EE60039: Probability and Random Processes for Signals and Systems
Instructor: Prof. Ashish R. Hota

Logistics:

- **Class Timing**: Monday: 12 noon - 12:55pm; Tuesday: 10am - 11:55am,
- **Venue**: NC 244
- **Instructor Email**: ashish.hota@ieee.org. Use EE60039 in Subject Line.
- **Course Website**: http://www.facweb.iitkgp.ac.in/~ahota/prob.html

Syllabus:

Module A: Introduction to Probability and Random Variables. 4 **Weeks. Main Reference: Chapters 1-5 of Wasserman.**

1. Probability Space. Independence. Conditional Probability. [Chapter 1 of Hajek, Chapter 2-5 of Chan, Chapter 1 of Gallager]
2. Random Variables and Vectors. Discrete and Continuous Distributions. [Chapter 1 of Hajek, Chapter 2-5 of Chan, Chapter 1 of Gallager]
3. Expectation, Moments, Characteristic Functions. [Chapter 1 of Hajek, Chapter 2-5 of Chan, Chapter 1 of Gallager]
4. Inequalities and Bounds. [Chapter 1-2 of Hajek, Chapter 6 of Chan, Chapter 1 of Gallager]
5. Convergence of Random Variables. Law of Large Numbers, Central Limit Theorem. [Chapter 2 of Hajek, Chapter 6 of Chan, Chapter 1 of Gallager]

Module B: Random Processes. 4 **Weeks.**

1. Definition, Discrete-time and Continuous-time Random Processes [Chapter 4 of Hajek, Chapter 10 of Chan]
2. Stationarity, Power Spectral Density, Second order Theory [Chapter 4, 8 of Hajek, Chapter 10 of Chan]

3. Gaussian Process [Chapter 3 of Hajek, Chapter 3 of Gallager]
4. Markov Chain, Classification of States, Limiting Distributions [Chapter 4 of Gallager]

Module C: Basics of Bayesian Estimation. 4 Weeks.

1. Maximum Likelihood, Maximum Aposteriori, Mean Square and Linear Mean Square Estimation [Chapter 5 of Hajek, Chapter 8 of Chan]
2. Conditional Expectation and Orthogonality [Chapter 3 of Hajek, Chapter 10 of Gallager]
3. Kalman Filters [Chapter 3 of Hajek]
4. Hidden Markov Models [Chapter 5 of Hajek]

Module D: Information, Entropy, and Divergence, 1 Week

Reference:

The subject will closely follow the treatment in the following texts.

1. Larry Wasserman, **All of Statistics**, Springer Texts in Statistics, 2004. Available at: https://link.springer.com/book/10.1007/978-0-387-21736-9
2. Bruce Hajek, **Random Process For Engineers**, Cambridge University Press, 2015. Available at: https://hajek.ece.illinois.edu/Papers/randomprocJuly14.pdf
3. Robert G. Gallager, **Stochastic Processes: Theory for Applications**, Cambridge University Press, 2013.
4. Stanley H. Chan, **Introduction to Probability for Data Science**, Michigan Publishing, 2021. Available at: https://probability4datascience.com/index.html
5. Jason Speyer and Walter Chung, **Stochastic Processes, Estimation and Control**, SIAM, 2008.

Evaluation Plan:

1. Midsem: 30%
2. Endsem: 50%
3. Homework and Class Performance: 20%

Probability Space

Notations:

- $\mathbb{N}$: set of natural numbers
- $\mathbb{R}$: set of real numbers
- $\mathbb{Z}$: set of integers
- $\mathbb{Q}$: set of rational numbers
- $\mathbb{R}^+ = \{x \in \mathbb{R} \mid x \geq 0\}$ and $\mathbb{Z}^+ = \{a \in \mathbb{Z} \mid a \geq 0\}$.
- For a set X, we denote the set of all its subsets by $\mathcal{P}(X)$

Definition 1. *Probability spaces are triplets of* $(\Omega, \mathcal{F}, \mathbb{P}(\cdot))$ *consisting*

- *Sample space: A set* Ω *that contains all possible outcomes.*
- *Events* $\mathcal{F}$*: This is a set consisting of* ***subsets*** *of* Ω *satisfying:*

  a. $\Omega \in \mathcal{F}$,

  b. ***Closed under complement****:* $E \in \mathcal{F}$ *implies* $E^c \in \mathcal{F}$*, and*

  c. ***Closed under countable union****: for any countably many subsets* $E_1, \ldots, E_k, \ldots \in \mathcal{F}$*, we have* $\cup_{k=1}^{\infty} E_k \in \mathcal{F}$.

- *Probability measure* $\mathbb{P}(\cdot)$*: is a function from* $\mathcal{F}$ *to* $[0, 1]$ *that satisfies:*

  i. $\mathbb{P}(\Omega) = 1$*, and*

  ii. For countably many subsets $\{E_k\}$ *in* $\mathcal{F}$ *that are mutually disjoint (i.e.,* $E_i \cap E_j = \emptyset$ *for all* $i \neq j$*), we have*

$$\mathbb{P}(\bigcup_{k=1}^{\infty} E_k) = \sum_{k=1}^{\infty} \mathbb{P}(E_k).$$

Note

- Any $\mathcal{F}$ that satisfies the properties a,b, and c is called a σ**-algebra** over Ω, and $(\Omega, \mathcal{F})$ is called a *measurable space*.

Examples

- Toss of a coin: $\Omega = \{H, T\}$
- Roll of a dice: $\Omega = \{1, 2, 3, 4, 5, 6\}$
- Waiting time for the next bus: $\Omega = \{t \geq 0\}$
- Each event is a subset of Ω.
- Event is "yes/no questions that can be answered after the experiment is conducted and the outcome is known"
- Example of measurable space: $\Omega = \{0, 1\}$, and $\mathcal{F} = \{\emptyset, \{1\}, \{2\}, \{1, 2\}\}$.
- Example of measurable space: In general, power set $\mathcal{P}(\Omega)$ is a σ-algebra for any Ω.
- Example of measurable space: $\{\emptyset, \Omega\}$ is a σ-algebra for any Ω.
- What if Ω is uncountable such as $\mathbb{R}^n$? Fortunately, for $\mathbb{R}^n$ there exists a σ-algebra that we can define meaningful measures (such as uniform) in $\mathbb{R}^n$, namely the *Borel* σ-algebra.
- Probability measure $\mathbb{P}$ **measures the size of a set (an event)**.
- Example: Does the following define a probability space?
  $\Omega_1 = \{1, 2, 3\}, \mathcal{F}_1 = \{\phi, \Omega, \{1\}, \{2, 3\}\}$
  $\mathbb{P}_1[\phi] = 0.5, \mathbb{P}_1[\{1\}] = 0.3$
  $\mathbb{P}_1[\Omega] = 0.2, \mathbb{P}_1[\{2, 3\}] = 0.9$.
- Countable Union: Example: (i) $A_i = \left[-1, 1 - \frac{1}{i}\right], i = 1, 2 \ldots$
  Specifically, $A_1 = [-1, 0], A_2 = [-1, 0.5], \ldots, A_{10} = [-1, 0.9]$
  $\bigcup_{i=1}^{\infty} A_i = \{a \mid a \in A_i \text{ for some finite } i\} = [-1, 1]$.
- Homework:
  $B_n[0, 1 - \frac{1}{n})$, $\bigcap_{n=1}^{\infty} B_n =$?
  $C_n = [0, 1 + \frac{1}{n})$, $\bigcap_{n=1}^{\infty} C_n =$?

