भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Probability and Stochastic Processes

Lecture 28: Law of Iterated Expectations, Conditional Expectation as an MMSE Estimator, Generating Functions: Probability Generating Function, Moment Generating Function, Characteristic Function

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

14 November 2025

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

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 measurable and such that $\mathbb{E}[g(X)]$ is well defined, then

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

Example

Let $X_1, X_2, \ldots$ be IID random variables. Let $\mathbb{E}[X_1]$ be well-defined.
Let N be a discrete random variable taking values in $\mathbb{N}$ and independent of $\{X_1, X_2, \ldots\}$, and let $\mathbb{E}[N]$ be well-defined.
Compute $\mathbb{E}[S_N]$ assuming it is well-defined, where $S_N = \sum_{n=1}^{N} X_n$.

- Using the law of iterated expectations, we may write
$$\mathbb{E}[S_N] = \mathbb{E}\Big[\mathbb{E}[S_N \mid N]\Big].$$
- For any $n \in \mathbb{N}$, we have
$$\mathbb{E}[S_N \mid \{N = n\}] = \mathbb{E}[S_n \mid \{N = n\}] \stackrel{(a)}{=} \mathbb{E}[S_n] = n \cdot \mathbb{E}[X_1],$$
where (a) above follows from the fact that $N \perp\!\!\!\perp \{X_1, X_2, \ldots\}$
- Then, we have
$$\mathbb{E}[S_N \mid N] = N \cdot \mathbb{E}[X_1],$$
from which it follows that
$$\mathbb{E}[S_N] = \mathbb{E}\Big[\mathbb{E}[S_N \mid N]\Big] = \mathbb{E}[N] \cdot \mathbb{E}[X_1].$$

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

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.).

Solution – 1

- For any $y \in \mathbb{N} \cup \{0\}$, we have

$$\mathbb{E}[X \mid \{Y = y\}] = \frac{2^y}{2} + \frac{(-2^y)}{2} = 0.$$

- Hence, it follows that

$$\mathbb{E}[X|Y] = 0.$$

- Using the law of iterated expectations, we get

$$\mathbb{E}[X] = \mathbb{E}\Big[\mathbb{E}[X|Y]\Big] = 0.$$

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Solution – $2, 3$

- The set of values that X takes is $\{\pm 2, \pm 4, \pm 8, \ldots\}$
- For any $k \in \mathbb{N}$,

$$\mathbb{P}(\{X = 2^k\}) = \mathbb{P}(\{X = 2^k\} \mid \{Y = k\}) \cdot \mathbb{P}(\{Y = k\}) = \frac{1}{2} \cdot \left(\frac{1}{2}\right)^k = \frac{1}{2^{k+1}} = \mathbb{P}(\{X = -2^k\}).$$

- Then, we have

$$\mathbb{E}[X_+] = \sum_{k \in \mathbb{N}} 2^k \cdot \frac{1}{2^{k+1}} = +\infty = \mathbb{E}[X_-]$$

- Thus, $\mathbb{E}[X]$ is not well-defined

The discrepancy in the answers of parts 1 and 2 is because of applying the law of iterated expectations **without checking** first whether $\mathbb{E}[X]$ is well-defined.

Conditional Expectation as MMSE Estimator

Conditional Expectation as the Projection

Proposition (Conditional Expectation as Projection)

Let $g : \mathbb{R} \to \mathbb{R}$ be **measurable**, and let X, Y be random variables.
Suppose that $X, Y, g(Y) \in \mathcal{L}^2(\Omega, \mathscr{F}, \mathbb{P})$. Then,

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

Equivalently, we have

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

Proof of Proposition

- By Hölder's inequality, we have

$$\Big|\mathbb{E}[X\,g(Y)]\Big| \le \mathbb{E}\Big[|X\,g(Y)|\Big] \le \sqrt{\mathbb{E}[X^2]\cdot\mathbb{E}\Big[\big(g(Y)\big)^2\Big]} < +\infty.$$

  This implies that $\mathbb{E}[X\,g(Y)]$ is well-defined

- Using the law of iterated expectations, we may write

$$\mathbb{E}[X\,g(Y)] = \mathbb{E}\Big[\mathbb{E}[X\,g(Y) \mid Y]\Big].$$

- For any $y \in \mathbb{R}$,

$$\mathbb{E}[X\,g(Y) \mid \{Y = y\}] = \mathbb{E}[X\,g(y) \mid \{Y = y\}] = g(y)\cdot\mathbb{E}[X \mid \{Y = y\}],$$

  from which it follows that

$$\mathbb{E}[X\,g(Y) \mid Y] = g(Y)\cdot\mathbb{E}[X \mid Y].$$

Tidbits on Conditional Expectation

- Let $\sigma(Y)$ denote the σ-algebra generated by Y
- Let $\mathcal{S}_{\sigma(Y)} \subset \mathscr{F}$ denote the collection of all random variables which are measurable with respect to $\sigma(Y)$
- That is,

$$\mathcal{S}_{\sigma(Y)} := \left\{ Z : \quad Z^{-1}(B) \in \sigma(Y) \text{ for every } B \in \mathscr{B}(\mathbb{R}) \right\}.$$

- If $g : \mathbb{R} \to \mathbb{R}$ is **measurable**, then

$$Z = g(Y) \in \mathcal{S}_{\sigma(Y)}.$$

- Conversely, if $Z \in \mathcal{S}_{\sigma(Y)}$, then

$$Z = g(Y) \qquad \text{for some measurable function } g$$

- $\mathbb{E}[X|Y] \in \mathcal{S}_{\sigma(Y)}$

Conditional Expectation as the MMSE Estimator

Theorem (Conditional Expectation as the MMSE Estimator)

Suppose that $X, Y \in \mathcal{L}^2(\Omega, \mathscr{F}, \mathbb{P})$.
The conditional expectation $\mathbb{E}[X|Y]$ is the **minimum mean-squared error (MMSE)** estimator for X given Y, i.e.,

$$\mathbb{E}[(X - \mathbb{E}[X|Y])^2] = \inf_{Z \in \mathcal{S}_{\sigma(Y)}} \mathbb{E}[(X - Z)^2].$$

Proof: Fix an arbitrary $Z \in \mathcal{S}_{\sigma(Y)}$. By definition, there exists g measurable such that $Z = g(Y)$.

