

---

# Statistics 212: Lecture 3 (Feb. 3, 2025)

## Introduction to Martingales

---

Instructor: Mark Sellke

Scribe: Diego Luca González Gauss, Enrico Antonio Jose Yao Yao-Bate

### 1 Odds and ends of conditional probability

We begin by proving a few more useful properties of conditional expectation.

**Proposition 1.1.** *If  $X \geq 0$  a.s., then  $\mathbb{E}(X|\mathcal{G}) \geq 0$  a.s.*

*Proof.* Define  $S := \{\mathbb{E}(X|\mathcal{G}) < 0\}$ . If  $P(S) > 0$ , then

$$0 \leq \int_S X dP = \int_S \mathbb{E}(X|\mathcal{G}) dP < 0,$$

a contradiction. □

**Corollary 1.2.** *If  $X \geq Y$ , then  $\mathbb{E}(X|\mathcal{G}) \geq \mathbb{E}(Y|\mathcal{G})$  a.s.*

**Proposition 1.3** (Tower rule). *Let  $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \mathcal{F}_3$ , with  $X$   $\mathcal{F}_3$ -measurable. Then*

$$\mathbb{E}[X|\mathcal{F}_1] = \mathbb{E}[\mathbb{E}[X|\mathcal{F}_2]|\mathcal{F}_1]$$

*Proof.* For all  $S \in \mathcal{F}_1 \subseteq \mathcal{F}_2$ ,

$$\int_S X dP = \int_S \mathbb{E}(X|\mathcal{F}_2) dP = \int_S \mathbb{E}(\mathbb{E}(X|\mathcal{F}_2)|\mathcal{F}_1) dP,$$

since  $\mathbb{E}(X|\mathcal{F}_2)$  is a  $\mathcal{F}_1$ -measurable set. □

To make the  $\mathcal{F}_3$  seem less mysterious, you can take  $\mathcal{F}_3 = \mathcal{F}$  as an example. Then what the above proposition just says is that for any measurable  $X$ , the relation above is true.

#### 1.1 Well, what about conditional probabilities and distributions?

Given that we've just rigorously defined conditional expectations, it is natural to then wonder how we can endow conditional probabilities and distributions with similarly rigorous definitions.

**Definition 1.4** (Conditional Probability).  $P(A|\mathcal{G}) := \mathbb{E}[I_A|\mathcal{G}]$ , where  $I_A$  is an indicator r.v.

This suggests that understanding conditional probability will conveniently follow from understanding conditional expectation.

**Definition 1.5** (Conditional Distribution). Suppose  $X$  is in a complete separable metric space, like  $C([0, 1])$  from the homework, or just  $\mathbb{R}$ . Then, for each  $q \in A$ , where  $A$  is the associated countable dense set, we can consider  $\mathbb{E}[X \leq q|\mathcal{G}]$ . This gives us our conditional distribution  $\text{Law}(X|\mathcal{G})$ .

## 2 Martingales

We now proceed to developing martingale theory. First, we give again the definition of a martingale, as well as some other useful constructions.

**Definition 2.1** (Filtration). A filtration  $\mathcal{F}_n$  is a sequence of nested  $\sigma$ -algebras  $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \dots$

We say that  $(X_t)_{t \in \mathbb{N}}$  is adapted to  $(\mathcal{F}_t)_{t \in \mathbb{N}}$  if  $X_t$  is  $\mathcal{F}_t$ -measurable for all  $t$ . As a notational convention, we let  $X_0$  be deterministic, and  $\mathcal{F}_0 = \{\emptyset, \Omega\}$  the trivial sigma-algebra.

**Definition 2.2** (Martingale). A martingale is a sequence  $(X_t)_{t \in \mathbb{N}}$  of r.v.s satisfying for all  $t$ ,

- (a)  $E|X_t| < \infty$
- (b)  $X_t$  is adapted to  $\mathcal{F}_t$
- (c)  $E(X_{t+1} | \mathcal{F}_t) = X_t$

**Definition 2.3** (Stopping time).  $\tau \in \mathbb{Z}_{\geq 0}$  is a stopping time if for all  $t$ , we have  $\{\tau \leq t\} \in \mathcal{F}_t$ .

**Theorem 2.4.** A stopped martingale is a martingale. That is, given  $X_t$  martingale and  $\tau$  stopping time, then  $X_{t \wedge \tau}$  is a martingale, where  $t \wedge \tau := \min\{t, \tau\}$ .