Elementary Properties implied by probability axioms

Let $A, B \in \mathcal{F}$. Then, the following properties are true.

1. $A \subseteq B$, then $\mathbb{P}(A) \leq \mathbb{P}(B)$.
2. $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.
3. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) \leq \mathbb{P}(A) + \mathbb{P}(B)$.
4. $\mathbb{P}\left(\bigcup_{i=1}^{N} A_i\right) \leqslant \sum_{i=1}^{N} \mathbb{P}(A_i)$ for every N, including $N = \infty$. (Union bound)

Conditional Probability and Independence

Definition 2. *Consider a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$*. Consider events* A *and* B *with* $\mathbb{P}(B) > 0$*. The* **conditional probability of** A **given** B *is*

$$\mathbb{P}(A \mid B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Definition 3. *A countable collection of events* $\{A_1, A_2, \ldots\}$ *are said to be mutually independent, if any finite collection from the above,* $\{A'_1, A'_2 \ldots A'_k\}$ *satisfies*

$$\mathbb{P}\left(A'_1 \cap A'_2 \ldots \cap A'_k\right) = \mathbb{P}\left(A'_1\right) \cdot \mathbb{P}\left(A'_2\right) \cdots \mathbb{P}\left(A'_k\right).$$

Notes:

- If A, B are independent, $\mathbb{P}(B) > 0$, then
  $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B) \Rightarrow \mathbb{P}(A \mid B) = \mathbb{P}(A)$.
  Knowledge that event B is true gives you no further information about occurence of A.
- Suppose A and B are disjoint can they be independent?
  No. Disjoint is the strongest form of dependence. Occurence of one event rules out the occurence of the other.

Baye's Law

Proposition 1. *Let Ω be the set of outcomes. Let $\{A_1, A_2, \ldots, A_k\}$ form a partition of Ω, and let B be another event. Then,*

- $\{A_1 \cap B, A_2 \cap B, \ldots, A_k \cap B\}$ *also form a partition of B.*
- *Law of Total Probability:*

$$\mathbb{P}(B) = \sum_{i=1}^{k} \mathbb{P}(A_i \cap B) = \sum_{i=1}^{k} \mathbb{P}(B \mid A_i)\, \mathbb{P}(A_i).$$

- *Baye's Law:*

$$\mathbb{P}(A_i \mid B) = \frac{\mathbb{P}(A_i \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B \mid A_i)\, \mathbb{P}(A_i)}{\sum_{j=1}^{k} \mathbb{P}(B \mid A_j)\, \mathbb{P}(A_j)}.$$

> Problem: Consider a disease that affects one out of every 1000 individuals. There is a test that detects the disease with 99% accuracy, that is, it classifies a healthy individual as having the disease with 1% chance, and a sick individual as healthy with 1% chance. Then,
>
> 1. What is the probability that a randomly chosen individual will test positive by the test?
> 2. Given that a person tests positive, what is the probability that he or she has the disease?

Homework: Repeat the above when detection accuracy is 99.9%, 99.99% and 99.999%.

Random Variable

Definition 4. *Let $(\Omega, \mathcal{F}, \mathbb{P}(\cdot))$ be a probability space. The mapping $X : \Omega \to \mathbb{R}$ is called a random variable if the pre-image of any interval $(-\infty, a]$ belongs to $\mathcal{F}$, i.e.*

$$X^{-1}((-\infty, a]) \in \mathcal{F} \qquad \textit{for } \textbf{every } a \in \mathbb{R}, \tag{1}$$

where

$$X^{-1}(B) := \{\omega \in \Omega \mid X(\omega) \in B\}.$$

Note: Functions that satisfy this property are called measurable functions. Measurability is a property of the function X and the σ-algebra.

Example: Let $\Omega = \{HH, TH, HT, TT\}$. Consider two σ-algebras defined on Ω.

- $\mathcal{F}_1 = \{\phi, \Omega, \{HH\}, \{HT, TH, TT\}\}$
- $\mathcal{F}_2 = \{\phi, \Omega, \{TT\}, \{HH, HT, TH\}\}$

Consider a function $Y : \Omega \to \mathbb{R}$ such that $Y(HH) = 1, Y(TH) = 1, Y(HT) = 1$ and $Y(TT) = 0$, i.e, $Y = 1$ when at least one win toss is head, and 0, otherwise.

- Is Y a random variable with respect to $\mathcal{F}_1$?
- Is Y a random variable with respect to $\mathcal{F}_2$?

A random variable is neither random, nor is it a variable. The function X itself is deterministic. Randomness is due to uncertainty regarding which outcome $\omega \in \Omega$ is true. Once the outcome ω is determined, the value $X(\omega)$ is also determined.

(Important) Indicator Random Variable

Definition 5. *(Indicator Function) For a set $E \subseteq \Omega$, define the indicator function of E as*

$$\mathbf{1}_E(\omega) = \begin{cases} 1 & \text{if } \omega \in E, \\ 0 & \text{if } \omega \notin E. \end{cases}$$

> Show that $\mathbf{1}_E(\omega)$ is a random-variable if and only if (iff) $E \in \mathcal{F}$.

Let us determine the pre-images for different $a \in \mathbb{R}$. We have three cases;

1. if $a < 0$, $\mathbf{1}_E^{-1}((-\infty, a]) =?$,
2. if $0 \leq a < 1$, then $\mathbf{1}_E^{-1}((-\infty, a]) =?$, and
3. if $1 \leq a$, then $\mathbf{1}_E^{-1}((-\infty, a]) =?$.

What do we conclude from here?

Probability Distribution of a Random Variable

Definition 6. *Consider a probability space* $(\Omega, \mathcal{F}, \mathbb{P}(\cdot))$ *and a random variable* $X : \Omega \to \mathbb{R}$ *defined on this space. The* **probability distribution function** *of* X *is a function* $F_X : \mathbb{R} \to [0, 1]$ *defined as*

$$F_X(\alpha) := \mathbb{P}(\{\omega : X(\omega) \leq \alpha\}) := \mathbb{P}(X^{-1}(-\infty, \alpha]) := \texttt{Prob}(X \leq \alpha).$$

Example: Let $\Omega = \{H, T\}, \mathcal{F} = 2^{\Omega}, \mathbb{P}\{H\} = \mathbb{P}\{T\} = \frac{1}{2}$. Consider two random variables.

- $Y_1(H) = 1, Y_1(T) = 0$.
- $Y_2(H) = 0, Y_2(T) = 1$.

Find the distribution functions F_{Y_1} and F_{Y_2}.

Properties of Distribution Function:

1. F_X is non-decreasing, i.e, if $\alpha_1, \leq \alpha_2, F_X(\alpha_1) \leq F_X(\alpha_2)$.
2. $\lim_{\alpha \to \infty} F_X(\alpha) = 1, \quad \lim_{\alpha \to -\infty} F_X(\alpha) = 0$.
3. F_X is right continuous, i.e., $F_X(\alpha) = \lim_{\epsilon \to 0^+} F_X(\alpha + \epsilon)$.

F_X **is called the cumulative distribution function (CDF).**

Example

Let $\Omega = \{1, 2, 3\}$ and $X : \Omega \to \mathbb{R}$ such that $X(1) = 0.5, X(2) = 0.7, X(3) = 0.7$.

