

# STATS310B – Lecture 3

Sourav Chatterjee  
Scribed by Michael Howes

01/11/22

## Contents

1	Jensen's inequality	1
2	Martingales	2
3	Stopping times	4

## 1 Jensen's inequality

Jensen's inequality is an important result about conditional expectations and convex functions. It states:

**Theorem 1** (Jensen's inequality). *Let  $X$  be an integrable random variable taking values in some interval  $I$ . Let  $\phi : I \rightarrow \mathbb{R}$  be a convex function such that  $\phi(X)$  is integrable. Then for any sub- $\sigma$ -algebra  $\mathcal{G}$ , we have*

$$\phi(\mathbb{E}(X|\mathcal{G})) \leq \mathbb{E}(\phi(X)|\mathcal{G}),$$

*almost surely.*

To prove Jensen's inequality, we will first prove some lemmas about convex functions.

**Lemma 1.** *Let  $\mathcal{A}$  be a collection of convex functions  $f : I \rightarrow \mathbb{R}$ . The function*

$$g(x) = \sup_{f \in \mathcal{A}} f(x),$$

*is also convex.*

*Proof.* Let  $x, y \in I$  and let  $t \in [0, 1]$ . For each  $f \in \mathcal{A}$  we have

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y) \leq tg(x) + (1-t)g(y).$$

Thus,

$$g(tx + (1-t)y) = \sup_{f \in \mathcal{A}} f(tx + (1-t)y) \leq tg(x) + (1-t)g(y) \quad \square$$

This lemma has the immediate corollary,

**Corollary 1.** *The supremum of a collection of affine functions is a convex function.*

We will next see that above corollary has a converse.

**Lemma 2.** Let  $\phi : I \rightarrow \mathbb{R}$  be a convex function on some interval  $I$ . Then there exists sequences  $a_n, b_n \in \mathbb{R}$  such that

$$\phi(x) = \sup_{n \in \mathbb{N}} a_n x + b_n.$$

*Proof.* We first will state and prove a fact. If  $\phi$  is convex and  $y \in I$ , then there exists a line  $x \mapsto a_y x + b_y$  such that  $\phi(y) = a_y y + b_y$  and  $\phi(x) \geq a_y x + b_y$  for all  $x \in I$ . That is, there exists a line through  $(y, \phi(y))$  that sits below the graph of  $\phi$ . The justification is as follows. By convexity, the function  $\frac{\phi(y+t) - \phi(y)}{t}$  is a decreasing function of  $t$ . Thus, the limits

$$c = \lim_{t \nearrow 0} \frac{\phi(y+t) - \phi(y)}{t}, \quad \text{and} \quad d = \lim_{t \searrow 0} \frac{\phi(y+t) - \phi(y)}{t},$$

both exist and  $c \leq d$ . Thus, we can take  $a_y$  to be any value in  $[c, d]$  and let  $b_y = \phi(y) - a_y y$ . The line  $x \mapsto a_y x + b_y$  sits below the graph  $\phi$  by the convexity of  $\phi$ .

Now define  $g : I \rightarrow \mathbb{R}$  by

$$g(x) = \sup_{y \in \mathbb{Q} \cap I} a_y x + b_y.$$

The function  $g$  is the supremum of a countable collection of affine functions and thus  $g$  is convex. In particular  $g$  and  $\phi$  are both continuous. Furthermore,  $\phi(x) = g(x)$  on  $\mathbb{Q} \cap I$  and so  $\phi = g$  which completes the proof.  $\square$

We are now ready to prove Jensen's inequality.

*Proof of Jensen's inequality.* By Lemma 2, we have

$$\phi(x) = \sup_{n \in \mathbb{N}} a_n x + b_n.$$

Thus, for every  $n$ ,

$$\mathbb{E}(\phi(X)|\mathcal{G}) \geq \mathbb{E}(a_n X + b_n|\mathcal{G}) = a_n \mathbb{E}(X|\mathcal{G}) + b_n,$$

almost surely. One can show that  $\mathbb{E}(X|\mathcal{G}) \in I$  almost surely since  $I$  is an interval. Thus, by taking a supremum over  $n$ , we have

$$\mathbb{E}(\phi(X)|\mathcal{G}) \geq \sup_{n \in \mathbb{N}} a_n \mathbb{E}(X|\mathcal{G}) + b_n = \phi(\mathbb{E}(X|\mathcal{G})),$$

almost surely.  $\square$

## 2 Martingales

**Definition 1.** Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space. A *filtration* is a monotone sequence of sub- $\sigma$ -algebras,

$$\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \dots \subseteq \mathcal{F}.$$

**Example 1.** If we have a sequence of random variables  $(X_n)_{n=1}^\infty$ , then the sequence  $\mathcal{F}_n = \sigma(X_1, X_2, \dots, X_n)$  is a filtration.