*Proof.* Clearly  $X_{t \wedge \tau}$  is  $\mathcal{F}_t$ -measurable. Fix any  $S \in \mathcal{F}_t$ . We want to show that  $E[X_{(t+1) \wedge \tau} | \mathcal{F}_t] = X_{t \wedge \tau}$ . To this end, decompose  $S = (S \cap \{\tau \leq t\}) \cup (S \cap \{\tau > t\})$ , which we'll call  $S_1$  and  $S_2$ , respectively. Then

$$\begin{aligned} \int_{S_1} X_{t+1 \wedge \tau} - X_{t \wedge \tau} dP &= \int X_t - X_t dP = 0 \\ \int_{S_2} X_{t+1 \wedge \tau} - X_{t \wedge \tau} dP &= \int_{S_2} X_{t+1} - X_t dP = 0, \end{aligned}$$

with the second equality in the second line because of property (c) of martingale  $X_t$ . □

### 2.1 An optional stopping theorem for arbitrary stopping times

Recall first the martingale convergence theorem and the optional stopping theorem from STAT 210.

**Theorem 2.5** (Martingale Convergence Theorem). If  $X_t$  martingale and  $\sup_{t \geq 0} E|X_t| < \infty$ , then  $X_\infty := \lim_{t \rightarrow \infty} X_t$  exists a.s.

**Theorem 2.6** (OST). Suppose  $X_t$  is a martingale w.r.t  $\mathcal{F}_t$ , and  $\tau$  is a stopping time. If  $\tau \leq T < \infty$  is almost surely bounded, then

$$EX_\tau = EX_0. \tag{1}$$

Note: using a limiting argument, one can establish (1) under alternative conditions such as:

- $P(\tau < \infty) = 1$  and  $|X_n| \leq c$  a.s. for all  $n$  for some constant  $c$
- $E(\tau) < \infty$  and  $|X_n - X_{n+1}| \leq c$  for all  $n$  and some constant  $c$ .

In both, the idea is to argue that  $\lim_{n \rightarrow \infty} E|X_n - X_\infty| = 0$ .

Looking at the above theorems, it is natural to wonder if OST can be applied to  $X_t$  without any of the annoying conditions. For example, it would be nice if  $E[X_\infty] = X_0$ . Unfortunately, the example below illustrates that this will not always be possible.

*Example* (Gambler's ruin). Let  $X_0 = 1$  and recursively define  $X_t$  by

$$X_{t+1}/X_t = \begin{cases} 2 & \text{with probability } 1/2 \\ 0 & \text{with probability } 1/2 \end{cases}.$$

Then  $X_t$  is a martingale, but  $E[X_\tau] = 0$  a.s., and therefore  $X_\tau = 0$  a.s., if  $\tau$  is the first time  $t$  that  $X_t = 0$ .

Thus, the natural question to ask is under what conditions are we guaranteed  $\mathbb{E}[X_\infty] = X_0$ . It is easy to see that a *sufficient* condition is  $X_t \xrightarrow{L^1} X_\infty$ . Indeed, then we would have

$$\lim_{n \rightarrow \infty} |\mathbb{E}[X_\infty] - \mathbb{E}[X_0]| \leq \lim_{n \rightarrow \infty} |\mathbb{E}[X_\infty - X_0]| = 0.$$

In fact **because** we already know  $X_n \rightarrow X_\infty$  almost surely, we will see that  $L^1$  convergence is equivalent to uniform integrability. The rest of the class will be devoted to understanding uniform integrability.

**Definition 2.7** (Uniformly Integrable Set of Random Variables).  $S \subseteq L^1(\Omega, \mathcal{F}, P)$  is U.I. if  $\lim_{c \rightarrow \infty} \sup_{X \in S} \mathbb{E}[|X| \cdot \mathbb{1}_{|X| \geq c}] = 0$ . Equivalently, for all  $\epsilon > 0$ , we have some  $c$  such that  $\sup_{X \in S} \mathbb{E}[|X| \cdot \mathbb{1}_{|X| \geq c}] < \epsilon$ .

*Example.* The gambler's ruin is not U.I.

*Proof.* Fix  $c > 0$ , and choose  $t$  s.t.  $2^t > c$ . Then  $\mathbb{E}[|X_t| \cdot \mathbb{1}_{|X_t| \geq c}] = 1$ . □

**Theorem 2.8.** If  $X_t \in L^1$  for all  $t \geq 0$  and  $X_t \xrightarrow{a.s.} X_\infty$ , then the following are equivalent:

- (a)  $\{X_t : t \geq 1\}$  is U.I.
- (b)  $X_\infty \in L^1$  and  $X_t \xrightarrow{L^1} X_\infty$
- (c)  $X_\infty \in L^1$  and  $\mathbb{E}|X_t| \rightarrow \mathbb{E}|X_\infty|$

The proof is deferred to the next class; however, we will introduce some useful lemmas to this end.

