भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Probability and Stochastic Processes

Lecture 25: Expectations of Functions of Random Variables, Expectations of Continuous Random Variables, Variance, Correlation, Covariance, Cauchy–Schwarz Inequality

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

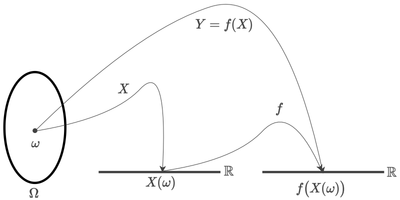
04 November 2025

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Expectations of Functions of Random Variables

Key Question: How to compute $\mathbb{E}[f(X)]$?

Expectation Over Different Spaces



Theorem (Expectations of Functions of Random Variables)

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$.
Let X be a RV, and let $f : \mathbb{R} \to \mathbb{R}$ be **measurable**. Let $Y = f(X)$.
Then, we have

$$\mathbb{E}[Y] = \mathbb{E}[f(X)] = \int_{\Omega} f(X(\omega))\, \mathrm{d}\mathbb{P}(\omega) = \int_{\mathbb{R}} f(x)\, \mathrm{d}\mathbb{P}_X(x) = \int_{\mathbb{R}} y\, \mathrm{d}\mathbb{P}_Y(y).$$

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Proof of Theorem

- Suppose f is **simple** with a **finite range**, say $\text{Range}(f) = \{y_1, \ldots, y_n\}$
- Then, $Y = f(X)$ is a **simple** RV having the canonical representation

$$Y(\omega) = f(X(\omega)) = \sum_{i=1}^{n} y_i \, \mathbf{1}_{A_i}(\omega), \qquad A_i = \{Y = y_i\} = \{f(X) = y_i\} = X^{-1}\Big(f^{-1}(\{y_i\})\Big),$$

  where $y_1, \ldots, y_n \geq 0$ are distinct and $A_1, \ldots, A_n$ partition Ω
- We then have

$$\mathbb{E}[f(X)] = \int_\Omega f(X(\omega)) \, \mathrm{d}\mathbb{P}(\omega) = \sum_{i=1}^{n} y_i \, \mathbb{P}(A_i) = \sum_{i=1}^{n} y_i \, \mathbb{P}(\{f(X) = y_i\}). \tag{1}$$

- $f : \mathbb{R} \to \mathbb{R}$ can also be represented in canonical form as

$$f(x) = \sum_{i=1}^{n} y_i \, \mathbf{1}_{B_i}(x), \qquad B_i = \{x' : f(x') = y_i\} = f^{-1}(\{y_i\}),$$

  where $B_1, \ldots, B_n$ partition $\mathbb{R}$
- Then, we have

$$\int_\mathbb{R} f(x) \, \mathrm{d}\mathbb{P}_X(x) = \sum_{i=1}^{n} y_i \, \mathbb{P}_X(B_i) = \sum_{i=1}^{n} y_i \, \mathbb{P}(\{X \in B_i\}) = \sum_{i=1}^{n} y_i \, \mathbb{P}(X \in f^{-1}(\{y_i\})) = \sum_{i=1}^{n} y_i \, \mathbb{P}(\{f(X) = y_i\})). \tag{2}$$

- Trivially, the function $g : \mathbb{R} \to \mathbb{R}$ defined as $g(y) = \sum_{i=1}^{n} y_i \, \mathbf{1}_{\{y_i\}}(y)$ is a simple function, and its expectation with respect to $\mathbb{P}_Y$ is given by

$$\int_\mathbb{R} g(y) \, \mathrm{d}\mathbb{P}_Y(y) = \int_\mathbb{R} y \, \mathrm{d}\mathbb{P}_Y(y) = \sum_{i=1}^{n} y_i \, \mathbb{P}_Y(\{y_i\}) = \sum_{i=1}^{n} y_i \, \mathbb{P}(\{Y = y_i\}) = \sum_{i=1}^{n} y_i \, \mathbb{P}(\{f(X) = y_i\}). \tag{3}$$

