Stochastic Processes

Lecture Notes

Chapter 1: Introduction to Random Variables

Payam Taebi

Fall 2024
https://stoch-sut.github.io/

Contents

1 Introduction

Welcome to Chapter 1 of the Stochastic Processes course. This chapter introduces the fundamental concepts of random variables, their distributions, and essential properties that form the basis for understanding more complex stochastic processes. We will explore both discrete and continuous random variables, delve into probability mass and density functions, and examine key theorems that underpin probability theory.

2 Random Variables (RV)

2.1 Definition

A random variable X is a function $X : \Omega \to \mathbb{R}$ where Ω is the sample space. It assigns a real number to each outcome in the sample space, allowing us to quantify uncertainty.

2.2 Types of Random Variables

- **Discrete RV**: Takes countable values (e.g., the outcome of a dice roll).
  - **Probability Mass Function (PMF)**: $P(X = x_i) = p(x_i)$
- **Continuous RV**: Takes uncountable values within an interval (e.g., height of individuals).
  - **Probability Density Function (PDF)**: $f_X(x)$
- **Real vs. Complex RV**:
  - **Real**: $X : \Omega \to \mathbb{R}$
  - **Complex**: $X : \Omega \to \mathbb{C}$

3 Density and Distribution Functions

3.1 Probability Mass Function (PMF)

Discrete RV:

$$P(X = x_i) = p(x_i)$$

Example: Fair Dice

Example 1. *For a fair six-sided die, the PMF is:*

$$p(x) = \frac{1}{6}, \quad x = 1, 2, 3, 4, 5, 6$$

Solution: Each outcome from 1 to 6 is equally likely, hence the probability for each outcome is $\frac{1}{6}$. □

3.2 Probability Density Function (PDF)

Continuous RV:

$$P(a \leq X \leq b) = \int_a^b f_X(x)\,dx$$

Common PDFs:

- **Uniform Distribution**:

$$f_X(x) = \begin{cases} \frac{1}{U-L} & L \leq x \leq U \\ 0 & \text{otherwise} \end{cases}$$

- **Gaussian (Normal) Distribution**:

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

- **Binomial Distribution**:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n$$

- **Poisson Distribution**:

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, \ldots$$

- **Exponential Distribution**:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Example: Exponential Distribution

Example 2. *Let X be an exponential random variable with rate parameter $\lambda = 2$. Find $P(X \leq 1)$.*

Solution: Using the Cumulative Distribution Function (CDF) of the exponential distribution:

$$P(X \leq 1) = 1 - e^{-\lambda \cdot 1} = 1 - e^{-2} \approx 0.8647$$

□

3.3 Cumulative Distribution Function (CDF)

Definition:

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(t)\,dt$$

Properties:

- Non-decreasing
- Right-continuous
- $\lim_{x \to -\infty} F_X(x) = 0$
- $\lim_{x \to \infty} F_X(x) = 1$

4 Expected Value and Moments

4.1 Expectation

Standard Form:

$$E[X] = \int_{-\infty}^{\infty} x f_X(x)\, dx \quad \text{(Continuous)}$$

$$E[X] = \sum_i x_i p(x_i) \quad \text{(Discrete)}$$

Alternative Forms:
For Non-Negative Random Variables ($X \geq 0$):

$$E[X] = \int_0^{\infty} P(X > a)\, da$$

General Form (Allowing X to Take Negative Values):

$$E[X] = \int_{-\infty}^{0} P(X < a)\, da + \int_0^{\infty} P(X > a)\, da$$

Explanation:

- This representation decomposes the expected value into contributions from the lower and upper tails of the distribution.
- It is particularly useful in scenarios where evaluating $E[X]$ directly is challenging, but tail probabilities $P(X > a)$ and $P(X < a)$ are easier to compute or estimate.

Example: Expected Value Using Tail Probabilities

Example 3. *Let X be a non-negative continuous random variable with PDF $f_X(x) = \lambda e^{-\lambda x}$ for $x \geq 0$ (Exponential Distribution). Compute $E[X]$ using the tail probability method.*

Solution: Using the alternative form for non-negative random variables:

$$E[X] = \int_0^{\infty} P(X > a)\, da = \int_0^{\infty} e^{-\lambda a}\, da = \frac{1}{\lambda}$$

For $\lambda = 2$:

$$E[X] = \frac{1}{2}$$

□

4.2 Linearity of Expectation

$$E[aX + b] = aE[X] + b$$

4.3 Higher Moments

$$M_n = E[X^n]$$

4.4 Variance

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

4.5 Function of RV

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\,dx$$

Example: Expectation of a Function

Example 4. *Let* $X \sim \mathcal{N}(0,1)$. *Compute* $E[X^2]$.