**Lemma 2.9.**  $S \subseteq L^1(\Omega, \mathcal{F}, P)$  is U.I. if and only if

- (a)  $S$  is  $L^1$ -bounded, i.e.  $\sup_{X \in S} \mathbb{E}[|X|] < \infty$
- (b) For all  $\epsilon > 0$ , there exists some  $\delta > 0$  such that for all  $X \in S$  and  $A \in \mathcal{F}$  with  $P(A) \leq \delta$ , we get

$$\mathbb{E}[|X| \cdot \mathbb{1}_A] \leq \epsilon$$

*Proof.* To see that UI implies a), we can choose  $(\epsilon, c)$  following the definition of UI. Then for all  $X \in S$ ,

$$\mathbb{E}[|X|] = \mathbb{E}[|X| \cdot \mathbb{1}_{|X| \leq c}] + \mathbb{E}[|X| \cdot \mathbb{1}_{|X| > c}] \leq c + \epsilon < \infty.$$

To see that UI implies b), we can fix some  $\epsilon > 0$  and choose  $c > 0$  for UI, and then take  $\delta \leq \epsilon/c$  so that

$$\mathbb{E}[|X| \cdot \mathbb{1}_A] = \mathbb{E}[|X| \cdot \mathbb{1}_{A \cap \{|X| \leq c\}}] + \mathbb{E}[|X| \cdot \mathbb{1}_{A \cap \{|X| > c\}}] \leq cP(A) + \epsilon \leq c\delta + \epsilon \leq 2\epsilon.$$

Lastly we can assume a) and b) and prove that this implies UI. Fixing  $\epsilon > 0$ , we can invoke b) to take some  $\delta$ , and then choose  $c$  to ensure UI. Actuating this, (a) lets us fix  $M := \sup_{X \in S} \mathbb{E}[|X|] < \infty$ . By Markov's inequality,

$$P(|X| \geq c) \leq M/c \leq \delta.$$

Then letting  $A = \{|X| \geq c\}$ , we get  $P(A) \leq \delta$  which implies  $\mathbb{E}[|X| \cdot \mathbb{1}_A] \leq \epsilon$  by (b). □

**Lemma 2.10.** If  $X \in L^1(\Omega, \mathcal{F}, P)$ , then  $S = \{\mathbb{E}[X|\mathcal{G}], \mathcal{G} \subseteq \mathcal{F}\}$  is U.I. (for  $\mathcal{F}$  a  $\sigma$ -algebra; this is motivated by the case where if  $X = X_\infty$ , then  $S = \{X_t : t \geq 0\}$ ).

To prove this lemma, we need to invoke one other recharacterization of U.I..

**Proposition 2.11.** *S is U.I. if and only if for all  $\epsilon$  there exists  $c$  such that  $\sup_{X \in S} \mathbb{E}(|X| - c)_+ \leq \epsilon$ .*

*Proof.* We get the following inequality (this can be motivated graphically for intuition):

$$(|X| - 2c)_+ \leq |X| \cdot 1_{|X| \geq 2c} \leq 2 \cdot (|X| - c)_+, \quad \forall X \in \mathbb{R}.$$

Now, assuming U.I., we can realize that our middle term will vanish, and so by our inequality the left-hand side will vanish as well. Then taking  $c = c'/2$  and considering the left-hand side we get  $(|X| - c)_+$  vanishes, so our right-hand side will vanish for  $c'$ ; but then we get  $\sup_{X \in S} \mathbb{E}[2 \cdot (|X| - c)_+] = \sup_{X \in S} 2\mathbb{E}(|X| - c)_+$  is bounded as chosen, so  $(|X| - c)_+ \leq \epsilon$  for chosen epsilon too and we have our equivalent definition.

In the other direction, let us assume  $\sup_{X \in S} \mathbb{E}(|X| - c)_+ \leq \epsilon$  in our equivalent definition, we can realize the right-hand side of our inequality vanishes, and so the middle term does as well. But the middle term vanishing is equivalent to U.I. and we're done. □

Now that we have this new definition for U.I., we can easily prove Lemma 2.12.

*Proof of 2.12.* Apply Jensen's inequality to  $\phi(X) = (|X| - c)_+$  to get

$$\mathbb{E}[\phi(\mathbb{E}[X|\mathcal{G}])] \leq \mathbb{E}[\phi(X)] = \mathbb{E}[|X| - c]_+ \leq \epsilon.$$

Realize that the leftmost term is equivalent to  $\mathbb{E}[(|Y| - c)_+]$  for  $Y = \mathbb{E}[X|\mathcal{G}]$ . Thus, in our equivalent definition of UI, any  $(\epsilon, c)$  that works for  $X$  also works for  $\mathbb{E}[X|\mathcal{G}]$ . (Note: this doesn't mean any  $(\epsilon, c)$  in the *original* definition of UI also works for  $\mathbb{E}[X|\mathcal{G}]$ . But checking through the proof, it means we just need to adjust by factors of 2.) □