Lecture 12: Martingales and Azuma's Inequality

Disclaimer

- This is a very informal treat of the concept of Martingales
- In particular, the inuitions are specific to discrete spaces
- Interested readers are referred to study σ-algebras for a more formal treatment of this material

In this Lecture

- Martingales
- Specific to Discrete Sample Spaces
- Specifically, Doob's Martingale
- Azuma's Inequality

σ-Field

- Let Ω be a (discrete) sample space with probability distribution p

Definition (σ-Field)

A σ-filed $\mathcal{F}$ on Ω is a collection of subsets of Ω such that the following constraints are satisfied

1. $\mathcal{F}$ contains $\emptyset$ and Ω, and
2. $\mathcal{F}$ is closed under unions, intersections, and complementation.

Example

- For example $\mathcal{F}_0 = \{\emptyset, \Omega\}$ is a σ-field
- Suppose $\Omega = \{0,1\}^n$
- Let $\mathcal{F}_1 = \mathcal{F}_0 \cup \{\ 0\{0,1\}^{n-1},\ 1\{0,1\}^{n-1}\ \}$. Note that $\mathcal{F}_1$ is also a σ-field
- Let $\mathcal{F}_2 = \left\{ S\{0,1\}^{n-2} \colon S \subseteq \{00, 01, 10, 11\}\ \right\}$. We use the convention that if $S = \emptyset$ then $S\{0,1\}^{n-2} = \emptyset$. So, $\mathcal{F}_2$ has 16 elements, and $\mathcal{F}_1 \subseteq \mathcal{F}_2$. It is easy to verify that $\mathcal{F}_2$ is a σ-field
- In general

$$\mathcal{F}_k = \left\{\ S\{0,1\}^{n-k} \colon S \subseteq \{\omega_1, \dots, \omega_k \colon \omega_i \in \{0,1\},\ \text{for all } i \in \{1, \dots, k\}\}\right\}$$

Smalled Set Containing and Element

- Let $x \in \Omega$
- Consider a σ-field $\mathcal{F}$ on Ω
- The smallest set in $\mathcal{F}$ containing x is the intersection of all sets in $\mathcal{F}$ that contain x. Formally, it is the following set

$$\mathcal{F}(x) := \bigcap_{\substack{S \in \mathcal{F} \\ x \in S}} S$$

- For example, let $n = 5$, $x = 01001$, and consider the σ-field $\mathcal{F}_2$ on Ω. In this case, the smallest set $\mathcal{F}_2(x)$ in $\mathcal{F}_2$ that contains x is $01\{0,1\}^{n-2}$.

$\mathcal{F}$-Measurable

- Let $f : \Omega \to \mathbb{R}$ be a function

Definition ($\mathcal{F}$-Measurable)

The function f is $\mathcal{F}$-measurable if, for all $y \in \mathcal{F}(x)$, we have $f(x) = f(y)$, where $\mathcal{F}(x)$ is the smallest subset in $\mathcal{F}$ containing x

- For example, let $n = 5$ and consider the σ-field $\mathcal{F}_2$ on Ω
- As we has seen, we have $\mathcal{F}_2(x) = x_1 x_2 \{0, 1\}^{n-2}$, where x_1 and x_2 are, respectively, the first and the second bits of x
- Let $f(x)$ be the total number of 1s in the first two coordinates of x. This function f is $\mathcal{F}_2$-measurable
- Let $f(x)$ be the expected value of 1s over all strings whose first two bits are $x_1 x_2$. This function f is also $\mathcal{F}_2$-measurable
- Let $f(x)$ be the total number of 1s in the first three bits of x. This function is <u>not</u> $\mathcal{F}_2$-measurable

Conditional Expectation

- Let p be a probability distribution over the sample space Ω
- Let $\mathcal{F}$ be a σ-field on Ω
- Let $f : \Omega \to \mathbb{R}$ be a function
- We define the conditional expectation as a function $\mathbb{E}\left[f|\mathcal{F}\right] : \Omega \to \mathbb{R}$ defined as follows

$$\mathbb{E}\left[f|\mathcal{F}\right](x) := \frac{1}{\sum_{y\in\mathcal{F}(x)} p(y)} \sum_{y\in\mathcal{F}(x)} f(y)p(y)$$

- We emphasize that f need not be $\mathcal{F}$-measurable to define the expectation in this manner!
- Note that $\mathbb{E}\left[f|\mathcal{F}\right](x) = \mathbb{E}\left[f|\mathcal{F}\right](y)$, for all $y \in \mathcal{F}(x)$