Proof of Theorem

- Suppose $f : \mathbb{R} \to \mathbb{R}$ is **non-negative**
- In this case, $Y = f(X)$ is a non-negative random variable
- Let $\{f_n\}$ be the quantization sequence associated with f, i.e.,

$$\forall\, x \in \mathbb{R}, \qquad f_n(x) = \sum_{\ell=0}^{n\,2^n-1} \frac{\ell}{2^n}\, \mathbf{1}_{\left\{\frac{\ell}{2^n} \le f < \frac{\ell+1}{2^n}\right\}}(x) + n\,\mathbf{1}_{\{f \ge n\}}(x), \qquad \lim_{n\to\infty} f_n(x) = f(x).$$

- Setting $Y_n = f_n(X)$ for each n, we observe that $Y_1, Y_2, \ldots$ are **simple** RVs with

$$\forall\, \omega \in \Omega, \quad Y_1(\omega) \le Y_2(\omega) \le \cdots, \qquad \lim_{n\to\infty} Y_n(\omega) = \lim_{n\to\infty} f_n(X(\omega)) = f(X(\omega)) = Y(\omega).$$

- We then have

$$\int_\Omega Y\, \mathrm{d}\mathbb{P} = \int_\Omega f(X)\, \mathrm{d}\mathbb{P} \overset{\text{MCT}}{=} \lim_{n\to\infty} \int_\Omega Y_n\, \mathrm{d}\mathbb{P} = \lim_{n\to\infty} \int_\Omega f_n(X)\, \mathrm{d}\mathbb{P} = \lim_{n\to\infty} \int_{\mathbb{R}} f_n(x)\, \mathrm{d}\mathbb{P}_X(x) \overset{\text{MCT}}{=} \int_{\mathbb{R}} f(x)\, \mathrm{d}\mathbb{P}_X(x).$$

Expectations of Continuous Random Variables

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Continuous Random Variable – Recap

Definition (Continuous Random Variable)

A random variable X is said to be continuous if $\mathbb{P}_X \ll \lambda$, $\quad \lambda$: **Lebesgue measure** on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$.

Completing the Radon–Nikodym Theorem:

There exists a **non-negative, measurable function**, say $f : \mathbb{R} \to [0, +\infty)$, such that

$$\mathbb{P}_X(B) = \int_B f(x) \, \mathrm{d}\lambda(x), \qquad B \in \mathscr{B}(\mathbb{R}).$$

The function f is called the **probability density function (PDF)** of X.

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Expectation of a Continuous Random Variable

Theorem (Expectation for Continuous Random Variables)

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Let $X : \Omega \to \mathbb{R}$ be a continuous random variable w.r.t. with PDF f_X. Suppose that $f : \mathbb{R} \to \mathbb{R}$ is **measurable**, and suppose that $\mathbb{E}[f(X)]$ is well-defined. Then,

$$\mathbb{E}[f(X)] = \int_{\mathbb{R}} f(x) f_X(x) \, \mathrm{d}\lambda(x).$$

In particular, for $f(x) = x$, we have

$$\mathbb{E}[X] = \int_{\mathbb{R}} x f_X(x) \, \mathrm{d}\lambda(x).$$

Proof of Theorem

- Suppose f is **simple** with a **finite range**, say $\text{Range}(f) = \{y_1, \ldots, y_n\}$
- f has the canonical representation

$$f(x) = \sum_{i=1}^{n} y_i \, \mathbf{1}_{B_i}(x), \qquad B_i = \{x' : f(x') = y_i\} = f^{-1}(\{y_i\}),$$

  where $B_1, \ldots, B_n$ partition $\mathbb{R}$
- Then, we have

$$\mathbb{E}[f(X)] = \int_{\mathbb{R}} f(x) \, \mathrm{d}\mathbb{P}_X(x) = \sum_{i=1}^{n} y_i \, \mathbb{P}_X(B_i) \overset{\text{R.N.Thm}}{=} \sum_{i=1}^{n} y_i \int_{B_i} f_X(x) \, \mathrm{d}\lambda(x) = \sum_{i=1}^{n} \int_{\mathbb{R}} y_i \, \mathbf{1}_{B_i}(x) f_X(x) \, \mathrm{d}\lambda(x) \quad = \int$$

