\mathcal{F}_\bullet$-super-martingale etc.

Observe that from the definition of super-martingale, it follows that $\mathbf{E}[X_n \mid \mathcal{F}_m] \le X_m$ for any $m < n$. If the index set is $\mathbb{R}_+$, then the right way to define a super-martingales is to ask for $\mathbf{E}[X_t \mid \mathcal{F}_s] \le X_s$ for any $s < t$ (since, the "previous time point" $t-1$ does not make sense!).

Exercise 12. *Given $X = (X_n)_{n\in\mathbb{N}}$, the*

Unlike say Markov chains, the definition of martingales does not appear to put too strong a restriction on the distributions of X_n, it is only on a few conditional expectations. Nevertheless, very power theorems can be proved at this level of generality, and there are any number of examples to justify making a definition whose meaning is not obvious on the surface. In this section we give classes of examples.

Example 13 (Random walk)**.** *Let ξ_n be independent random variables with finite mean and let $\mathcal{F}_n = \sigma\{\xi_1, \dots, \xi_n\}$ (so $\mathcal{F}_0 = \{\emptyset, \Omega\}$). Define $X_0 = 0$ and $X_n = \xi_1 + \ldots + \xi_n$ for $n \ge 1$. Then, X is $\mathcal{F}_\bullet$-adapted, X_n have finite mean, and*

$$\begin{aligned}\mathbf{E}[X_n \mid \mathcal{F}_{n-1}] &= \mathbf{E}[X_{n-1} + \xi_n \mid \mathcal{F}_{n-1}] \\ &= \mathbf{E}[X_{n-1} \mid \mathcal{F}_{n-1}] + \mathbf{E}[\xi_n \mid \mathcal{F}_{n-1}] \\ &= X_{n-1} + \mathbf{E}[\xi_n]\end{aligned}$$

since $X_{n-1} \in \mathcal{F}_{n-1}$ and ξ_n is independent of $\mathcal{F}_{n-1}$. Thus, if $\mathbf{E}[\xi_n]$ is positive for all n, then X is a sub-martingale; if $\mathbf{E}[\xi_n]$ is negative for all n, then X is a super-martingale; if $\mathbf{E}[\xi_n] = 0$ for all n, then X is a martingale.

Example 14 (Product martingale)**.** *Let ξ_n be independent, non-negative random variables and let $X_n = \xi_1\xi_2 \ldots \xi_n$ and $X_0 = 1$. Then, with $\mathcal{F}_n = \sigma\{\xi_1, \dots, \xi_n\}$, we see that X is $\mathcal{F}_\bullet$-adapted and $\mathbf{E}[X_n]$ exists (equals the product of $\mathbf{E}[\xi_k]$, $k \le n$). Lastly,*

$$\mathbf{E}[X_n \mid \mathcal{F}_{n-1}] = \mathbf{E}[X_{n-1}\xi_n \mid \mathcal{F}_{n-1}] = X_{n-1}\mu_n$$

where $\mu_n = \mathbf{E}[\xi_n]$. Hence, if $\mu_n \ge 1$ for all n, then X is a sub-martingale, if $\mu_n = 1$ for all n, then X is a martingale, and if $\mu_n \le 1$ for all n, then X is a super-martingale.

In particular, replacing ξ_n by ξ_n/μ_n, we see that $Y_n := \frac{X_n}{\mu_1 \cdots \mu_n}$ is a martingale.

Example 15 (Doob martingale)**.** *Here is a very general way in which any (integrable) random variable can be put at the end of a martingale sequence. Let X be an integrable random variable on $(\Omega, \mathcal{F}, \mathbf{P})$ and let $\mathcal{F}_\bullet$ be any filtration. Let $X_n = \mathbf{E}[X \mid \mathcal{F}_n]$. Then, (X_n) is $\mathcal{F}_\bullet$-adapted, integrable and*

$$\mathbf{E}[X_n \mid \mathcal{F}_{n-1}] = \mathbf{E}\left[\mathbf{E}\left[X \mid \mathcal{F}_n\right] \mid \mathcal{F}_{n-1}\right] = \mathbf{E}[X \mid \mathcal{F}_{n-1}] = X_{n-1}$$

by the tower property of conditional expectation. Thus, (X_n) is a martingale. Such martingales got by conditioning one random variable w.r.t. an increasing family of sigma-algebras is called a Doob martingale[4].

Often $X = f(\xi_1, \ldots, \xi_m)$ is a function of independent random variables ξ_k, and we study X be sturying the evolution of $\mathbf{E}[X \mid \xi_1, \ldots, \xi_k]$, revealing the information of xi_ks, one by one. This gives X as the end-point of a Doob martingale. The usefulness of this construction will be clear in a few lectures.

Example 16 (Increasing process). *Let A_n, $n \geq 0$, be a sequence of random variables such that $A_0 \leq A_1 \leq A_2 \leq \ldots a.s.$ Assume that A_n are integrable. Then, if $\mathcal{F}_\bullet$ is any filtration to which A is adapted, then*

$$\mathbf{E}[A_n \mid \mathcal{F}_{n-1}] - A_{n-1} = \mathbf{E}[A_n - A_{n-1} \mid \mathcal{F}_{n-1}] \geq 0$$

by positivity of conditional expectation. Thus, A is a sub-martingale. Similarly, a decreasing sequence of random variables is a super-martingale[5].

Example 17 (Branching process). *Let $L_{n,k}$, $n \geq 1$, $k \geq 1$, be i.i.d. random variables taking values in $\mathbb{N} = \{0, 1, 2, \ldots\}$. We define a sequence Z_n, $n \geq 0$ by setting $Z_0 = 1$ and*

$$Z_n = \begin{cases} L_{n,1} + \ldots + L_{n,Z_{n-1}} & \text{if } Z_{n-1} \geq 1, \\ 0 & \text{if } Z_{n-1} = 0. \end{cases}$$

This is the formal definition of the generation sizes of a branching process.

Informally, a branching process is a random tree, also called a Galton-Watson tree, *which has one individual in the 0th generation. That individual has $L_{1,1}$ offsprings all of who belong to the 1st generation. Each of them, independently, have offsprings (according to the distribution of L), and these individuals comprise the second generation. And so on, the process continues till some generation becomes empty or if that does not happen, it continues for ever. What we call Z_n is just the nth generation size, forgetting the tree structure. The basic question about branching processes is whether there is a positive probability for the tree to survive forever (we shall answer this later).*

Returning to Z_n, let $\mathcal{F}_n = \sigma\{L_{m,k} : m \leq n, k \geq 1\}$ so that $Z_n \in \mathcal{F}_n$. Assume that $\mathbf{E}[L] = m < \infty$. Then, (see the exercise below to justify the steps)

$$\begin{aligned} \mathbf{E}[Z_n \mid \mathcal{F}_{n-1}] &= \mathbf{E}[\mathbf{1}_{Z_{n-1} \geq 1}(L_{n,1} + \ldots + L_{n,Z_{n-1}}) \mid \mathcal{F}_{n-1}] \\ &= \mathbf{1}_{Z_{n-1} \geq 1} Z_{n-1} m \\ &= Z_{n-1} m. \end{aligned}$$

Thus, $\frac{1}{m^n} Z_n$ is a martingale.

[4] J. Doob was the one who defined the notion of martingales and discovered most of the basic general theorems about them that we shall see. To give a preview, one fruitful question will be to ask if a given martingale sequence is in fact a Doob martingale.

[5] An interesting fact that we shall see later is that any sub-martingale is a sum of a martingale and an increasing process. This seems reasonable since a sub-martingale increases on average while a martingale stays constant on average.

Exercise 18. *If N is a $\mathbb{N}$-valued random variable independent of ξ_m, $m \geq 1$, and ξ_m are i.i.d. with mean μ, then $\mathbf{E}[\sum_{k=1}^{N} \xi_k \mid N] = \mu N$.*

Example 19 (Pólya's urn scheme)**.** *An urn has $b_0 > 0$ black balls and $w_0 > 0$ white balls to start with. A ball is drawn uniformly at random and returned to the urn with an additional new ball of the same colour. Draw a ball again and repeat. The process continues forever. A basic question about this process is what happens to the contents of the urn? Does one colour start dominating, or do the proportions of black and white equalize?*

In precise notation, the above description may be captured as follows. Let U_n, $n \geq 1$, be i.i.d. Uniform$[0,1]$ random variables. Let $b_0 > 0$, $w_0 > 0$, be given. Then, define $B_0 = b_0$ and $W_0 = w_0$. For $n \geq 1$, define (inductively)

$$\xi_n = \mathbf{1}\left(U_n \leq \frac{B_{n-1}}{B_{n-1} + W_{n-1}}\right), \quad B_n \;\; = B_{n-1} + \xi_n, \quad W_n = W_{n-1} + (1 - \xi_n).$$

Here, ξ_n is the indicator that the nth draw is a black, B_n and W_n stand for the number of black and white balls in the urn before the $(n+1)$st draw. It is easy to see that $B_n + W_n = b_0 + w_0 + n$ (since one ball is added after each draw).

Let $\mathcal{F}_n = \sigma\{U_1, \ldots, U_n\}$ so that ξ_n, B_n, W_n are all $\mathcal{F}_n$ measurable. Let $X_n = \frac{B_n}{B_n + W_n} = \frac{B_n}{b_0 + w_0 + n}$ be the proportion of balls after the nth draw (X_n is $\mathcal{F}_n$-measurable too). Observe that

$$\mathbf{E}[B_n \mid \mathcal{F}_{n-1}] = B_{n-1} + \mathbf{E}[\mathbf{1}_{U_n \leq X_{n-1}} \mid \mathcal{F}_{n-1}] = B_{n-1} + X_{n-1} = \frac{b_0 + w_0}{b_0 + w_0 + n - 1} B_{n-1}.$$

Thus,

$$\begin{aligned} \mathbf{E}[X_n \mid \mathcal{F}_{n-1}] &= \frac{1}{b_0 + w_0 + n} \mathbf{E}[B_n \mid \mathcal{F}_{n-1}] \\ &= \frac{1}{b_0 + w_0 + n - 1} B_{n-1} \\ &= X_{n-1} \end{aligned}$$

showing that (X_n) is a martingale.

New martingales out of old: Let $(\Omega, \mathcal{F}, \mathcal{F}_\bullet, \mathbf{P})$ be a filtered probability space.

▶ Suppose $X = (X_n)_{n \geq 0}$ is a $\mathcal{F}_\bullet$-martingale and $\varphi : \mathbb{R} \to \mathbb{R}$ is a convex function. If $\varphi(X_n)$ has finite expectation for each n, then $(\varphi(X_n))_{n \geq 0}$ is a sub-martingale. If X was a sub-martingale to start with, and if φ is increasing and convex, then $(\varphi(X_n))_{n \geq 0}$ is a sub-martingale.

Proof. $\mathbf{E}[\varphi(X_n) \mid \mathcal{F}_{n-1}] \geq \varphi(\mathbf{E}[X_n \mid \mathcal{F}_{n-1}])$ by conditional Jensen's inequality. If X is a martingale, then the right hand side is equal to $\varphi(X_{n-1})$ and we get the sub-martingale property for $(\varphi(X_n))_{n \geq 0}$.

