ECE 250: Stochastic Processes: Week #6&7

Instructor: Behrouz Touri

Outline:

- Conditional Expectation
- Martingales and S-Martingales
- Martingale Convergence Theorem
- Applications

Martingale Theory

- Motivation:
  1. Stochastic Gradient Descent, why should it work?
  2. Polarization dynamics?
  3. Polya Urn (type) model/dynamics
  4. ...

Conditional Expectation: the undergrad way

- Similar to a random variable X, the joint CDF of random variables $\mathbf{X} = (X_1, \ldots, X_n)$ is defined by:

$$F_{\mathbf{X}}(x) = \Pr(X_1 \leq x_1 \cap X_2 \leq x_2 \cap \cdots \cap X_n \leq x_n),$$

  where $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ is a given vector.

- We say that the random vector $\mathbf{X}$ is (or random variables $X_1, \ldots, X_n$ are) jointly continuously distributed (or have joint probability density function) if there exists a function $f_X : \mathbb{R}^n \to \mathbb{R}^+$ such that

$$F_{\mathbf{X}}(x) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_X(z_1, \ldots, z_n) dz_n \ldots dz_1.$$

- Examples:

  – i.i.d. Gaussian random variables $X_1, \ldots, X_n$ where

$$f_X(x_1, \ldots, x_n) = \frac{1}{\sqrt{(2\pi\sigma^2)^n}} e^{-\sum_{i=1}^n \frac{(x_i-\mu)^2}{2\sigma^2}}.$$

  – More generally a Gaussian random vector $\mathbf{X}$ with mean vector μ and a covariance matrix C is a random vector with the joint PDF

$$f_X(x_1, \ldots, x_n) = \frac{1}{\sqrt{(2\pi)^n \det(C)}} e^{-\frac{1}{2}(x-\mu)^T C^{-1}(x-\mu)}.$$

Conditional Expectation: the undergrad way cont.

- For random variables X, Y that are jointly continuous, we define:

$$\mathbb{E}[X \mid Y = y] = \int_{-\infty}^{\infty} x \frac{f_{X,Y}(x, y)}{f_y(y)} dx,$$

when $f_Y(y) \neq 0$ and either $\int_0^\infty x \frac{f_{X,Y}(x,y)}{f_y(y)} dx < \infty$ or $\int_{-\infty}^0 x \frac{f_{X,Y}(x,y)}{f_y(y)} dx > -\infty$.

- In addition to being hard to compute (except for very structured random variables) as it requires joint distribution, non-zero y, etc.

- Conditional expectation of jointly discrete random variables suffer from similar problems.

Conditional Expectation: some definitions

- Some σ-algebra definitions:
  - Let $(\Omega, \mathcal{F}_o, \Pr(\cdot))$ be the **underlying (main) probability space**.
  - We say that $\mathcal{F} \subseteq \mathcal{F}_o$ is a sub-σ algebra if it is a subset of $\mathcal{F}$ that is a σ-algebra itself.
  - We say that a sequence of sub-σ algebras $\{\mathcal{F}_k\}$ (of $\mathcal{F}_o$) is a *filtration* (of $\mathcal{F}_o$) if $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \mathcal{F}_3 \subseteq \cdots \subseteq \mathcal{F}_o$.
- Some random variable definitions:
  - We say that a random variable is measurable with respect to $\mathcal{F} \subseteq \mathcal{F}_o$ if $X^{-1}((-\infty, a]) \in \mathcal{F}$ for all a (or equivalently, $X^{-1}(B) \in \mathcal{F}$ for all $B \in \mathcal{B}$).
  - For a random variable X, we define $\sigma(X)$ to be the smallest σ-algebra that contains all $X^{-1}((-\infty, a])$ for all a. We refer to $\sigma(X)$ as the σ-algebra generated by X.
  - Similarly, for random variables $X_1, \ldots, X_k$, we define $\sigma(X_1, \ldots, X_k)$ to be the smallest σ-algebra containing all $X_i^{-1}((-\infty, a])$ for all $a \in \mathbb{R}$ and all $1 \leq i \leq k$.
  - Recall: For a random variable X, we say that it is measurable with respect to $\mathcal{F}$ if $X^{-1}((-\infty, a]) \in \mathcal{F}$ for all $a \in \mathbb{R}$.
- For a random process $\{X_k\}$ if we define:

$$\mathcal{F}_k = \sigma(X_1, \ldots, X_k),$$

  then $\{\mathcal{F}_k\}$ is a filtration. This is called the natural filtration for the random process ($\{X_k\}$).
- We say that a random process $\{X_k\}$ is adapted to filtration $\{\mathcal{F}_k\}$ if X_k is measurable with respect to $\mathcal{F}_k$.
- Example: Every random process is adapted to its natural filtration.

Conditional Expectation

- Definition: For a random variable X and a (sub) σ-algebra $\mathcal{F}$, we say that Z is the conditional expectation of X given $\mathcal{F}$ and denote it by $Z = \mathbb{E}[X \mid \mathcal{F}]$ if:

  1. Z is measurable with respect to $\mathcal{F}$, and
  2. $\mathbb{E}[X\mathbf{1}_A] = \mathbb{E}[Z\mathbf{1}_A]$ (or equivalently $\mathbb{E}[(X - Z)\mathbf{1}_A] = 0$) for all $A \in \mathcal{F}$.

  Important: Almost always we use the above definition with $\mathcal{F} = \sigma(X_1, \ldots, X_k)$ for some random variables $X_1, \ldots, X_k$.

- Very important fact: A random variable Z is measurable with respect to $\mathcal{F} = \sigma(X_1, \ldots, X_n)$ if and only if $Z = h(X_1, \ldots, X_n)$ for some deterministic measurable function $h(\cdot)$.

- In other words, for a random variable X and a sequence $X_1, \ldots, X_k$, $\mathbb{E}[X \mid X_1, \ldots, X_k]$ is some deterministic function h of $X_1, \ldots, X_k$ such that $\mathbb{E}[(X - h(X_1, \ldots, X_k))\mathbf{1}_A] = 0$ for all $A \in \sigma(X_1, \ldots, X_k)$.

- Implication: if $X, X_1, \ldots, X_k$ are independent, $\mathbb{E}[X \mid X_1, \ldots, X_k] = \mathbb{E}[X]$ (here $\mathbb{E}[X]$ should be viewed as $\mathbb{E}[X]\mathbf{1}_\Omega$).

- Fact: For any r.v. X with $\mathbb{E}[|X|] < \infty$ and any σ-algebra $\mathcal{F} \subseteq \mathcal{F}_o$, $\mathbb{E}[X \mid \mathcal{F}]$ exists and it is almost surely unique.

- For random variables X, Y, we define $\mathbb{E}[Y \mid X] := \mathbb{E}[Y \mid \sigma(X)]$.

Conditional Expectation: Example

Example: Let $\Omega = \{1, \ldots, 6\}$ and $\mathcal{F}_o = \mathcal{P}(\Omega)$ and $X = \mathbf{1}_{\{1,3,5\}}$ and Y, be the random variable with $Y(i) = i$. Assume that $\mathbf{Pr}(\{i\}) = \frac{1}{6}$ and $\mathbf{Pr}(A) = \sum_{i \in A} \mathbf{Pr}(\{i\})$ for any $A \subseteq \Omega$.

