भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Probability and Stochastic Processes

Lecture 27: Conditional Expectations, MMSE Estimation

Karthik P. N.
Assistant Professor, Department of AI
Email: pnkarthik@ai.iith.ac.in

11 November 2025

Vector Space of Random Variables

- Fix $p \geq 1$, and let

$$\mathcal{L}^p(\Omega, \mathscr{F}, \mathbb{P}) := \Big\{ X : \; \mathbb{E}[|X|^p] < +\infty \Big\}.$$

- $\mathcal{L}^p$ is a vector space over $\mathbb{R}$
  - $X, Y \in \mathcal{L}^p \implies X + Y \in \mathcal{L}^p$ (Minkowski's inequality)
  - $X \in \mathcal{L}^p, \quad \alpha \in \mathbb{R} \implies \alpha X \in \mathcal{L}^p$
- The function $\|\cdot\|_p : \mathcal{L}^p(\Omega, \mathscr{F}, \mathbb{P}) \to [0, \; +\infty)$ defined as

$$\|X\|_p := \left(\mathbb{E}[|X|^p]\right)^{\frac{1}{p}}$$

is a **semi-norm**

  - **(Homogeneity):** $\|\alpha X\|_p = |\alpha| \, \|X\|_p$ for all $\alpha \in \mathbb{R}$ and $X \in \mathcal{L}^p(\Omega, \mathscr{F}, \mathbb{P})$
  - **(Triangle Inequality):** $\|X + Y\|_p \leq \|X\|_p + \|Y\|_p$ for all $X, Y \in \mathcal{L}^p(\Omega, \mathscr{F}, \mathbb{P})$
  - **(Non-Negativity):** $\|X\|_p \geq 0$ for all $X \in \mathcal{L}^p(\Omega, \mathscr{F}, \mathbb{P})$
  - **(Definiteness):** $\|X\|_p = 0 \implies \mathbb{P}(\{X = 0\}) = 1$

Indian Institute of Technology Hyderabad
भारतीय प्रौद्योगिकी संस्थान हैदराबाद

An Equivalence Relation and Equivalence Classes

- The relation $\overset{R}{\sim}$ on $\mathcal{L}^p(\Omega, \mathscr{F}, \mathbb{P}) \times \mathcal{L}^p(\Omega, \mathscr{F}, \mathbb{P})$ defined as

$$X \overset{R}{\sim} Y \qquad \Longleftrightarrow \qquad \mathbb{P}(\{X = Y\}) = 1$$

  is an **equivalence relation**

- The relation $\overset{R}{\sim}$ partitions $\mathcal{L}^p(\Omega, \mathscr{F}, \mathbb{P})$ into **equivalence classes**
- The equivalence class of $X \in \mathcal{L}^p(\Omega, \mathscr{F}, \mathbb{P})$ is defined as

$$[X] := \Big\{ Y \in \mathcal{L}^p(\Omega, \mathscr{F}, \mathbb{P}) : \;\; \mathbb{P}(\{X = Y\}) = 1 \Big\}.$$

- If $Y \in [X]$, then

$$(\|X\|_p)^p = \mathbb{E}[|X|^p] = \mathbb{E}[|X|^p \cdot \mathbf{1}_{\{X=Y\}}] = \mathbb{E}[|Y|^p \cdot \mathbf{1}_{\{X=Y\}}] = \mathbb{E}[|Y|^p] = (\|Y\|_p)^p.$$

  That is, **all random variables belonging to an equivalence class have identical norm**

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Turning Semi-Norm into a Norm on Equivalence Classes

- Define a new function n_p on equivalence classes of $\mathcal{L}^p(\Omega, \mathscr{F}, \mathbb{P})$ as

$$n_p([X]) := \|X\|_p, \qquad X \in \mathcal{L}^p(\Omega, \mathscr{F}, \mathbb{P}).$$

- n_p is a **norm** on the equivalence classes of $\mathcal{L}^p(\Omega, \mathscr{F}, \mathbb{P})$
  - **(Homogeneity):** For all $\alpha \in \mathbb{R}$ and $X \in \mathcal{L}^p(\Omega, \mathscr{F}, \mathbb{P})$,

$$n_p([\alpha X]) = \|\alpha X\|_p = |\alpha| \cdot \|X\|_p = |\alpha| \cdot n_p([X]).$$