If X was only a sub-martingale, then $\mathbf{E}[X_n \mid \mathcal{F}_{n-1}] \geq X_{n-1}$ and hence the increasing property of φ is required to conclude that $\varphi(\mathbf{E}[X_n \mid \mathcal{F}_{n-1}]) \geq \varphi(X_{n-1})$. ■

► If $t_0 < t_1 < t_2 < \dots$ is a subsequence of natural numbers, and X is a martingale/sub-martingale/super-martingale, then $X_{t_0}, X_{t_1}, X_{t_2}, \dots$ is also a martingale/sub-martingale/super-martingale. This property is obvious. But it is a very interesting question that we shall ask later as to whether the same is true if t_i are random times.

If we had a continuos time-martingale $X = (X_t)_{t\geq 0}$, then again $X(t_i)$ would be a discrete time martingale for any $0 < t_1 < t_2 < \dots$. Results about continuous time martingales can in fact be deduced from results about discrete parameter martingales using this observation and taking closely spaced points t_i. If we get to continuous-time martingales at the end of the course, we shall explain this fully.

► Let X be a martingale and let $H = (H_n)_{n\geq 1}$ be a predictable sequence. This just means that $H_n \in \mathcal{F}_{n-1}$ for all $n \geq 1$. Then, define $(H.X)_n = \sum_{k=1}^{n} H_k(X_k - X_{k-1})$. Assume that $(H.X)_n$ is integrable for each n (true for instance if H_n is a bounded random variable for each n). Then, $(H.X)$ is a martingale. If X was a sub-martingale to start with, then $(H.X)$ is a sub-martingale provided H_n are non-negative, in addition to being predictable.

Proof. $\mathbf{E}[(H.X)_n - (H.X)_{n-1} \mid \mathcal{F}_{n-1}] = \mathbf{E}[H_n(X_n - X_{n-1}) \mid \mathcal{F}_{n-1}] = H_n\mathbf{E}[X_n - X_{n-1} \mid \mathcal{F}_{n-1}]$. If X is a martingale, the last term is zero. If X is a sub-martingale, then $\mathbf{E}[X_n - X_{n-1} \mid \mathcal{F}_{n-1}] \geq 0$ and because H_n is assumed to be non-negative, the sub-martingale property of $(H.X)$ is proved. ■

7. Stopping times

Let $(\Omega, \mathcal{F}, \mathcal{F}_\bullet, \mathbf{P})$ be a filtered probability space. Let $T : \Omega \to \mathbb{N} \cup \{+\infty\}$ be a random variable. If $\{T \leq n\} \in \mathcal{F}_n$ for each $n \in \mathbb{N}$, we say that T is a *stopping time*.

Equivalently we may ask for $\{T = n\} \in \mathcal{F}_n$ for each n. The equivalence with the definition above follows from the fact that $\{T = n\} = \{T \leq n\} \setminus \{T \leq n-1\}$ and $\{T = n\} = \cup_{k=0}^{n}\{T = k\}$. The way we defined it makes sense also for continuous time. For example, if $(\mathcal{F}_t)_{t\geq 0}$ is a filtration and $T : \Omega \to [0, +\infty]$ is a random variable, then we say that T is a stopping time if $\{T \leq t\} \in \mathcal{F}_t$ for all $t \geq 0$.

Example 20. *Let X_k be random variables on a common probability space and let $\mathcal{F}^X$ be the natural filtration generated by them. If $A \in \mathcal{B}(\mathbb{R})$ and $\tau_A = \min\{n \geq 0 : X_n \in A\}$, then τ_A is a stopping time. Indeed, $\{\tau_A = n\} = \{X_0 \notin A, \dots, X_{n-1} \notin A, X_n \in A\}$ which clearly belongs to $\mathcal{F}_n$.*

On the other hand, $\tau'_A := \max\{n : X_n \notin A\}$ is not a stopping time as it appears to require future knowledge. One way to make this precise is to consider $\omega_1, \omega_2 \in \Omega$ such that $\tau'_A(\omega_1) = 0 < \tau'_A(\omega_2)$ but $X_0(\omega_1) = X_0(\omega_2)$. I we can find such ω_1, ω_2, then any event in $\mathcal{F}_0$ contains both of them or neither. But $\{\tau'_A \leq 0\}$ contains ω_1 but not ω_2, hence it cannot be in $\mathcal{F}_0$. In a general probability space we cannot guarantee the existence of ω_1, ω_2 (for example Ω may contain only one point or X_k may be constant random variables!), but in sufficiently rich spaces it is possible. See the exercise below.

Exercise 21. *Let $\Omega = \mathbb{R}^{\mathbb{N}}$ with $\mathcal{F} = \mathcal{B}(\mathbb{R}^{\mathbb{N}})$ and let $\mathcal{F}_n = \sigma\{\Pi_0, \Pi_1, \ldots, \Pi_n\}$ be generated by the projections $\Pi_k : \Omega \to \mathbb{R}$ defined by $\Pi_k(\omega) = \omega_k$ for $\omega \in \Omega$. Give an honest proof that τ'_A defined as above is not a stopping time (let A be a proper subset of $\mathbb{R}$).*

Suppose T, S are two stopping times on a filtered probability space. Then $T \wedge S$, $T \vee S$, $T + S$ are all stopping times. However cT and $T - S$ need not be stopping times (even if they take values in $\mathbb{N}$). This is clear, since $\{T \wedge S \leq n\} = \{T \leq n\} \cup \{S \leq n\}$ etc. More generally, if $\{T_m\}$ is a countable family of stopping times, then $\max_m T_m$ and $\min_m T_m$ are also stopping times.

Small digression into continuous time: We shall use filtrations and stopping times in the Brownian motion class too. There the index set is continuous and complications can arise. For example, let $\Omega = C[0, \infty)$, $\mathcal{F}$ its Borel sigma-algebra, $\mathcal{F}_t = \sigma\{Pi_s : s \leq t\}$. Now define $\tau, \tau' : C[0, \infty) \to [0, \infty)$ by $\tau(\omega) = \inf\{t \geq 0 : \omega(t) \geq 1\}$ and $\tau'(\omega) = \inf\{t \geq 0 : \omega(t) > 1\}$ where the infimum is interpreted to be $+\infty$ is the set is empty. In this case, τ is an $\mathcal{F}_\bullet$-stopping time but τ' is not (why?). In discrete time there is no analogue of this situation. When we discuss this in Brownian motion, we shall enlarge the sigma-algebra $\mathcal{F}_t$ slightly so that even τ' becomes a stopping time. This is one of the reasons why we do not always work with the natural filtration of a sequence of random variables.

The sigma algebra at a stopping time: If T is a stopping time for a filtration $\mathcal{F}_\bullet$, then we want to define a sigma-algebra $\mathcal{F}_T$ that contains all information up to and including the random time T.

To motivate the idea, assume that $\mathcal{F}_n = \sigma\{X_0, \ldots, X_n\}$ for some sequence $(X_n)_{n\geq 0}$. One might be tempted to define $\mathcal{F}_T$ as $\sigma\{X_0, \ldots, X_T\}$ but a moment's thought shows that this does not make sense as written since T itself depends on the sample point.

What one really means is to partition the sample space as $\Omega = \sqcup_{n\geq 0}\{T = n\}$ and on the portion $\{T = n\}$ we consider the sigma-algebra generated by $\{X_0, \ldots, X_n\}$. Putting all these together we get a sigma-algebra that we call $\mathcal{F}_T$. To summarize, we say that $A \in \mathcal{F}_T$ if and only if $A \cap \{T = n\} \in \mathcal{F}_n$ for each $n \geq 0$. Observe that this condition is equivalent to asking for $A \cap \{T \leq n\} \in \mathcal{F}_n$ for each $n \geq 0$ (check!). Thus, we arrive at the definition

$$\mathcal{F}_T := \{A \in \mathcal{F} : A \cap \{T \leq n\} \in \mathcal{F}_n\} \quad \text{for each } n \geq 0.$$

Remark 22. *When working in continuous time, the partition $\{T = t\}$ is uncountable and hence not a good one to work with. As defined, $\mathcal{F}_T = \{A \in \mathcal{F} : A \cap \{T \leq t\} \in \mathcal{F}_t\}$ for each $t \geq 0$.*

We make some basic observations about $\mathcal{F}_T$.

(1) $\mathcal{F}_T$ is a sigma-algebra. Indeed,

$$(A \cap \{T \le n\})^c = \{T \le n\} \setminus (A \cap \{T \le n\}),$$

$$(\bigcup_{k\ge 1} A_k) \cap \{T \le n\} = \bigcup_{k \ge 1} (A \cap \{T \le n\}).$$

From these it follows that $\mathcal{F}_T$ is closed under complements and countable unions. As $\Omega \in \mathcal{F}_T$ since $\{T \le n\} \in \mathcal{F}_n$, we see that $\Omega \in \mathcal{F}_n$. Thus $\mathcal{F}_T$ is a sigma-algebra.

(2) T is $\mathcal{F}_T$-measurable. To show this we just need to show that $\{T \le m\} \in \mathcal{F}_T$ for any $m \ge 0$. But that is true because for every $n \ge 0$ we have

$$\{T \le m\} \cap \{T \le n\} = \{T \le m \wedge n\} \in \mathcal{F}_{m \wedge n} \subseteq \mathcal{F}_n.$$

A related observation is given as exercise below.

(3) If T, S are stopping times and $T \le S$, then $\mathcal{F}_T \subseteq \mathcal{F}_S$. To see this, suppose $A \in \mathcal{F}_T$. Then $A \cap \{T \le n\} \in \mathcal{F}_n$ for each n.

Consider $A \cap \{S \le n\} = A \cap \{S \le n\} \cap \{T \le n\}$ since $\{S \le n\} \subseteq \{T \le n\}$. But $A \cap \{S \le n\} \cap \{T \le n\}$ can be written as $(A \cap \{T \le n\}) \cap \{S \le n\}$ which belongs to $\mathcal{F}_n$ since $A \cap \{T \le n\} \in \mathcal{F}_n$ and $\{S \le n\} \in \mathcal{F}_n$.

Exercise 23. *Let $X = (X_n)_{n \ge 0}$ be adapted to the filtration $\mathcal{F}_\bullet$ and let T be a $\mathcal{F}_\bullet$-stopping time. Then X_T is $\mathcal{F}_T$-measurable.*

For the sake of completeness: In the last property stated above, suppose we only assumed that $T \le S$ *a.s.* Can we still conclude that $\mathcal{F}_T \subseteq \mathcal{F}_S$? Let $C = \{T > S\}$ so that $C \in \mathcal{F}$ and $\mathbf{P}(C) = 0$. If we try to repeat the proof as before, we end up with

$$A \cap \{S \le n\} = [(A \cap \{T \le n\}) \cap \{S \le n\}] \ \cup \ (A \cap C).$$

The first set belongs to $\mathcal{F}_n$ but there is no assurance that $A \cap C$ does, since we only know that $C \in \mathcal{F}$.

One way to get around this problem (and many similar ones) is to complete the sigma-algebras as follows. Let $\mathcal{N}$ be the collection of all null sets in $(\Omega, \mathcal{F}, \mathbf{P})$. That is,

$$\mathcal{N} = \{A \subseteq \Omega : \exists\, B \in \mathcal{F} \text{ such that } B \supseteq A \text{ and } \mathbf{P}(B) = 0\}.$$

Then define $\bar{\mathcal{F}}_n = \sigma\{\mathcal{F}_n \cup \mathcal{N}\}$. This gives a new filtration $\bar{\mathcal{F}}_\bullet = (\bar{\mathcal{F}}_n)_{n \ge 0}$ which we call the completion of the original filtration (strictly speaking, this completion depended on $\mathcal{F}$ and not merely on $\mathcal{F}_\bullet$. But we can usually assume without loss of generality that $\mathcal{F} = \sigma\{\cup_{n \ge 0} \mathcal{F}_n\}$ by decreasing $\mathcal{F}$ if necessary. In that case, it is legitimate to call $\bar{\mathcal{F}}_\bullet$ the completion of $\mathcal{F}_\bullet$ under $\mathbf{P}$).

