Contents

Chapter 1

Random Variables and Distributions

1.1 Probability Primer (quick but complete)

Definition

A probability space is $(\Omega, \mathcal{F}, \mathbb{P})$ where $\mathcal{F}$ is a σ-algebra on Ω and $\mathbb{P} : \mathcal{F} \to [0,1]$ is countably additive with $\mathbb{P}(\Omega) = 1$.

Basic rules.

$$\mathbb{P}(A^c) = 1 - \mathbb{P}(A), \quad \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B), \quad \mathbb{P}\Big(\bigcup_i A_i\Big) \leq \sum_i \mathbb{P}(A_i).$$

Proposition

Total probability and Bayes. If $(B_i)_i$ is a partition with $\mathbb{P}(B_i) > 0$,

$$\mathbb{P}(A) = \sum_i \mathbb{P}(A \mid B_i)\mathbb{P}(B_i), \qquad \mathbb{P}(B_k \mid A) = \frac{\mathbb{P}(A \mid B_k)\mathbb{P}(B_k)}{\sum_i \mathbb{P}(A \mid B_i)\mathbb{P}(B_i)}.$$

1.2 Random Variables, CDF, PDF/PMF, Support

Definition

A real random variable is a measurable $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$. Its distribution is the pushforward measure $\mathbb{P}_X(B) = \mathbb{P}(X \in B)$. The **CDF** is $F_X(x) = \mathbb{P}(X \leq x)$.

Properties of the CDF: nondecreasing, right-continuous; $F_X(-\infty) = 0$, $F_X(\infty) = 1$. If F_X is (a.e.) differentiable, the **PDF** is $f_X(x) = F_X'(x) \geq 0$ with $\int_{\mathbb{R}} f_X = 1$. For discrete X, the **PMF** is $p_X(x) = \mathbb{P}(X = x)$ with $\sum_x p_X(x) = 1$.

Proposition

Expectation (LOTUS) and variance. For measurable g,

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}} g(x)\, dF_X(x) = \begin{cases} \sum_x g(x)\, p_X(x), & \text{discrete,} \\ \int g(x)\, f_X(x)\, dx, & \text{continuous.} \end{cases}$$

Mean $\mu = \mathbb{E}[X]$, variance $\mathrm{Var}(X) = \mathbb{E}[(X-\mu)^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$.

Proposition

Core algebra. For $a, b \in \mathbb{R}$:

$$\mathbb{E}[aX+b] = a\,\mathbb{E}[X] + b, \qquad \mathrm{Var}(aX+b) = a^2\,\mathrm{Var}(X).$$

If $X \perp\!\!\!\perp Y$, then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ and $\mathrm{Var}(X+Y) = \mathrm{Var}\,X + \mathrm{Var}\,Y$.

1.3 Most-Used Distributions (pdf/cdf/mean/variance)

Discrete

Bernoulli (p)

$\mathbb{P}(X=1) = p$, $\mathbb{P}(X=0) = 1-p$; $\quad \mathbb{E}[X] = p$, $\mathrm{Var}(X) = p(1-p)$.

Binomial (n, p)

$\mathbb{P}(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$, $k = 0, \ldots, n$; $\quad \mathbb{E}[X] = np$, $\mathrm{Var}(X) = np(1-p)$.

Geometric (p) (trials until first success)

$\mathbb{P}(X=k) = (1-p)^{k-1} p$, $k = 1, 2, \ldots$; $\quad \mathbb{E}[X] = 1/p$, $\mathrm{Var}(X) = (1-p)/p^2$; *memoryless.*

Poisson (λ)

$\mathbb{P}(X=k) = e^{-\lambda}\lambda^k/k!$, $k = 0, 1, 2, \ldots$; $\quad \mathbb{E}[X] = \mathrm{Var}(X) = \lambda$.

Continuous

Uniform (a, b)

$$f(x) = \frac{1}{b-a} \text{ on } (a,b); \quad F(x) = \frac{x-a}{b-a} \text{ on } (a,b); \quad \mathbb{E}[X] = \frac{a+b}{2}, \ \mathrm{Var}(X) = \frac{(b-a)^2}{12}.$$

Exponential (λ)

$f(x) = \lambda e^{-\lambda x}$ for $x > 0$; $\quad F(x) = 1 - e^{-\lambda x}$; $\quad \mathbb{E}[X] = 1/\lambda$, $\mathrm{Var}(X) = 1/\lambda^2$; *memoryless.*

Gamma (k, θ) (shape–scale)

$$f(x) = \frac{x^{k-1} e^{-x/\theta}}{\Gamma(k)\theta^k} \text{ for } x > 0; \quad \mathbb{E}[X] = k\theta, \ \mathrm{Var}(X) = k\theta^2.$$

Normal (μ, σ^2)

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\Big(-\frac{(x-\mu)^2}{2\sigma^2}\Big); \quad \mathbb{E}[X] = \mu, \ \mathrm{Var}(X) = \sigma^2; \quad F(x) = \Phi\big(\frac{x-\mu}{\sigma}\big).$$

Beta (α, β) on $(0,1)$

$$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}; \quad \mathbb{E}[X] = \frac{\alpha}{\alpha+\beta}, \ \mathrm{Var}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

Chi-square (k)

If $Z_i \sim \mathcal{N}(0,1)$ i.i.d., then $\sum_{i=1}^k Z_i^2 \sim \chi^2(k)$; $\quad \mathbb{E}[X] = k$, $\mathrm{Var}(X) = 2k$.