Filtration

- Let Ω be a sample space with probability distribution p

Definition (Filtration)

A sequence of σ-fields $\mathcal{F}_0, \mathcal{F}_1, \ldots, \mathcal{F}_n$ is a filtration if

$$\{\emptyset, \Omega\} = \mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_n$$

Beginning of "Intuition Slides"

Sample Space

- As time processes, new information about the sample is revealed to us
- At time 1, we learn the value of ω_1 of the random variable $\mathbb{X}_1$
- At time 2, we learn the value of ω_2 of the random variable $\mathbb{X}_2$
- And so on. At time t, we learn the value ω_t of the random variable $\mathbb{X}_t$
- By the end of time n, we know the value ω_n of the last random variable $\mathbb{X}_n$
- At this point, $f(\mathbb{X}_1, \ldots, \mathbb{X}_n)$ can be calculated, where f is a function that we are interested in

Examples

- Balls and Bins. At time i we find out the bin ω_i that the balls i goes into
- Coin tosses. At time i we find out the outsome ω_i of the i-th coin toss
- Hypergeometric Series. At time i we find out the color ω_i of the i-th ball drawn from the jar (where sampling is being carries out without replacement)
- Bounded Difference Function. At time i we find out the outcome ω_i of the i-th variable of the input of the function f

Filtration

- In a filtration, the σ-field $\mathcal{F}_k$ represents the knowledge we have after knowing the outcomes $(\omega_1, \ldots, \omega_k)$
- For instance, the σ-field $\mathcal{F}_0$ represents "we know nothing about the sample"
- For instance, the σ-field $\mathcal{F}_n$ represents "we know everything about the sample"

Tree Representation

- Think of a rooted tree
- For every internal node, the outgoing edges represent the various possible outcomes in the next time step
- Leaves represent that the entire sample is already known
- The sequence of outcomes $(\omega_1, \ldots, \omega_n)$ represents a "root-to-leaf" path
- Consider a filtration $\{\emptyset, \Omega\} = \mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_n$. The set $\mathcal{F}_k(x)$ corresponding to this root-to-leaf path is the depth-k node on this path

Measurable with respect to a σ-field

- Consider a filtration $\{\emptyset, \Omega\} = \mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_n$
- A random variable $\mathbb{F}_k = f(\mathbb{X}_1, \ldots, \mathbb{X}_n)$ will be measurable with respect to the σ-field $\mathcal{F}_k$ if the value of $f(\mathbb{X}_1, \ldots, \mathbb{X}_n)$ depends only on $(\omega_1, \ldots, \omega_k)$

End of "Intuition Slides"

Martingales

Definition (Martingale Sequence)

Let $\{\emptyset, \Omega\} = \mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_n$ be a filtration. The sequence $(\mathbb{F}_1, \ldots, \mathbb{F}_n)$ forms a martingale with respect to this filtration if

1. $\mathbb{F}_i$ is $\mathcal{F}_i$-measurable, for $1 \leqslant i \leqslant n$, and
2. $\mathbb{E}\left[\mathbb{F}_{t+1} | \mathcal{F}_t\right] = (\mathbb{F}_t | \mathcal{F}_t)$, for $0 \leqslant t < n$.

- Note that given $\mathcal{F}_t = (\omega_1, \ldots, \omega_t)$, the value of $\mathbb{F}_t$ is fixed. So, we can write $\mathbb{E}\left[\mathbb{F}_t | \mathcal{F}_t\right](x)$ in short as $(\mathbb{F}_t | \mathcal{F}_t)(x)$
- Note that given $\mathcal{F}_t = (\omega_1, \ldots, \omega_t)$, the outcome of $\mathbb{F}_{t+1}$ is not yet fixed and is (possibly) random
- The second equation in the definition is an "equality of two functions." It means that $\mathbb{E}\left[\mathbb{F}_{t+1} | \mathcal{F}_t\right](x)$ is equal to $(\mathbb{F}_t | \mathcal{F}_t)(x)$ for all $x \in \Omega$.