It is to be noted that after enlargement, $\mathcal{F}_\bullet$-adapted processes remain adapted to $\bar{\mathcal{F}}_\bullet$, stopping times for $\mathcal{F}_\bullet$ remain stopping times for $\bar{\mathcal{F}}_\bullet$, etc. Since the enlargement is only by $\mathbf{P}$-null sets, it can be see that $\mathcal{F}_\bullet$-super-martingales remain $\bar{\mathcal{F}}_\bullet$-super-martingales, etc. Hence, there is no loss in working in the completed sigma algebras.

Henceforth we shall simply assume that our filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_\bullet, \mathbf{P})$ is such that all $\mathbf{P}$-null sets in $(\Omega, \mathcal{F}, \mathbf{P})$ are contained in $\mathcal{F}_0$ (and hence in $\mathcal{F}_n$ for all n). Let us say that $\mathcal{F}_\bullet$ is complete to mean this.

Exercise 24. *Let T, S be stopping times with respect to a complete filtration $\mathcal{F}_\bullet$. If $T \le S$ $a.s$ (w.r.t. $\mathbf{P}$), show that $\mathcal{F}_T \subseteq \mathcal{F}_S$.*

Exercise 25. *Let $T_0 \le T_1 \le T_2 \le \ldots$ $(a.s.)$ be stopping times for a complete filtration $\mathcal{F}_\bullet$. Is the filtration $(\mathcal{F}_{T_k})_{k\ge 0}$ also complete?*

8. Optional stopping or sampling

Let $X = (X_n)_{n\ge 0}$ be a super-martingale on a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_\bullet, \mathbf{P})$. We know that (1) $\mathbf{E}[X_n] \le \mathbf{E}[X_0]$ for all $n \ge 0$ and (2) $(X_{n_k})_{k\ge 0}$ is a super-martingale for any subsequence $n_0 < n_1 < n_2 < \ldots$.

Optional stopping theorems are statements that assert that $\mathbf{E}[X_T] \le \mathbf{E}[X_0]$ for a stopping time T. *Optional sampling theorems* are statements that assert that $(X_{T_k})_{k\ge 0}$ is a super-martingale for an increasing sequence of stopping times $T_0 \le T_1 \le T_2 \le \ldots$. Usually one is not careful to make the distinction and OST could refer to either kind of result.

Neither of these statements is true without extra conditions on the stopping times. But they are true when the stopping times are bounded, as we shall prove in this section. In fact, it is best to remember only that case, and derive more general results whenever needed by writing a stopping time as a limit of bounded stopping times. For example, $T \wedge n$ are bounded stopping times and $T \wedge n \stackrel{a.s.}{\to} T$ as $n \to \infty$.

Now we state the precise results for bounded stopping times.

Theorem 26 (Optional stopping theorem)**.** *Let $(\Omega, \mathcal{F}, \mathcal{F}_\bullet, \mathbf{P})$ be a filtered probability space and let T be a stopping time for $\mathcal{F}_\bullet$. If $X = (X_n)_{n\ge 0}$ is a $\mathcal{F}_\bullet$-super-martingale, then $(X_{T\wedge n})_{n\ge 0}$ is a $\mathcal{F}_\bullet$-super-martingale. In particular, $\mathbf{E}[X_{T\wedge n}] = \mathbf{E}[X_0]$ for all $n \ge 0$.*

If T is a bounded stopping time, that is $T \le N$ $a.s.$ for some N. Taking $n = N$ in the theorem we get the following corollary.

Corollary 27. *If T is a bounded stopping time and X is a super-martingale, then $\mathbf{E}[X_T] \le \mathbf{E}[X_0]$.*

Here and elsewhere, we just state the result for super-martingales. From this, the reverse inequality holds for sub-martingales (by applying the above to $-X$) and hence equality holds for martingales.

Proof of Theorem 26. Let $H_n = \mathbf{1}_{n \le T}$. Then $H_n \in \mathcal{F}_{n-1}$ because $\{T \ge n\} = \{T \le n-1\}^c$ belongs to $\mathcal{F}_{n-1}$. By the observation earlier, $(H.X)_n$, $n \ge 0$, is a super-martingale. But $(H.X)_n = X_{T \wedge n}$ and this proves that $(X_{T\wedge n})_{n\ge 0}$ is an $\mathcal{F}_\bullet$-super-martingale. Then of course $\mathbf{E}[X_{T\wedge n}] \le \mathbf{E}[X_0]$. ■

We already showed how to get the corollary from Theorem 26. Now we state the optional sampling theorem.

Theorem 28 (Optional sampling theorem)**.** *Let $(\Omega, \mathcal{F}, \mathcal{F}_\bullet, \mathbf{P})$ be a filtered probability space and let $X = (X_n)_{n\ge 0}$ be a $\mathcal{F}_\bullet$-super-martingale. Let T_n, $n \ge 0$ be bounded stopping times for $\mathcal{F}_\bullet$ such that $T_0 \le T_1 \le T_2 \le \dots$ a.s. Then, $(X_{T_k})_{k\ge 0}$ is a super-martingale with respect to the filtration $(\mathcal{F}_{T_k})_{k \ge 0}$.*

Proof. Since X is adapted to $\mathcal{F}_\bullet$, it follows that X_{T_k} is $\mathcal{F}_{T_k}$-measurable. Further, if $|T_k| \le N_k$ w.p.1. for a fixed number N_k, then $|X_{T_k}| \le |X_0| + \dots + |X_{N_k}$ which shows the integrability of X_{T_k}. The theorem will be proved if we show that if $S \le T \le N$ where S, T are stopping times and N is a fixed number, then

$$\mathbf{E}[X_T \mid \mathcal{F}_S] \le X_S \;\; a.s. \tag{2}$$

Since X_S and $\mathbf{E}[X_T \mid \mathcal{F}_S]$ are both $\mathcal{F}_S$-measurable, (2) follows if we show that $\mathbf{E}[(X_T - X_S)\mathbf{1}_A] \le 0$ for every $A \in \mathcal{F}_S$.

Now fix any $A \in \mathcal{F}_S$ and define $H_k = \mathbf{1}_{S+1 \le k \le T}\mathbf{1}_A$. This is the indicator of the event $A \cap \{S \le k-1\} \cap \{T \ge k\}$. Since $A \in \mathcal{F}_S$ we see that $A \cap \{S \le k-1\} \in \mathcal{F}_{k-1}$ while $\{T \ge k\} = \{T \le k-1\}^c \in \mathcal{F}_{k-1}$. Thus, H is predictable. In words, this is the betting scheme where we bet 1 rupee on each game from time $S+1$ to time T, but only if A happens (which we know by time S). By the gambling lemma, we conclude that $\{(H.X)_n\}_{n\ge 1}$ is a super-martingale. But $(H.X)_n = (X_{T\wedge n} - X_{S \wedge n})\mathbf{1}_A$. Put $n = N$ and get $\mathbf{E}[(X_T - X_S)\mathbf{1}_A] \ge 0$ since $(H.X)_0 = 0$. Thus (2) is proved. ■

An alternate proof of Theorem 28 is outlined below.

Second proof of Theorem 28. As in the first proof, it suffices to prove (2).

First assume that $S \le T \le S+1$ $a.s.$ Let $A \in \mathcal{F}_S$. On the event $\{S = T\}$ we have $X_T - X_S = 0$. Therefore,

$$\begin{aligned} \int_A (X_T - X_S)dP &= \int\limits_{A \cap \{T = S+1\}} (X_{S+1} - X_S)dP \\ &= \sum_{k=0}^{N-1} \int\limits_{A\cap\{S=k\}\cap\{T=k+1\}} (X_{k+1} - X_k)dP. \end{aligned} \tag{3}$$

For fixed k, we see that $A \cap \{S = k\} \in \mathcal{F}_k$ since $A \in \mathcal{F}_S$ and $\{T = k+1\} = \{T \le k\}^c \cap \{S = k\} \in \mathcal{F}_k$ because $T \le S+1$. Therefore, $A \cap \{S = k\} \cap \{T = k+1\} \in \mathcal{F}_k$ and the super-martingale property of X implies that $\int_B (X_{k+1} - X_k)dP \le 0$ for any $B \in \mathcal{F}_k$. Thus, each term in (3) is non-positive. Hence $\int_A X_S dP \ge \int_A X_T dP$ for every $A \in \mathcal{F}_T$. This just means that $\mathbf{E}[X_S \mid \mathcal{F}_T] \le X_T$. This completes the proof assuming $T \le S \le T+1$.

In general, since $S \leq T \leq N$, let $S_0 = S$, $S_1 = T \wedge (S+1)$, $S_2 = T \wedge (S+2), \ldots, S_n = T \wedge (S+N)$ so that each S_k is a stopping time, $S_N = S$, and for each k we have $S_k \leq S_{k+1} \leq S_k + 1$ $a.s.$ Deduce from the previous case that $\mathbf{E}[X_T \mid \mathcal{F}_S] \leq X_S$ $a.s.$ ■

We end this section by giving an example to show that optional sampling theorems can fail if the stopping times are not bounded.

Example 29. *Let ξ_i be i.i.d. Ber$_{\pm}(1/2)$ random variables and let $X_n = \xi_1 + \ldots + \xi_n$ (by definition $X_0 = 0$). Then X is a martingale. Let $T_1 = \min\{n \geq 1 : X_n = 1\}$.*

A theorem of Pólya asserts that $T_1 < \infty$ w.p.1. But $X_{T_1} = 1$ $a.s.$ while $X_0 = 0$. Hence $\mathbf{E}[X_{T_1}] \neq \mathbf{E}[X_0]$, violating the optional stopping property (for bounded stopping times we would have had $\mathbf{E}[X_T] = \mathbf{E}[X_0]$). In gambling terminology, if you play till you make a profit of 1 rupee and stop, then your expected profit is 1 (an not zero as optional stopping theorem asserts).

If $T_j = \min\{n \geq 0 : X_n = j\}$ for $j = 1, 2, 3, \ldots$, then it again follows from Pólya's theorem that $T_j < \infty$ $a.s.$ and hence $X_{T_j} = j$ $a.s.$ Clearly $T_0 < T_1 < T_2 < \ldots$ but $X_{T_0}, X_{T_1}, X_{T_2}, \ldots$ is not a super-martingale (in fact, being increasing it is a sub-martingale!).

This example shows that applying optional sampling theorems blindly without checking conditions can cause trouble. But the boundedness assumption is by no means essential. Indeed, if the above example is tweaked a little, optional sampling is restored.

Example 30. *In the previous example, let $-A < 0 < B$ be integers and let $T = \min\{n \geq 0 : X_n = -A \text{ or } X_n = B\}$. Then T is an unbounded stopping time. In gambling terminology, the gambler has capital A and the game is stopped when he/she makes a profit of B rupees or the gambler goes bankrupt. If we set $B = 1$ we are in a situation similar to before, but with the somewhat more realistic assumption that the gambler has finite capital.*

By the optional sampling theorem $\mathbf{E}[X_{T \wedge n}] = 0$. By a simple argument (or Pólya's theorem) one can prove that $T < \infty$ w.p.1. Therefore, $X_{T \wedge n} \stackrel{a.s.}{\to} X_T$ as $n \to \infty$. Further, $|X_{T \wedge n}| \leq B + A$ from which by DCT it follows that $\mathbf{E}[X_{T \wedge n}] \to \mathbf{E}[X_T]$. Therefore, $\mathbf{E}[X_T] = 0$. In other words optional stopping property is restored.