Useful relationships.

- Binomial(n, p) with $n \to \infty$, $p \to 0$, $np \to \lambda \Rightarrow$ Poisson(λ) (Poisson limit).
- Sums: independent Poissons add; independent Gammas with same scale add; Normal sums stay Normal.
- Affine change: if X has density f_X, then $Y = aX + b$ has $f_Y(y) = \frac{1}{|a|} f_X\big(\frac{y-b}{a}\big)$ and $\mathbb{E}[Y] = a\,\mathbb{E}[X] + b$, $\mathrm{Var}(Y) = a^2\,\mathrm{Var}(X)$.

1.4 Joint Distributions, Marginals, Conditionals, Independence

Definition

For (X, Y) continuous, the joint density $f_{X,Y}$ satisfies $f_{X,Y} \geq 0$ and $\iint f_{X,Y} = 1$. **Marginals**:

$$f_X(x) = \int f_{X,Y}(x, y)\, dy, \qquad f_Y(y) = \int f_{X,Y}(x, y)\, dx.$$

For discrete variables, replace integrals by sums.

Definition

Conditional distribution. If $f_X(x) > 0$,

$$f_{Y|X}(y \mid x) = \frac{f_{X,Y}(x, y)}{f_X(x)}, \qquad \mathbb{E}[Y \mid X = x] = \int y\, f_{Y|X}(y \mid x)\, dy.$$

Definition

Independence. $X \perp\!\!\!\perp Y$ iff $f_{X,Y}(x, y) = f_X(x)f_Y(y)$ (continuous) or $p_{X,Y}(x, y) = p_X(x)p_Y(y)$ (discrete).

Proposition

Total expectation and total variance.

$$\mathbb{E}[X] = \mathbb{E}\big[\mathbb{E}[X \mid Y]\big], \qquad \mathrm{Var}(X) = \mathbb{E}[\,\mathrm{Var}(X \mid Y)] + \mathrm{Var}(\mathbb{E}[X \mid Y]).$$

Probabilities involving several variables. For independent continuous X, Y,

$$\mathbb{P}(X < Y) = \int_{-\infty}^{\infty} f_X(x)\, F_Y(x)\, dx = \int f_Y(y)\, [1 - F_X(y)]\, dy.$$

More generally, $\mathbb{P}((X, Y) \in A) = \iint_A f_{X,Y}(x, y)\, dx\, dy$.

1.5 Distributions of Functions of Random Variables

One variable (transport theorem)

Theorem

If $Y = \phi(X)$ with ϕ continuously differentiable and strictly monotone, then

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

Many-to-one case. If ϕ is m-to-1 on relevant parts of the support,

$$f_Y(y) = \sum_{i=1}^{m} f_X(x_i(y)) \left|\frac{d}{dy} x_i(y)\right| \quad \text{where } \phi(x_i(y)) = y.$$

Example

$X \sim \text{Unif}(-1,1)$, $Y = X^2$. For $y \in (0,1)$, the inverse branches are $x_\pm(y) = \pm\sqrt{y}$, so

$$f_Y(y) = f_X(\sqrt{y})\frac{1}{2\sqrt{y}} + f_X(-\sqrt{y})\frac{1}{2\sqrt{y}} = \frac{1}{2}\cdot\frac{1}{2\sqrt{y}} + \frac{1}{2}\cdot\frac{1}{2\sqrt{y}} = \frac{1}{2\sqrt{y}}.$$

Two variables (Jacobian)

Theorem

If $(U,V) = T(X,Y)$ is a C^1 bijection with inverse $(x(u,v), y(u,v))$ and Jacobian $J_{T^{-1}}$, then

$$f_{U,V}(u,v) = f_{X,Y}(x(u,v), y(u,v)) \left|\det J_{T^{-1}}(u,v)\right|.$$

Example

Polar coordinates. $X = R\cos\Theta$, $Y = R\sin\Theta$ has $|\det J| = r$. If $X, Y \sim \mathcal{N}(0,1)$ i.i.d., then

$$f_{R,\Theta}(r,\theta) = \frac{1}{2\pi} e^{-r^2/2}\, r,$$

so R and Θ are independent, R is Rayleigh ($re^{-r^2/2}$ for $r > 0$), and $\Theta \sim \text{Unif}(0, 2\pi)$.

Common constructions.