Solution: Since X is a standard normal random variable:

$$E[X^2] = \mathrm{Var}(X) + (E[X])^2 = 1 + 0 = 1$$

□

5 Joint and Conditional Distributions

5.1 Joint Probability Functions

Joint PDF/PMF:

$$f_{X,Y}(x,y) = P(X = x, Y = y) \quad \text{(Discrete)}$$

$$f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y} \quad \text{(Continuous)}$$

5.2 Marginal Distributions

From Joint PDF:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx$$

5.3 Conditional Distributions

Definition:

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

Bayes' Theorem:

$$f_{Y|X}(y|x) = \frac{f_X(x|y) f_Y(y)}{f_X(x)}$$

5.4 Independence

Condition:

$$X \text{ and } Y \text{ are independent} \iff f_{X,Y}(x,y) = f_X(x) f_Y(y)$$

5.5 Transformation of Variables

Function of Joint RVs:

$$f_{U,V}(u,v) = f_{X,Y}(x(u,v), y(u,v))\,|\det J|$$

where J is the Jacobian matrix:

$$J = \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{pmatrix}$$

Example: Transformation of Variables

Example 5. *Let $U = X + Y$ and $V = X - Y$, where X and Y are independent standard normal random variables. Find the joint PDF of U and V.*

Solution: First, express X and Y in terms of U and V:

$$X = \frac{U+V}{2}, \quad Y = \frac{U-V}{2}$$

Compute the Jacobian determinant:

$$J = \begin{pmatrix} \frac{\partial X}{\partial U} & \frac{\partial X}{\partial V} \\ \frac{\partial Y}{\partial U} & \frac{\partial Y}{\partial V} \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{pmatrix}$$

$$\det J = \left(\frac{1}{2}\right)\left(-\frac{1}{2}\right) - \left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = -\frac{1}{4} - \frac{1}{4} = -\frac{1}{2}$$

Taking absolute value:

$$|\det J| = \frac{1}{2}$$

Since X and Y are independent standard normals:

$$f_{X,Y}(x,y) = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}}$$

Thus, the joint PDF of U and V is:

$$f_{U,V}(u,v) = f_{X,Y}\left(\frac{u+v}{2}, \frac{u-v}{2}\right) \cdot \frac{1}{2} = \frac{1}{2\pi} e^{-\frac{\left(\frac{u+v}{2}\right)^2 + \left(\frac{u-v}{2}\right)^2}{2}} \cdot \frac{1}{2} = \frac{1}{4\pi} e^{-\frac{u^2+v^2}{4}}$$

□

6 Linear Correlation

6.1 Covariance

Definition:

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

6.2 Correlation Coefficient

Definition:

$$\rho_{XY} = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y}$$

where $\sigma_X = \sqrt{\mathrm{Var}(X)}$ and $\sigma_Y = \sqrt{\mathrm{Var}(Y)}$

6.3 Properties

- $\rho_{XY} = 1$: Perfect positive linear relationship
- $\rho_{XY} = -1$: Perfect negative linear relationship
- $\rho_{XY} = 0$: No linear relationship (not necessarily independent)

Example: Correlation Calculation

Example 6. *Let X and Y be two random variables with:*

$$E[X] = 2, \quad E[Y] = 3, \quad E[XY] = 11, \quad Var(X) = 4, \quad Var(Y) = 9$$

Compute the correlation coefficient ρ_{XY}.

Solution: First, compute the covariance:

$$\text{Cov}(X,Y) = E[XY] - E[X]E[Y] = 11 - (2)(3) = 5$$

Next, compute the standard deviations:

$$\sigma_X = \sqrt{4} = 2, \quad \sigma_Y = \sqrt{9} = 3$$

Finally, compute the correlation coefficient:

$$\rho_{XY} = \frac{5}{(2)(3)} = \frac{5}{6} \approx 0.8333$$

□

7 Important Theorems

7.1 Central Limit Theorem (CLT)

Statement:

- Let $X_1, X_2, \ldots, X_n$ be i.i.d. RVs with $E[X_i] = \mu$ and $\text{Var}(X_i) = \sigma^2$.
- Define $S_n = \sum_{i=1}^{n} X_i$.
- Then, as $n \to \infty$,
$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} \mathcal{N}(0,1)$$

Exception:

- Cauchy Distribution does not satisfy CLT conditions.