- Find the smallest σ-algebra on Ω such that X is a random variable.
- Let $\mathbb{P}(\{1\}) = 0.3$. Find the distribution F_X.

Random Vectors and Random Processes

- *Random Vectors*: Any mapping $X : \Omega \to \mathbb{R}^n$ with $X(\omega) = (X_1(\omega), X_2(\omega), \ldots, X_n(\omega))$ is called a random vector if X_i is a random variable for all $i = 1, \ldots, n$.
- *Random Process*: An infinitely indexed collection $\{X_\alpha\}_{\alpha \in I}$ of random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ is called a random process.
- If the index set I is a discrete set (usually $I = \mathbb{Z}^+$), the random process is called a discrete-time random process. When $I = \mathbb{R}$ or $I = \mathbb{R}^+$, the random process is called a continuous-time random process.

More generally, a random variable X maps one probability space $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$ to another $(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$ in a systematic manner such that

- $X : \Omega_1 \to \Omega_2$,
- for any event $E_2 \in \mathcal{F}_2$, its pre-image $\{\omega \in \Omega_1 | X(\omega) \in E_2\} \in \mathcal{F}_1$, and
- for any event $E_2 \in \mathcal{F}_2$, $\mathbb{P}_2(E_2) := \mathbb{P}_1(\{\omega \in \Omega_1 | X(\omega) \in E_2\})$ is called the *induced measure*.

For a real-valued random variable $X : \Omega \to \mathbb{R}$, the corresponding σ-algebra on $\mathbb{R}$ is called the *Borel σ-algebra* and the induced measure gives rise to the distribution function.

Discussion on Random Variables

- When $|\Omega|$ is finite, we can define the collection of events $\mathcal{F} = 2^{\Omega}$.
- However, then $|\Omega|$ is (uncountably) infinite, there are several technical difficulties that arise in defining $\mathcal{F} = 2^{\Omega}$.
- When $\Omega = \mathbb{R}$, we use a specific σ-algebra as the set of events.

Definition 7. *The Borel σ- algebra, denoted $\mathbb{B}(\mathbb{R})$, is the smallest σ- algebra that contains all sets of the form $(-\infty, \alpha]$ for every $\alpha \in \mathbb{R}$.*

- Define $\mathcal{F} := \{(-\infty, \alpha] | \alpha \in \mathbb{R}\}$. Is $\mathcal{F}$ a σ- algebra ?
- $\mathbb{B}(\mathbb{R}) = \sigma\{\mathcal{F}\}$ is the σ- algebra generated by the sets contained in $\mathcal{F}$.
- Thus, a real-valued random variable $X : \Omega \to \mathbb{R}$ is a mapping from an arbitrary probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to $(\mathbb{R}, \mathbb{B}(\mathbb{R}), \mathbb{P}_X)$ where the induced measure $\mathbb{P}_X$ is characterized in terms of the distribution function F_X.

Discrete Random Variable

- A random variable is **discrete** when X takes a finite or countable number of values.
- Suppose X takes values in the set $\{x_1, x_2, ...x_n\}$. Then,

$$\mathbb{P}(X = x_1) = \mathbb{P}(\{\omega \in \Omega | X(\omega) = x_1\}) =: p_X(x_1), \dots$$

$$\mathbb{P}(X = x_n) = \mathbb{P}(\{\omega \in \Omega | X(\omega) = x_n\}) =: p_X(x_n).$$

  where p_X is called the **probability mass function**.
- The quantities $\{p_X(x_1)....p_X(x_n)\}$ satisfy $p_X(x_i) \geq 0$ and $\sum_{i=1}^{n} p_X(x_i) = 1$.
- The distribution function F_X is a stair case function.
- Example: Bernoulli Random variable (p):
  $X = 1$ with probability p and $X = 0$ with probability $1 - p$.
- Binomial r.v (n, p):
  Outcome of n coin tosses where each coin toss comes Head with probability p.
  $\Omega = \{HH...H...T, ...THHT..., TT...T\}$
  $X : \Omega \to \mathbb{R}$ gives the number of Heads in n coin tosses.
  E.g., $X(HH...H) = n$, $X(HTT...T) = 1$, and so on.
  Can we express a Binomial r.v in terms of a collection of Bernoulli r.v.s?
- Homework: write a program to plot the pmf and distribution of a Binomial r.v. for $n = 25$ and $p \in \{0, 0.2, 0.4, 0.6, 0.8, 1\}$.

Continuous Random Variable

- A random variable X is continuous if there exists a function $f_X : \mathbb{R} \to [0, \infty]$ such that for every $\alpha \in \mathbb{R}$, we have

$$F_X(\alpha) = \mathbb{P}(\{\omega \in \Omega | X(\omega) \leq \alpha\}) = \int_{-\infty}^{\alpha} f_X(x) dx.$$

- f_X: is called the **probability density function (pdf)** of r.v. X.
- If F_X is differentiable at α, $f_X(\alpha) = \frac{dF_X(x)}{dx} |_{x=\alpha}$.
- Example: X is uniformly distributed between $[a, b]$. The pdf is given by

$$f_X(\alpha) = \begin{cases} \frac{1}{b-a}, & \text{when} \quad \alpha \in [a, b], \\ 0, & \text{otherwise.} \end{cases}$$

  Determine the distribution function F_X.

- Example: Let X be a r.v which takes value 0 with probability 0.5. Otherwise, it is uniformly dist. between 0.5 to 1.
  1. Plot $F_X(\alpha)$ for $\alpha \in \mathbb{R}$.
  2. Find $f_X(x)$ such that $\int_{-\infty}^{\alpha} f_X(x) dx = F_X(\alpha)$.
- Exponential r.v. $(\lambda > 0)$: The pdf is given by The pdf is given by

$$f_X(\alpha) = \begin{cases} \lambda e^{-\lambda \alpha}, & \text{when} \quad \alpha \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$

  Determine the distribution function F_X.

- Gaussian r.v. $(\mu, \sigma > 0)$: The pdf $f_X(\alpha) = \frac{1}{\sqrt{2\pi}\sigma} e^{\left(\frac{-(\alpha-\mu)^2}{2\sigma^2}\right)}$ for $\alpha \in \mathbb{R}$.

Properties of probability density function

For a continuous random variable X, its pdf satisfies the following properties.

1. $f_X(x) \geq 0$, for every $x \in \mathbb{R}$.
2. $\int_{-\infty}^{\infty} f_X(x)dx = F_X(\infty) = 1$.
3. $f_X(x)$ is not a probability; if can be take values larger than 1 at some points.
4. $F_X(x+\epsilon) - F_X(x) = \int_{-\infty}^{x+\epsilon} f_X(x)dx - \int_{-\infty}^{x} f_X(x)dx = \int_{x}^{x+\epsilon} f_X(x)dx$.
5. $\mathbb{P}(a \leq X \leq b) = F_X(b) - F_X(a) = \int_a^b f_X(x)dx$.

Note: If the CDF of a r.v. X is continuous at some α, then $\mathbb{P}(X = \alpha) = 0$.

Expectation of a Random Variable

- A r.v. X is called a *simple random variable* if it takes finite number of possible values, i.e.,

$$X(\omega) = \begin{cases} a_1, & \text{if} \quad \omega \in A_1 \\ a_2, & \text{if} \quad \omega \in A_2, \ldots \\ a_n, & \text{if} \quad \omega \in A_n. \end{cases}$$

  For this simple r.v X, we define $\mathbb{E}[X] := \sum_{i=1}^{n} a_i \mathbb{P}(A_i) \in \mathbb{R}$.