- *Sum (convolution)*: $Z = X + Y$ (indep. continuous) has $f_Z(z) = \int f_X(x) f_Y(z-x)\,dx$. In transform language, MGFs/CFs multiply.
- *Max/min (i.i.d. continuous)*: $F_{\max}(x) = [F(x)]^n$, $F_{\min}(x) = 1 - [1 - F(x)]^n$.
- *Ratio/scaling*: If $X > 0$ and $Y = cX$, then $f_Y(y) = \frac{1}{|c|} f_X(y/c)$. Ratios like $X/(X+Y)$ often lead to Beta laws (see Example 3 below).

1.6 Worked Examples

Worked Solution

Example 1 (sum of uniforms). $X, Y \sim \text{Unif}(0,1)$ independent; $Z = X + Y$. Then

$$f_Z(z) = \int_0^1 \mathbf{1}_{0<x<1}\, \mathbf{1}_{0<z-x<1}\, dx = \begin{cases} z, & 0 < z < 1, \\ 2 - z, & 1 \le z < 2, \\ 0, & \text{else.} \end{cases}$$

Worked Solution

Example 2 (who is larger?). $X \sim \mathrm{Exp}(\lambda)$, $Y \sim \mathrm{Exp}(\mu)$ independent. Then

$$\mathbb{P}(X < Y) = \int_0^\infty f_X(x)\mathbb{P}(Y > x)\,dx = \int_0^\infty \lambda e^{-\lambda x}\, e^{-\mu x}\,dx = \frac{\lambda}{\lambda+\mu}.$$

Worked Solution

Example 3 (Gamma–Beta trick). $X \sim \Gamma(\alpha,\theta)$, $Y \sim \Gamma(\beta,\theta)$ independent. Let $U = X+Y$, $V = \frac{X}{X+Y}$. The Jacobian of $(x,y) \mapsto (u,v)$ is $|J| = u$. Then

$$f_{U,V}(u,v) = \frac{u^{\alpha+\beta-1}e^{-u/\theta}}{\Gamma(\alpha)\Gamma(\beta)\theta^{\alpha+\beta}}\, v^{\alpha-1}(1-v)^{\beta-1},$$

which factorizes: $U \sim \Gamma(\alpha+\beta,\theta)$ and $V \sim \mathrm{Beta}(\alpha,\beta)$, *independent.*

Worked Solution

Example 4 (absolute value). $X \sim \mathcal{N}(0,1)$, $Y = |X|$. Many-to-one transform with branches $\pm y$ yields

$$f_Y(y) = f_X(y) + f_X(-y) = \frac{2}{\sqrt{2\pi}}e^{-y^2/2}, \quad y > 0.$$

1.7 Practice: Compute and Reason (with solutions)

Exercises

E1. (Basics) Show that F_X is right-continuous and $\lim_{x\to-\infty} F_X(x) = 0$, $\lim_{x\to\infty} F_X(x) = 1$.

E2. (Expectation) If $X \sim \mathrm{Unif}(a,b)$, compute $\mathbb{E}[X^k]$ for $k \in \mathbb{N}$.

E3. (Joint $\to$ marginal/conditional) $f_{X,Y}(x,y) = c(x+y)$ on $0 < x < 1$, $0 < y < 1$. Find c, f_X, f_Y, $\mathbb{E}[X]$, $\mathbb{E}[Y]$, $\mathrm{Cov}(X,Y)$, and $f_{Y|X}$.

E4. (Convolution) $X, Y \sim \mathrm{Exp}(\lambda)$ independent. Find the pdf of $Z = X+Y$ and its mean/-variance.

E5. (Transport) $X \sim \mathrm{Unif}(-1,1)$ and $Y = X^2$. Find F_Y and f_Y.

E6. (Who is larger?) $X \sim \mathrm{Exp}(2)$, $Y \sim \mathrm{Exp}(3)$ independent. Compute $\mathbb{P}(X < Y)$.

E7. (Order stats) $U_i \sim \mathrm{Unif}(0,1)$ i.i.d. Find $F_{\max}$ and $\mathbb{E}[\max_i U_i]$.

Solution sketches

Worked Solution

(1) From measure properties of distribution functions.

(2) $\mathbb{E}[X^k] = \dfrac{1}{b-a}\int_a^b x^k dx = \dfrac{b^{k+1}-a^{k+1}}{(k+1)(b-a)}$.

(3) $1 = \int_0^1\int_0^1 c(x+y)dydx \Rightarrow c = 1$. $f_X(x) = x + \frac{1}{2}$, $f_Y(y) = y + \frac{1}{2}$. $\mathbb{E}[X] = \mathbb{E}[Y] = 7/12$, $\mathbb{E}[XY] = 1/4$, Cov $= -13/144$. $f_{Y|X}(y \mid x) = \frac{x+y}{x+1/2}$.

(4) $Z \sim \Gamma(2, 1/\lambda)$ with $f_Z(z) = \lambda^2 z e^{-\lambda z}$ $(z > 0)$, $\mathbb{E}[Z] = 2/\lambda$, $\mathrm{Var}(Z) = 2/\lambda^2$.

(5) $F_Y(y) = \mathbb{P}(|X| \leq \sqrt{y}) = \frac{\sqrt{y}-(-\sqrt{y})}{2} = \sqrt{y}$ on $(0,1)$; $f_Y(y) = \frac{1}{2\sqrt{y}}$.