Example

- Consider tossing a coin that gives head with probability p, and tails with probability $(1 - p)$, independently n times
- $\mathcal{F}_t$ is the outcome of the first t coin tosses
- Let $\mathbb{S}_t$ represent the number of heads in the first t coin tosses
- Note that $\mathbb{S}_t(x)$ is fixed given $\mathcal{F}_t(x)$, where $x \in \Omega$
- Note that $(\mathbb{S}_{t+1}|\mathcal{F}_t)(y) = (\mathbb{S}_t|\mathcal{F}_t)(y) + 1$ with probability p (for a random y that is consistent with $\mathcal{F}_t(x)$), else $(\mathbb{S}_{t+1}|\mathcal{F}_t)(y) = (\mathbb{S}_t|\mathcal{F}_t)(y)$
- Therefore, $\mathbb{E}\left[\mathbb{S}_{t+1}|\mathcal{F}_t\right](x) = (\mathbb{S}_t|\mathcal{F}_t)(x) + p$
- So, $(\mathbb{S}_1, \dots, \mathbb{S}_n)$ is not a martingale sequence with respect to the filtration $\{\emptyset, \Omega\} = \mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_n$

Example

- Let f be a function and we consider a filtration $\{\emptyset, \Omega\} = \mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_n$
- Let $\mathbb{F}_t$ be the following random varible

$$\mathbb{F}_t(x) = \mathbb{E}\left[f(\omega_1, \ldots, \omega_t, \mathbb{X}_{t+1}, \ldots, \mathbb{X}_n)\right],$$

  where $\omega_1, \ldots, \omega_t$ are the first t outcomes of $x \in \Omega$
- First, prove that $\mathbb{F}_t$ is $\mathcal{F}_t$ measurable
- Next, prove that $(\mathbb{F}_0, \ldots, \mathbb{F}_n)$ is a martingale with respect to the filtration $\{\emptyset, \Omega\} = \mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_n$

Martingale Difference Sequence

- Let $\{\emptyset, \Omega\} = \mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_n$ be a filtration
- Let $(\mathbb{F}_0, \ldots, \mathbb{F}_n)$ be a martingale difference sequence with respect to the filtration above
- Let $\mathbb{Y}_0 = \mathbb{F}_0$, and $\mathbb{Y}_{t+1} = \mathbb{F}_{t+1} - \mathbb{F}_t$, for $0 \leqslant t < n$
- Intuition: $\mathbb{Y}_{t+1}$ measures the increase in $\mathbb{Y}_{t+1}$ from $\mathbb{Y}_t$
- Note that $\mathbb{E}\left[\mathbb{Y}_{t+1} | \mathcal{F}_t\right] = 0$

Definition (Azuma's Inequality)

Suppose $(\mathbb{Y}_0, \ldots, \mathbb{Y}_n)$ be a martingale difference sequence with respect to the filtration $\{\emptyset, \Omega\} = \mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_n$. Suppose $a_{t+1} \leqslant (\mathbb{Y}_{t+1} | \mathcal{F}_t)(x) \leqslant b_{t+1}$, for $0 \leqslant t < n$. Then

$$\mathbb{P}\left[\sum_{i=1}^{n} \mathbb{Y}_t \geqslant t\right] \leqslant \exp\left(-\frac{2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right)$$

Proof Outline

- We are interested in computing

$$\mathbb{E}\left[\exp\left(h\sum_{i=1}^{n}\mathbb{Y}_i\right)\right] = \mathbb{E}\left[\exp\left(h\sum_{i=1}^{n-1}\mathbb{Y}_i\right)\exp(h\mathbb{Y}_n)\right]$$
$$\leqslant EX\exp\left(h\sum_{i=1}^{n-1}\mathbb{Y}_i\right)\exp(p_n\mathrm{e}^{a_n}+q_n\mathrm{e}^{b_n}),$$

  where $p_n + q_n = 1$ and $p_n a_n + q_n a_n = 0$.

- Inductively, we get

$$\mathbb{E}\left[\exp\left(h\sum_{i=1}^{n}\mathbb{Y}_i\right)\right] \leqslant \prod_{i=1}^{n}(p_i\mathrm{e}^{a_i}+q_i\mathrm{e}^{b_i})$$

- Rest of the proof is identical to the Hoeffding's Bound proof

Differences from Hoeffding's Bound Proof

- The distribution $\mathbb{Y}_{t+1}$ can depend on the outcomes $(\omega_1, \ldots, \omega_t)$
- But the only restrictions are that $\mathbb{E}\left[\mathbb{Y}_{t+1}|\mathcal{F}_t\right] = 0$ and the outcomes of $(\mathbb{Y}_{t+1}|\mathcal{F}_t)(x)$ are in the range $[a_{t+1}, b_{t+1}]$
- Prove: The Bounded difference inequality using Azuma's Inequality
- Prove: The concentration of the Hypergeometric distribution using Azuma's Inequality