- Indicator r.v for event A is a simple random variable with $\mathbb{E}[\mathbf{1}_A] = \mathbb{P}(A)$.

- For a non-negative random variable X,
  - there exists a sequence of simple random variables $\{X_1, X_2, \ldots\}$ which converges to X.
  - the expectation of each simple r.v in the sequence $\{\mathbb{E}[X_1], \mathbb{E}[X_2], \ldots\}$ can be computed as above, and this sequence of real number is convergent,
  - $\mathbb{E}[X] := \lim_{n\to\infty} \mathbb{E}[X_n]$.
- For discrete r.v: $\mathbb{E}[X] = \sum_{i=1}^{n} x_i \mathbb{P}(X = x_i) = \sum_{i=1}^{n} x_i p_X(x_i)$.
- For continuous r.v: $\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x) dx$
- General notation: $\mathbb{E}[X] = \int x dF_X(x) = \int_{\Omega} X(\omega) d\mathbb{P}(\omega)$.

- Note: $\mathbb{E}[X] \in \mathbb{R}$, i.e., expectation of a random variable is a deterministic scalar without any randomness in it.

Properties of Expectation

Definition 8. *For two random variables X and Y, we define*

- $X = Y$ *almost surely (a.s.) if* $P[\{\omega \in \Omega | X(\omega) = Y(\omega)\}] = 1$.
- $X \leq Y$ *almost surely (a.s.) if* $P[\{\omega \in \Omega | X(\omega) \leq Y(\omega)\}] = 1$.

Properties of Expectation:

- Linearity: For two random variables X, and Y,

  $$\mathbb{E}[\alpha X + \beta Y] = \alpha \mathbb{E}[X] + \beta \mathbb{E}[Y] \quad \text{for any} \quad \alpha, \beta \in \mathbb{R}.$$

  Equivalently, $\mathbb{E}[\alpha X] = \alpha \mathbb{E}[X]$, & $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$.

- If $X = Y$ a.s, then $\mathbb{E}[X] = \mathbb{E}[Y]$.

- If $X \leq Y$ a.s, then $\mathbb{E}[X] \leq \mathbb{E}[Y]$.

Function of random variables

- Let X be a random variable. Then $Y = g(X)$ is a random variable if the function g is *measurable*, i.e., for any $\alpha \in \mathbb{R}$, the inverse map $Y^{-1}\big((-\infty, \alpha]\big)$ belongs to the Borel σ-algebra over $\mathbb{R}$.

- All continuous functions are measurable. In fact, almost all functions we encounter satisfies this property.

- Example: If X is a random variable, so are $\sin(X), \log(X), X^k$, and so on.

- **Law of the unconscious statistician (LOTUS)**: If $Y = g(X)$, then

$$\mathbb{E}[Y] = \int y dF_Y(y) = \int_{-\infty}^{\infty} g(x) dF_X(x).$$

  There is no need to find the distribution of Y. Thus, for a continuous r.v. X with density f_X, we have

$$\mathbb{E}[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x) dx$$
$$\mathbb{E}[X^k] = \int_{-\infty}^{\infty} x^k f_X(x) dx$$
$$\mathbb{E}[\sin(X)] = \int_{-\infty}^{\infty} \sin(x) f_X(x) dx.$$

- $\mathbb{E}[X^k]$ is called the k-th moment of X.

- **Variance** of a r.v. X is defined as $\mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$.

Characteristic Function

Characteristic function of a r.v X is defined as

$$C_X(h) = \mathbb{E}[e^{ihX}], \qquad \text{where} \quad i = \sqrt{-1}.$$

- For a continuous r.v, $C_X(h) = \int_{-\infty}^{\infty} e^{ihX} f_X(x)dx$.
- $C_X(0) = \mathbb{E}[1] = 1$.
- $\frac{dC_X(h)}{dh} = \int_{-\infty}^{\infty} (ix)e^{ihx} f_X(x)dx$.
- $\frac{dC_X(h)}{dh}|_{h=0} = \int_{-\infty}^{\infty} (ix) f_X(x)dx = i\mathbb{E}[X]$.
- How about higher order derivatives?

Random Vector

- A random vector $X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}$ such that each X_i, $1 \leq i \leq n$ is a r.v..

- Joint distribution function (CDF) $F_X : \mathbb{R}^n \to [0, 1]$ is defined as

$$\begin{aligned} F_X(c_1, c_2, ...c_n) &= \mathbb{P}[\{\omega \in \Omega | X_1(\omega) \leq c_1, X_2(\omega) \leq c_2, \ldots, X_n(\omega) \leq c_n\}] \\ &= \mathbb{P}[\cap_{i=1}^n \{\omega \in \Omega | X_i(\omega) \leq c_i\}]. \end{aligned}$$

- The random variables $X_1, X_2, ...X_n$ are jointly continuous if there exists a function $f_X : \mathbb{R}^n \;\to\; \mathbb{R}_{\geq 0}$

$$F_X(c_1, c_2, \ldots, c_n) = \int_{-\infty}^{c_1} \int_{-\infty}^{c_2} \ldots \int_{-\infty}^{c_n} f_X(x_1, x_2, ...x_n) dx_1 dx_2 ....dx_n.$$

- Random vector $X = [X_1, \ldots, X_n]^\top$ is jointly discrete if each X_i is a joint discrete random variable. Joint pmf is defined as

$$p_X(c_1, c_2, ...c_n) = \mathbb{P}(\{\omega \in \Omega | X_i(\omega) = c_i, \; 1 \leq i \leq n\}).$$

- Joint Characteristic Function: For a continuous random vector X,

$$\begin{aligned} C_X(h_1, h_2, ...h_n) &= \mathbb{E}[e^{i(h_1 X_1 + h_2 X_2 + ...h_n X_n)}] \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} .... \int_{-\infty}^{\infty} e^{i(h_1 x_1 + h_2 x_2 + ...h_n x_n)} f_X(x_1, x_2, ...x_n) dx_1 dx_2 dx_n. \end{aligned}$$

- Expectation: $\mathbb{E}[X] = \begin{bmatrix} \mathbb{E}[X_1] \\ \mathbb{E}[X_2] \\ \vdots \\ \mathbb{E}[X_n] \end{bmatrix} \in \mathbb{R}^n$.

Computing Marginal Distributions

If joint distribution/ density/ mass function is given, we can compute the distribution/ density/ PMF of each individual constituent random variable.

- Joint distribution $F_X(c_1, c_2, \ldots, c_n) = \mathbb{P}[\cap_{i=1}^n \{\omega | X_i(\omega) \le c_i\}]$.

