Lecture 12: Martingales and Azuma’s Inequality
Disclaimer

- This is a very informal treat of the concept of Martingales
- In particular, the intuitions are specific to discrete spaces
- Interested readers are referred to study $\sigma$-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
Let $\Omega$ be a (discrete) sample space with probability distribution $p$

**Definition ($\sigma$-Field)**

A $\sigma$-field $\mathcal{F}$ on $\Omega$ is a collection of subsets of $\Omega$ such that the following constraints are satisfied:

1. $\mathcal{F}$ contains $\emptyset$ and $\Omega$, and
2. $\mathcal{F}$ is closed under unions, intersections, and complementation.
For example $F_0 = \{\emptyset, \Omega\}$ is a $\sigma$-field.

Suppose $\Omega = \{0, 1\}^n$.

Let $F_1 = F_0 \cup \{0\{0, 1\}^{n-1}, 1\{0, 1\}^{n-1}\}$. Note that $F_1$ is also a $\sigma$-field.

Let $F_2 = \left\{ S\{0, 1\}^{n-2} : S \subseteq \{00, 01, 10, 11\} \right\}$. We use the convention that if $S = \emptyset$ then $S\{0, 1\}^{n-2} = \emptyset$. So, $F_2$ has 16 elements, and $F_1 \subseteq F_2$. It is easy to verify that $F_2$ is a $\sigma$-field.

In general

$F_k = \left\{ S\{0, 1\}^{n-k} : S \subseteq \{\omega_1, \ldots, \omega_k : \omega_i \in \{0, 1\}, \text{ for all } i \in \{1, \ldots, k\}\} \right\}$
Let $x \in \Omega$

Consider a $\sigma$-field $\mathcal{F}$ on $\Omega$

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

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

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

Concentration
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 $\sigma$-field $\mathcal{F}_2$ on $\Omega$.

As we has seen, we have $\mathcal{F}_2(x) = x_1x_2\{0, 1\}^{n-2}$, 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 $x_1x_2$. 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 **not** $\mathcal{F}_2$-measurable.
Let $p$ be a probability distribution over the sample space $\Omega$
Let $F$ be a $\sigma$-field on $\Omega$
Let $f : \Omega \to \mathbb{R}$ be a function
We define the conditional expectation as a function $E[f|F] : \Omega \to \mathbb{R}$ defined as follows
\[
E[f|F](x) := \frac{1}{\sum_{y \in F(x)} p(y)} \sum_{y \in F(x)} f(y)p(y)
\]
We emphasize that $f$ need not be $F$-measurable to define the expectation in this manner!
Note that $E[f|F](x) = E[f|F](y)$, for all $y \in F(x)$
Filtration

Let $\Omega$ be a sample space with probability distribution $p$.

**Definition (Filtration)**

A sequence of $\sigma$-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”
As time processes, new information about the sample is revealed to us

- At time 1, we learn the value of $\omega_1$ of the random variable $X_1$
- At time 2, we learn the value of $\omega_2$ of the random variable $X_2$
- And so on. At time $t$, we learn the value $\omega_t$ of the random variable $X_t$
- By the end of time $n$, we know the value $\omega_n$ of the last random variable $X_n$
- At this point, $f(X_1, \ldots, X_n)$ 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 $\omega_i$ that the ball $i$ goes into.
- Coin tosses. At time $i$ we find out the outcome $\omega_i$ of the $i$-th coin toss.
- Hypergeometric Series. At time $i$ we find out the color $\omega_i$ of the $i$-th ball drawn from the jar (where sampling is being carried out without replacement).
- Bounded Difference Function. At time $i$ we find out the outcome $\omega_i$ of the $i$-th variable of the input of the function $f$. 

Concentration
In a filtration, the $\sigma$-field $\mathcal{F}_k$ represents the knowledge we have after knowing the outcomes $(\omega_1, \ldots, \omega_k)$.

For instance, the $\sigma$-field $\mathcal{F}_0$ represents “we know nothing about the sample”.