Gambler's ruin problem: Let X be simple symmetric random walks as before. What is the probability that X hits B before $-A$? With $T = \min\{n \geq 0 : X_n = -A \text{ or } X_n = B\}$ we already showed that $\mathbf{E}[X_T] = 0$. But we also know that $T < \infty$ w.p.1. (why?). Hence, if $\alpha = \mathbf{P}\{X_T = B\}$ then $1 - \alpha = \mathbf{P}\{X_T = -A\}$ and

$$0 = \mathbf{E}[X_T] = \alpha B - (1 - \alpha) A$$

which gives $\alpha = \frac{A}{A+B}$.

Exercise 31. *Let ξ_i be i.i.d. with $\xi_1 = +1$ w.p. p and $\xi_1 = -1$ w.p. $q = 1 - p$. Let $X_n = \xi_1 + \ldots + \xi_n$. Find the probability that X hits B before $-A$ (for $A, B > 0$, of course).*

9. DISCRETE DIRICHLET PROBLEM

As an application of optional sampling theorems, we study the discrete Dirichlet problem.

Let $G = (V, E)$ be a finite connected graph with vertex set V and edge set E. Let A be a proper subset of V. A function $f : V \to \mathbb{R}$ is said to be super-harmonic on $V \setminus A$ is $f(x) \geq \frac{1}{d(x)} \sum_{y:y\sim x} f(y)$ for each $x \in V \setminus A$. Here $d(x)$ is the degree of the vertex x.

In words, the value of f at any vertex (not in A) is at least as large as the average value at its neighbours (some of which may lie in A). Similarly, we say that f is sub-harmonic on $V \setminus A$ is $-f$ is super-harmonic and that f is harmonic if it is super-harmonic as well as sub-harmonic.

Dirichlet problem:: Given $\varphi : A \to \mathbb{R}$, does there exist a function $f : V \to \mathbb{R}$ such that $f\big|_A = \varphi$ and f is harmonic on $V \setminus A$? If yes, it is unique?

Electrical networks: Imagine that G is an electric network where each edge is replaced by a unit resistor. The vertices in A are connected to batteries and the voltages at these points are maintained at $\varphi(x)$, $x \in A$. Then, electric current flows through the network and at each vertex a voltage is established. Call these voltages $f(x)$, $x \in V$. Clearly $f\big|_A = \varphi$ since we control the voltages there. Kirchoff's law says that at any vertex $x \notin A$, the net incoming current is zero. But the current from y to x is precisely $f(x) - f(y)$ (by Ohm's law). Thus, Kirchoff's law implies that $0 = \sum_{y:y\sim x} f(x) - f(y)$ which is precisely the same as saying that f is harmonic in $V \setminus A$.

Thus, from a physical perspective, the existence of a solution to the Dirichlet problem is clear! It is not hard to do it mathematically either. The equations for f give $|V \setminus A|$ many linear equations for the variables $(f(x))_{x \in V\setminus A}$. If we check that the linear combinations form a non-singular matrix, we get existence as well as uniqueness. This can be done. But instead we give a probabilistic construction of the solution using random walks.

Existence: Let X_n be simple random walk on G. This means that it is a markov chain with the state space V and transition matrix $p(x, y) = \frac{1}{d(x)}$ if $y \sim x$ and $p(x, y) = 0$ otherwise. Let $T = \min\{n \geq 0 : X_n \in A\}$. Then $T < \infty$ w.p.1. (why?) and hence we may define $f(x) := \mathbf{E}_x[\varphi(X_T)]$ (we put x in the subscript of $\mathbf{E}$ to indicate the starting location). For $x \in A$ clearly $T = 0$ $a.s.$and hence $f(x) = \varphi(x)$. We only need to check the harmonicity of f on $V \setminus A$. This is easy to see by conditioning on the first step, X_1.

$$\mathbf{E}_x[\varphi(X_T)] = \sum_{y:y\sim x} \frac{1}{d(x)} \mathbf{E}_x[\varphi(X_T)\big|X_1 = y] = \frac{1}{d(x)} \sum_{y:y\sim x} \mathbf{E}_y[\varphi(X_T)] = \frac{1}{d(x)} \sum_{y:y\sim x} f(y).$$

In the second equality, we used the fact that $x \notin A$. Thus, f is harmonic in $V \setminus A$.

Uniqueness: We show uniqueness by invoking the optional sampling theorem. Let f be a solution to the Dirichlet problem. Define $M_k = f(X_{T\wedge k})$. We claim that M is a martingale (for any starting point of X). Indeed,

$$\mathbf{E}[M_{k+1} \mid \mathcal{F}_k] = \mathbf{E}[f(X_{T\wedge(k+1)}) \mid X_k] = \begin{cases} f(X_k) & \text{if } T \le k \\ \frac{1}{d(X_k)} \sum_{v:v\sim X_k} f(v) & \text{if } T > k \end{cases}$$

If $T > k$ then $X_k \in V \setminus A$ and hence $\frac{1}{d(X_k)} \sum_{v:v\sim X_k} f(v)$ is equal to $f(X_k)$ which is M_k. Therefore, in all cases, $\mathbf{E}[M_{k+1} \mid \mathcal{F}_k] = M_k$.

T is not a bounded stopping time, by f is a bounded function and hence M is uniformly bounded. Hence, from $M_k \overset{a.s.}{\to} f(X_T)$ we get that $\mathbf{E}_x[f(X_T)] = \lim \mathbf{E}_x[M_k] = \mathbf{E}_x[M_0] = f(x)$. But $X_T \in A$ and hence $f(x) = \mathbf{E}_x[\varphi(X_T)]$. Thus, any solution to the Dirichlet problem is equal to the solution we constructed. This proves uniqueness.

Example 32. *If $V = \{-a, -a+1, \ldots, b\}$ is a subgraph of $\mathbb{Z}$ (edges between i and $i+1$), then it is easy to see that the harmonic functions are precisely of the form $f(k) = \alpha k + \beta$ for some $\alpha, \beta \in \mathbb{R}$ (check!).*

Let $A = \{-a, b\}$ and let $\varphi(-a) = 0$, $\varphi(b) = 1$. The unique harmonic extension of this is clearly $f(k) = \frac{k+a}{b+a}$. Hence, $\mathbf{E}_0[\varphi(X_T)] = \frac{a}{b+a}$. But $\varphi(X_T) = \mathbf{1}_{X_T=b}$ and thus we get back the Gambler's ruin probability $\mathbf{P}_0\{$hit b before $-a\} = \frac{a}{a+b}$.

10. Maximal inequality

Kolmogorov's proof of his famous inequality was perhaps the first proof using martingales, although the term did not exist then!

Lemma 33 (Kolmogorov's maximal inequality)**.** *Let ξ_k be independent random variables with zero means and finite variances. Let $S_n = \xi_1 + \ldots + \xi_n$. Then,*

$$\mathbf{P}\left\{\max_{k\le n} |S_k| \ge t\right\} \le \frac{1}{t^2} Var(S_n).$$

Proof. We know that $(S_k)_{k\ge 0}$ is a martingale and $(S_k^2)_{k\ge 0}$ is a sub-martingale. Let $T = \min\{k : |S_k| \ge t\} \wedge n$ (i.e., T is equal to n or to the first time S exits $(-t,t)$, whichever is earlier). Then T is a bounded stopping time and $T \le n$. By OST, S_T, S_n is a sub-martingale and thus $\mathbf{E}[S_T^2] \le \mathbf{E}[S_n^2]$. By Chebyshev's inequality,

$$\mathbf{P}\left\{\max_{k\le n} |S_k| \ge t\right\} = \mathbf{P}\{S_T^2 \ge t^2\} \le \frac{1}{t^2}\mathbf{E}[S_T^2] \le \frac{1}{t^2}\mathbf{E}[S_n^2].$$

Thus the inequality follows. ∎

This is an amazing inequality that controls the supremum of the entire path $S_0, S_1, \ldots, S_n$ in terms of the end-point alone! It takes a little thought to realize that the inequality $\mathbf{E}[S_T^2] \le \mathbf{E}[S_n^2]$ is not a paradox. One way to understand it is to realize that if the path goes beyond $(-t,t)$, then there is a significant probability for the end point to be also large. This intuition is more clear in certain precursors to Kolmogorov's maximal inequality. In the following exercise you will prove one such, for symmetric, but not necessarily integrable, random variables.

Exercise 34. *Let ξ_k be independent symmetric random variables and let $S_k = \xi_1 + \ldots + \xi_k$. Then for $t > 0$, we have*

$$\mathbf{P}\left\{\max_{k\le n} S_k \ge t\right\} \le 2\mathbf{P}\left\{S_n \ge t\right\}.$$

Hint: *Let T be the first time k when $S_k \ge t$. Given everything up to time $T = k$, consider the two possible future paths formed by $(\xi_{k+1}, \ldots, \xi_n)$ and $(-\xi_{k+1}, \ldots, -\xi_n)$. If $S_T \ge t$, then clearly for at least one of these two continuations, we must have $S_n \ge t$. Can you make this reasoning precise and deduce the inequality?*

For a general super-martingale or sub-martingale, we can write similar inequalities that control the running maximum of the martingale in terms of the end-point.

Lemma 35 (Doob's inequalities)**.** *Let X be a super-martingale. Then for any $t > 0$ and any $n \ge 1$,*

(1) $\mathbf{P}\{\max_{k\le n} X_k \ge t\} \le \frac{1}{t}\mathbf{E}[X_0] + \mathbf{E}[(X_n)_-]$,

(2) $\mathbf{P}\{\min_{k\le n} X_k \le -t\} \le \frac{1}{t}\mathbf{E}[(X_n)_-]$.

Proof. Let $T = \min\{k : X_k \ge n\} \wedge n$. By OST $\{X_1, X_T\}$ is a super-martingale and hence $\mathbf{E}[X_T] \le \mathbf{E}[X_1]$. But

$$\begin{aligned}\mathbf{E}[X_T] &= \mathbf{E}[X_T \mathbf{1}_{X_T\ge t}] + \mathbf{E}[X_T \mathbf{1}_{X_T<t}] \\ &= \mathbf{E}[X_T \mathbf{1}_{X_T\ge t}] + \mathbf{E}[X_n \mathbf{1}_{X_T<t}] \\ &\ge \mathbf{E}[X_T \mathbf{1}_{X_T\ge t}] + \mathbf{E}[(X_n)_-]\end{aligned}$$

since $\mathbf{E}[X_n \mathbf{1}_A] \ge -\mathbf{E}[(X_n)_-]$ for any A. Thus, $\mathbf{E}[X_T \mathbf{1}_{X_T\ge t}] \le \mathbf{E}[X_1] + \mathbf{E}[(X_n)_-]$. Now use Chebyshev's inequality to write $\mathbf{P}\{X_T \ge t\} \le \frac{1}{t}\mathbf{E}[X_T \mathbf{1}_{X_T\ge t}]$ to get the first inequality.