Proof of Theorem

- Suppose that f is **non-negative**
- Let $\{f_n\}$ be the quantization sequence associated with f, i.e.,

$$\forall\, x \in \mathbb{R}, \qquad f_n(x) = \sum_{\ell=0}^{n\,2^n-1} \frac{\ell}{2^n}\,\mathbf{1}_{\left\{\frac{\ell}{2^n} \leq f < \frac{\ell+1}{2^n}\right\}}(x) + n\,\mathbf{1}_{\{f \geq n\}}(x), \qquad \lim_{n\to\infty} f_n(x) = f(x).$$

- Let $Y_n = f_n(X)$ for each $n \in \mathbb{N}$, and let $Y = f(X)$
- Then, we observe that $Y_1, Y_2, \ldots$ is a sequence of **simple** RVs, and

$$\forall\, \omega \in \Omega, \quad Y_1(\omega) \leq Y_2(\omega) \leq \cdots, \qquad \lim_{n\to\infty} Y_n(\omega) = \lim_{n\to\infty} f_n(X(\omega)) = f(X(\omega)) = Y(\omega).$$

- Then, we have

$$\mathbb{E}[f(X)] = \int_{\mathbb{R}} f(x)\,\mathrm{d}\mathbb{P}_X(x) = \int_{\mathbb{R}} \lim_{n\to\infty} f_n(x)\,\mathrm{d}\mathbb{P}_X(x) \overset{\text{MCT}}{=} \lim_{n\to\infty} \int_{\mathbb{R}} f_n(x)\,\mathrm{d}\mathbb{P}_X(x) = \lim_{n\to\infty} \mathbb{E}[f_n(X)] = \lim_{n\to\infty} \int_{\mathbb{R}} f_n($$

Examples

- Suppose $X \sim \text{Exponential}(\mu)$. Compute $\mathbb{E}[X]$ and $\mathbb{E}[X^2]$.

- Suppose $X \sim \mathcal{N}(\mu, \sigma^2)$. Compute $\mathbb{E}[X]$, $\mathbb{E}[X^2]$, and $\mathbb{E}[(X-\mu)^3]$.

- **(Cauchy Distribution)**

  Suppose X has the PDF
  $$f_X(x) = \frac{c}{1+x^2}, \qquad x \in \mathbb{R}.$$
  Compute $\mathbb{E}[X]$.

- Suppose $X \sim \text{Uniform}[a, b]$. Compute $\mathbb{E}[X]$.

Important Formula for Expectation of Non-Negative RVs

Lemma (Expectation of Non-Negative Discrete/Continuous Random Variables)

Suppose X is a **non-negative** random variable.

1. If X is discrete and takes values in $\{0, 1, 2, \ldots\}$, then

$$\mathbb{E}[X] = \sum_{n=0}^{\infty} \mathbb{P}(\{X > n\}) = \sum_{n=1}^{\infty} \mathbb{P}(\{X \geq n\})$$

(See Homework 7, Question 8).

2. If X is continuous, then

$$\mathbb{E}[X] = \int_0^{\infty} \mathbb{P}(\{X > x\})\, \mathrm{d}x.$$

Proof of 2:

- Suppose f_X denotes the PDF of X
- Then,

$$\int_0^{\infty} \mathbb{P}(\{X > x\})\, \mathrm{d}x = \int_0^{\infty} \int_x^{\infty} f_X(t)\, \mathrm{d}t\, \mathrm{d}x = \int_0^{\infty} \int_0^{t} f_X(t)\, \mathrm{d}x\, \mathrm{d}t = \int_0^{\infty} t\, f_X(t)\, \mathrm{d}t = \mathbb{E}[X].$$

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Variance, Covariance, and Correlation

Variance

Definition (Variance)

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$.
Let X be random variable. Let $\mathbb{E}[X]$ be well-defined (i.e., not of the form $\infty - \infty$).
The **variance** of X is defined as

$$\mathrm{Var}(X) := \mathbb{E}[(X - \mathbb{E}[X])^2].$$

Remarks:

- $(X - \mathbb{E}[X])^2$ is a non-negative random variable, so its expectation is well-defined
- $\mathrm{Var}(X) \geq 0$
- The quantity $\sigma_X = \sqrt{\mathrm{Var}(X)}$ is called the **standard deviation** of X

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

A Result on Zero Variance

Lemma (On Zero Variance)