(6) $\mathbb{P}(X < Y) = \dfrac{2}{2+3} = \dfrac{2}{5}$.

(7) $F_{\max}(x) = x^n$ on $(0,1)$, so $f_{\max}(x) = nx^{n-1}$ and $\mathbb{E}[\max] = n/(n+1)$.

What to master from this chapter: how to read and use CDF/PDF/PMF; the shape and parameters of core distributions; marginalization and conditioning; independence; computing probabilities involving several variables; and transforming variables (1D transport and 2D Jacobians).

Chapter 2

Distributions and the Transport Theorem

2.1 Random Variables and Distributions

Definition

A **random variable** is a measurable function $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$. Its **distribution** is the measure $\mathbb{P}_X(B) = \mathbb{P}(X \in B)$ for all $B \in \mathcal{B}$.

2.1.1 CDF and PDF

The cumulative distribution function (CDF) is

$$F_X(x) = \mathbb{P}(X \leq x),$$

with properties F_X nondecreasing, right-continuous, $F_X(-\infty) = 0$, $F_X(\infty) = 1$. If F_X is differentiable, its derivative $f_X(x) = F_X'(x)$ is the **probability density function (PDF)**.

For a discrete variable, $p_X(x) = \mathbb{P}(X = x)$ and $F_X(x) = \sum_{t \leq x} p_X(t)$.

2.1.2 Expectation and Variance

For an integrable function g, **LOTUS** (Law of the Unconscious Statistician) states:

$$\mathbb{E}[g(X)] = \int g(x) f_X(x) dx,$$

for continuous X, or $\sum_x g(x) p_X(x)$ for discrete X.

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

2.2 The Transport Theorem (Change of Variables)

Theorem

Let X be a random variable with PDF f_X and let $Y = \phi(X)$ be a continuously differentiable, monotone function. Then the PDF of Y is

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

Sketch. If ϕ is strictly increasing, $\mathbb{P}(Y \le y) = \mathbb{P}(X \le \phi^{-1}(y)) = F_X(\phi^{-1}(y))$. Differentiating both sides gives the formula. □

Example

Linear Transformations: If $Y = aX + b$ with $a \neq 0$, then

$$F_Y(y) = F_X\left(\frac{y-b}{a}\right), \quad f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right).$$

Example

Polynomial Transformation: Let $X \sim \mathrm{Unif}(0,1)$ and $Y = X^2$. Then $F_Y(y) = P(Y \le y) = P(X \le \sqrt{y}) = \sqrt{y}$ for $0 < y < 1$. Hence $f_Y(y) = \frac{1}{2\sqrt{y}}$.

Example

Trigonometric Transformation: If $\Theta \sim \mathrm{Unif}(0, 2\pi)$ and $X = \cos\Theta$, then the PDF of X is

$$f_X(x) = \frac{1}{\pi\sqrt{1-x^2}}, \quad x \in (-1, 1).$$

2.3 Multivariate Transport Theorem

Theorem

If (X, Y) is a continuous random vector with joint PDF $f_{X,Y}$ and $(U, V) = T(X, Y)$ is a one-to-one, differentiable transformation with inverse T^{-1}, then

$$f_{U,V}(u, v) = f_{X,Y}(x(u, v), y(u, v)) \left| \det J_{T^{-1}}(u, v) \right|,$$

where $J_{T^{-1}}$ is the Jacobian matrix of T^{-1}.

Example

Polar Transformation: If (X,Y) has joint PDF $f_{X,Y}(x,y)$, define $R=\sqrt{X^2+Y^2}$, $\Theta=\arctan(Y/X)$. Then

$$|\det J| = r, \quad f_{R,\Theta}(r,\theta) = f_{X,Y}(r\cos\theta, r\sin\theta)\, r.$$

If $(X,Y)\sim\mathcal{N}(0,1)$ i.i.d., then $f_{R,\Theta}(r,\theta)=\frac{1}{2\pi}e^{-r^2/2}r$, so R and Θ are independent.

2.4 Worked Examples

Worked Solution

Example 1 — Absolute Value: Let $X\sim\mathcal{N}(0,1)$, $Y=|X|$. Find f_Y.

$$f_Y(y) = f_X(y) + f_X(-y) = \frac{2}{\sqrt{2\pi}}e^{-y^2/2},\ y>0.$$

Worked Solution

Example 2 — Exponential Scaling: Let $X\sim\mathrm{Exp}(1)$ and $Y=2X$. Then $f_Y(y)=\frac{1}{2}e^{-y/2}$ for $y>0$.

Worked Solution

Example 3 — Sum of Two Uniforms: If $X,Y\sim\mathrm{Unif}(0,1)$ independent, then $Z=X+Y$ has PDF

$$f_Z(z)=\begin{cases} z, & 0<z<1,\\ 2-z, & 1\le z<2,\\ 0, & \text{otherwise.}\end{cases}$$

2.5 Exercises

E1. $X\sim\mathrm{Unif}(-1,1)$. Find the PDF of $Y=X^2$.