For the second inequality, define $T = \min\{k : X_k \le -t\} \wedge n$. By OST $\{(X_T), (X_n)\}$ is a super-martingale and hence $\mathbf{E}[X_T] \ge \mathbf{E}[X_n]$. But

$$\begin{aligned}\mathbf{E}[X_T] &= \mathbf{E}[X_T \mathbf{1}_{X_T\le -t}] + \mathbf{E}[X_n \mathbf{1}_{X_T\ge -t}] \\ &\le -t\mathbf{P}\{X_T \le -t\} + \mathbf{E}[(X_n)_+].\end{aligned}$$

Hence $\mathbf{P}\{X_T \le -t\} \le \frac{1}{t}\{\mathbf{E}[(X_n)_+] - \mathbf{E}[X_n]\} = \frac{1}{t}\mathbf{E}[(X_n)_-]$. ■

It is clear how to write the corresponding inequalities for sub-martingales. For example, if $X_0, \ldots, X_n$ is a sub-martingale, then

$$\mathbf{P}\left\{\max_{k\le n} X_k \ge t\right\} \le \frac{1}{t}\mathbf{E}[(X_n)_+] \text{ for any } \lambda > 0.$$

If ξ_i are independent with zero mean and finite variances and $S_n = \xi_1 + \ldots + \xi_n$ is the corresponding random walk, then the above inequality when applied to the sub-martingale S_k^2 reduces to Kolmogorov's maximal inequality.

Maximal inequalities are useful in proving the Cauchy property of partial sums of a random series with independent summands. Here is an exercise.

Exercise 36. *Let ξ_n be independent random variables with zero means. Assume that $\sum_n Var(\xi_n) < \infty$. Show that $\sum_k \xi_k$ converges almost surely. [Extra: If interested, extend this to independent ξ_ks taking values in a separable Hilbert space H such that $\mathbf{E}[\langle \xi_k, u\rangle] = 0$ for all $u \in H$ and such that $\sum_n \mathbf{E}[\|\xi_n\|^2] < \infty$.]*

11. Doob's up-crossing inequality

For a real sequence $x_0, x_1, \ldots, x_n$ and any $a < b$, define the number of up-crossings of the sequence over the interval $[a, b]$ as the maximum number k for which there exist indices $0 \le i_1 < j_1 < i_2 < j_2 < \ldots < i_k < j_k \le n$ such that $x_{i_r} \le a$ and $x_{j_r} \ge b$ for all $r = 1, 2, \ldots, k$. Intuitively it is the number of times the sequence crosses the interval in the upward direction. Similarly we can define the number of down-crossings of the sequence (same as the number of up-crossings of the sequence $(-x_k)_{0\le k\le n}$ over the interval $[-b, -a]$).

Lemma 37 (Doob's up-crossing inequality)**.** *Let $X_0, \ldots, X_n$ be a sub-martingale. Let $U_n[a, b]$ denote the number of up-crossings of the sequence $X_0, \ldots, X_n$ over the interval $[a, b]$. Then,*

$$\mathbf{E}[U_n[a,b] \mid \mathcal{F}_0] \le \frac{\mathbf{E}[(X_n - a)_+ \mid \mathcal{F}_0] - (X_0 - a)_+}{b - a}.$$

What is the importance of this inequality? It will be in showing the convergence of martingales or super-martingales under some mild conditions. In continuous time, it will yield regularity of paths of martingales (existence of right and left limits).

The basic point is that a real sequence $(x_n)_n$ converges if and only if the number of up-crossings of the sequence over any interval is finite. Indeed, if $\liminf x_n < a < b < \limsup x_n$, then the sequence has infinitely many up-crossings and down-crossings over $[a, b]$. Conversely, if $\lim x_n$ exists, then the sequence is Cauchy and hence over any interval $[a, b]$ with $a < b$, there can be only finitely many up-crossings. In these statements the limit could be $\pm\infty$.

Proof. First assume that $X_n \ge 0$ for all n and that $a = 0$. Let $T_0 = 0$ and define the stopping times

$$T_1 := \min\{k \ge T_0 : X_k = 0\},\ T_3 := \min\{k \ge T_2 : X_k = 0\}, \quad \ldots$$

$$T_2 := \min\{k \ge T_1 : X_k \ge b\},\ T_4 := \min\{k \ge T_3 : X_k \ge b\}, \quad \ldots$$

where the minimum of an empty set is defined to be n. T_i are strictly increasing up to a point when T_k becomes equal to n and then the later ones are also equal to n. In what follows we only need T_k for $k \le n$ (thus all the sums are finite sums).

$$X_n - X_0 = \sum_{k\ge 0} X(T_{2k+1}) - X(T_{2k}) + \sum_{k\ge 1} X(T_{2k}) - X(T_{2k-1})$$
$$\ge \sum_{k\ge 0}(X(T_{2k+1}) - X(T_{2k})) + bU_n[0,b].$$

The last inequality is because for each k for which $X(T_{2k}) \ge b$, we get one up-crossing and the corresponding increment $X(T_{2k}) - X(T_{2k-1}) \ge b$.

Now, by the optional sampling theorem (since $T_{2k+1} \ge T_{2k}$ are both bounded stopping times), we see that

$$\mathbf{E}[X(T_{2k+1}) - X(T_{2k}) \mid \mathcal{F}_0] = \mathbf{E}\left[\mathbf{E}[X(T_{2k+1}) - X(T_{2k}) \mid \mathcal{F}_{T_{2k}}] \mid \mathcal{F}_0\right] \ge 0.$$

Therefore, $\mathbf{E}[X_n - X_0 \mid \mathcal{F}_0] \ge (b-a)\mathbf{E}[U_n[a,b] \mid \mathcal{F}_0]$. This gives the up-crossing inequality when $a = 0$ and $X_n \ge 0$.

In the general situation, just apply the derived inequality to the sub-martingale $(X_k - a)_+$ (which crosses $[0, b-a]$ whenever X crosses $[a,b]$) to get

$$\mathbf{E}[(X_n - a)_+ \mid \mathcal{F}_0] - (X_0 - a)_+ \ge (b-a)\mathbf{E}[U_n[a,b] \mid \mathcal{F}_0]$$

which is what we claimed. ■

The break up of $X_n - X_0$ over up-crossing and down-crossings was okay, but how did the expectations of increments over down-crossings become non-negative? There is a distinct sense of something suspicious about this! The point is that $X(T_3) - X(T_2)$, for example, is not always non-negative. If X never goes below a after T_2, then it can be positive too. Indeed, the sub-martingale property somehow ensures that this positive part off sets the $-(b-a)$ increment that would occur if $X(T_3)$ did go below a.

We invoked OST in the proof. Optional sampling was in turn proved using the gambling lemma. It is an instructive exercise to write out the proof of the up-crossing inequality directly using the gambling lemma (start betting when below a, stop betting when reach above b, etc.).

12. Convergence theorem for super-martingales

Now we come to the most important part of the theory.

Theorem 38 (Super-martingale convergence theorem). *Let X be a super-martingale on $(\Omega, \mathcal{F}, \mathcal{F}_\bullet, \mathbf{P})$. Assume that $\sup_n \mathbf{E}[(X_n)_-] < \infty$.*

(1) Then, $X_n \overset{a.s.}{\to} X_\infty$ for some integrable (hence finite) random variable X_∞.

(2) In addition, $X_n \to X_\infty$ in L^1 if and only if $\{X_n\}$ is uniformly integrable. If this happens, we also have $\mathbf{E}[X_\infty \mid \mathcal{F}_n] \le X_n$ for each n.

In other words, when a super-martingale does not explode to $-\infty$ (in the mild sense of $\mathbf{E}[(X_n)_-]$ being bounded), it must converge almost surely!

Proof. Fix $a < b$. Let $D_n[a,b]$ be the number of down-crossings of $X_0, \ldots, X_n$ over $[a,b]$. By applying the up-crossing inequality to the sub-martingale $-X$ and the interval $[-b,-a]$, and taking expectations, we get

$$\begin{aligned}\mathbf{E}[D_n[a,b]] &\leq \frac{\mathbf{E}[(X_n-b)_-] - \mathbf{E}[(X_0-b)_-]}{b-a} \\ &\leq \frac{1}{b-a}(\mathbf{E}[(X_n)_-] + |b|) \leq \frac{1}{b-a}(M+|b|)\end{aligned}$$

where $M = \sup_n \mathbf{E}[(X_n)_-]$. Let $D[a,b]$ be the number of down-crossings of the whole sequence (X_n) over the interval $[a,b]$. Then $D_n[a,b] \uparrow D[a,b]$ and hence by MCT we see that $\mathbf{E}[D[a,b]] < \infty$. In particular, $D[a,b] < \infty$ w.p.1.

Consequently, $\mathbf{P}\{D[a,b] < \infty \text{ for all } a < b, \quad a,b \in \mathbb{Q}\} = 1$. Thus, X_n converges w.p.1., and we define X_∞ as the limit (for ω in the zero probability set where the limit does not exist, define X_∞ as 0). Thus $X_n \stackrel{a.s.}{\to} X_\infty$.

We observe that $\mathbf{E}[|X_n|] = \mathbf{E}[X_n] + 2\mathbf{E}[(X_n)_-] \leq \mathbf{E}[X_0] + 2M$. By Fatou's lemma, $\mathbf{E}[|X_\infty|] \leq \liminf \mathbf{E}[|X_n|] \leq 2M + \mathbf{E}[X_0]$. Thus X_∞ is integrable.

This proves the first part. The second part is very general - whenever $X_n \stackrel{a.s.}{\to} X$, we have L^1 convergence if and only if $\{X_n\}$ is uniformly integrable. Lastly, $\mathbf{E}[X_{n+m} \mid \mathcal{F}_n] \leq X_n$ for any $n, m \geq 1$. Let $m \to \infty$ and use L^1 convergence of X_{n+m} to X_∞ to get $\mathbf{E}[X_\infty \mid \mathcal{F}_n] \leq X_n$.

This completes the proof. ∎

A direct corollary that is often used is

Corollary 39. *A non-negative super-martingale converges almost surely to a finite random variable.*

13. Convergence theorem for martingales

Now we deduce the consequences for martingales.

Theorem 40 (Martingale convergence theorem)**.** *Let $X = (X_n)_{n\geq 0}$ be a martingale with respect to $\mathcal{F}_\bullet$. Assume that X is L^1-bounded.*

(1) Then, $X_n \stackrel{a.s.}{\to} X_\infty$ for some integrable (in particular, finite) random variable X_∞.

(2) In addition, $X_n \stackrel{L^1}{\to} X_\infty$ if and only if X is uniformly integrable. In this case, $\mathbf{E}[X_\infty \mid \mathcal{F}_n] = X_n$ for all n.

(3) If X is L^p bounded for some $p > 1$, then $X_n \stackrel{L^p}{\to} X_\infty$.

Observe that for a martingale the condition of L^1-boundedness, $\sup_n \mathbf{E}[|X_n|] < \infty$, is equivalent to the weaker looking condition $\sup_n \mathbf{E}[(X_n)_-] < \infty$, since $\mathbf{E}[|X_n|] - 2\mathbf{E}[(X_n)_-] = \mathbf{E}[X_n] = \mathbf{E}[X_0]$ is a constant.

Proof. The first two parts of the proof are immediate since a martingale is also a super-martingale. To conclude $\mathbf{E}[X_\infty \mid \mathcal{F}_n] = X_n$, we apply the corresponding inequality in the super-martingale convergence theorem to both X and to $-X$.