**Definition 2.** A sequence of random variables  $(X_n)_{n \geq 0}$  is *adapted* to a filtration  $(\mathcal{F}_n)_{n \geq 0}$  if for every  $n \geq 0$ ,  $X_n$  is  $\mathcal{F}_n$ -measurable.

**Example 2.** Given any sequence  $(X_n)_{n \geq 0}$ . The sequence  $(X_n)_{n \geq 0}$  is adapted to the filtration  $\mathcal{F}_n = \sigma(X_0, X_1, \dots, X_n)$ .

**Definition 3.** Let  $(\mathcal{F}_n)_{n \geq 0}$  be a filtration. A *martingale* is a sequence of random variables  $(X_n)_{n \geq 0}$  adapted to  $(\mathcal{F}_n)_{n \geq 0}$ , such that

1. For every  $n$ ,  $\mathbb{E}|X_n| < \infty$ .
2. For every  $n$ ,  $\mathbb{E}(X_{n+1}|\mathcal{F}_n) = X_n$ , almost surely.

**Example 3.** Let  $X_1, X_2, \dots$  be independent with  $\mathbb{E}[X_i] = 0$ . Let  $S_0 = 0$  and  $S_n = \sum_{i=1}^n X_i$  for  $n \geq 1$ . Then  $(S_n)_{n \geq 0}$  is a martingale with respect to the filtration  $\mathcal{F}_n = \sigma(X_i, 1 \leq i \leq n)$ . Note that

$$|S_n| \leq \sum_{i=1}^n |X_i|.$$

Thus,  $S_n$  is integrable for each  $n$ . The random variable  $S_n$  is also  $\mathcal{F}_n$  measurable. Furthermore,

$$\begin{aligned} \mathbb{E}(S_{n+1}|\mathcal{F}_n) &= \mathbb{E}(X_{n+1} + S_n|\mathcal{F}_n) \\ &= \mathbb{E}(X_{n+1}|\mathcal{F}_n) + \mathbb{E}(S_n|\mathcal{F}_n) \\ &= \mathbb{E}[X_{n+1}] + S_n \\ &= S_n. \end{aligned}$$

More generally if  $X_i$  is integrable with mean  $\mu_i$ , then  $S_n = \sum_{i=1}^n X_i - \mu_i$  is a martingale.

**Example 4.** Let  $X_1, \dots, X_n$  be independent with  $\mathbb{E}|X_i|^2 < \infty$ . Suppose that  $\mathbb{E}[X_i] = 0$  and  $\mathbb{E}[X_i^2] = \sigma^2$  for all  $i$ . Let,

$$Z_n = \left( \sum_{i=1}^n X_i \right)^2 - n\sigma^2 = S_n^2 - n\sigma^2,$$

and  $Z_0 = 0$ . Then  $Z_n$  is a martingale with respect to  $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$ .

Note first that  $\mathbb{E}[|Z_n|] \leq \mathbb{E}[S_n^2] + n = 2n$  and so each  $Z_n$  is integrable. Also,

$$\begin{aligned} \mathbb{E}(Z_{n+1}|\mathcal{F}_n) &= \mathbb{E}((S_n + X_{n+1})^2|\mathcal{F}_n) - (n+1)\sigma^2 \\ &= \mathbb{E}(S_n^2|\mathcal{F}_n) + 2\mathbb{E}(S_n X_{n+1}|\mathcal{F}_n) + \mathbb{E}(X_{n+1}^2|\mathcal{F}_n) - (n+1)\sigma^2 \\ &= S_n^2 + 2S_n \mathbb{E}(X_{n+1}|\mathcal{F}_n) + \mathbb{E}(X_{n+1}^2|\mathcal{F}_n) - (n+1)\sigma^2 \\ &= S_n^2 + 2S_n \mathbb{E}[X_{n+1}] + \mathbb{E}[X_{n+1}^2] - (n+1)\sigma^2 \\ &= S_n^2 - n\sigma^2. \end{aligned}$$