E2. If $X\sim\mathrm{Exp}(\lambda)$, find the distribution of $Y=aX+b$.

E3. Let $X\sim\mathrm{Unif}(0,\pi)$, $Y=\sin X$. Compute f_Y.

E4. If (X,Y) has joint PDF $f(x,y)=e^{-(x+y)}$ for $x,y>0$, find the PDF of $(U,V)=(X+Y, X/(X+Y))$.

E5. (Challenge) Derive the CDF and PDF of the maximum of n i.i.d. $\mathrm{Unif}(0,1)$.

Worked Solution

Solutions Sketch: (1) $Y \in (0,1)$, $f_Y(y) = \frac{1}{2\sqrt{y}}$. (2) Linear transform formula. (3) $f_Y(y) = \frac{1}{\pi\sqrt{1-y^2}}$ for $y \in (0,1)$. (4) Use Jacobian $|J| = u$. (5) $f_{X_{(n)}}(x) = nx^{n-1}$.

Chapter 3

Expectation and Variance

3.1 Definition of Expectation

Definition

For a real random variable X with CDF F_X, the **expectation** of a measurable function $g(X)$ is

$$\mathbb{E}[g(X)] = \begin{cases} \displaystyle\sum_x g(x)\, p_X(x), & \text{if } X \text{ is discrete,} \\ \displaystyle\int_{-\infty}^{\infty} g(x)\, f_X(x)\, dx, & \text{if } X \text{ is continuous.} \end{cases}$$

Proposition

(LOTUS — Law of the Unconscious Statistician) For any measurable g,

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}} g(x)\, dF_X(x).$$

Example

If $X \sim \text{Unif}(0,1)$, then $\mathbb{E}[X^2] = \int_0^1 x^2 dx = \frac{1}{3}$.

3.2 Basic Properties

Proposition

Let $a, b \in \mathbb{R}$. Then:

$$\mathbb{E}[aX + b] = a\mathbb{E}[X] + b, \qquad \operatorname{Var}(aX + b) = a^2 \operatorname{Var}(X).$$

If X, Y are independent: $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$, hence $\operatorname{Var}(X + Y) = \operatorname{Var} X + \operatorname{Var} Y$.

Definition

The **variance** of X is $\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$. It measures dispersion around the mean.

Proposition

Useful identities:

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

The **correlation** is $\rho(X,Y) = \mathrm{Cov}(X,Y)/(\sigma_X \sigma_Y) \in [-1,1]$.

Example

If $X \sim \mathrm{Exp}(\lambda)$, then $\mathbb{E}[X] = 1/\lambda$ and $\mathrm{Var}(X) = 1/\lambda^2$.

3.3 Conditional Expectation

Definition

The **conditional expectation** of Y given X is the random variable $\mathbb{E}[Y|X]$ satisfying

$$\mathbb{E}[g(X)\mathbb{E}[Y|X]] = \mathbb{E}[g(X)Y] \quad \text{for all measurable } g.$$

For discrete (X,Y), $\mathbb{E}[Y|X=x] = \sum_y y\,\mathbb{P}(Y=y|X=x)$. For continuous (X,Y), $\mathbb{E}[Y|X=x] = \int y\, f_{Y|X}(y|x)\,dy$.

Proposition

Tower Property: $\mathbb{E}[\mathbb{E}[Y|X]] = \mathbb{E}[Y]$. **Law of Total Variance:**

$$\mathrm{Var}(Y) = \mathbb{E}[\mathrm{Var}(Y|X)] + \mathrm{Var}(\mathbb{E}[Y|X]).$$

Example

If $X \sim \mathrm{Unif}(0,1)$ and $Y|X=x \sim \mathrm{Exp}(x)$, then

$$\mathbb{E}[Y|X=x] = 1/x, \quad \mathbb{E}[Y] = \int_0^1 \frac{1}{x}\,dx = \infty$$

(diverges — expectation does not always exist).

3.4 Key Inequalities

Proposition

Markov's Inequality. If $X \geq 0$ and $a > 0$, then

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

Proposition

Chebyshev's Inequality. For any X with finite variance and $t > 0$,

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq t) \leq \frac{\mathrm{Var}(X)}{t^2}.$$

Proposition

Jensen's Inequality. For convex ϕ,

$$\phi(\mathbb{E}[X]) \leq \mathbb{E}[\phi(X)].$$

Equality holds iff X is a.s. constant or ϕ is linear.

3.5 Worked Examples

Worked Solution

Example 1 — Expectation of a nonlinear transform. Let $X \sim \mathrm{Unif}(0,1)$, find $\mathbb{E}[\sqrt{X}]$ and $\mathrm{Var}(\sqrt{X})$.

$$\mathbb{E}[\sqrt{X}] = \int_0^1 x^{1/2} dx = \frac{2}{3}, \quad \mathbb{E}[X] = \frac{1}{2}, \quad \mathbb{E}[X^{1/2}] = \frac{2}{3}.$$

Then $\mathbb{E}[(\sqrt{X})^2] = \mathbb{E}[X] = 1/2$ and $\mathrm{Var}(\sqrt{X}) = 1/2 - (2/3)^2 = 1/18$.