For the third part, if X is L^p bounded, then it is uniformly integrable and hence $X_n \to X_\infty$ $a.s.$and in L^1. To get L^p convergence, consider the non-negative sub-martingale $\{|X_n|\}$ and let $Y = \sup_n |X_n|$. From Lemma 41 we conclude that $Y \in L^p$. Now, $|X_n - X_\infty|^p \stackrel{a.s.}{\to} 0$ and $|X_n - X_\infty|^p \le 2^{p-1}(|X_n|^p + (X^*)^p)$ by the inequality $|a+b|^p \le 2^{p-1}(|a|^p + |b|^p)$ by the convexity of $x \mapsto |x|^p$. Thus, $|X_n = X_\infty|^p$ is dominated by $2^p Y^p$ which is integrable. Dominated convergence theorem shows that $\mathbf{E}[|X_n - X_\infty|^p] \to 0$. ■

We used the following lemma in the above proof.

Lemma 41. *Let $(Y_n)_{n\ge 0}$ be an L^p-bounded non-negative sub-martingale. Then $Y^* := \sup_n Y_n$ is in L^p and in fact $\mathbf{E}[(Y^*)^p] \le C_p \sup_n \mathbf{E}[Y_n^p]$.*

Proof. Let $Y_n^* = \max_{k\le n} Y_k$. Fix $\lambda > 0$ and let $T = \min\{k \ge 0 : Y_k \ge \lambda\}$. By the optional sampling theorem, for any fixed n, the sequence of two random variables $\{Y_{T\wedge n}, Y_n\}$ is a sub-martingale. Hence, $\int_A Y_n dP \ge \int_A Y_{T\wedge n} dP$ for any $A \in \mathcal{F}_{T\wedge n}$. Let $A = \{Y_{T\wedge n} \ge \lambda\}$ so that $\mathbf{E}[Y_n \mathbf{1}_A] \ge \mathbf{E}[Y_{T\wedge n}\mathbf{1}_{Y_{T\wedge n}\ge\lambda}] \ge \lambda \mathbf{P}\{Y_n^* \ge \lambda\}$. On the other hand, $\mathbf{E}[Y_n\mathbf{1}_A] \le \mathbf{E}[Y_n \mathbf{1}_{Y^*\ge\lambda}]$ since $Y_n^* \le Y^*$. Thus, $\lambda\mathbf{P}\{Y_n^* > \lambda\} \le \mathbf{E}[Y_n\mathbf{1}_{Y^*\ge\lambda}]$.

Let $n \to \infty$. Since $Y_n^* \uparrow Y^*$, we get

$$\lambda\mathbf{P}\{Y^* > \lambda\} \le \limsup_{n\to\infty} \lambda\mathbf{P}\{Y_n^* \ge \lambda\} \le \limsup_{n\to\infty} \mathbf{E}[Y_n\mathbf{1}_{Y^*\ge\lambda}] = \mathbf{E}[Y_\infty \mathbf{1}_{Y^*\ge\lambda}].$$

where Y_∞ is the a.s. and L^1 limit of Y_n (exists, because $\{Y_n\}$ is L^p bounded and hence uniformly integrable). To go from the tail bound to the bound on pth moment, we use the identity $\mathbf{E}[(Y^*)^p] = \int_0^\infty p\lambda^{p-1}\mathbf{P}\{Y^* \ge \lambda\}d\lambda$ valid for any non-negative random variable in place of Y^*. Using the tail bound, we get

$$\begin{aligned}\mathbf{E}[(Y^*)^p] &\le \int_0^\infty p\lambda^{p-2}\mathbf{E}[Y_\infty\mathbf{1}_{Y^*\ge\lambda}]d\lambda \le \mathbf{E}\left[\int_0^\infty p\lambda^{p-2} Y_\infty \mathbf{1}_{Y^*\ge\lambda} d\lambda\right] \quad \text{(by Fubini's)} \\ &= \frac{p}{p-1}\mathbf{E}[Y_\infty \cdot (Y^*)^{p-1}].\end{aligned}$$

Let q be such that $\frac{1}{q} + \frac{1}{p} = 1$. By Hölder's inequality, $E[Y_\infty \cdot (Y^*)^{p-1}] \le \mathbf{E}[Y_\infty^p]^{\frac{1}{p}} \mathbf{E}[(Y^*)^{q(p-1)}]^{\frac{1}{q}}$. Since $q(p-1) = p$, this gives us $\mathbf{E}[(Y^*)^p]^{1-\frac{1}{q}} \le \frac{p}{p-1}\mathbf{E}[Y_\infty^p]^{\frac{1}{p}}$. But $\mathbf{E}[Y_\infty^p] \le C_p \mathbf{E}[Y_\infty^p]$. The latter is bounded by $C_p \liminf \mathbf{E}[Y_n^p] \le C_p \liminf \mathbf{E}[Y_n^p] \le C_p \sup_n \mathbf{E}[Y_n^p]$ by virtue of Fatou's lemma. ■

One way to think of the martingale convergence theorem is that we have extended the martingale from the index set $\mathbb{N}$ to $\mathbb{N} \cup \{+\infty\}$ retaining the martingale property. Indeed, the given martingale sequence is the Doob martingale given by the limit variable X_∞ with respect to the given filtration.

While almost sure convergence is remarkable, it is not strong enough to yield useful conclusions. Convergence in L^1 or L^p for some $p \geq 1$ are much more useful.

14. REVERSE MARTINGALES

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space. Let $\mathcal{F}_i$, $i \in I$ be sub-sigma algebras of $\mathcal{F}$ indexed by a partially ordered set $(I, \leq)$ such that $\mathcal{F}_i \subseteq \mathcal{F}_j$ whenever $i \leq j$. Then, we may define a martingale or a sub-martingale etc., with respect to this "filtration" $(\mathcal{F}_i)_{i \in I}$. For example, a martingale is a collection of integrable random variables X_i, $i \in I$ such that $\mathbf{E}[X_j \mid \mathcal{F}_i] = X_i$ whenever $i \leq j$.

If the index set is $-\mathbb{N} = \{0, -1, -2, \ldots\}$ with the usual order, we say that X is a reverse martingale or a reverse sub-martingale etc.

What is different about reverse martingales as compared to martingales is that our questions will be about the behaviour as $n \to -\infty$, towards the direction of decreasing information. It turns out that the results are even cleaner than for martingales!

Theorem 42 (Reverse martingale convergence theorem)**.** *Let $X = (X_n)_{n \leq 0}$ be a reverse martingale. Then $\{X_n\}$ is uniformly integrable. Further, there exists a random variable $X_{-\infty}$ such that $X_n \to X_{-\infty}$ almost surely and in L^1.*

Proof. Since $X_n = \mathbf{E}[X_0 \mid \mathcal{F}_n]$ for all n, the uniform integrability follows from Exercise 43.

Let $U_n[a, b]$ be the number of down-crossings of $X_n, X_{n+1}, \ldots, X_0$ over $[a, b]$. The up-crossing inequality (applied to $X_n, \ldots, X_0$ over $[a, b]$) gives $\mathbf{E}[U_n[a, b]] \leq \frac{1}{b-a}\mathbf{E}[(X_0 - a)_+]$. Thus, the expected number of up-crossings $U_\infty[a, b]$ by the full sequence $(X_n)_{n \leq 0}$ has finite expectation, and hence is finite w.p.1.

As before, w.p.1., the number of down-crossings over any interval with rational end-points is finite. Hence, $\lim_{n \to -\infty} X_n$ exists almost surely. Call this $X_{-\infty}$. Uniform integrability shows that convergence also takes place in L^1. ■

The following exercise was used in the proof.

Exercise 43. *Let X be a random variable on $(\Omega, \mathcal{F}, \mathbf{P})$. Then the collection $\{\mathbf{E}[X \mid \mathcal{G}] : \mathcal{G} \subseteq \mathcal{F}\}$ is uniformly integrable.*

What about reverse super-martingales or reverse sub-martingales? Although we shall probably have no occasion to use this, here is the theorem which can be proved on the same lines.

Theorem 44. *Let $(X_n)_{n \leq 0}$ be a reverse super-martingale. Assume that $\sup_n \mathbf{E}[X_n] < \infty$. Then $\{X_n\}$ is uniformly integrable and X_n converges almost surely and in L^1 to some random variable $X_{-\infty}$.*

Proof. Exercise. ■

This covers almost all the general theory that we want to develop. The rest of the course will consist in milking these theorems to get many interesting consequences.

15. Application: Lévy's forward and backward laws

Let X be an integrable random variable on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$.

Question 1: If $\mathcal{F}_n$, $n \geq 0$, is an increasing sequence of sigma-algebras, then what happens to the sequence $\mathbf{E}[X \mid \mathcal{F}_n]$ as $n \to \infty$?

Question 2: If $\mathcal{G}_n$, $n \geq 0$ is a decreasing sequence of sigm-algebras, then what happens to $\mathbf{E}[X \mid \mathcal{G}_n]$ as $n \to \infty$.

Note that the question here is different from conditional MCT. The random variable is fixed and the sigm-algebras are changing. A natural guess is that the limit might be $\mathbf{E}[X \mid \mathcal{F}_\infty]$ and $\mathbf{E}[X \mid \mathcal{G}_\infty]$ respectively, where $\mathcal{F}_\infty = \sigma\{\bigcup_n \mathcal{F}_n\}$ and $\mathcal{G}_\infty = \bigcap_n \mathcal{G}_n$. We shall prove that these guesses are correct.

Forward case: The sequence $X_n = \mathbf{E}[X \mid \mathcal{F}_n]$ is a martingale because of the tower property $\mathbf{E}[\mathbf{E}[X \mid \mathcal{F}_n] \mid \mathcal{F}_m] = \mathbf{E}[X \mid \mathcal{F}_m]$ for $m < n$. Recall that such martingales are called Doob martingales.

Being conditional expectations of a given X, the martingale is uniformly integrable and hence X_n converges $a.s.$and in L^1 to some X_∞. We claim that $X_\infty = \mathbf{E}[X \mid \mathcal{F}_\infty]$ $a.s.$.

Indeed, X_n is $\mathcal{F}_\infty$-measurable for each n and hence the limit X_∞ is $\mathcal{F}_\infty$-measurable (since the convergence is almost sure, there is a null set issue which might make it necessary to either complete the sigma-algebras, or you may interpret it as saying that X_∞ can be modified on a set of zero probability to make it $\mathcal{F}_\infty$-measurable).

Define the measure μ and ν on $\mathcal{F}_\infty$ by $\mu(A) = \int_A X dP$ and $\nu(A) = \int_A X_\infty dP$ for $A \in \mathcal{F}_\infty$. What we want to show is that $\mu(A) = \nu(A)$ for all $A \in \mathcal{F}_\infty$. If $A \in \mathcal{F}_m$ for some m, then $\nu(A) = \int_A X dP$ since $X_m = \mathbf{E}[X \mid \mathcal{F}_m]$. But it is also true that $\nu(A) = \int_A X_n dP$ for $n > m$ since $X_m = \mathbf{E}[X_n \mid \mathcal{F}_m]$. Let $n \to \infty$ and use L^1 convergence to conclude that $\nu(A) = \mu(A)$.

Thus, $\bigcup_n \mathcal{F}_n$ is a π-system on which μ and ν agree. By the $\pi\lambda$ theorem, they agree of $\mathcal{F}_\infty = \sigma\{\bigcup_n \mathcal{F}_n\}$. This completes the proof that $\mathbf{E}[X \mid \mathcal{F}_n] \overset{a.s.,\, L^1}{\longrightarrow} \mathbf{E}[X \mid \mathcal{F}_\infty]$.