1. What is $\sigma(X)$?
2. What is $\mathbb{E}[Y \mid X]$?

Conditional Expectation: Properties and some Examples

- *Measurable observation*: $\mathbb{E}[X \mid \mathcal{F}] = X$ if X is measurable w.r.t $\mathcal{F}$.
- *Linearity*: $\mathbb{E}[\alpha X + \beta Y \mid \mathcal{F}] = \alpha\mathbb{E}[X \mid \mathcal{F}] + \beta\mathbb{E}[Y \mid \mathcal{F}]$.
- *Product Rule*: $\mathbb{E}[XY \mid \mathcal{F}] = \mathbb{E}[X \mid \mathcal{F}]Y$ for any Y that is measurable with respect to $\mathcal{F}$.
- *Independence*: If X is independent of $\mathcal{F}$, then $\mathbb{E}[X \mid \mathcal{F}] = \mathbb{E}[X]$.
- When definable, the undergrad definition of conditional expectation coincides with this definition.
- Tower rule (or law of total probability):

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

  for any σ-algebras $\mathcal{F}_1 \subseteq \mathcal{F}_2$.
- *Monotone Convergence Theorem*: If $X_1 \leq X_2 \leq \ldots$ and $X_k \to X$ almost surely, $\mathbb{E}[X_k \mid \mathcal{F}] \to \mathbb{E}[X \mid \mathcal{F}]$.
- Simple random walk: Let X_k be i.i.d. random variables with zero mean. Let $\mathcal{F}_k = \sigma(X_1, \ldots, X_k)$. Then, we have:

$$\begin{aligned}
\mathbb{E}[S_{k+1} \mid X_1, \ldots, X_k] &= \mathbb{E}[S_{k+1} \mid \mathcal{F}_k] \\
&= \mathbb{E}[X_{k+1} + S_k \mid \mathcal{F}_k] \\
(\text{linearity}\to) &= \mathbb{E}[X_{k+1} \mid \mathcal{F}_k] + \mathbb{E}[S_k \mid \mathcal{F}_k] \\
(\text{measurability and independence}\to) &= S_k.
\end{aligned}$$

Martingale: Definition

- Definition: We say that a random process $\{X_k\}$ adapted to a filtration $\{\mathcal{F}_k\}$ is

  a. a *martingale* if

  $$\mathbb{E}[X_{k+1} \mid \mathcal{F}_k] = X_k,$$

  for all $k \geq 0$,

  b. a *super-martingale* if

  $$\mathbb{E}[X_{k+1} \mid \mathcal{F}_k] \leq X_k,$$

  for all $k \geq 0$, and

  c. a *sub-martingale* if

  $$\mathbb{E}[X_{k+1} \mid \mathcal{F}_k] \geq X_k,$$

  for all $k \geq 0$.

- Note that if $\{X_k\}$ is a sub-martingale, $\{-X_k\}$ is a super-martingale.

- If the filtration is not given, the filtration is assumed to be the natural filtration of X_k. In other words, we simply say that $\{X_k\}$ is a martingale if

  $$\mathbb{E}[X_{k+1} \mid X_1, \ldots, X_k] = \mathbb{E}[X_{k+1} \mid \sigma(X_1, \ldots, X_k)] = X_k,$$

  for all $k \geq 1$.

Martingale: Examples

- **Simple symmetric random walk**: $S_n = \sum_{k=1}^{n} X_k$ where $\{X_k\}$ are i.i.d. random variables with zero mean.
- **Polar code dynamics**:
  - let Z_k be an i.i.d. binary random process with $\mathbf{Pr}(Z_k = 0) = \frac{1}{2} = \mathbf{Pr}(Z_k = 1)$. Let $X_1 \in (0, 1)$ and

$$X_{k+1} = Z_{k+1}(2X_k - X_k^2) + (1 - Z_{k+1})X_k^2.$$

  - Note that X_k is some function of $Z_2, \ldots, Z_k$ and X_1. Therefore, Z_{k+1} is independent of X_k and