Worked Solution

Example 2 — Covariance and independence. Let X, Y be independent $\mathrm{Unif}(0,1)$. Then $\mathrm{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 1/4 - 1/4 = 0$. Independence $\Rightarrow$ zero covariance.

Worked Solution

Example 3 — Conditional expectation as regression. Suppose (X,Y) have joint density $f_{X,Y}(x,y) = 2$ on $0 < y < x < 1$. Then $f_{Y|X}(y|x) = 1/x$ for $0 < y < x$, and $\mathbb{E}[Y|X = x] = x/2$. The regression line is $\mathbb{E}[Y|X] = X/2$.

Worked Solution

Example 4 — Chebyshev bound. If X has mean μ and variance σ^2, then $\mathbb{P}(|X-\mu| \geq 2\sigma) \leq 1/4$ (at most 25

3.6 Exercises

E1. Prove $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$ directly from the integral definition.

E2. For $X \sim \mathrm{Exp}(\lambda)$, compute $\mathbb{E}[X^2]$ and deduce $\mathrm{Var}(X)$.

E3. Let $X \sim \mathrm{Unif}(0,1)$, find $\mathbb{E}[\ln X]$ and $\mathrm{Var}(\ln X)$.

E4. Verify the law of total variance for the model $Y|X \sim \mathrm{Bernoulli}(X)$, $X \sim \mathrm{Unif}(0,1)$.

E5. Show that Jensen's inequality implies $\mathbb{E}[e^X] \geq e^{\mathbb{E}[X]}$ for all real X.

Worked Solution

Solution sketches: (1) Expand integrals linearly. (2) $\mathbb{E}[X^2] = 2/\lambda^2$, $\mathrm{Var} = 1/\lambda^2$. (3) $\mathbb{E}[\ln X] = \int_0^1 \ln x \, dx = -1$, $\mathbb{E}[(\ln X)^2] = 2$, $\mathrm{Var} = 1$. (4) $\mathrm{Var}(Y) = \mathbb{E}[X(1-X)] + \mathrm{Var}(X) = \frac{1}{6}$. (5) Follows from convexity of e^x.

Chapter 4

Convergence and Limit Theorems

4.1 Modes of Convergence

Definition

Let (X_n) be a sequence of random variables and X another random variable.

- **Almost sure (a.s.) convergence:** $X_n \to X$ a.s. if $\mathbb{P}(\lim_{n\to\infty} X_n = X) = 1$.
- **Convergence in probability:** $X_n \xrightarrow{p} X$ if for all $\varepsilon > 0$, $\mathbb{P}(|X_n - X| > \varepsilon) \to 0$.
- **Convergence in distribution:** $X_n \xrightarrow{d} X$ if $F_{X_n}(x) \to F_X(x)$ for all continuity points x of F_X.

Proposition

Implications.

$$X_n \xrightarrow{a.s.} X \;\Rightarrow\; X_n \xrightarrow{p} X \;\Rightarrow\; X_n \xrightarrow{d} X.$$

The converses are false in general.

Example

Let X_n be the average of n i.i.d. Bernoulli(1/2) variables. Then $X_n \xrightarrow{p} 1/2$ by the Law of Large Numbers (next section).

4.2 The Law of Large Numbers (LLN)

Theorem

Weak Law of Large Numbers (WLLN). Let $X_1, \ldots, X_n$ be i.i.d. with mean μ and finite variance σ^2. Then

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p} \mu.$$

Sketch. Since $\mathbb{E}[\bar{X}_n] = \mu$ and $\mathrm{Var}(\bar{X}_n) = \sigma^2/n$, Chebyshev's inequality gives $\mathbb{P}(|\bar{X}_n - \mu| > \varepsilon) \leq \sigma^2/(n\varepsilon^2) \to 0$. □

Theorem

Strong Law of Large Numbers (SLLN). If (X_i) are i.i.d. with $\mathbb{E}[|X_i|] < \infty$, then

$$\bar{X}_n \xrightarrow{a.s.} \mathbb{E}[X_1].$$

Interpretation: both LLNs formalize that empirical averages stabilize to their expected value.

4.3 Central Limit Theorem (CLT)

Theorem

Classical CLT. If $X_1, \ldots, X_n$ are i.i.d. with mean μ and variance $\sigma^2 > 0$, then

$$Z_n = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{d} \mathcal{N}(0,1).$$

Idea. Use moment generating functions. $M_{Z_n}(t) = \left(M_{(X_i-\mu)/\sigma}(t/\sqrt{n})\right)^n \to e^{t^2/2}$, the MGF of $\mathcal{N}(0,1)$. □

Example

If $X_i \sim \mathrm{Bernoulli}(p)$, $\bar{X}_n$ has mean p and variance $p(1-p)/n$. Then

$$\mathbb{P}(\bar{X}_n \leq x) \approx \Phi\left(\frac{x-p}{\sqrt{p(1-p)/n}}\right)$$

for large n, where Φ is the standard normal CDF.