- Marginal distribution of the second constituent random variable

$$
\begin{aligned}
F_{X_2}(c_2) &= \mathbb{P}[\{\omega \in \Omega | X_2(\omega) \le c_2\}] \\
&= \mathbb{P}[\cap_{i=1, i\neq 2}^n \{\omega | X_i(\omega) \le \infty\} \cap \{\omega \in \Omega | X_2(\omega) \le c_2\}] \\
&= \lim_{c_1\to\infty} \lim_{c_3\to\infty} \ldots \lim_{c_n\to\infty} F_X(c_1, c_2, \ldots, c_n)
\end{aligned}
$$

- Suppose joint density $f_X(c_1, c_2, \ldots, c_n)$ is given, Find $f_{X_2}(c_2)$. Recall that

$$
\begin{aligned}
F_X(c_1, c_2, \ldots, c_n) &= \int_{x_1=-\infty}^{c_1} \int_{x_2=-\infty}^{c_2} \ldots \int_{x_n=-\infty}^{c_n} f_X(x_1, x_2, \ldots, x_n) dx_1 dx_2 \ldots dx_n \\
F_{X_2}(c_2) &= \lim_{c_i\to\infty, i\neq 2} F_X(c_1, c_2, \ldots, c_n) \\
&= \int_{x_1=-\infty}^{\infty} \int_{x_2=-\infty}^{c_2} \ldots \int_{x_n=-\infty}^{\infty} f_X(x_1, x_2, \ldots, x_n) dx_1 dx_2 \ldots dx_n \\
&= \int_{x_2=-\infty}^{c_2} \left[\int_{x_1=-\infty}^{\infty} \ldots \int_{x_n=-\infty}^{\infty} f_X(x_1 \ldots x_n) dx_1 dx_3 \ldots dx_n \right] dx_2 \\
&=: \int_{x_2=-\infty}^{c_2} f_{X_2}(x_2) dx_2
\end{aligned}
$$

Example

Consider a random vector $\begin{bmatrix} X \\ Y \end{bmatrix}$ with joint density

$$f_{XY}(x,y) = \begin{cases} x + cy^2, & x \in [0,1], y \in [0,1] \\ 0, & \text{otherwise.} \end{cases}$$

- Find the value of c.
- Find marginal densities $f_X(x)$ and $f_Y(y)$.
- Find the cumulative distribution function $F_{XY}(c_1, c_2)$.
- Compute $\mathbb{P}[0 \leq X \leq \frac{1}{2},\ 0 \leq Y \leq \frac{1}{2}]$ using the density and the cumulative distribution function.

Independence of Random Variables

A collection of random variables $\{X_1, X_2, \ldots, X_n\}$ are said to be mutually independent if for any collection of Borel subsets (events on $\mathbb{R}$) $\{A_1, A_2, \ldots, A_n\}$ the underlying events $\{\omega \in \Omega | X_1(\omega) \in A_1\}, \{\omega \in \Omega | X_2(\omega) \in A_2\} \ldots$ are mutually independent.

We have the following equivalent conditions that are easier to verify.

- Joint CDF satisfies the following property.

$$\begin{aligned} F_X(c_1, c_2, \ldots, c_n) &= \mathbb{P}[\cap_{i=1}^n \{\omega | X_i(\omega) \leq c_i\}] \\ &= \Pi_{i=1}^n \mathbb{P}[\{\omega | X_i(\omega) \leq c_1\}] \\ &= F_{X_1}(c_1) \times F_{X_2}(c_2) \times \ldots \times F_{X_n}(c_n) \end{aligned}$$

- For a discrete set of random variables, independence is equivalent to joint pmf satisfying

$$p_X(c_1, \ldots, c_n) = p_{X_1}(c_1) \times \ldots \times p_{X_n}(c_n).$$

- For a continuous set of random variables, independence is equivalent to joint pdf satisfying

$$f_X(c_1, \ldots, c_n) = f_{X_1}(c_1) \times \ldots \times f_{X_n}(c_n).$$

- Joint characteristic function satisfies

$$C_X(h_1, \ldots, h_n) = C_{X_1}(h_1) \times \ldots \times C_{X_n}(h_n) \quad \forall \{h_1 h_2 \ldots h_n\}.$$

  Only checking $C_X(h, h, \ldots, h) = C_{X_1}(h) \times \ldots \times C_{X_n}(h)$ is not enough to conclude. that X_i's are independent.

- $\mathbb{E}[X_1 X_2 \ldots X_n] = \mathbb{E}[X_1] \times \ldots \times \mathbb{E}[X_n]$.

- More generally, for any collection of bounded continuous functions $\{g_1, g_2, \ldots, g_n\}$, $\mathbb{E}[g_1(X_1) g_2(X_2) \ldots g_n(X_n)] = \mathbb{E}[g_1(X_1)] \times \ldots \times \mathbb{E}[g_n(X_n)]$.

Practice Problems

Let X and Y have joint density

$$f_{XY}(x,y) = \begin{cases} 2e^{-(x+2y)}, & \text{if} x > 0, y > 0, \\ 0, & \text{otherwise.} \end{cases}$$

Determine whether X and Y are independent.

Consider a random variable X with cumulative distribution function given by:

$$F_X(x) = \begin{cases} 1 - 3^{-\lfloor x \rfloor}, & x \geq 0, \\ 0, & \text{otherwise,} \end{cases}$$

where $\lfloor x \rfloor$ is the floor of x, i.e., the largest integer smaller than or equal to x. Is X a discrete or continuous random variable? Compute $\mathbb{P}[X = 2]$ and $\mathbb{P}[X > 2]$.

Let X and Y be two independent random variables, each having uniform distribution over the range $[0, 1]$. Let $Z = \max(X, Y)$ and $W = \min(X, Y)$.

1. Determine the CDF and expectation of Z.
2. Determine the CDF and expectation of W.
3. Determine the covariance $\text{cov}(Z, W)$.

Correlation and Covariance

Correlation between two random variables X and Y is defined as $\mathbb{E}[XY]$.

Let X and Y be discrete random variables that take values as $X \in \{x_1, x_2 .....x_n\}$ and $Y \in \{y_1, y_2 ......y_m\}$. Let the joint pmf be $p_{ij} = \mathbb{P}(X = x_i, Y = y_j)$. Then,

$$\mathbb{E}[XY] = \sum_{i=1}^{n} \sum_{j=1}^{m} x_i y_j p_{ij} = x^T P y, \quad \text{where,}$$

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}, \quad P = \begin{bmatrix} p_{11} & p_{12} \dots p_{1m} \\ p_{21} & p_{22} \dots p_{2m} \\ & \vdots \\ p_{n1} & p_{n2} \dots p_{nm} \end{bmatrix}.$$

Covariance between two random variables X and Y is

$$\begin{aligned} \texttt{cov}(X, Y) &= \mathbb{E}\big[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\big] \\ &= \mathbb{E}\big[XY - X\mathbb{E}[Y] - Y\mathbb{E}[X] + \mathbb{E}[X]\mathbb{E}[Y]\big] \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \end{aligned}$$

Correlation Coefficient

$$\rho_{X,Y} = \frac{\texttt{cov}(X, Y)}{\sqrt{\texttt{var}(X)}\sqrt{\texttt{var}(Y)}}, \qquad -1 \leq \rho_{XY} \leq 1.$$

Inner product interpretation

$$x^T y = \sum_{i=1}^{n} x_i y_i = ||x|| \;\; ||y|| \cos\theta \implies \cos\theta = \frac{x^T y}{||x|| \; ||y||}.$$

> If $Y = aX + b$, then determine $\rho_{X,Y}$.

Properties of Covariance

Covariance satisfies the following properties.