Example: CLT Application

Example 7. *Let $X_1, X_2, \ldots, X_{100}$ be i.i.d. exponential random variables with $\lambda = 1$. Approximate the distribution of $S_{100} = \sum_{i=1}^{100} X_i$ using CLT.*

Solution: Each X_i has $E[X_i] = 1$ and $\text{Var}(X_i) = 1$. By CLT:

$$\frac{S_{100} - 100}{10} \xrightarrow{d} \mathcal{N}(0,1)$$

Thus, S_{100} is approximately normally distributed with mean 100 and variance 100, i.e., $S_{100} \sim \mathcal{N}(100, 100)$. □

7.2 Law of Large Numbers (LLN)

Weak Law:

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \mu \quad \text{as } n \to \infty$$

Strong Law:

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{\text{a.s.}} \mu \quad \text{as } n \to \infty$$

7.3 Chebyshev's Inequality

Statement:

$$P(|X - E[X]| \geq k\sigma) \leq \frac{1}{k^2}, \quad k > 0$$

Alternative Forms:

- For any $\epsilon > 0$:

$$P(|X - \mu| \geq \epsilon) \leq \frac{\text{Var}(X)}{\epsilon^2}$$

Example: Chebyshev's Inequality Application

Example 8. *Let X be a random variable with $E[X] = 10$ and $Var(X) = 25$. Find an upper bound for $P(|X - 10| \geq 10)$ using Chebyshev's inequality.*

Solution: Using Chebyshev's inequality:

$$P(|X - 10| \geq 10) \leq \frac{25}{10^2} = \frac{25}{100} = 0.25$$

□

7.4 Cauchy-Schwarz Inequality

Statement:

$$|E[XY]| \leq \sqrt{E[X^2]E[Y^2]}$$

Equality Condition:

- Occurs if and only if X and Y are linearly dependent.

Example: Cauchy-Schwarz Inequality

Example 9. *Let X and Y be random variables with $E[X] = E[Y] = 0$, $E[X^2] = 4$, $E[Y^2] = 9$, and $E[XY] = 6$. Verify the Cauchy-Schwarz inequality and determine if equality holds.*

Solution: Compute $|E[XY]|$ and $\sqrt{E[X^2]E[Y^2]}$:

$$|E[XY]| = |6| = 6$$

$$\sqrt{E[X^2]E[Y^2]} = \sqrt{4 \times 9} = \sqrt{36} = 6$$

Since $|E[XY]| = \sqrt{E[X^2]E[Y^2]}$, equality holds, implying that X and Y are linearly dependent.
□

8 Stochastic Processes

8.1 Definition

A stochastic process $\{X(t)|t \in T\}$ is a collection of random variables indexed by time t, where T can be discrete or continuous. It models systems or phenomena that evolve over time under uncertainty.

8.2 Examples

- Random Walk
- Poisson Process
- Gaussian Process
- Stock Prices

8.3 Properties

- **Stationarity**:
  - **Strict Sense**: All finite-dimensional distributions are invariant under time shifts.
  - **Wide Sense (Weak Stationarity)**:
    - $E[X(t)] = \mu$ (constant)
    - $\text{Cov}(X(t), X(t+\tau))$ depends only on τ.
- **Ergodicity**: Time averages equal ensemble averages.
- **Markov Property**: Future states depend only on the present state, not on the history.

8.4 Mean and Variance Functions

Mean:

$$m(t) = E[X(t)]$$

Variance:

$$\text{Var}(X(t)) = E[X(t)^2] - (E[X(t)])^2$$

8.5 Joint Distributions

For $X(t_1), X(t_2), \ldots, X(t_n)$:

$$f_{X(t_1),X(t_2),\ldots,X(t_n)}(x_1, x_2, \ldots, x_n)$$

8.6 Examples of Stochastic Processes

8.6.1 Example 1: Linear Stochastic Process

Example 10. *Consider the stochastic process:*

$$X(t) = At + b$$

where:

- *A is a Gaussian RV, $A \sim \mathcal{N}(0,1)$*
- *b is a constant*

Find the PDF of $X(t)$, and compute its mean and variance.

Solution: Since $A \sim \mathcal{N}(0,1)$, $X(t)$ is also normally distributed:

$$X(t) \sim \mathcal{N}(b, t^2)$$

Thus, the PDF of $X(t)$ is:

$$f_X(x,t) = \frac{1}{\sqrt{2\pi t^2}} \exp\left(-\frac{(x-b)^2}{2t^2}\right)$$

Mean and Variance:

$$E[X(t)] = b$$
$$\mathrm{Var}(X(t)) = t^2$$

□

8.6.2 Example 2: Oscillatory Stochastic Process

Example 11. *Consider the stochastic process:*

$$X(t) = a\cos(\omega_0 t) + \phi$$

where:

- *a is a constant*
- *ϕ is uniformly distributed in $(0, 2\pi]$*

Find the PDF of $X(t)$, and compute its mean and variance.