  - **(Triangle Inequality):** For all $X, Y \in \mathcal{L}^p(\Omega, \mathscr{F}, \mathbb{P})$,

$$n_p([X+Y]) = \|X+Y\|_p \leq \|X\|_p + \|Y\|_p = n_p([X]) + n_p([Y]).$$

  - **(Non-Negativity and definiteness):** For all $X \in \mathcal{L}^p(\Omega, \mathscr{F}, \mathbb{P})$,

$$n_p([X]) = \|X\|_p \geq 0, \qquad n_p([X]) = 0 \iff [X] = [0].$$

$\mathcal{L}^2(\Omega, \mathscr{F}, \mathbb{P})$ in More Depth

- Consider the space $\mathcal{L}^2(\Omega, \mathscr{F}, \mathbb{P}) = \left\{X : \ \mathbb{E}[X^2] < +\infty\right\}$.
- Using monotonicity property,

$$X \in \mathcal{L}^2(\Omega, \mathscr{F}, \mathbb{P}) \quad \Longrightarrow \quad \mathbb{E}[X^2] < +\infty \quad \Longrightarrow \quad \left|\mathbb{E}[X]\right| < +\infty$$

- Define a function $\langle \cdot, \cdot \rangle : \mathcal{L}^2(\Omega, \mathscr{F}, \mathbb{P}) \times \mathcal{L}^2(\Omega, \mathscr{F}, \mathbb{P}) \to \mathbb{R}$ as

$$\langle X, Y \rangle := \mathrm{Cov}(X, Y), \qquad X, Y \in \mathcal{L}^2(\Omega, \mathscr{F}, \mathbb{P}).$$

- $\langle \cdot, \cdot \rangle$ is a **semi-inner product**:
  - **(Symmetry):** $\langle X, Y \rangle = \mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X) = \langle Y, X \rangle$
  - **(Linearity):**

$$\langle aX + bY, Z \rangle = \mathrm{Cov}(aX + bY, Z) = a\,\mathrm{Cov}(X, Z) + b\,\mathrm{Cov}(Y, Z) = a\,\langle X, Z \rangle + b\,\langle Y, Z \rangle.$$

  - (**Non-negativity** + **Semi-definiteness**):

$$\langle X, X \rangle \geq 0, \qquad \langle X, X \rangle = 0 \quad \Longleftrightarrow \quad \mathbb{P}(\{X = 0\}) = 1.$$

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

$\mathcal{L}^2(\Omega, \mathscr{F}, \mathbb{P})$ + its Equivalence Classes

- Consider the space $\mathcal{L}^2(\Omega, \mathscr{F}, \mathbb{P})$ along with its equivalence classes
- If $X' \in [X]$ and $Y' \in [Y]$, then

$$\langle X, Y\rangle = \text{Cov}(X, Y) = \text{Cov}(X', Y') = \langle X', Y'\rangle.$$

- Define a new function IP_2 as

$$\mathrm{IP}_2([X], [Y]) = \langle X, Y\rangle, \qquad X, Y \in \mathcal{L}^2(\Omega, \mathscr{F}, \mathbb{P}).$$

- Then, IP_2 is a well-defined notion of inner **inner product**
- For any $X, Y \in \mathcal{L}^2(\Omega, \mathscr{F}, \mathbb{P})$ such that $\mathbb{E}[X] = 0 = \mathbb{E}[Y]$,

$$\rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \cdot \sqrt{\text{Var}}(Y)}} = \frac{\mathrm{IP}_2(X, Y)}{\|X\|_2 \cdot \|Y\|_2} = \cos(\theta),$$

where θ is the “angle” between the “vectors” X, Y in $\mathcal{L}^2(\Omega, \mathscr{F}, \mathbb{P})$

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Conditional Expectations

Case 1: X, Y Jointly Discrete

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$.
Suppose X and Y are **jointly discrete**, with $\mathbb{P}_{X,Y}(E) = 1$ for some countable $E \subset \mathbb{R}^2$