**Example 5.** Let  $X_1, X_2, \dots$  be i.i.d. random variables such that for some  $\theta \in \mathbb{R}$ ,  $m(\theta) = \mathbb{E}[e^{\theta X_i}] < \infty$ . Then

$$M_n = \frac{e^{\theta S_n}}{m(\theta)^n},$$

is a martingale with respect to  $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$ . Note that

$$\mathbb{E}[|M_n|] = \frac{1}{m(\theta)^n} \mathbb{E}[e^{\theta S_n}] = \frac{1}{m(\theta)^n} \prod_{i=1}^n \mathbb{E}[e^{\theta X_i}] = 1.$$

Furthermore,

$$\begin{aligned} \mathbb{E}(M_{n+1}|\mathcal{F}_{n+1}) &= \mathbb{E}\left(\frac{e^{\theta X_{n+1}}}{m(\theta)} M_n | \mathcal{F}_n\right) \\ &= M_n \mathbb{E}\left(\frac{e^{\theta X_{n+1}}}{m(\theta)} | \mathcal{F}_n\right) \\ &= M_n \mathbb{E}\left[\frac{e^{\theta X_{n+1}}}{m(\theta)}\right] \\ &= M_n. \end{aligned}$$

### 3 Stopping times

**Definition 4.** Let  $\{F_n\}$  be a filtration. A *stopping time* is a random variable  $T$  such that  $T$  takes values in  $\{0, 1, \dots\} \cup \{\infty\}$  and for every  $n \in \{0, 1, \dots\}$ , then event  $\{T = n\} \in \mathcal{F}_n$ .

For a filtration  $\{F_n\}$ , we can think of each  $\sigma$ -algebra  $\mathcal{F}_n$  as the information available at time  $n$ . The condition  $\mathcal{F}_n \subseteq \mathcal{F}_{n+1}$  means that we gain more information over a time. A stopping time is then a procedure where the decision to stop at time  $n$  depends only on the information available at time  $n$ .

**Example 6.** Let  $\{X_n\}_{n \geq 0}$  be a sequence of random variables adapted to a filtration  $\{\mathcal{F}_n\}_{n \geq 0}$ . Take any Borel set  $A \subseteq \mathbb{R}$ , let  $T = \inf\{n \geq 0 : X_n \in A\}$  where  $\inf \emptyset = \infty$ . Then  $T$  is a stopping time since

$$\{T = n\} = \{X_n \in A\} \cap \left( \bigcap_{i=1}^{n-1} \{X_i \notin A\} \right) \in \mathcal{F}_n.$$

**Example 7.** Let  $\{S_n\}_{n \geq 0}$  be a simply symmetric random walk (SSRW). That is,  $S_0 = 0$  and  $S_n = X_1 + \dots + X_n$  where  $X_i$  are i.i.d. and  $\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = -1) = \frac{1}{2}$ . Take any integers  $a < 0 < b$  and define  $T = \inf\{n : S_n = a \text{ or } S_n = b\}$ . By the previous example,  $T$  is a stopping time.

Related to stopping times, is the concept of a stopped  $\sigma$ -algebra.

**Definition 5.** Given a filtration  $\{\mathcal{F}_n\}_{n \geq 0}$  and a stopping time  $T$ . The *stopped  $\sigma$ -algebra* is defined to be

$$\mathcal{F}_T = \{A \in \mathcal{F} : A \cap \{T = n\} \in \mathcal{F}_n \text{ for all } n\}.$$

**Remark 1.** The stopped  $\sigma$ -algebra is indeed a  $\sigma$ -algebra. Since

1.  $\emptyset \in \mathcal{F}_T$  trivially.
2. If  $A \in \mathcal{F}_T$ , then  $A^c \cap \{T = n\} = \{T = n\} \setminus (A \cap \{T = n\})$  and  $\{T = n\} \in \mathcal{F}_n$  since  $T$  is a stopping time.
3. If  $A_i \in \mathcal{F}_T$ , then  $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}_T$  since  $(\bigcup_{i=1}^{\infty} A_i) \cap \{T = n\} = \bigcup_{i=1}^{\infty} A_i \cap \{T = n\}$ .

**Example 8.** Consider the SSRW from the previous example with  $T = \min\{n : S_n = a \text{ or } S_n = b\}$ . The event  $A = \{S_n \geq 0, \text{ for all } n \leq T\}$  is in  $\mathcal{F}_T$ .