भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

# Probability and Stochastic Processes

## Conditional Expectations, Law of Iterated Expectations, Conditional Expectation as an MMSE Estimator

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

28/30 October 2024

# Conditional Expectations

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

# Conditional Expectation

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$.
Let $X, Y$ be random variables with respect to $\mathscr{F}$.

**Objective**

To define the following quantities:

- $\mathbb{E}[X|\{Y = y\}]$, for any $y \in \mathbb{R}$.
- $\mathbb{E}[X|Y]$.

Programme:
We shall define the above quantities by considering $X$ discrete/continuous, and $Y$ discrete/continuous.

# Case 1: $X, Y$ Jointly Discrete

Let $X, Y$ have the joint PMF $p_{X,Y}$.

- Step 1: Conditional PMF of $X$, conditioned on the event $\{Y = y\}$:

$$p_{X|Y=y}(x) = \frac{p_{X,Y}(x, y)}{p_Y(y)}, \qquad x \in \mathbb{R}.$$

- Step 2: The quantity $\mathbb{E}[X|\{Y = y\}]$ is defined as the expectation with respect to the conditional PMF $p_{X|Y=y}$, i.e.,

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

# Case 1: $X, Y$ Jointly Discrete

Let $X, Y$ have the joint PMF $p_{X,Y}$.

- Step 3: Define the function $\psi_1 : \mathbb{R} \to \mathbb{R}$ as

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

- Step 4: 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 $\{-1, 0, 1\}$.
Suppose that

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

Determine $\mathbb{E}[Y|X]$.
Exercise: Determine $\mathbb{E}[X|Y]$.

# Case 2: $X, Y$ Jointly Continuous

Let $X, Y$ have the joint PDF $f_{X,Y}$.

- Step 1: Conditional PDF of $X$, conditioned on the event $\{Y = y\}$:

$$f_{X|Y=y}(x) = \frac{f_{X,Y}(x, y)}{f_Y(y)}, \qquad x \in \mathbb{R}.$$

- Step 2: The quantity $\mathbb{E}[X|\{Y = y\}]$ is defined as the expectation with respect to the conditional PDF $f_{X|Y=y}$, i.e.,

$$\mathbb{E}[X|\{Y = y\}] := \int_{-\infty}^{+\infty} x \cdot f_{X|Y=y}(x).$$

# Case 2: $X, Y$ Jointly Continuous

Let $X, Y$ have the joint PDF $p_{X,Y}$.

- Step 3: Define the function $\psi_2 : \mathbb{R} \to \mathbb{R}$ as

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

- Step 4: The quantity $\mathbb{E}[X|Y]$ is simply defined as

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

# Example

- Let $X$ and $Y$ have the joint PDF

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

What is $\mathbb{E}[Y|X]$?

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

# Case $3$: $X$ Continuous, $Y$ Discrete

- Step 1: Conditional CDF of $X$, conditioned on the event $\{Y = y\}$:

$$F_{X|\{Y=y\}}(x) = \mathbb{P}(\{X \le x\}|\{Y = y\}), \qquad x \in \mathbb{R}.$$

# Case $3$: $X$ Continuous, $Y$ Discrete

- Step 1: Conditional CDF of $X$, conditioned on the event $\{Y = y\}$:

$$F_{X|\{Y=y\}}(x) = \mathbb{P}(\{X \leq x\}|\{Y = y\}), \qquad x \in \mathbb{R}.$$

- Step 2: Get the conditional PDF of $X$, conditioned on the event $\{Y = y\}$:

$$h_y(x) = \frac{d}{dx} F_{X|\{Y=y\}}(x).$$

# Case 3: $X$ Continuous, $Y$ Discrete

- Step 1: Conditional CDF of $X$, conditioned on the event $\{Y = y\}$:

$$F_{X|\{Y=y\}}(x) = \mathbb{P}(\{X \leq x\}|\{Y = y\}), \qquad x \in \mathbb{R}.$$

- Step 2: Get the conditional PDF of $X$, conditioned on the event $\{Y = y\}$:

$$h_y(x) = \frac{d}{dx} F_{X|\{Y=y\}}(x).$$

- Step 3: The quantity $\mathbb{E}[X|\{Y = y\}]$ is defined as the expectation with respect to the conditional PDF $h_y(x)$, i.e.,