- $E_1 := \{x \in \mathbb{R} : \exists y \in \mathbb{R} \text{ such that } (x,y) \in E\}, \quad E_2 = \{y \in \mathbb{R} : \exists x \in \mathbb{R} \text{ such that } (x,y) \in E\}$
- **Conditional PMF conditioned on A:**

$$p_{X|A}(x) = \frac{\mathbb{P}(\{X = x\} \cap A)}{\mathbb{P}(A)}, \qquad A \in \mathscr{F} : \mathbb{P}(A) > 0.$$

- Then, we have

$$\mathbb{E}[X|A] := \sum_{x \in E_1} x \cdot p_{X|A}(x)$$

- **Conditional PMF conditioned on $\{Y = y\}$:** For any $y \in E_2$ such that $p_Y(y) > 0$,

$$p_{X|Y=y}(x) = \frac{p_{X,Y}(x,y)}{p_Y(y)}, \qquad \mathbb{E}[X|\{Y=y\}] = \sum_{x \in E_1} x \cdot p_{X|Y=y}(x)$$

Case 1: X, Y Jointly Discrete

- Define the function $\psi_1 : \mathbb{R} \to \mathbb{R}$ as

$$\psi_1(y) := \begin{cases} \mathbb{E}[X|\{Y = y\}], & p_Y(y) > 0, \\ 0, & p_Y(y) = 0. \end{cases}$$

Conditional Expectation $\mathbb{E}[X|Y]$

The quantity $\mathbb{E}[X|Y]$ is simply defined as

$$\mathbb{E}[X|Y] = \psi_1(Y).$$

Example

Suppose that X takes values uniformly in the set $\{-1, 0, 1\}$. Determine $\mathbb{E}[Y|X]$ if

$$p_{Y|X=x}(y) = \frac{1}{2}\mathbf{1}_{\{|y-x|=1\}}.$$

- The set of values that Y takes is $E = \{-2, -1, 0, 1, 2\}$
- Given $\{X = x\}$, we observe that Y takes values $x - 1$ and $x + 1$ with probability $1/2$ each
- Therefore, we have

$$\psi_1(x) = \mathbb{E}[Y|\{X = x\}] = \sum_{y \in E} y \cdot p_{Y|X=x}(y) = (x-1)\cdot\frac{1}{2} + (x+1)\cdot\frac{1}{2} = x.$$

- We therefore have

$$\mathbb{E}[Y|X] = \psi_1(X) = X.$$

X and Y Jointly Continuous

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$.
Suppose X and Y are **jointly continuous** with joint PDF $f_{X,Y}$.

- **Conditional PDF conditioned on A:**

$$f_{X|A}(x) = \frac{\mathrm{d}}{\mathrm{d}x} F_{X|A}(x) = \frac{\mathrm{d}}{\mathrm{d}x} \frac{\mathbb{P}(\{X \leq x\} \cap A)}{\mathbb{P}(A)}, \qquad A \in \mathscr{F} : \mathbb{P}(A) > 0.$$

- Then, we have

$$\mathbb{E}[X|A] := \int_{-\infty}^{\infty} x \cdot f_{X|A}(x)\, \mathrm{d}x.$$

- **Conditional PDF conditioned on $\{Y = y\}$:** For any $y \in \mathbb{R}$ such that $f_Y(y) > 0$, we have

$$f_{X|Y=y}(x) = \frac{f_{X,Y}(x, y)}{f_Y(y)}, \qquad \mathbb{E}[X|\{Y = y\}] = \int_{-\infty}^{\infty} x \cdot f_{X|Y=y}(x)\, \mathrm{d}x.$$

Case 2: X, Y Jointly Continuous

- Define the function $\psi_2 : \mathbb{R} \to \mathbb{R}$ as

$$\psi_2(y) := \begin{cases} \mathbb{E}[X|\{Y = y\}], & f_Y(y) > 0, \\ 0, & f_Y(y) = 0. \end{cases}$$