4.4 Delta Method

Theorem

If $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$ and g is differentiable at θ_0, then

$$\sqrt{n}(g(\hat{\theta}_n) - g(\theta_0)) \xrightarrow{d} \mathcal{N}(0, [g'(\theta_0)]^2\sigma^2).$$

Example

If $\bar{X}_n$ estimates μ with variance σ^2/n, then $g(\bar{X}_n) = \bar{X}_n^2$ satisfies

$$\sqrt{n}(\bar{X}_n^2 - \mu^2) \xrightarrow{d} \mathcal{N}(0, 4\mu^2\sigma^2).$$

4.5 Slutsky's Theorem and Continuous Mapping

Proposition

Slutsky's Theorem. If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} c$ (constant), then:

$$X_n + Y_n \xrightarrow{d} X + c, \quad X_n Y_n \xrightarrow{d} cX, \quad X_n / Y_n \xrightarrow{d} X/c \ (c \neq 0).$$

Proposition

Continuous Mapping Theorem. If $X_n \xrightarrow{d} X$ and g is continuous, then $g(X_n) \xrightarrow{d} g(X)$.

4.6 Worked Examples

Worked Solution

Example 1 — LLN for Exponentials. If $X_i \sim \text{Exp}(1)$, then $\mathbb{E}[X_i] = 1$ and $\text{Var}(X_i) = 1$. By WLLN, $\bar{X}_n \xrightarrow{p} 1$. By CLT, $\sqrt{n}(\bar{X}_n - 1) \xrightarrow{d} \mathcal{N}(0, 1)$.

Worked Solution

Example 2 — Normal Approximation of Binomial. If $S_n \sim \text{Binomial}(n, p)$, then

$$\frac{S_n - np}{\sqrt{np(1-p)}} \xrightarrow{d} \mathcal{N}(0, 1).$$

Worked Solution

Example 3 — Delta Method. If $X_i \sim \text{Exp}(\lambda)$ and $\hat{\lambda} = 1/\bar{X}_n$, then $\sqrt{n}(\hat{\lambda} - \lambda) \xrightarrow{d} \mathcal{N}(0, \lambda^2)$. By the Delta Method, for $g(x) = 1/x$,

$$\sqrt{n}(g(\bar{X}_n) - g(1/\lambda)) \xrightarrow{d} \mathcal{N}(0, \lambda^2).$$

4.7 Exercises

E1. Let $X_1, \ldots, X_n$ i.i.d. with mean μ, variance σ^2. Prove $\mathbb{E}[\bar{X}_n] = \mu$, $\text{Var}(\bar{X}_n) = \sigma^2/n$.

E2. Verify the WLLN using Chebyshev's inequality for $X_i \sim \text{Exp}(1)$.

E3. Use the CLT to approximate $\mathbb{P}(S_{100} \geq 60)$ when $S_{100} \sim \text{Binomial}(100, 0.5)$.

E4. Apply the Delta Method to $g(x) = \ln x$ for $\bar{X}_n$ estimating $\mu > 0$.

E5. (Challenge) State a version of the multivariate CLT for $\mathbf{X}_i \in \mathbb{R}^d$.

Worked Solution

Solutions Sketch: (1) Direct from linearity. (2) $\mathbb{P}(|\bar{X}_n - 1| > \varepsilon) \leq 1/(n\varepsilon^2)$. (3) Mean $= 50$, SD $= \sqrt{25} = 5$, so $\mathbb{P}(S_{100} \geq 60) \approx 1 - \Phi(2) = 0.0228$. (4) $g'(\mu) = 1/\mu$, hence asymptotic variance $(\sigma^2/\mu^2 n)$. (5) Multivariate: $\sqrt{n}(\bar{\mathbf{X}}_n - \boldsymbol{\mu}) \xrightarrow{d} \mathcal{N}(0, \Sigma)$.

Worked Solution

Takeaways:

- LLN — averages converge to expectation (consistency).
- CLT — fluctuations around the mean are Gaussian.
- Delta Method — propagate asymptotic normality through smooth transformations.

Chapter 5

Summary and Formula Sheet

5.1 Probability Basics

Worked Solution

CDF:	$F_X(x) = \mathbb{P}(X \leq x)$, nondecreasing, right-continuous.
PDF:	$f_X(x) = F'_X(x)$, $f_X \geq 0$, $\int f_X = 1$.
Discrete:	$p_X(x) = \mathbb{P}(X = x)$, $\sum_x p_X(x) = 1$.
Expectation:	$\mathbb{E}[g(X)] = \int g(x)f_X(x)dx$ or $\sum g(x)p_X(x)$.
Variance:	$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$.
Covariance:	$\mathrm{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$.
Correlation:	$\rho = \mathrm{Cov}(X, Y)/(\sigma_X\sigma_Y) \in [-1, 1]$.

