Lecture 09: Martingales and Azuma's Inequality

Disclaimer

- This is a very informal treatment of the concept of Martingales
- In particular, the intuitions are specific to discrete spaces
- Inquisitive 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 sample space with probability distribution p

Definition (σ-Field)

A σ-field $\mathcal{F}$ on Ω is a collection of subsets of Ω that

1. Contains $\emptyset$ and Ω,
2. 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 $\mathcal{F}_1 = \mathcal{F}_0 \cup \{\ 0\{0,1\}^{n-1},\ 1\{0,1\}^{n-1}\}$. This is also a σ-field
- Let $\mathcal{F}_2 = \{\ S\{0,1\}^{n-2} \colon S \subseteq \{00, 01, 10, 11\}\ \}$. We use the convention: 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, $\mathcal{F}_i = \{\ S\{0,1\}^{n-k} \colon S \subseteq \{\omega_1 \dots \omega_k \colon \omega_i \in \{0,1\}, \text{ for all } i \in \{1, \dots, k\}\}\ \}$.

Smallest Set Containing an 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 $\mathcal{F}_2$ containing 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 had 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 second bit of x
- Let $f(x)$ be the total number of 1s in the first two coordinates of x. This function 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 is $\mathcal{F}_2$-measurable
- Let $f(x)$ be the total number of 1s in the first three coordinates of x. This function is *not* $\mathcal{F}_2$-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 $y \in \mathcal{F}(x)$

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

Intuition Slides

Sample Space

1. As time progresses, new information is revealed to us
2. At time 1, we learn the value ω_1 of the random variable $\mathbb{X}_1$
3. At time 2, we learn the value ω_2 of the random variable $\mathbb{X}_2$
4. And so on. At time t, we learn the value ω_t of the random variable $\mathbb{X}_t$
5. By the end of time n, we know the value ω_n of the last random variable $\mathbb{X}_n$
6. At this point of time, $f(\mathbb{X}_1, \ldots, \mathbb{X}_n)$ 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 ball i goes into.
- Coin tosses. At time i we find out the outcome 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 draw from the jar (where sampling is being carried out without replacement).
- Bounded Difference Function. At time i we find out the outcome ω_i of the i-th variable of the function f.

Filtration

- The filtration $\mathcal{F}_k$ is the tuple of outcomes $(\omega_1, \ldots, \omega_k)$
- The filtration $\mathcal{F}_0$ represents no outcome is known
- The filtration $\mathcal{F}_n$ represents that all outcomes are known

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 all outcomes are already known
- The sequence of outcomes $(\omega_1, \ldots, \omega_n)$ represents a root to leaf path
- The filtration $\mathcal{F}_k$ corresponding to this path is the depth-k node on this path

Measurable with respect to a Filtration

- A random variable $\mathbb{X}_k$ will be measurable with respect to a filtration $\mathcal{F}_k$ if the value of $\mathbb{X}_k$ depends only on $(\omega_1, \ldots, \omega_k)$

Definition (Martingale Sequence)

Let $\{\emptyset, \Omega\} = \mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_n$ be a filtration. The sequence $(\mathbb{X}_1, \ldots, \mathbb{X}_n)$ forms a martingale with respect to this filtration if $\mathbb{X}_i$ is $\mathcal{F}_i$-measurable, for $1 \leqslant i \leqslant n$, and

$$\mathbb{E}\left[\mathbb{X}_{t+1} | \mathcal{F}_t\right] = (\mathbb{X}_t | \mathcal{F}_t),$$

for $0 \leqslant t < n$.

- Note that given $\mathcal{F}_t = (\omega_1, \ldots, \omega_t)$, the value of $\mathbb{X}_t$ is fixed
- Note that given $\mathcal{F}_t = (\omega_1, \ldots, \omega_t)$, the outcome of $\mathbb{X}_{t+1}$ is not yet fixed and is a random

Example

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

Example

- Consider the random variable $\mathbb{T}_t = \mathbb{S}_t - tp$
- Note that $\mathbb{T}_t$ is fixed given $\mathcal{F}_t$
- Note that $(\mathbb{T}_{t+1}|\mathcal{F}_t) = (\mathbb{S}_t|\mathcal{F}_t) + 1 - (t+1)p$ with probability p, and $(\mathbb{T}_{t+1}|\mathcal{F}_t) = (\mathbb{S}_t|\mathcal{F}_t) - (t+1)p$ with probability $(1-p)$
- Therefore, $\mathbb{E}\left[\mathbb{T}_{t+1}|\mathcal{F}_t\right] = (\mathbb{S}_t|\mathcal{F}_t) + p - (t+1)p = (\mathbb{T}_t|\mathcal{F}_t)$
- So, the sequence $(\mathbb{T}_0, \mathbb{T}_1, \dots, \mathbb{T}_n)$ is a martingale sequence with respect to the filtration $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_n$

Example

- Let f be a function and $\mathcal{F}_t$ is the filtration where the first t arguments to f have been fixed
- Let $\mathbb{F}_t$ be the random variable

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

- Prove: $(\mathbb{F}_0, \ldots, \mathbb{F}_n)$ is a martingale with respect to the filtration $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_n$
- This martingale is called: Doob's Martingale

Martingale Difference Sequence

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

Theorem (Azuma's Inequality)

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

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

Proof Outline

- We only show how to bound

$$\mathbb{E}\left[\exp\left(h\sum_{i=1}^{n}\mathbb{Y}_i\right)\right]$$

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

- We are interested in computing

$$\mathbb{E}\left[\exp\left(h\sum_{i=1}^{n}\mathbb{Y}_i\right)\right] = \mathbb{E}\left[\exp\left(h\sum_{i=1}^{n-1}\mathbb{Y}_i\right)\exp\left(h\mathbb{Y}_n\right)\right]$$

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

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

- Inductively,

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

Proof Outline III

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

Difference from Hoeffding's Bound

- 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 $(\mathbb{Y}_{t+1} | \mathcal{F}_t)$ 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