Conditional Expectation $\mathbb{E}[X|Y]$

The quantity $\mathbb{E}[X|Y]$ is simply defined as

$$\mathbb{E}[X|Y] = \psi_2(Y).$$

Example

Determine $\mathbb{E}[Y|X]$ for the case when X and Y have the joint PDF

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{x}, & 0 \leq y \leq x \leq 1, \\ 0, & \text{otherwise.} \end{cases}$$

- For any fixed $x \in [0,1]$, we have

$$f_{Y|X=x}(y) = \frac{1}{x}, \qquad y \in [0,x].$$

- Therefore, we have

$$\psi_2(x) = \mathbb{E}[Y|\{X=x\}] = \frac{x}{2}.$$

- Then, it follows that

$$\mathbb{E}[Y|X] = \psi_2(X) = \frac{X}{2}.$$

Case 3: X Continuous, Y Discrete

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$.
Suppose X is a **continuous** RV with PDF f_X and Y is a **discrete** RV with PMF p_Y.

- **Conditional PDF conditioned on A**:

$$f_{X|A}(x) = \frac{\mathrm{d}}{\mathrm{d}x} F_{X|A}(x) = \frac{\mathrm{d}}{\mathrm{d}x} \frac{\mathbb{P}(\{X \le x\} \cap A)}{\mathbb{P}(A)}, \qquad A \in \mathscr{F} : \mathbb{P}(A) > 0.$$

- Then, we have

$$\mathbb{E}[X|A] := \int_{-\infty}^{\infty} x \cdot f_{X|A}(x)\, \mathrm{d}x.$$

- **Conditional PDF conditioned on $\{Y = y\}$**: For any $y \in \mathbb{R}$ such that $p_Y(y) > 0$, we have

$$f_{X|Y=y}(x) = \frac{\mathrm{d}}{\mathrm{d}x} F_{X|Y=y}(x) = \frac{\mathrm{d}}{\mathrm{d}x} \frac{\mathbb{P}(\{X \le x\} \cap \{Y = y\})}{\mathbb{P}(\{Y = y\})}, \qquad \mathbb{E}[X|\{Y = y\}] = \int_{-\infty}^{\infty} x \cdot f_{X|Y=y}(x)\, \mathrm{d}x.$$

Case 3: X Continuous, Y Discrete

- Define the function $\psi_3 : \mathbb{R} \to \mathbb{R}$ as

$$\psi_3(y) := \begin{cases} \mathbb{E}[X|\{Y = y\}], & p_Y(y) > 0, \\ 0, & p_Y(y) = 0. \end{cases}$$

Conditional Expectation $\mathbb{E}[X|Y]$

The quantity $\mathbb{E}[X|Y]$ is simply defined as

$$\mathbb{E}[X|Y] = \psi_3(Y).$$

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Case 4: X Discrete, Y Continuous

Suppose X is a **discrete** RV with PMF p_X and Y is a **continuous** RV with PDF f_Y.

- There exists **countable** $E \subset \mathbb{R}$ such that $\mathbb{P}_X(E) = 1$
- **Conditional PMF conditioned on A:**

$$p_{X|A}(x) = \frac{\mathbb{P}(\{X = x\} \cap A)}{\mathbb{P}(A)}, \qquad A \in \mathscr{F} : \ \mathbb{P}(A) > 0.$$

- We then have

$$\mathbb{E}[X|A] := \sum_{x \in E} x \cdot p_{X|A}(x).$$

- **Conditional PMF conditioned on $\{Y = y\}$:**

$$p_{X|Y=y}(x) = \frac{f_{Y|X=x}(y)\, p_X(x)}{\underbrace{\sum_{x \in E} f_{Y|X=x}(y)\, p_X(x)}_{f_Y(y)}}, \qquad x \in \mathbb{R}, \ \ y : f_Y(y) > 0.$$

- Define $\mathbb{E}[X|\{Y = y\}]$ as

$$\mathbb{E}[X|\{Y = y\}] = \sum_{x \in E} x \cdot p_{X|Y=y}(x).$$

Case 3: X Continuous, Y Discrete

- Define the function $\psi_4 : \mathbb{R} \to \mathbb{R}$ as

$$\psi_4(y) := \begin{cases} \mathbb{E}[X|\{Y = y\}], & f_Y(y) > 0, \\ 0, & f_Y(y) = 0. \end{cases}$$