5.2 Transformation and Transport Theorem

Worked Solution

1D:	$f_Y(y) = f_X(\phi^{-1}(y))\,\lvert\frac{d}{dy}\phi^{-1}(y)\rvert$.
Many-to-one:	$f_Y(y) = \sum_i f_X(x_i)\,\lvert(\phi^{-1})'_i(y)\rvert$.
2D (Jacobian):	$f_{U,V}(u, v) = f_{X,Y}(x(u, v), y(u, v))\,\lvert\det J_{T^{-1}}\rvert$.
Example:	Polar coords: $X = R\cos\Theta$, $Y = R\sin\Theta$, $\lvert J\rvert = r$.

5.3 Expectation and Variance Properties

Worked Solution

Linearity	$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$
Scaling	$\mathrm{Var}(aX + b) = a^2 \mathrm{Var}(X)$
Sum of indep.	$\mathrm{Var}(X + Y) = \mathrm{Var}\, X + \mathrm{Var}\, Y$
Total Variance	$\mathrm{Var}(X) = \mathbb{E}[\mathrm{Var}(X\|Y)] + \mathrm{Var}(\mathbb{E}[X\|Y])$
Tower Property	$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X\|Y]]$

5.4 Inequalities

Worked Solution

Markov:	$\mathbb{P}(X \geq a) \leq \mathbb{E}[X]/a$ for $X \geq 0$.
Chebyshev:	$\mathbb{P}(\|X - \mu\| \geq t) \leq \mathrm{Var}(X)/t^2$.
Jensen:	$\phi(\mathbb{E}[X]) \leq \mathbb{E}[\phi(X)]$ for convex ϕ.
Cauchy–Schwarz:	$\|\mathbb{E}[XY]\| \leq \sqrt{\mathbb{E}[X^2]\mathbb{E}[Y^2]}$.

5.5 Modes of Convergence

Worked Solution

Almost sure	$\mathbb{P}(\lim X_n = X) = 1$
In probability	$\mathbb{P}(\|X_n - X\| > \varepsilon) \to 0$
In distribution	$F_{X_n}(x) \to F_X(x)$ (at continuity points)
Relation	a.s. $\Rightarrow$ in p $\Rightarrow$ in d

5.6 Limit Theorems

Worked Solution

WLLN:	$\bar{X}_n \xrightarrow{p} \mu$.
SLLN:	$\bar{X}_n \xrightarrow{a.s.} \mu$.
CLT:	$\sqrt{n}(\bar{X}_n - \mu)/\sigma \xrightarrow{d} \mathcal{N}(0, 1)$.
Delta Method:	$\sqrt{n}(g(\hat{\theta}_n) - g(\theta_0)) \to \mathcal{N}(0, [g'(\theta_0)]^2\sigma^2)$.
Slutsky:	If $X_n \to_d X$, $Y_n \to_p c$, then $X_n + Y_n \to_d X + c$.
Continuous Mapping:	$g(X_n) \to_d g(X)$ if g continuous.

5.7 Classical Distributions (Quick Reference)

Worked Solution

Distribution	**PDF / PMF**	**Mean, Variance**
hline Uniform(a, b)	$1/(b-a)$	$(a+b)/2$, $(b-a)^2/12$
Bernoulli(p)	$p^x(1-p)^{1-x}$	p, $p(1-p)$
Binomial(n, p)	$\binom{n}{k}p^k(1-p)^{n-k}$	np, $np(1-p)$
Poisson(λ)	$e^{-\lambda}\lambda^k/k!$	λ, λ
Exponential(λ)	$\lambda e^{-\lambda x}$ $(x>0)$	$1/\lambda$, $1/\lambda^2$
Normal(μ, σ^2)	$\frac{1}{\sqrt{2\pi\sigma^2}}e^{-(x-\mu)^2/2\sigma^2}$	μ, σ^2
Gamma(k, θ)	$x^{k-1}e^{-x/\theta}/(\Gamma(k)\theta^k)$	$k\theta$, $k\theta^2$
Beta(α, β)	$\frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}$	$\alpha/(\alpha+\beta)$, $\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

5.8 MGF and CF Reference

Worked Solution

MGF: $M_X(t) = \mathbb{E}[e^{tX}]$, when finite.
CF: $\varphi_X(t) = \mathbb{E}[e^{itX}]$, always exists.
MGF of Normal(μ, σ^2): $e^{\mu t + \frac{1}{2}\sigma^2 t^2}$.
MGF of Exponential(λ): $\frac{\lambda}{\lambda - t}$, $t < \lambda$.
MGF of Bernoulli(p): $(1-p) + pe^t$.
MGF of Poisson(λ): $\exp[\lambda(e^t - 1)]$.

5.9 Convergence Toolbox

Worked Solution

Chebyshev LLN: $\mathbb{P}(|\bar{X}_n - \mu| > \varepsilon) \le \sigma^2/(n\varepsilon^2)$.
MGF proof of CLT: $M_{Z_n}(t) \to e^{t^2/2}$.
Slutsky usage: Replace estimated variance by true one in test stats.
Delta method: Propagate asymptotic normality via derivative.

With these tables and the three core chapters mastered, you're fully equipped for the midterm — both theory and computation.