- We have

$$\begin{aligned}\mathbb{E}[(X-Z)^2] &= \mathbb{E}[(X-g(Y))^2] = \mathbb{E}\left[(X - \mathbb{E}[X|Y] + \mathbb{E}[X|Y] - g(Y))^2\right] \\ &= \mathbb{E}\left[(X - \mathbb{E}[X|Y])^2\right] + \underbrace{\mathbb{E}\left[(\mathbb{E}[X|Y] - g(Y))^2\right]}_{\text{Term 1}} + 2\,\underbrace{\mathbb{E}\left[(X - \mathbb{E}[X|Y])(\mathbb{E}[X|Y] - g(Y))\right]}_{\text{Term 2}}\end{aligned}$$

- From previous proposition, we have Term $2 = 0$
- Also, Term $1 \geq 0$, with equality if and only if $Z = \mathbb{E}[X|Y]$

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Generating Functions

Probability Generating Function (PGF)

Definition (Probability Generating Function)

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Let X be an **integer-valued** random variable.
The **probability generating function (PGF)** of X is defined as

$$G_X(z) := \mathbb{E}[z^X] = \sum_{k \in \mathbb{Z}} z^k \, p_X(k), \qquad z \in \mathbb{R}.$$

- The **region of convergence (ROC)** of a PGF is defined as the set

$$\mathrm{ROC}(G_X) = \Big\{ z \in \mathbb{R} : |G_X(z)| < +\infty \Big\}.$$

- For any $z \in \mathbb{R}$ such that $|z| \leq 1$, we have

$$|G_X(z)| = \Big| \sum_{k \in \mathbb{Z}} z^k \, p_X(k) \Big| \leq \sum_{k \in \mathbb{Z}} |z|^k \, p_X(k) \leq \sum_{k \in \mathbb{Z}} p_X(k) = 1,$$

thus proving that

$$\{z \in \mathbb{R} : \ |z| \leq 1\} \ \subseteq \ \mathrm{ROC}(G_X).$$

Examples

- If $X \sim \text{Poisson}(\lambda)$, then
$$G_X(z) = e^{\lambda(z-1)}, \qquad z \in \mathbb{R}.$$
- If $X \sim \text{Geometric}(p)$, then
$$G_X(z) = \frac{pz}{1-(1-p)z}, \qquad |z| < \frac{1}{1-p}.$$

Properties of PGF

- $G_X(1) = 1$
- $\frac{\mathrm{d}}{\mathrm{d}z} G_X(z)\Big|_{z=1} = \mathbb{E}[X]$
- More generally, for any $k \in \mathbb{N}$,

$$\frac{\mathrm{d}^k}{\mathrm{d}z^k} G_X(z)\Big|_{z=1} = \mathbb{E}[X(X-1)\cdots(X-k+1)].$$

- $X \perp\!\!\!\perp Y \implies G_{X+Y}(z) = G_X(z) \cdot G_Y(z)$. Furthermore,

$$\mathrm{ROC}(G_{X+Y}) = \mathrm{ROC}(G_X) \cap \mathrm{ROC}(G_Y).$$

- Let $Y = \sum_{i=1}^{N} X_i$, where $X_1, X_2, \ldots$ are IID integer-valued and N is independent of $\{X_1, X_2, \ldots\}$ and taking values in $\mathbb{N}$. Then, for any $z \in \mathbb{R}$ such that $G_Y(z)$ is well-defined, we have

$$G_Y(z) = G_N(G_{X_1}(z)).$$

Moment Generating Function (MGF)

Moment Generating Function (MGF)

Definition (Moment Generating Function)

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Let X be a random variable.
The **moment generating function (MGF)** of X is a function $M_X : \mathbb{R} \to [0, +\infty]$ defined as

$$M_X(t) := \mathbb{E}[e^{tX}], \qquad t \in \mathbb{R}.$$

The **region of convergence (ROC)** of MGF is defined as the set

$$\mathrm{ROC}(M_X) = \left\{ t \in \mathbb{R} : \; M_X(t) < +\infty \right\}.$$

Examples

- If $X \sim \text{Exponential}(\mu)$, then

$$M_X(t) = \begin{cases} \frac{\mu}{\mu - t}, & t < \mu, \\ +\infty, & t \geq \mu. \end{cases}$$

- If $X \sim \mathcal{N}(0, 1)$, then

$$M_X(t) = e^{\frac{t^2}{2}}, \qquad t \in \mathbb{R}.$$

- If $f_X(x) = \frac{1}{\pi} \frac{1}{1+x^2}$, $x \in \mathbb{R}$, then

$$M_X(t) = \begin{cases} 1, & t = 0, \\ +\infty, & t \neq 0. \end{cases}$$