- $\texttt{cov}(X, X) = \texttt{var}(X)$.
- $\texttt{cov}(X, Y) = \texttt{cov}(Y, X)$.
- $\texttt{cov}(aX, Y) = a\texttt{cov}(X, Y)$.
- $\texttt{cov}(X + c, Y) = \texttt{cov}(X, Y)$.
- $\texttt{cov}(X + Z, Y) = \texttt{cov}(X, Y) + \texttt{cov}(Z, Y)$.
- More generally,

$$\texttt{cov}(\sum_{i=1}^{n} a_i X_i, \sum_{j=1}^{m} b_j Y_j) = \sum_{i=1}^{n} \sum_{j=1}^{m} b_j a_i \texttt{cov}(X_i, Y_j).$$

Two random variables X and Y are said to be **uncorrelated** if $\texttt{cov}(X, Y) = 0$. If X and Y are independent, then they are uncorrelated. However, the converse is not true.

Covariance Matrix of a Random Vector

For a random vector $X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}$, the covariance matrix contains the covariance of each pair of constituent random variables.

$$\begin{aligned}
\texttt{cov}(X, X) = \texttt{cov}(X) &= \begin{bmatrix} \texttt{cov}(X_1) & \texttt{cov}(X_1, X_2) \ldots \texttt{cov}(X_1, X_n) \\ & \vdots \\ \texttt{cov}(X_n, X_1) & \texttt{cov}(X_n, X_2) \ldots \texttt{cov}(X_n) \end{bmatrix} \in \mathbb{R}^{n \times n} \\
&= \mathbb{E} \begin{bmatrix} X_1 - \mathbb{E}[X_1] & (X_1 - \mathbb{E}[X_1]) \ (X_2 - \mathbb{E}[X_2]) \ldots (X_n - \mathbb{E}[X_n]) \\ X_2 - \mathbb{E}[X_2] & \\ \vdots & \\ X_n - \mathbb{E}[X_n] & \end{bmatrix} \\
&= \mathbb{E}[(X - \mathbb{E}[X]) \ \ (X - \mathbb{E}[X])^\top].
\end{aligned}$$

For two random vectors $X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}$, and $Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_m \end{bmatrix}$, the (cross)-covariance matrix is given by

$$\begin{aligned}
\texttt{cov}(X, Y) &= \begin{bmatrix} \texttt{cov}(X_1, Y_1) & \texttt{cov}(X_1, Y_2) \ldots \texttt{cov}(X_1, Y_m) \\ & \vdots \\ \texttt{cov}(X_n, Y_1) & \texttt{cov}(X_n, Y_2) \ldots \texttt{cov}(X_n, Y_m) \end{bmatrix} \in \mathbb{R}^{n \times m} \\
&= \mathbb{E}[(X - \mathbb{E}[X]) \ \ (Y - \mathbb{E}[Y])^\top].
\end{aligned}$$

Sum of IID Random Variables

- In many applications, we need to repeat the experiment to generate more samples. The outcome of every experiment is random and the experiments are independent.

- Let X_i be the random variable that represents the outcome of i-th experiment.

- The collection $\{X_i\}_{i=1,2,\dots,N}$ is said to be independent and identically distributed (IID) is each X_i has the same distribution and the random variables in the collection are mutually independent.

- Suppose $\mathbb{E}[X_i] = \mu$ and $\texttt{var}(X_i) = \sigma^2$. Let $S = \sum_{i=1}^{n} X_i$. Determine the expectation and variance of S.

- Determine the characteristic function of S from the characteristic function of X_i.

Solution

Let $S = \sum_{i=1}^n X_i$ and $\bar{S} := \frac{1}{n}\sum_{i=1}^n X_i$.

$\mathbb{E}[S] = \mathbb{E}\big[\sum_{i=1}^n X_i\big] = \sum_{i=1}^n \mathbb{E}[X_i] = n\mu$.

$\mathbb{E}[\overline{S}] = \mu$.

The variance of the sum is given by

$$\begin{aligned}
\texttt{var}(S) = \texttt{var}\Big(\sum_{i=1}^n X_i\Big) &= \mathbb{E}\big[(S - \mathbb{E}[S])^2\big] \\
&= \mathbb{E}\big[(\sum_{i=1}^n X_i - n\mu)^2\big] \\
&= \mathbb{E}\big[(\sum_{i=1}^n (X_i - \mu))^2\big] \\
&= \mathbb{E}\big[\sum_{i=1}^n (X_i - \mu)^2 + \underline{\sum_{i \neq j} (X_i - \mu)(X_j - \mu)}\big] \\
&= n\sigma^2,
\end{aligned}$$

since when two r.v.s X_i and X_j are independent,

$$\mathbb{E}\big[(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])\big] = \mathbb{E}[(X_i - \mathbb{E}[X_i])] \times \mathbb{E}[X_j - \mathbb{E}[X_j]] = 0.$$

Now: $\texttt{var}(\bar{S}) = (\frac{1}{n})^2 \texttt{var}(S) = \frac{\sigma^2}{n}$

More generally, if $\texttt{var}(X) = \sigma^2$, then $\texttt{var}(cX) = c^2\sigma^2$.

Gaussian Random Variable

A Gaussian random variable X is characterized by two parameters: mean (μ) and variance σ^2, and is denoted $\mathcal{N}(\mu, \sigma^2)$. The distribution is defined below.

- If $\sigma = 0$, then $\mathbb{P}[X = \mu] = 1$ and $\mathbb{P}[X \neq \mu] = 0$.
- If $\sigma > 0$, it is a continuous random variable with density and CDF

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad \Psi(c) = \int_{-\infty}^{c} f_X(x) dx.$$

Consequently,

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = 1.$$

> The characteristic function of $X \sim \mathcal{N}(\mu, \sigma^2)$ is given by
>
> $$\Phi_X(h) = e^{i\mu h - \frac{h^2\sigma^2}{2}}.$$

Most derivations involving Gaussian random variables and vectors leverage characteristic function.

> Suppose $X_1, X_2, \ldots, X_n$ be a collection of Gaussian random variables and are independent. Then, show that $Z := \sum_{i=1}^{n} a_i X_i$ is a Gaussian random variable.

Jointly Gaussian Random Variables

Definition 9. *A collection of random variables $(X_t)_{t\in T}$ is called jointly Gaussian if every finite linear combination is Gaussian. In particular, X is a Gaussian random vector if its constituent random variables are jointly Gaussian.*

Two random vectors $X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}$ *and* $Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_m \end{bmatrix}$ *are jointly Gaussian if the collection $\{X_1, \ldots, X_n, Y_1, \ldots, Y_m\}$ is jointly Gaussian.*

A Gaussian random vector $X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}$ is characterized by two quantities:

$$\text{mean:} \quad \mu_X = \mathbb{E}[X] = \begin{bmatrix} \mathbb{E}[X_1] \\ \mathbb{E}[X_2] \\ \vdots \\ \mathbb{E}[X_n] \end{bmatrix} \in \mathbb{R}^n \quad \text{and}$$

$$\text{covariance matrix:} \quad C_X \in \mathbb{R}^{n\times n} \quad \text{with} \quad (C_X)_{i,j} = \mathsf{cov}(X_i, X_j).$$

Show that the joint characteristic function of Gaussian random vector X is given by

$$\Phi_X(h) = e^{ih^\top \mu_X - \frac{h^\top C_X h}{2}}.$$

Properties of Gaussian Random Vectors

If $X_1, X_2, \ldots, X_n$ are jointly Gaussian, then each X_i is Gaussian.

If each of $X_1, X_2, \ldots, X_n$ are individually Gaussian and independent, then the collection is jointly Gaussian.

