Advanced Probability

LMD3 Course Notes

General Information

- **Course:** Advanced Probability
- **Unit:** Fundamental
- **Credits:** 6
- **Coefficient:** 4
- **Prerequisites:** Real Analysis (1,2,3), Probability 1

Objectives

This course provides a detailed study of the main concepts and methods of probability, including:

- Probability of events
- Laws and moments of random variables
- Conditioning and regression
- Transforms of random variables
- Gaussian distributions

Course Outline

Motivation

The real world is full of uncertainty. Weather, measurement errors, financial markets, genetics, and even everyday decisions all involve randomness. **Probability theory** provides the mathematical framework we need to model and analyze this uncertainty. It allows us to make predictions, quantify risks, and build models that are used in fields as diverse as science, engineering, economics, and data science.

In this module, we will study the foundations of probability at an advanced level, with the goal of developing both intuition and rigorous mathematical tools.

1. Fundamental Review on Random Variables

Just like in algebra or calculus, we deal with variables that take specific values. In probability, however, we work with **random variables**, which do not take a fixed value, but instead have a probability of taking certain values.

Each random variable has important **numerical characteristics**, such as the mean and the standard deviation. These parameters already tell us a lot about the distribution.

We will also study several key distributions, each describing different kinds of events:

- **Bernoulli distribution:** models a system with two outcomes (e.g., heads/tails, success/failure).
- **Binomial distribution:** the sum of independent Bernoulli random variables. This distribution is very important and widely used.
- **Normal (Gaussian) distribution:** one of the most important distributions in probability and statistics, used in many fields.
- **Poisson distribution:** models the number of occurrences of rare events in a fixed time or space interval.
- **Exponential distribution:** describes waiting times between events in a Poisson process.

2. Functions of Random Variables

We will then study important functions of random variables:

- Expectation, variance, and standard deviation,
- Moment generating functions (MGFs),
- Characteristic functions.

3. Modes of Convergence

Different types of convergence for sequences of random variables will be introduced, and the relationships between them will be studied.

4. Limit Theorems

This part covers the cornerstone results of probability:

- Weak Law of Large Numbers (WLLN),
- Strong Law of Large Numbers (SLLN),
- Central Limit Theorem (CLT).

The CLT is particularly important: it tells us that if we take a large number of independent random variables and look at their average, then this average behaves approximately like a Gaussian random variable, regardless of the original distribution.

5. Random Vectors

Finally, we extend the theory to random vectors:

- Probability laws of random vectors,
- Numerical characteristics (expectation, covariance matrix),
- Moment and characteristic functions,
- Conditional expectation,
- Multivariate normal distribution,
- Convergence and the multivariate Central Limit Theorem.

Review of Basic Concepts

Sample Space: The set of all possible outcomes of an experiment is called the *sample space*, denoted by Ω.

Probability of an event:

$$P(A) = \frac{\text{number of ways } A \text{ can happen}}{\text{total number of possible outcomes}}$$

Examples:

- Coin flips: If I flip a coin twice,

  $$\Omega = \{HH, HT, TH, TT\}$$

  Let $A =$ "at least one head". Then $A = \{HH, HT, TH\}$ and $P(A) = 3/4$. Let $B =$ "no heads". Then $B = \{TT\}$ and $P(B) = 1/4$.

- Dice rolls: If I roll two dice, then

  $$\Omega = \{11, 12, \ldots, 66\}, \quad |\Omega| = 36$$

  Let $A =$ "at least one die is a 5". Then

  $$A = \{15, 25, 35, 45, 55, 65, 51, 52, 53, 54, 56\}, \quad |A| = 11,$$

  so $P(A) = 11/36$.

Axioms of Probability:

1. $0 \leq P(E) \leq 1$ for any event E,
2. $P(\Omega) = 1$,
3. For any sequence of mutually exclusive events $E_1, E_2, \ldots,$

$$P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} P(E_i), \quad n = 1, 2, \ldots$$

Conditional Probability and Independence

Conditional Probability: Given two events A and B with $P(B) > 0$, the conditional probability of A given B is defined as

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

Example: Suppose we roll a fair die. Let

$$A = \{\text{even number}\} = \{2, 4, 6\}, \quad B = \{\text{number greater than 3}\} = \{4, 5, 6\}.$$

Then

$$P(A \cap B) = P(\{4, 6\}) = \tfrac{2}{6}, \quad P(B) = \tfrac{3}{6},$$

so

$$P(A \mid B) = \frac{2/6}{3/6} = \tfrac{2}{3}.$$

—

Independence of Events

Definition: Two events A and B are said to be *independent* if

$$P(A \cap B) = P(A) \cdot P(B).$$

Example: Flip two fair coins. Let A = "first coin is heads" and B = "second coin is heads." Then

$$P(A) = P(B) = \tfrac{1}{2}, \quad P(A \cap B) = P(\{HH\}) = \tfrac{1}{4}.$$

Since $P(A \cap B) = (1/2)(1/2)$, the events A and B are independent.