$$\mathbb{E}[X|\{Y = y\}] := \int_{-\infty}^{+\infty} x \cdot h_y(x).$$

# Case $3$: $X$ Continuous, $Y$ Discrete

- Step 4: Define the function $\psi_3 : \mathbb{R} \to \mathbb{R}$ as

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

# Case 3: $X$ Continuous, $Y$ Discrete

- Step 4: Define the function $\psi_3 : \mathbb{R} \to \mathbb{R}$ as

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

- Step 5: The quantity $\mathbb{E}[X|Y]$ is simply defined as

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

## Case 4: $X$ Discrete, $Y$ Continuous

- Step 1: The joint probability

$$i(x, y) = \mathbb{P}(\{X = x\} \cap \{Y \leq y\}).$$

## Case 4: $X$ Discrete, $Y$ Continuous

- Step 1: The joint probability

$$i(x, y) = \mathbb{P}(\{X = x\} \cap \{Y \leq y\}).$$

- Step 2: Construction of the conditional PMF of $X$, conditioned on the event $\{Y = y\}$:

$$i_y(x) = \frac{1}{f_Y(y)} \cdot \frac{d}{dy} i(x, y), \qquad x \in \mathbb{R}.$$

## Case 4: $X$ Discrete, $Y$ Continuous

- Step 1: The joint probability

$$i(x,y) = \mathbb{P}(\{X = x\} \cap \{Y \leq y\}).$$

- Step 2: Construction of the conditional PMF of $X$, conditioned on the event $\{Y = y\}$:

$$i_y(x) = \frac{1}{f_Y(y)} \cdot \frac{d}{dy} i(x,y), \qquad x \in \mathbb{R}.$$

- Step 3: The quantity $\mathbb{E}[X|\{Y = y\}]$ is defined as the expectation with respect to the conditional PMF $i_y(x)$, i.e.,

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

# Case 4: $X$ Discrete, $Y$ Continuous

- Step 4: Define the function $\psi_4 : \mathbb{R} \to \mathbb{R}$ as

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

# Case 4: $X$ Discrete, $Y$ Continuous

- Step 4: Define the function $\psi_4 : \mathbb{R} \to \mathbb{R}$ as

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

- Step 5: The quantity $\mathbb{E}[X|Y]$ is simply defined as

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

# Example

- Let $Y \sim \mathcal{N}(0,1)$. Suppose that the conditional PMF of $X$, conditioned on the event $\{Y=y\}$, is

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

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

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

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

# Law of Iterated Expectations

# Law of Iterated Expectations

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$.
Let $X$ and $Y$ be random variables w.r.t. $\mathscr{F}$.

**Theorem (Law of Iterated Expectations)**

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

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

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

# Proof – $X, Y$ Jointly Discrete

$$\mathbb{E}[\mathbb{E}[g(X)|Y]] = \sum_{y} \mathbb{E}[g(X)|\{Y = y\}]\, p_Y(y)$$

# Proof – $X, Y$ Jointly Discrete

$$\mathbb{E}[\mathbb{E}[g(X)|Y]] = \sum_{y} \mathbb{E}[g(X)|\{Y = y\}]\, p_Y(y)$$

$$= \sum_{y:p_Y(y)>0} \mathbb{E}[g(X)|\{Y = y\}]\, p_Y(y)$$

# Proof – $X, Y$ Jointly Discrete

$$\mathbb{E}[\mathbb{E}[g(X)|Y]] = \sum_{y} \mathbb{E}[g(X)|\{Y = y\}]\, p_Y(y)$$

$$= \sum_{y:p_Y(y)>0} \mathbb{E}[g(X)|\{Y = y\}]\, p_Y(y)$$

$$= \sum_{y:p_Y(y)>0} \sum_{x} g(x)\, p_{X|\{Y=y\}}(x)\, p_Y(y)$$

# Proof – $X, Y$ Jointly Discrete

$$\mathbb{E}[\mathbb{E}[g(X)|Y]] = \sum_{y} \mathbb{E}[g(X)|\{Y = y\}]\, p_Y(y)$$

$$= \sum_{y:p_Y(y)>0} \mathbb{E}[g(X)|\{Y = y\}]\, p_Y(y)$$

$$= \sum_{y:p_Y(y)>0} \sum_{x} g(x)\, p_{X|\{Y=y\}}(x)\, p_Y(y)$$

$$= \sum_{y:p_Y(y)>0} \sum_{x} g(x)\, \frac{p_{X,Y}(x,y)}{p_Y(y)}\, p_Y(y)$$

# Proof – $X, Y$ Jointly Discrete

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

# Proof – $X, Y$ Jointly Discrete

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

# Proof – $X, Y$ Jointly Discrete

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

# 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 independent of $\{X_n\}_{n=1}^{\infty}$.
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.).

Indian Institute of Technology Hyderabad

# Conditioning on Events – 1

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$.
Let $X$ be a random variable w.r.t. $\mathscr{F}$.
Let $A \in \mathscr{F}$ with $\mathbb{P}(A) > 0$.