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$.
Let X be random variable. Let $\left|\mathbb{E}[X]\right| < +\infty$. Then,

$$\mathrm{Var}(X) = 0 \quad \Longleftrightarrow \quad \mathbb{P}(\{X = c\}) = 1 \quad \text{for some constant } c \in \mathbb{R}.$$

Proof ($\Longleftarrow$):

- Suppose $\mathbb{P}(\{X = c\}) = 1$ for some $c \in \mathbb{R}$
- Then, we have

$$\mathbb{E}[X] = \mathbb{E}[X \cdot \mathbf{1}_{\{X=c\}}] = \mathbb{E}[c \cdot \mathbf{1}_{\{X=c\}}] = c \cdot \mathbb{P}(\{X = c\}) = c.$$

- Therefore,

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[(X - c)^2] = \mathbb{E}[(X - c)^2 \cdot \mathbf{1}_{\{X=c\}}] = \mathbb{E}[(c - c)^2 \cdot \mathbf{1}_{\{X=c\}}] = 0.$$

A Result on Zero Variance

Lemma (On Zero Variance)

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$.

Let X be random variable. Let $\left|\mathbb{E}[X]\right| < +\infty$. Then,

$$\mathrm{Var}(X) = 0 \quad \Longleftrightarrow \quad \mathbb{P}(\{X = c\}) = 1 \quad \text{for some constant } c \in \mathbb{R}.$$

Proof ($\implies$):

- Suppose $\mathrm{Var}(X) = 0$
- Then, we have

$$\mathbb{E}[(X - \mathbb{E}[X])^2] = 0 \implies \mathbb{P}(\{(X - \mathbb{E}[X])^2 = 0\}) = 1 \implies \mathbb{P}(\{X = \mathbb{E}[X]\}) = 1.$$

An Alternative Expression for Variance

Alternative Expression for Variance

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$.
Let $X : \Omega \to \mathbb{R}$ be random variable. Let $\mathbb{E}[X]$ be well defined (i.e., not of the form $\infty - \infty$).

1. If $\left|\mathbb{E}[X]\right| = +\infty$, then $\mathrm{Var}(X) = +\infty$.
2. If $\left|\mathbb{E}[X]\right| < +\infty$, then
$$\mathrm{Var}(X) = \mathbb{E}[X^2] - \left(\mathbb{E}[X]\right)^2.$$
In this case, we always have
$$\mathbb{E}[X^2] \geq (\mathbb{E}[X])^2.$$

Examples

- Compute the variance of $X \sim \mathrm{Ber}(p)$.
- What is the variance of $X \sim \mathrm{Poisson}(\lambda)$?
- What is the variance of $X \sim \mathrm{Unif}([a, b])$?
- What is the variance of $X \sim \mathrm{Exponential}(\mu)$?
- What is the variance of $X \sim \mathcal{N}(\mu, \sigma^2)$?
- Give an example of a random variable X for which $\left|\mathbb{E}[X]\right| < +\infty$, but $\mathrm{Var}(X) = +\infty$.

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Covariance

Definition (Covariance)

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

$$\mathrm{Cov}(X, Y) := \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])],$$

provided the expectation on the right-hand side is well defined (i.e., not $\infty - \infty$).
Furthermore,

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

provided the right-hand side is not of the form $\infty - \infty$.

Remarks:

- $\mathrm{Cov}(X, Y)$ can be negative, positive, or zero
- If $Y = X$, then

$$\mathrm{Cov}(X, X) = \mathbb{E}[X^2] - \Big(\mathbb{E}[X]\Big)^2 = \mathrm{Var}(X).$$

Uncorrelated Random Variables

Definition (Uncorrelated Random Variables)

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$.
Let X, Y be random variables. Let $\mathbb{E}[X], \mathbb{E}[Y]$ be well defined (i.e., not of the form $\infty - \infty$).
X and Y are said to be **uncorrelated** if

$$\mathrm{Cov}(X, Y) = 0.$$

Uncorrelatedness and Independence

Theorem (Uncorrelatedness and Independence)

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

$$\mathrm{Cov}(X, Y) = 0.$$

The **converse is not true in general**.