If X is a Gaussian random vector, and $Y = AX + b$ where A is a given matrix and b is a given vector of suitable dimensions, then show that Y is a Gaussian random vector, and find its mean and covariance.

If a collection of jointly Gaussian random variables are uncorrelated, then they are independent.

Let X be a Gaussian random vector and V be another Gaussian random vector uncorrelated with X. Let $Y = AX + V$ where A is a given matrix. Find the mean and covariance of Y. Is Y Gaussian? Does the answer change when $\mathbb{E}[V] = 0$.

Inequalities and Bounds

Union bound: If A_1, A_2 .... A_n are events, $\mathbb{P}(\cup_{i=1}^{n} A_i) \;\; \leq \;\; \sum_{i=1}^{n} \mathbb{P}(A_i)$ (Equality holds when A_i are disjoint)

Markov's Inequality: Let X be a non negative r.v. Then, for any $\epsilon > 0$,

$$\mathbb{P}(X \geq \epsilon) \;\leq \frac{\mathbb{E}[X]}{\epsilon}.$$

Note: This bound is useful for large values of ϵ. In particular, if $\epsilon < \frac{1}{\mathbb{E}[X]}$, then $\frac{\mathbb{E}[X]}{\epsilon} > 1$ which is trivial.

Main idea: $Y \leq X \Rightarrow \mathbb{E}[Y] \leq \mathbb{E}[X]$. Define

$$Y = \begin{cases} \epsilon, & \text{when} \quad X \geq \epsilon, \\ 0, & \text{otherwise.} \end{cases}$$

Is $Y \leq X?$, $\quad \mathbb{E}[Y] =?$

Chebyshev's Inequality: For any random variable X, with $\mathbb{E}[X] = \mu$, and any $\epsilon > 0$,

$$\mathbb{P}[|X - \mu| \geq \epsilon] \leq \frac{\mathtt{var}(X)}{\epsilon^2}.$$

Proof: Apply Markov's inequality to $Y = (X - \mu)^2$

Application: $\lim_{n\to\infty} \mathbb{P}[|\overline{S} - \mu| \geq \epsilon] \leq \frac{\mathtt{var}(\overline{S})}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} = 0.$

Inequalities and Bounds

Hoeffding Inequality: Let $X_1, X_2, ...X_n$ be independent random variables with $X_i \in [a_i, b_i]$. Let $S_n = \sum_{i=1}^n X_i$. Then,

$$\mathbb{P}\Big[S_n - \mathbb{E}[S_n] \geq \epsilon\Big] \leq e^{-\frac{2\epsilon^2}{\sum_{i=1}^N (b_i - a_i)^2}},$$

$$\mathbb{P}\Big[S_n - \mathbb{E}[S_n] \leq -\epsilon\Big] \leq e^{-\frac{2\epsilon^2}{\sum_{i=1}^N (b_i - a_i)^2}}.$$

If X_i's are i.i.d. with $a_i = 0, b_i = 1$, then

$$\mathbb{P}\left[\mid \frac{S_n}{n} - \mathbb{E}[X_1] \mid \geq \epsilon\right] \leq 2e^{-2\epsilon^2 n}.$$

Discuss: Confidence interval using Hoeffding and Chebyshev.

Cauchy Schwartz Inequality: For two random variables X and Y,

$$(\mathbb{E}[XY])^2 \leq \mathbb{E}[X^2]\mathbb{E}[Y^2]$$

Proof: Define $Z = (sX + Y)^2 \geq 0$. Then, for every $s \in \mathbb{R}$,

$$\begin{aligned}
&\mathbb{E}[Z] \geq 0 \\
\implies &\mathbb{E}[s^2X^2 + 2sXY + Y^2] \geq 0 \\
\implies &s^2\mathbb{E}[X^2] + 2s\mathbb{E}[XY] + \mathbb{E}[Y^2] \geq 0
\end{aligned}$$

Define: $h(s) := s^2\mathbb{E}[X^2] + 2s\mathbb{E}[XY] + \mathbb{E}[Y^2]$. Since $h(s) \geq 0$ for all $s \in \mathbb{R}$, it does not have distinct real roots. From $b^2 - 4ac \leq 0$ formula for quadratic functions, we obtain the inequality.

Corollary: Correlation coefficient lies in $[-1, 1]$.

Chernoff Bound

Note that $\mathbb{P}(X \geq \epsilon) = \mathbb{P}(e^{tX} \geq e^{t\epsilon})$ for any $t > 0$ since $x \geq y \Leftrightarrow e^x \geq e^y$. From Markov's inequality, we have

$$\begin{aligned}
\mathbb{P}(X \geq \epsilon) &= \mathbb{P}(e^{tX} \geq e^{t\epsilon}) \\
&\leq \frac{\mathbb{E}[e^{tX}]}{e^{t\epsilon}} \qquad \text{for every} \quad t > 0 \\
&\leq \min_{t>0} \left[e^{-t\epsilon}\mathbb{E}[e^{tX}]\right] \\
&= \min_{t>0} \left[e^{-t\epsilon} m_X(t)\right] \\
&= \min_{t>0} \left[e^{\log(m_X(t)-t\epsilon)}\right] \\
&= e^{-\left[\max_{t>0}(t\epsilon - \log(m_X(t))\right]},
\end{aligned}$$

where $\mathbb{E}[e^{tX}] = m_X(t)$ is called the *moment generating function* of X.

Let $X \sim$ Binomial r.v (n,p) with probability mass function given by $\mathbb{P}(X = k) = (nk)p^k(1-p)^{n-k}$, with $k = \{0, 1, 2.....n\}$. Find upper bounds on $\mathbb{P}(X \geq q)$ using Markov, Chebyshev and Chernoff bounds.

Homework: Plot the true probability and the bounds.

Distribution of sum of two random variables

Let X_1 and X_2 be two continuous random variables,
Let us try to find distribution and density of $\underline{X_1 + X_2}$ when

- X_1, X_2 are arbitrary
- X_1, X_2 are independent
- X_1, X_2 are IID.

Convergence of Sequences

Definition 10. *A sequence of real numbers* $(x_n)_{n\in\mathbb{N}} := (x_1, x_2, \ldots, x_n, \ldots)$ *with each* $x_i \in \mathbb{R}$, *is said to converge to* $x^* \in \mathbb{R}$ *if of every* $\epsilon > 0$, *there exists* n_ϵ *such that* $|x_n - x^*| < \epsilon$ *for every* $x \geq n_\epsilon$. *Then, we write* $\lim_{n\to\infty} x_n = x^*$.

Note: The above definition requires us to first conjecture a limit point x^*, which may not always be trivial.

Definition 11. *A sequence* $(x_n)_{n\in\mathbb{N}}$ *is a* <u>*Cauchy sequence*</u> *if*

$$\lim_{n,m\to\infty} |x_n - x_m| = 0.$$

<u>Proposition:</u> If a sequence $(x_n)_{n\in\mathbb{N}}$ is a Cauchy sequence, then it converges to some finite limit.

<u>Example:</u> Let $x_n = \frac{1}{n}$, i.e., the sequence $(x_n)_{n\in\mathbb{N}} = (1, \frac{1}{2}, \frac{1}{3} \ldots)$. What is a possible value of x^*? Is this sequence a Cauchy sequence?