## Proposition (Conditioning on an Event)

- Suppose that $X$ is discrete. Then,

$$\mathbb{E}[X|A] = \sum_x x\, p_{X|A}(x),$$

where $p_{X|A}(x)$ is defined as

$$p_{X|A}(x) = \frac{\mathbb{P}(\{X = x\} \cap A)}{\mathbb{P}(A)}.$$

# Conditioning on Events – 2

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$.
Let $X$ be a random variable w.r.t. $\mathscr{F}$.
Let $A \in \mathscr{F}$ with $\mathbb{P}(A) > 0$.

**Proposition (Conditioning on an Event)**

- Suppose that $X$ is continuous. Then,

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

where $f_{X|A}(x)$ is defined as

$$f_{X|A}(x) = \frac{\mathrm{d}}{\mathrm{d}x} F_{X|A}(x).$$

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

# Miscellaneous Problems on Conditional Expectations

# Miscellaneous Problems

- Let $Y \sim \mathrm{Geometric}(p)$ for some $p > 0$.
  Compute $\mathbb{E}[Y]$ using the law of iterated expectations.

# Miscellaneous Problems

- Let $Y \sim \text{Geometric}(p)$ for some $p > 0$.
  Compute $\mathbb{E}[Y]$ using the law of iterated expectations.

- Suppose that $X \sim \text{Exponential}(\lambda)$ for some $\lambda > 0$.
  Compute $\mathbb{E}[X|\{X > 1\}]$.

# Miscellaneous Problems

- Let $Y \sim \text{Geometric}(p)$ for some $p > 0$.
  Compute $\mathbb{E}[Y]$ using the law of iterated expectations.

- Suppose that $X \sim \text{Exponential}(\lambda)$ for some $\lambda > 0$.
  Compute $\mathbb{E}[X|\{X > 1\}]$.

- Let $X$ and $Y$ be jointly continuous with the joint PDF

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{2}\, y\, e^{-xy}, & x > 0,\ 0 < y < 2, \\ 0, & \text{otherwise.} \end{cases}$$

  Compute $\mathbb{E}[e^{X/2}|Y]$.

# Miscellaneous Problems

- What is $\mathbb{E}[g(X)|X]$?

- If $X \perp\!\!\!\perp Y$, what is $\mathbb{E}[X|Y]$?

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

# Conditional Expectation as MMSE Estimator

# Conditional Expectation as MMSE Estimator

Recall: if $|\mathbb{E}[X]| < +\infty$, then

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

More generally, the following holds.

# Conditional Expectation as MMSE Estimator

Recall: if $|\mathbb{E}[X]| < +\infty$, then

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

More generally, the following holds.

**Proposition**

Let $f$ be Borel-measurable, and let $|\mathbb{E}[Xf(Y)]| < +\infty$. Then,

$$\mathbb{E}[Xf(Y)] = \mathbb{E}[f(Y)\,\mathbb{E}[X|Y]].$$

The previous result simply follows by taking $f \equiv 1$.

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

# Conditional Expectation as MMSE Estimator

Recall: if $|\mathbb{E}[X]| < +\infty$, then

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

More generally, the following holds.

**Proposition**

Let $f$ be Borel-measurable, and let $|\mathbb{E}[Xf(Y)]| < +\infty$. Then,

$$\mathbb{E}[Xf(Y)] = \mathbb{E}[f(Y)\,\mathbb{E}[X|Y]].$$

The previous result simply follows by taking $f \equiv 1$.

$$\mathbb{E}[(X - \mathbb{E}[X|Y])\,f(Y)] = 0.$$

# Conditional Expectation as MMSE Estimator

Recall: if $|\mathbb{E}[X]| < +\infty$, then

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

More generally, the following holds.

**Proposition**

Let $f$ be Borel-measurable, and let $|\mathbb{E}[Xf(Y)]| < +\infty$. Then,

$$\mathbb{E}[Xf(Y)] = \mathbb{E}[f(Y)\,\mathbb{E}[X|Y]].$$

The previous result simply follows by taking $f \equiv 1$.

$$\mathbb{E}[(X - \mathbb{E}[X|Y])\,f(Y)] = 0.$$

$$\mathrm{Cov}(X - \mathbb{E}[X|Y], f(Y)) = 0.$$

# MMSE

**Theorem (Conditional Expectation as the MMSE Estimator)**

The conditional expectation $\mathbb{E}[X|Y]$ is the minimum mean-squared error (MMSE) estimator for $X$ given $Y$, i.e., for any Borel-measurable function $h$,

$$\mathbb{E}[(X - \mathbb{E}[X|Y])^2] \leq \mathbb{E}[(X - h(Y))^2].$$