- **Proof:**

$$\mathbb{E}[X\,Y] = \int_{\mathbb{R}^2} x\,y\,\mathrm{d}\mathbb{P}_{X,Y}(x,y) \overset{X \perp\!\!\!\perp Y}{=} \int_{\mathbb{R}^2} x\,y\,\mathrm{d}\mathbb{P}_X(x)\,\mathrm{d}\mathbb{P}_Y(y) = \left(\int_{\mathbb{R}} x\,\mathrm{d}\mathbb{P}_X(x)\right)\cdot\left(\int_{\mathbb{R}} y\,\mathrm{d}\mathbb{P}_Y(y)\right) = \mathbb{E}[X]\,\mathbb{E}[Y]$$

- **Converse not true in general:** Let $X \sim \mathcal{N}(0,1)$, and let $Y = X^2$. Then,

$$\mathbb{E}[XY] = \mathbb{E}[X^3] = 0, \qquad \mathbb{E}[X]\,\mathbb{E}[Y] = 0 \cdot 1 = 0, \qquad \mathrm{Cov}(X,Y) = 0,$$

but $X \not\!\perp\!\!\!\perp Y$

Variance of Sum of Two Random Variables

Lemma (Variance of Sum of Two Random Variables)

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

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X,Y).$$

In particular, if X, Y are uncorrelated, then

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y).$$

Correlation Coefficient and Cauchy–Schwarz Inequality

Correlation Coefficient

Definition (Correlation Coefficient)

The correlation coefficient between X and Y is defined as

$$\rho_{X,Y} := \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X) \cdot \mathrm{Var}(Y)}},$$

whenever $\mathrm{Cov}(X, Y)$ is well-defined.

Remark: $\rho_{X,Y}$ can be positive, negative, or zero

The Cauchy–Schwarz Inequality

Theorem (Cauchy–Schwarz Inequality)

For any two random variables X and Y,

$$-1 \leq \rho_{X,Y} \leq 1.$$

Furthermore, the following hold.

1. If $\rho_{X,Y} = 1$, then there exists $a > 0$ such that

$$\mathbb{P}\left(\left\{\frac{Y - \mathbb{E}[Y]}{X - \mathbb{E}[X]} = a\right\}\right) = 1.$$

2. If $\rho_{X,Y} = -1$, then there exists $a < 0$ such that

$$\mathbb{P}\left(\left\{\frac{Y - \mathbb{E}[Y]}{X - \mathbb{E}[X]} = a\right\}\right) = 1.$$

Proof of CS Inequality

- Define $\widetilde{X}, \widetilde{Y}$ as

$$\widetilde{X} := X - \mathbb{E}[X], \qquad \widetilde{Y} := Y - \mathbb{E}[Y].$$

- The following holds:

$$\mathbb{E}\left[\left(\widetilde{X} - \frac{\mathbb{E}[\widetilde{X}\,\widetilde{Y}]}{\mathbb{E}[(\widetilde{Y})^2]}\,\widetilde{Y}\right)^2\right] \geq 0.$$

- Expanding the inner squared term and using linearity of expectations, we arrive at the CS inequality
- **Equality in CS inequality:**

$$\mathbb{E}\left[\left(\widetilde{X} - \frac{\mathbb{E}[\widetilde{X}\,\widetilde{Y}]}{\mathbb{E}[(\widetilde{Y})^2]}\,\widetilde{Y}\right)^2\right] = 0 \implies \mathbb{P}\left(\widetilde{X} = \frac{\mathbb{E}[\widetilde{X}\,\widetilde{Y}]}{\mathbb{E}[(\widetilde{Y})^2]}\,\widetilde{Y}\right) = 1.$$

- Let $a \in \mathbb{R}$ be defined as

$$a := \left(\frac{\mathbb{E}[\widetilde{X}\,\widetilde{Y}]}{\mathbb{E}[(\widetilde{Y})^2]}\right)^{-1} = \frac{\mathbb{E}[(\widetilde{Y})^2]}{\mathbb{E}[\widetilde{X}\,\widetilde{Y}]} = \frac{\mathrm{Var}(Y)}{\mathrm{Cov}(X,Y)} = \frac{\sqrt{\mathrm{Var}(Y)}}{\rho_{X,Y}\sqrt{\mathrm{Var}(X)}} \quad \begin{cases} > 0, & \rho_{X,Y} = 1, \\ < 0, & \rho_{X,Y} = -1. \end{cases}$$