<u>Convergence of Random Variables:</u>
Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with $\Omega = [0,1]$, and $\mathbb{P}\big[[a,b]\big] = \mathbb{P}\big[(a,b)\big] = b - a$ (uniform distribution). Let $(X_n)_{n\in\mathbb{N}}$ is a sequence of random variables defined on $(\Omega, \mathcal{F}, \mathbb{P})$ with

$$X_n(\omega) = \omega^n, \qquad \omega \in [0,1].$$

What do we mean by convergence of this sequence?

Almost Sure Convergence

Definition 12 (Almost Sure Convergence)**.** *A sequence of random variables* $(X_n)_{n\in\mathbb{N}}$ *converges almost surely to a random variable* X^* *if*

$$\mathbb{P}\big[\{\omega : \lim_{n\to\infty} X_n(\omega) = X^*(\omega)\}\big] = 1.$$

Equivalently, $\mathbb{P}\big[\{\omega : \lim_{n\to\infty} X_n(\omega) \neq X^*(\omega)\}\big] = 0.$

Note: For a given outcome ω, $\big(X_n(\omega)\big)$ is a sequence of real numbers.

Example: Does the sequence with $X_n(\omega) = \omega^n, \quad \omega \in [0,1]$ converge almost surely to some X^*?

Example: Consider a sequence of random variables defined as:

$$\begin{aligned}
X_1(\omega) &= 1, \quad \omega \in [0,1],\\
X_2(\omega) &= 1, \quad \omega \in [0,0.5],\\
X_3(\omega) &= 1, \quad \omega \in [0.5,1],\\
X_4(\omega) &= 1, \quad \omega \in [0,0.25],\\
X_5(\omega) &= 1, \quad \omega \in [0.25,0.5],\\
X_6(\omega) &= 1, \quad \omega \in [0,5,0.75], \qquad \text{and so on.}
\end{aligned}$$

Does this sequence converge almost surely to some X^*?

Let us determine the following quantities.
$\mathbb{P}(X_n \neq 0) =?$
$\mathbb{E}[X_n] =?$
Both the above quantities define a sequence of real numbers. Do those sequence converge?

Convergence in Probability and in Mean Square Sense

Definition 13. *A sequence of random variables* $(X_n)_{n\in\mathbb{N}}$ *converges to a r.v.* X^* *in probability, denoted* $X_n \ \to^P \ X^*$*, if*

$$\lim_{n\to\infty} \mathbb{P}[|X_n - X^*| \geq \epsilon] = 0 \qquad \textit{for every} \quad \epsilon > 0.$$

Definition 14. *A sequence of random variables* $(X_n)_{n\in\mathbb{N}}$*, with* $\mathbb{E}[X_n^2] < \infty \quad \forall n$*, converges to* X^* *in mean square sense if*

$$\lim_{n\to\infty} \mathbb{E}\big[(X_n - X^*)^2\big] = 0.$$

This is denoted by $X_n \ \to^{m.s} \ X^*$.

Example: Consider the following two sequence of random variables:

$$X_n(\omega) = \begin{cases} 1, & \omega \in [0, \frac{1}{n}] \\ 0, & \text{otherwise.} \end{cases}$$

$$Y_n(\omega) = \begin{cases} n, & \omega \in [0, \frac{1}{n}] \\ 0, & \text{otherwise.} \end{cases}$$

Determine if the above sequences converge almost surely, in probability and in mean-square sense.

Convergence in Distribution

Definition 15. *A sequence of random variables* $(X_n)_{n\in\mathbb{N}}$ *converges in distribution to r.v.* X^* *if either of the following are true.*

- $\lim_{n\to\infty} F_{X_n}(x) = F_{X^*}(x)$ *at all points of continuity of* F_{X^*}.
- $\lim_{n\to\infty} \mathbb{E}[e^{ihX_n}] = \mathbb{E}[e^{ihX^*}]$ *for all* $h \in \mathbb{R}$.
- $\mathbb{E}[g(X_n)] \to \mathbb{E}[g(X^*)]$ *for every bounded continuous function* g.

Example

Let X be a r.v with CDF

$$F_X(\alpha) = \begin{cases} \frac{\alpha}{\theta}, & \alpha \in [0,\theta], \\ 1, & \alpha \geq \theta, \\ 0, & \alpha \leq 0. \end{cases}$$

Let $X_1, X_2, \ldots, X_n$ be i.i.d with distribution F_X. Define a sequence

$$Y_k = \text{max}_{i\in\{1,2,\ldots k\}} X_i.$$

Show that the sequence $(Y_n)_{n\in\mathbb{N}}$ converges in distribution to a random variable Y^* whose distribution is given by

$$F_{Y^*}(\alpha) = \begin{cases} 1, & \alpha \geq \theta \\ 0, & \text{otherwise.} \end{cases}$$

- Convergence in probability implies convergence in distribution.
- Mean-square convergence implies convergence in probability.
- Almost sure convergence implies convergence in probability.

Cauchy Criterion

Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of r.v defined on $(\Omega, \mathcal{F}, \mathbb{P})$. Then,

- X_n converges almost surely to some random variable if

$$\mathbb{P}[\{\omega : \lim_{m,n\to\infty} |X_m(\omega) - X_n(\omega)| = 0\}] = 1.$$

- X_n converges in probability to some r.v if

$$\lim_{m,n\to\infty} \mathbb{P}[|X_m - X_n| > \epsilon] = 0.$$

- X_n converges in m.s sense to some r.v if

$$\lim_{n,m\to\infty} \mathbb{E}[(X_m - X_n)^2] = 0.$$

Limit Theorems

Theorem 1 (Law of Large Numbers). *Let $X_1, X_2.....$ be a sequence of random variables. Each X_i has mean μ_X, i.e., $\mathbb{E}[X_i] = \mu_X$. Define $S_n := \sum_{i=1}^{n} X_i$. Then,*

- $\frac{S_n}{n} \rightarrow^{a.s}_{m.s} \mu_X$ *if* $\texttt{var}(X_i) \leq C \quad \forall i \in \mathbb{N}$ *and* $\texttt{cov}(X_i, X_j) = 0 \quad \forall i \neq j$.
- *If $X_1, X_2, \ldots$ i.i.d, then $\frac{S_n}{n} \rightarrow^{p} \mu_X$ (Weak law of large numbers)*
- *If $X_1, X_2, \ldots$ i.i.d, then $\frac{S_n}{n} \rightarrow^{a.s} \mu_X$ (Strong law of large numbers).*

Note: What about the random variable $\frac{S_n}{n} - \mu_X$? What is its mean and variance?

Theorem 2 (Central Limit Theorem). *Let each X_i be i.i.d, with $\mathbb{E}[X_i] = \mu_X$ and $\texttt{var}(X_i) = \sigma^2$. Let $\mathcal{N}(\mu, \sigma^2)$ denote Gaussian distribution with mean μ and variance σ^2. Then,*

- $\left(\frac{S_n - \mu_X n}{\sqrt{n}}\right) \rightarrow^{d} \mathcal{N}(0, \sigma^2)$,
- $\sqrt{n}\left(\frac{\frac{S_n}{n} - \mu_X}{\sigma}\right) \rightarrow^{d} \mathcal{N}(0, 1)$,
- $\sqrt{n}\frac{S_n}{n} = \frac{S_n}{\sqrt{n}} \rightarrow^{d} \mathcal{N}(\mu_X, \sigma^2)$.