Conditional Expectation $\mathbb{E}[X|Y]$

The quantity $\mathbb{E}[X|Y]$ is simply defined as

$$\mathbb{E}[X|Y] = \psi_4(Y).$$

Example

Let $Y \sim \mathcal{N}(0,1)$, and

$$p_{X|\{Y=y\}}(x) = \frac{1}{2}\mathbf{1}_{\{|x-\text{sgn}(y)|=1\}},$$

where $\text{sgn}(y)$ denotes the sign of y, and is defined as

$$\text{sgn}(y) = \begin{cases} 1, & y > 0, \\ 0, & y = 0, \\ -1, & y < 0. \end{cases}$$

Compute $\mathbb{E}[X|Y]$ and $\mathbb{E}[Y|X]$.

- For any given Y, observe that

$$\psi_4(y) = \mathbb{E}[X|\{Y=y\}] = \frac{\text{sgn}(Y)+1}{2} + \frac{\text{sgn}(Y)-1}{2} = \text{sgn}(Y).$$

- Therefore, we have

$$\mathbb{E}[X|Y] = \psi_4(Y) = \text{sgn}(Y).$$

Law of Iterated Expectations

Theorem (Law of Iterated Expectations)

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Let X and Y be random variables.
Suppose that $\mathbb{E}[X]$ is well defined, i.e., not of the form $\infty - \infty$. Then,

$$\mathbb{E}[X] = \mathbb{E}\big[\mathbb{E}[X|Y]\big].$$

More generally, if $g : \mathbb{R} \to \mathbb{R}$ is a function such that $\mathbb{E}[g(X)]$ is well defined, then

$$\mathbb{E}[g(X)] = \mathbb{E}\big[\mathbb{E}[g(X)|Y]\big].$$

Proof – X, Y Jointly Discrete

$$
\begin{aligned}
\mathbb{E}[\mathbb{E}[g(X)|Y]] &= \sum_y \mathbb{E}[g(X)|\{Y = y\}]\, p_Y(y) \\
&= \sum_{y:p_Y(y)>0} \mathbb{E}[g(X)|\{Y = y\}]\, p_Y(y) \\
&= \sum_{y:p_Y(y)>0} \sum_x g(x)\, p_{X|Y=y}(x)\, p_Y(y) \\
&= \sum_{y:p_Y(y)>0} \sum_x g(x)\, \frac{p_{X,Y}(x,y)}{p_Y(y)}\, p_Y(y) \\
&= \sum_{y:p_Y(y)>0} \sum_x g(x)\, p_{X,Y}(x,y) = \sum_x g(x) \sum_{y:p_Y(y)>0} p_{X,Y}(x,y) \\
&= \sum_x g(x) \sum_{y:p_Y(y)>0} p_X(x) \\
&= \mathbb{E}[g(X)].
\end{aligned}
$$

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Example

Let $X_1, X_2, \ldots$ be i.i.d. random variables with $|\mathbb{E}[X_1]| < +\infty$.
Let N be a discrete random variable taking values in $\mathbb{N}$ and independent of $\{X_1, X_2, \ldots\}$.
Compute $\mathbb{E}[S_N]$, where $S_N = \sum_{i=1}^{N} X_i$.

Example + Caution!

Let Y be geometric with parameter $p = 0.5$.
Conditioned on $\{Y = y\}$, let X take the values $\pm 2^y$ with equal probability, i.e.,

$$p_{X|Y=y}(x) = \frac{1}{2}\,\mathbf{1}_{\{-2^y,\ 2^y\}}(x).$$

1. Compute $\mathbb{E}[X|Y]$, and use it to compute $\mathbb{E}[X]$.
2. Compute p_X and use it to compute $\mathbb{E}[X]$.
   In particular, show that it is different from the answer of part (1).
3. Explain the discrepancy in the answers of parts (1.) and (2.).