For instance, the $\sigma$-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
Consider a filtration \( \{\emptyset, \Omega\} = \mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_n \)

A random variable \( \mathcal{F}_k = f(X_1, \ldots, X_n) \) will be measurable with respect to the \( \sigma \)-field \( \mathcal{F}_k \) if the value of \( f(X_1, \ldots, X_n) \) depends only on \((\omega_1, \ldots, \omega_k)\)
End of “Intuition Slides”
Definition (Martingale Sequence)

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

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

Note that given \( \mathcal{F}_t = (\omega_1, \ldots, \omega_t) \), the value of \( \mathcal{F}_t \) is fixed. So, we can write \( \mathbb{E} [\mathcal{F}_t|\mathcal{F}_t] (x) \) in short as \( (\mathcal{F}_t|\mathcal{F}_t)(x) \).

Note that given \( \mathcal{F}_t = (\omega_1, \ldots, \omega_t) \), the outcome of \( \mathcal{F}_{t+1} \) is not yet fixed and is (possibly) random.

The second equation in the definition is an "equality of two functions." It means that \( \mathbb{E} [\mathcal{F}_{t+1}|\mathcal{F}_t] (x) \) is equal to \( (\mathcal{F}_t|\mathcal{F}_t)(x) \) for all \( x \in \Omega \).
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 $S_t$ represent the number of heads in the first $t$ coin tosses.

Note that $S_t(x)$ is fixed given $\mathcal{F}_t(x)$, where $x \in \Omega$.

Note that $(S_{t+1}|\mathcal{F}_t)(y) = (S_t|\mathcal{F}_t)(y) + 1$ with probability $p$ (for a random $y$ that is consistent with $\mathcal{F}_t(x)$), else $(S_{t+1}|\mathcal{F}_t)(y) = (S_t|\mathcal{F}_t)(y)$.

Therefore, $\mathbb{E}[S_{t+1}|\mathcal{F}_t](x) = (S_t|\mathcal{F}_t)(x) + p$.

So, $(S_1, \ldots, S_n)$ is not a martingale sequence with respect to the filtration $\{\emptyset, \Omega\} = \mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_n$. 

Concentration
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 variable 
\[
\mathbb{F}_t(x) = \mathbb{E}\left[ f(\omega_1, \ldots, \omega_t, X_{t+1}, \ldots, X_n) \right],
\]
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 \((\mathcal{F}_0, \ldots, \mathcal{F}_n)\) be a martingale difference sequence with respect to the filtration above.
- Let \( Y_0 = \mathcal{F}_0 \), and \( Y_{t+1} = \mathcal{F}_{t+1} - \mathcal{F}_t \), for \( 0 \leq t < n \).
- Intuition: \( Y_{t+1} \) measures the increase in \( Y_{t+1} \) from \( Y_t \).
- Note that \( \mathbb{E} [Y_{t+1} | \mathcal{F}_t] = 0 \).
Azuma’s Inequality

Definition (Azuma’s Inequality)

Suppose \((Y_0, \ldots, Y_n)\) 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 \(a_{t+1} \leq (Y_{t+1} | \mathcal{F}_t)(x) \leq b_{t+1}\), for \(0 \leq t < n\). Then

\[
\mathbb{P} \left[ \sum_{i=1}^{n} Y_t \geq t \right] \leq \exp \left( - \frac{2t^2}{\sum_{i=1}^{n} (b_i - a_i)^2} \right)
\]
Proof Outline

- We are interested in computing
  \[
  E \left[ \exp \left( h \sum_{i=1}^{n} Y_i \right) \right] = E \left[ \exp \left( h \sum_{i=1}^{n-1} Y_i \right) \exp(hY_n) \right] \leq EX \exp \left( h \sum_{i=1}^{n-1} Y_i \right) \exp(p_n e^{a_n} + q_n e^{b_n}),
  \]

  where \( p_n + q_n = 1 \) and \( p_n a_n + q_n a_n = 0 \).
- Inductively, we get
  \[
  E \left[ \exp \left( h \sum_{i=1}^{n} Y_i \right) \right] \leq \prod_{i=1}^{n} (p_i e^{a_i} + q_i e^{b_i})
  \]
- Rest of the proof is identical to the Hoeffding’s Bound proof
The distribution $Y_{t+1}$ can depend on the outcomes $(ω_1, . . . , ω_t)$

But the only restrictions are that $E[ Y_{t+1} | F_t ] = 0$ and the outcomes of $(Y_{t+1} | F_t)(x)$ 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