Backward case: Write $X_{-n} = \mathbf{E}[X \mid \mathcal{G}_n]$ for $n \in \mathbb{N}$. Then X is a reverse martingale w.r.t the filtration $\mathcal{G}_{-n}$, $n \in \mathbb{N}$. By the reverse martingale convergence theorem, we get that X_n converges almost surely and in L^1 to some X_∞.

We claim that $X_\infty = \mathbf{E}[X \mid \mathcal{G}_\infty]$. Since X_∞ is $\mathcal{G}_n$ measurable for every n (being the limit of X_k, $k \ge n$), it follows that X_∞ is $\mathcal{G}_\infty$-measurable. Let $A \in \mathcal{G}_\infty$. Then $A \in \mathcal{G}_n$ for any n and hence $\int_A X dP = \int_A X_n dP$ which converges to $\int_A X_\infty dP$. Thus, $\int_A X dP = \int_A X_\infty dP$ for all $A \in \mathcal{F}_\infty$.

As a corollary of the forward law, we may prove Kolmogorov's zero-one law.

Theorem 45 (Kolmogoro'v zero-one law). *Let ξ_n, $n \ge 1$ be independent random variables and let $\mathcal{T} = \bigcap_n \sigma\{\xi_n, \xi_{n+1}, \ldots\}$ be the tail sigma-algebra of this sequence. Then $\mathbf{P}(A)$ is 0 or 1 for every $A \in \mathcal{T}$.*

Proof. Let $\mathcal{F}_n = \sigma\{\xi_1, \ldots, \xi_n\}$. Then $\mathbf{E}[\mathbf{1}_A \mid \mathcal{F}_n] \to \mathbf{E}[\mathbf{1}_A \mid \mathcal{F}_\infty]$ in L^1 and almost surely. But $\mathcal{F}_\infty = \sigma\{\xi_1, \xi_2, \ldots\}$. Thus if $A \in \mathcal{T} \subseteq \mathcal{F}_\infty$ then $\mathbf{E}[\mathbf{1}_A \mid \mathcal{F}_\infty] = \mathbf{1}_A$ $a.s.$On the other hand, $A \in \sigma\{xi_n, \xi_{n+1}, \ldots\}$ from which it follws that A is independent of $\mathcal{F}_n$ and hence $\mathbf{E}[\mathbf{1}_A \mid \mathcal{F}_n] = \mathbf{E}[\mathbf{1}_A] = \mathbf{P}(A)$. The conclusion is that $\mathbf{1}_A = \mathbf{P}(A)$ $a.s.$, which is possible if and only if $\mathbf{P}(A)$ equals 0 or 1. ■

16. Critical branching process

Let Z_n, $n \ge 0$ be the generation sizes of a Galton-Watson tree with offspring distribution $p = (p_k)_{k \ge 0}$. If $m = \sum_k k p_k$ is the mean, then Z_n / m^n is a martingale (we saw this earlier).

If $m < 1$, then $\mathbf{P}\{Z_n \ge 1\} \le \mathbf{E}[Z^n] = m^n \to 0$ and hence, the branching process becomes extinct w.p.1. For $m = 1$ this argument fails. We show using martingales that extinction happens even in this cases.

Theorem 46. *If $m = 1$ and $p_1 \ne 1$, then the branching process becomes extinct almost surely.*

Proof. If $m = 1$, then Z_n is a non-negative martingale and hence converges almost surely to some Z_∞. But Z_n is integer-valued. Thus,

$$Z_\infty = j \Leftrightarrow Z_n = j \text{ for all } n \ge n_0 \text{ for some } n_0.$$

But if $j \ne 0$ and $p_1 < 1$, then it is easy to see that $\mathbf{P}\{Z_n = j \text{ for all } n \ge n_0\} = 0$ (since conditional on $\mathcal{F}_{n-1}$, there is a probability of p_0^j that $Z_n = 0$). Thus, $Z_n = 0$ eventually. ■

In the supercritical case we know that there is a positive probability of survival. If you do not know this, prove it using the second moment method as follows.

Exercise 47. *By conditioning on $\mathcal{F}_{n-1}$ (or by conditioning on $\mathcal{F}_1$), show that (1) $\mathbf{E}[Z_n] = m^n$, (2) $\mathbf{E}[Z_n^2] \asymp (1+\sigma^2)m^{2n}$. Deduce that $\mathbf{P}\{Z_n > 0\}$ stays bounded away from zero. Conclude positive probability of survival.*

We also have the martingale Z_n/m^n. By the martingale convergence theorem $W := \lim Z_n/m^n$ exists, $a.s.$ On the event of extinction, clearly $W = 0$. On the event of survival, is it necessarily the case that $W > 0$ $a.s.$? If yes, this means that whenever the branching process survives, it does so by growing exponentially, since $Z_n \sim W \cdot m^n$. The answer is given by the famous Kesten-Stigum theorem.

Theorem 48 (Kesten-Stigum theorem). *Assume that $\mathbf{E}[L] > 1$ and that $p_1 \neq 1$. Then, $W > 0$ almost surely on the event of survival if and only if $\mathbf{E}[L \log_+ L] < \infty$.*

Probably in a later lecture we shall prove a weaker form of this, that if $\mathbf{E}[L^2] < \infty$, then $W > 0$ on the event of survival.

17. PÓLYA'S URN SCHEME

Initially the urn contain b black and w white balls. Let B_n be the number of black balls after n steps. Then $W_n = b + w + n - B_n$. We have seen that $X_n := B_n/(B_n + W_n)$ is a martingale. Since $0 \leq X_n \leq 1$, uniform integrability is obvious and $X_n \to X_\infty$ almost surely and in L^1. Since X_n are bounded, the convergence is also in L^p for every p. In particular, $\mathbf{E}[X_n^k] \to \mathbf{E}[X_\infty^k]$ as $n \to \infty$ for each $k \geq 1$.

Theorem 49. *X_∞ bas Beta(b, w) distribution.*

Proof. Let V_k be the colour of the kth ball drawn. It takes values 1 (for black) and 0 (for white). It is an easy exercise to check that

$$\mathbf{P}\{V_1 = \epsilon_1, \ldots, V_m = \epsilon_m\} = \frac{b(b+1)\ldots(b+r-1)w(w+1)\ldots(w+s-1)}{(b+w)(b+w+1)\ldots(b+w+n-1)}$$

if $r = \epsilon_1 + \ldots + \epsilon_m$ and $s = n - r$. The key point is that the probability does not depend on the order of ϵ_is. In other words, any permutation of $(V_1, \ldots, V_n)$ has the same distribution as $(V_1, \ldots, V_n)$, a property called *exchangeability*.

From this, we see that for any $0 \leq r \leq n$, we have

$$\mathbf{P}\{X_n = \frac{b+r}{b+w+n}\} = \binom{n}{r} \frac{b(b+1)\ldots(b+r-1)w(w+1)\ldots(w+(n-r)-1)}{(b+w)(b+w+1)\ldots(b+w+n-1)}.$$

In the simplest case of $b = w = 1$, the right hand side is $\frac{1}{n+1}$. That is, X_n takes the values $\frac{r+1}{n+2}$, $0 \leq r \leq n$, with equal probabilities. Clearly then $X_n \xrightarrow{d}$ Unif$[0,1]$. Hence, $X_\infty \sim$ Unif$[0,1]$. In general, we leave it as an exercise to show that X_∞ has Beta(b, w) distribution. ■

Here is a possibly clever way to avoid computations in the last step.

Exercise 50. *For each initial value of b, w, let $\mu_{b,w}$ be the distribution of X_∞ when the urn starts with b black and w white balls. Each $\mu_{b,w}$ is a probability measure on $[0, 1]$.*

1. *Show that $\mu_{b,w} = \frac{b}{b+w}\mu_{b+1,w} + \frac{w}{b+w}\mu_{b,w+1}$.*
2. *Check that Beta(b, w) distributions satisfy the above recursions.*
3. *Assuming $(b, w) \mapsto \mu_{b,w}$ is continuous, deduce that $\mu_{b,w} =$ Beta(b, w) is the only solution to the recursion.*

One can introduce many variants of Pólya's urn scheme. For example, whenever a ball is picked, we may add r balls of the same color and q balls of the opposite color. That changes the behaviour of the urn greatly and in a typical case, the proportions of black balls converges to a constant.

Here is a muti-color version which shares all the features of Pólya's urn above.

Multi-color Pólya's urn cheme: We have ℓ colors denoted $1, 2, \ldots, \ell$. Initially an urn contains $b_k > 0$ balls of color k (b_k need not be integers). At each step of the process, a ball is drawn uniformly at random from the urn, its color noted, and returned to the urn with another ball of the same color. Let $B_k(n)$ be the number of balls of jth color after n draws. Let ξ_n be the color of the ball drawn in the nth draw.

Exercise 51. *(1) Show that $\frac{1}{n+b_1+\ldots+b_\ell}(B_1(n), \ldots, B_\ell(n))$ converges almost surely (and in L^p for any p) to some random vector $(Q_1, \ldots, Q_\ell)$.*

(2) Show that $\xi_1, \xi_2, \ldots$ is an exchangeable sequence.

(3) For $b_1 = \ldots = b_\ell = 1$, show that $(Q_1, \ldots, Q_\ell)$ has Dirichlet$(1, 1, \ldots, 1)$ distribution. In general, it has Dirichlet$(b_1, \ldots, b_\ell)$ distribution.

This means that $Q_1 + \ldots + Q_\ell = 1$ and $(Q_1, \ldots, Q_{\ell-1})$ has density

$$\frac{\Gamma(b_1 + \ldots + b_\ell)}{\Gamma(b_1) \ldots \Gamma(b_\ell)} x_1^{b_1-1} \ldots x_{\ell-1}^{b_{\ell-1}-1} (1 - x_1 - \ldots - x_{\ell-1})^{b_\ell - 1}$$

on $\Delta = \{(x_1, \ldots, x_{\ell-1}) : x_i > 0$ for all i and $x_1 + \ldots + x_{\ell-1} < 1\}$.

18. LIOUVILLE'S THEOREM

Recall that a harmonic function on $\mathbb{Z}^2$ is a function $f : \mathbb{Z}^2 \to \mathbb{R}$ such that $f(x) = \frac{1}{4}\sum_{y:y\sim x} f(x, y)$ for all $x \in \mathbb{Z}^2$.

Theorem 52 (Liouville's theorem)**.** *If f is a non-constant harmonic function on $\mathbb{Z}^2$, then* $\sup f = +\infty$ *and* $\inf f = -\infty$.

Proof. If not, by negating and/or adding a constant we may assume that $f \geq 0$. Let X_n be simple random walk on $\mathbb{Z}^2$. Then $f(X_n)$ is a martingale. But a non-negative super-martingale converges almost surely. Hence $f(X_n)$ converges almost surely.

But Pólya's theorem says that X_n visits every vertex of $\mathbb{Z}^2$ infinitely often w.p.1. This contradicts the convergence of $f(X_n)$ unless f is a constant. ■

Observe that the proof shows that a non-constant super-harmonic function on $\mathbb{Z}^2$ cannot be bounded below. The proof uses recurrence of the random walk. But in fact the same theorem holds on $\mathbb{Z}^d$, $d \geq 3$, although the simple random walk is transient there.

For completeness, here is a quick proof of Pólya's theorem in two dimensions.