Solution: Since ϕ is uniformly distributed in $(0, 2\pi]$, $X(t)$ oscillates between $a\cos(\omega_0 t)$ and $a\cos(\omega_0 t) + 2\pi$.

Assuming ϕ is uniformly distributed over $(0, 2\pi)$ and considering the distribution of $X(t)$:

$$f_X(x,t) = \frac{1}{2\pi} \quad \text{for } x \in (a\cos(\omega_0 t), a\cos(\omega_0 t) + 2\pi)$$

Mean and Variance:

$$E[X(t)] = a\cos(\omega_0 t) + E\{\phi\} = a\cos(\omega_0 t) + \pi$$

$$\mathrm{Var}\{X(t)\} = \mathrm{Var}\{\phi\} = \frac{(2\pi)^2}{12} = \frac{\pi^2}{3}$$

□

9 Additional Formulas

9.1 Moment Generating Function (MGF)

Definition:

$$M_X(t) = E[e^{tX}]$$

Properties:

- **Linearity**: For constants a, b,

$$M_{aX+b}(t) = e^{bt} M_X(at)$$

- **Independence**: If X and Y are independent,

$$M_{X+Y}(t) = M_X(t) M_Y(t)$$

- **Derivatives**: The n-th derivative of $M_X(t)$ evaluated at $t = 0$ gives the n-th moment:

$$M_X^{(n)}(0) = E[X^n]$$

9.2 Characteristic Function

Definition:

$$\phi_X(\omega) = E[e^{i\omega X}]$$

where i is the imaginary unit and ω is a real number.

Properties:

- **Always Exists**: For any RV X, the characteristic function exists.
- **Uniqueness**: The characteristic function uniquely determines the distribution of X.
- **Independence**: If X and Y are independent,

$$\phi_{X+Y}(\omega) = \phi_X(\omega)\phi_Y(\omega)$$

- **Relation to MGF**:

$$\phi_X(\omega) = M_X(i\omega)$$

9.2.1 Characteristic Function of Normal Distribution

Example 12. *For $X \sim \mathcal{N}(\mu, \sigma^2)$, find the characteristic function $\phi_X(\omega)$.*

Solution: The characteristic function of a normal distribution is:

$$\phi_X(\omega) = e^{i\omega\mu - \frac{1}{2}\omega^2\sigma^2}$$

This formula shows that the characteristic function of a normal distribution is an exponential function involving the mean μ and variance σ^2. □

9.3 Generating Functions Comparison

MGF vs. Characteristic Function:

- **MGF**:
$$M_X(t) = E[e^{tX}]$$
  - Defined for real t.
  - May not exist for all t depending on the distribution.
- **Characteristic Function**:
$$\phi_X(\omega) = E[e^{i\omega X}]$$
  - Defined for all real ω.
  - Always exists for any RV X.

Relation Between MGF and Characteristic Function:

$$\phi_X(\omega) = M_X(i\omega)$$

This shows that the characteristic function is essentially the MGF evaluated at an imaginary argument.

9.4 Covariance Matrix (for Multivariate RVs)

Definition:
$$\Sigma = \begin{pmatrix} \mathrm{Var}(X) & \mathrm{Cov}(X,Y) \\ \mathrm{Cov}(Y,X) & \mathrm{Var}(Y) \end{pmatrix}$$

9.5 Transformation of a Single RV

If $Y = g(X)$, and g is monotonic:

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|$$

10 Key Concepts

10.1 Stationary Process

- **Strict Stationarity**: All joint distributions are invariant under time shifts.
- **Wide Sense Stationarity**: Mean is constant and covariance depends only on time difference.

10.2 Ergodicity

Time averages equal ensemble averages.

10.3 Markov Property

Future states depend only on the present state, not on the history.

11 Quick Reference: Common Distributions

Distribution	PMF/PDF	Parameters
Uniform	$f_X(x) = \frac{1}{U-L}$ for $L \leq x \leq U$	Lower L, Upper U
Gaussian	$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	Mean μ, Std Dev σ
Binomial	$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$	Trials n, Success p
Poisson	$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$	Rate λ
Exponential	$f_X(x) = \lambda e^{-\lambda x}$ for $x \geq 0$	Rate λ

Example: Quick Reference Usage

Example 13. *Identify the distribution and its parameters for the following scenario: A machine produces widgets with the number of defects following a Poisson distribution with an average rate of 3 defects per hour.*

Solution: The number of defects follows a Poisson distribution with parameter $\lambda = 3$. □

12 Conclusion

These lecture notes cover the fundamental concepts of Chapter 1 in Stochastic Processes, including random variables, probability distributions, expectation, covariance, important theorems, and an introduction to stochastic processes. For detailed explanations and further readings, please refer to your course materials or standard textbooks.