$$\begin{aligned}
\mathbb{E}[X_{k+1} \mid \mathcal{F}_k] &= \mathbb{E}[Z_{k+1}(2X_k - X_k^2) \mid \mathcal{F}_k] + \mathbb{E}[(1 - Z_{k+1})X_k^2 \mid \mathcal{F}_k] \\
&= \mathbb{E}[Z_{k+1}(2X_k - X_k^2) \mid \mathcal{F}_k] + \mathbb{E}[(1 - Z_{k+1})X_k^2 \mid \mathcal{F}_k] \\
&= \mathbb{E}[Z_{k+1} \mid \mathcal{F}_k](2X_k - X_k^2) + \mathbb{E}[(1 - Z_{k+1}) \mid \mathcal{F}_k]X_k^2 \\
&= \frac{1}{2}(2X_k - X_k^2) + \frac{1}{2}X_k^2 \\
&= X_k.
\end{aligned}$$

- **Polya Urn Model**: Let $X_k = \frac{\text{number red balls at } k}{\text{total number of balls at } k}$ with $X_2 = \frac{1}{2}$. Then,

$$\begin{aligned}
\mathbb{E}[X_{k+1} \mid \mathcal{F}_k] &= X_k \frac{kX_k + 1}{k + 1} + (1 - X_k)\frac{kX_k}{k + 1} \\
&= \frac{kX_k + X_k}{k + 1} \\
&= X_k.
\end{aligned}$$

Martingale: Properties

- If $\{X_k\}$ is a super-martingale (sub-martingale) then $\mathbb{E}[X_n \mid \mathcal{F}_k] \le X_k$ almost surely for all $n \ge k$. Similar result holds for sub and regular martingales (with the respective inequality sign).
- If $\{X_k\}$ is a martingale, then $\{\Phi(X_k)\}$ is a sub-martingale for any convex function. (similar result holds for concave functions and super-martingales).
- Example: (Simple random walk squared) $\{S_n^2\}$ is a sub-martingale for the symmetric random walk S_n in $\mathbb{R}$.

Doob's Martingale Convergence Theorem

Theorem 1. *(sub-martingale version) Suppose that* $\{X_k\}$ *is a sub-martingale such that* $\sup_k \mathbb{E}[X_k^+] < \infty$*. Then* X_k *is convergent to a limit* X *almost surely with* $\mathbb{E}[X] < \infty$*.*

Theorem 2. *(super-martingale version) Suppose that* $\{X_k\}$ *is a subper-martingale with* $\sup_k \mathbb{E}[X_k^-] < \infty$*. Then* X_k *is convergent to a limit* X *almost surely with* $\mathbb{E}[X] < \infty$*.*

Corollary: Any martingale with either bounded $\mathbb{E}[X_k^+]$ or $\mathbb{E}[X_k^-]$ is convergent. As a result, any non-negative super-martingale is convergent almost surely.

Doob's Martingale Convergence Theorem: Two Viewpoints

Theorem 3. *(super-martingale version) Suppose that $\{X_k\}$ is a subper-martingale with $\sup_k \mathbb{E}[X_k^-] < \infty$. Then X_k is convergent to a limit X almost surely with $\mathbb{E}[X] < \infty$.*

- *Extension of this important result*: For a non-increasing sequence $\{\gamma_k\}$, if it is bounded from bellow, then it is convergent.
- *Extension of Sum of Independent Random Variables*: Martingales can be viewed as sum of *dependent* random variables...

Doob's Martingale Convergence Theorem: Implications

- **Simple symmetric random walk**: Martingale but non-convergent. Why?
- **Polar code dynamics**:
  - let Z_k be an i.i.d. binary random process with $\mathbf{Pr}(Z_k = 0) = \frac{1}{2} = \mathbf{Pr}(Z_k = 1)$. Let $X_1 \in (0, 1)$ and
  $$X_{k+1} = Z_{k+1}(2X_k - X_k^2) + (1 - Z_{k+1})X_k^2.$$
  - Martingale and since $X_k \in (0, 1)$ almost surely, therefore, it is convergent almost surely.
- **Polya Urn Model**: Since $X_k \in (0, 1)$ almost surely and it is a martingale, it converges almost surely.