Exercise 53. *Let S_n be simple symmetric random walk on $\mathbb{Z}^2$ atarted at $(0,0)$.*

(1) *Show that $\mathbf{P}\{S_{2n}=(0,0)\} = \frac{1}{4^{2n}}\sum_{k=0}^{n}\frac{(2n)!}{k!^2(n-k)!^2}$ and that this expression reduces to $\left(\frac{1}{2^{2n}}\binom{2n}{n}\right)^2$.*

(2) *Use Stirling's formula to show that $\sum_n \mathbf{P}\{S_{2n}=(0,0)\}=\infty$.*

(3) *Conclude that $\mathbf{P}\{S_n=(0,0) \text{ i.o.}\}=1$.*

The question of existence of bounded or positive harmonic functions on a graph (or in the continuous setting) is important. Here are two things that we may cover if we get time.

- There are no bounded harmonic functions on $\mathbb{Z}^d$ (Blackwell).
- Let μ be a probability measure on $\mathbb{R}$ and let f be a harmonic function for the random walk with step distribution μ. This just means that f is continuous and $\int_{\mathbb{R}} f(x+a)d\mu(x)=f(a)$. Is f necessarily constant? We shall discuss this later (under the heading "Choquet-Deny theorem").

19. Strong law of large numbers

Theorem 54. *Let ξ_n, $n\ge 1$ be i.i.d. real-valued random variables with zero mean and let $S_n=\xi_1+\ldots+\xi_n$. Then $\frac{1}{n}S_n \stackrel{a.s.}{\to} 0$.*

Proof. Let $\mathcal{G}_n=\sigma\{S_n,S_{n+1},\ldots\}=\sigma\{S_n,\xi_{n+1},\xi_{n+2},\ldots\}$, a decreasing sequence of sigma-algebras. Hence $M_{-n}:=\mathbf{E}[\xi_1 \mid \mathcal{G}_n]$ is a reverse martingale and hence converges almsot surely and in L^1 to some $M_{-\infty}$.

But $\mathbf{E}[\xi_1 \mid \mathcal{G}_{n+1}]=\frac{1}{n}S_n$ (why?). Thus, $\frac{1}{n}S_n\to M_{-\infty}$ almost surely and in L^1. But the limit of $\frac{1}{n}S_n$ is clearly a tail random variable of ξ_ns and hence must be constant. Thus, $M_{-\infty}=\mathbf{E}[M_{-\infty}]=\lim\frac{1}{n}\mathbf{E}[S_n]=0$. In conclusion, $\frac{1}{n}S_n\stackrel{a.s.}{\to}0$. ■

20. Hewitt-Savage zero-one law

There are many zero-one laws in probability, asserting that a whole class of events are trivial. For a sequence of random variables, here are three important classes of such events.

Below, ξ_n, $n\ge 1$, are random variables on a common probability space $(\Omega,\mathcal{F},\mathbf{P})$ and taking values in $(X,\mathcal{F})$. Then $\xi=(\xi_n)_{n\ge 1}$ is a random variable taking values in $(X^{\mathbb{N}},\mathcal{F}^{\otimes\mathbb{N}})$. These definitions can be extended to two sided-sequences $(\xi_n)_{n\in\mathbb{Z}}$ easily.

(1) The *tail sigma-algebra* is defined as $\mathcal{T}=\cap_n\mathcal{T}_n$ where $\mathcal{T}_n=\sigma\{\xi_n,\xi_{n+1},\ldots\}$.

(2) The *exchangeable sigma-algebra* $\mathcal{S}$ is the sigma-algebra of those events that are invariant under finite permutations.

More precisely, let G be the sub-group (under composition) of all bijections $\pi:\mathbb{N}\to\mathbb{N}$ such that $\pi(n)=n$ for all but finitely many n. It is clear how G acts on $X^{\mathbb{N}}$:

$$\pi((\omega_n))=(\omega_{\pi(n)}).$$

Then

$$\mathcal{S} := \{\xi^{-1}(A) : A \in \mathcal{F}^{\otimes \mathbb{N}} \text{ and } \pi(A) = A \text{ for all } \pi \in G\}.$$

If G_n is the sub-group of $\pi \in G$ such that $\pi(k) = k$ for every $k > n$ and $\mathcal{S}_n := \{\xi^{-1}(A) : A \in \mathcal{F}^{\otimes \mathbb{N}}$ and $\pi(A) = A$ for all $\pi \in G_n\}$, then $\mathcal{S}_n$ are sigma-algebras that decrease to $\mathcal{S}$.

(3) The *translation-invariant sigma-algebra* $\mathcal{I}$ is the sigma-algebra of all events invariant under translations.

More precisely, let $\theta_n : X^{\mathbb{N}} \to X^{\mathbb{N}}$ be the translation map $[\theta_n(\omega)]_k = \omega_{n+k}$. Then, $\mathcal{I} = \{A \in \mathcal{F}^{\otimes \mathbb{N}} : \theta_n(A) = A \text{ for all } n \in \mathbb{N}\}$ (these are events invariant under the action of the semi-group $\mathbb{N}$).

Kolmogorov's zero-one law asserts that under and product measure $\mu_1 \otimes \mu_2 \otimes \ldots$, the tail s-gma-algebra is trivial. Ergodicity is the statement that $\mathcal{I}$ is trivial and it is true for i.i.d. product measures $\mu^{\otimes \mathbb{N}}$. The exchangeable sigma-algebra is also trivial under i.i.d. product measure, which is the result we prove in this section. First an example.

Example 55. *The event $A = \{\omega \in \mathbb{R}^{\mathbb{N}} : \lim \omega_n = 0\}$ is an invariant event. In fact, every tail event is an invariant event. But the converse is not true. For example,*

$$A = \{\omega \in \mathbb{R}^{\mathbb{N}} : \limsup_{n\to\infty}(\omega_1 + \ldots + \omega_n) \leq 0\}$$

is an invariant event but not a tail event. This is because $\omega = (-1, 1, -1, 1, -1, \ldots)$ belongs to A but $\omega' = (1, -1, -1, 1, -1, \ldots)$ got by permuting the first two co-ordinates, is not in A.

Theorem 56 (Hewitt-Savage 0-1 law)**.** *Let μ be a probability measure on $(X, \mathcal{F})$. Then the invariant sigma-algebra $\mathcal{S}$ is trivial under the product measure $\mu^{\otimes \mathbb{N}}$.*

In terms of random variables, we may state this as follows: Let ξ_n be i.i.d. random variables taking values in X. Let $f : X^{\mathbb{N}} \mapsto \mathbb{R}$ be a measurable function such that $f \circ \pi = f$ for all $\pi \in G$. Then, $f(\xi_1, \xi_2, \ldots)$ is almost surely a constant.

We give a proof using reverse martingale theorem. There are also more direct proofs.

Proof. For any integrable Y (that is measurable w.r.t $\mathcal{F}^{\otimes \mathbb{N}}$), the sequence $\mathbf{E}[Y \mid \mathcal{S}_n]$ is a reverse martingale and hence $\mathbf{E}[Y \mid \mathcal{S}_n] \overset{a.s.,\, L^1}{\longrightarrow} \mathbf{E}[Y \mid \mathcal{S}]$.

Now fix $k \geq 1$ and let $\varphi : X^k \to \mathbb{R}$ be a bounded measurable function. Take $Y = \varphi(X_1, \ldots, X_k)$. We claim that

$$\mathbf{E}[\varphi(X_1, \ldots, X_k) \mid \mathcal{S}_n] = \frac{1}{n(n-1)\ldots(n-k+1)} \sum_{\substack{1 \leq i_1, \ldots, i_k \leq n \\ \text{distinct}}} \varphi(X_{i_1}, \ldots, X_{i_k}).$$

To see this, observe that by symmetry (since $\mathcal{S}_n$ does no distinguish between $X_1, \ldots, X_n$), we have $\mathbf{E}[\varphi(X_{i_1}, \ldots, X_{i_k}) \mid \mathcal{S}_n]$ is the same for all distinct $i_1, \ldots, i_k \leq n$. When you add all these up, we get

$$\mathbf{E}\left[\sum_{\substack{1\le i_1,\dots,i_k\le n\\ \text{distinct}}} \varphi(X_{i_1},\dots,X_{i_k}) \,\Big|\, \mathcal{S}_n\right] = \sum_{\substack{1\le i_1,\dots,i_k\le n\\ \text{distinct}}} \varphi(X_{i_1},\dots,X_{i_k})$$

since the latter is clearly $\mathcal{S}_n$-measurable. There are $n(n-1)\dots(n-k+1)$ terms on the left, each of which is equal to $\mathbf{E}[\varphi(X_1,\dots,X_k)\,|\,\mathcal{S}_n]$. This proves the claim.

Together with the reverse martingale theorem, we have shown that

$$\frac{1}{n(n-1)\dots(n-k+1)} \sum_{\substack{1\le i_1,\dots,i_k\le n\\ \text{distinct}}} \varphi(X_{i_1},\dots,X_{i_k}) \overset{a.s.,\, L^1}{\longrightarrow} \mathbf{E}[\varphi(X_1,\dots,X_k)\,|\,\mathcal{S}].$$

The number of summands on the left in which X_1 participates is $k(n-1)(n-2)\dots(n-k+1)$. If $|\varphi|\le M_\varphi$, then the total contribution of all terms containing X_1 is at most

$$M_\varphi \frac{k(n-1)(n-2)\dots(n-k+1)}{n(n-1)(n-2)\dots(n-k+1)} \to 0$$

as $n\to\infty$. Thus, the limit is a function of $X_2,X_3,\dots$. By a similar reasoning, the limit is a tail-random variable for the sequence $X_1,X_2,\dots$. By Kolmogorov's zero-one law it must be a constant (then the constant must be its expectation). Hence,

$$\mathbf{E}[\varphi(X_1,\dots,X_k)\,|\,\mathcal{S}] = \mathbf{E}[\varphi(X_1,\dots,X_k)].$$

As this is true for every bounded measurable φ, we see that $\mathcal{S}$ is independent of $\sigma\{X_1,\dots,X_k\}$. As this is true for every k, $\mathcal{S}$ is independent of $\sigma\{X_1,X_2,\dots\}$. But $\mathcal{S}\subseteq\sigma\{X_1,X_2,\dots\}$ and therefore $\mathcal{S}$ is independent of itself. This implies that for any $A\in\mathcal{S}$ we must have $\mathbf{P}(A)=\mathbf{P}(A\cap A)=\mathbf{P}(A)^2$ which implies that $\mathbf{P}(A)$ equals 0 or 1. ■

21. Exchangeable random variables

Let ξ_n, $n\ge 1$, be any sequence of random variables. Let $S_n=\xi_1+\ldots+\xi_n$ and $\mathcal{G}_n=\sigma\{S_n,S_{n+1},\dots\}$ and $\mathcal{G}=\cap_n\mathcal{G}_n$.

Then $X_n := \mathbf{E}[\xi_1\,|\,\mathcal{G}_n]$ is a reverse-martingale. If $(\xi_n)_n$ is an exchangeable sequence, then $\mathbf{E}[\xi_k\,|\,\mathcal{G}_{n+1}]$ is the same for $1\le k\le n$ by symmetry (write out the argument clearly), and hence $X_n=\frac{1}{n}S_n$. By the reverse martingale theorem

$$\frac{1}{n}S_n \overset{a.s.,L^1}{\longrightarrow} \mathbf{E}[\xi_1\,|\,\mathcal{G}].$$

Although the limit of S_n/n is a tail random variable of the sequence $(\xi_n)_n$, outside the independent case, there is no reason for it to be constant.

xxxxxNeed to write this yetxxxxxxxxx