Generalization: A collection of events $E_1, E_2, \ldots, E_n$ is called *mutually independent* if, for every subset $\{i_1, \ldots, i_k\} \subseteq \{1, \ldots, n\}$,

$$P(E_{i_1} \cap \cdots \cap E_{i_k}) = P(E_{i_1}) \cdots P(E_{i_k}).$$

Remark 0.1. Disjoint *means two events cannot occur together, i.e.,* $A \cap B = \varnothing$*.* Independent *means the occurrence of one does not affect the probability of the other; the events may still happen together.*

—

From Events to Random Variables

So far, we have described uncertainty in terms of events within a sample space and assigned probabilities to these events. However, in practice, we often want to associate *numbers* with outcomes in order to perform calculations, compute averages, and apply mathematical tools.

This brings us to the notion of a **random variable**: a measurable function that assigns a real number to each outcome of the experiment. In other words, random variables translate the abstract language of outcomes and events into the numerical framework of mathematics.

Example: If the experiment is flipping a coin once, the sample space is

$$\Omega = \{\text{Heads}, \text{Tails}\}.$$

Define a random variable X by

$$X(\text{Heads}) = 1, \quad X(\text{Tails}) = 0.$$

Here, X converts qualitative outcomes (Heads/Tails) into quantitative values (1/0) that we can analyze with probability tools.

—

Definition of a σ-algebra

A collection $\mathcal{F}$ of subsets of a sample space Ω is called a **σ-algebra** if:

1. $\Omega \in \mathcal{F}$,
2. If $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$,
3. If $A_1, A_2, \dots \in \mathcal{F}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

Chapter 1

Fundamental Review on Random Variables

1.1 Introduction

The world around us is full of uncertainty, and probability is the mathematical tool we use to describe and analyze it. We cannot predict events like weather, measurement errors, or financial markets with certainty, but we can model their likelihood. In this course, we will learn how to turn uncertainty into something we can measure and work with.

Probability theory provides the mathematical framework for modeling uncertainty. Random variables and their probability distributions constitute the central tools of this framework. In this chapter, we review the essential notions that will serve as the foundation for advanced probability: numerical characteristics of random variables, common probability laws, and operations involving random variables.

1.1.1 Random Variables and Probability Laws

Simple Intuitive Definition

A random variable is just a rule that assigns a number to every possible outcome of an experiment.

Example 1.1. *If the experiment is flipping a coin:*

$$X = \begin{cases} 1 & \text{if the outcome is Heads,} \\ 0 & \text{if the outcome is Tails.} \end{cases}$$

Here, X is a random variable because it turns outcomes (Heads/Tails) into numbers (1/0).

Definition 1.1 (Random Variable)**.** *A* random variable *is a measurable function*

$$X : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R},$$

where

- *Ω is the* ***sample space*** *(the set of all possible outcomes),*
- *$\mathcal{F}$ is a* ***σ-algebra*** *of events (the collection of subsets of Ω on which probabilities are defined),*

- $\mathbb{P}$ *is the* ***probability measure****.*

We denote by

$$(X = x) = \{\, \omega \in \Omega \ : \ X(\omega) = x \,\}$$

for simplicity.

- *If* X *takes only finitely or countably many values, it is called* ***discrete****.* ***Examples:***
  - *Flip a coin 4 times. The sample space is*

    $$\Omega = \{HHHH, HHHT, HHTH, \dots, TTTT\}.$$

    Define X *as the number of heads. Then*

    $$X \in \{0, 1, 2, 3, 4\},$$

    which is a finite set.
  - *Number of arrivals in a queue:*

    $$X \in \{0, 1, 2, 3, \dots\},$$

    which is countable.
- *If* X *takes values in an interval of* $\mathbb{R}$ *with a density, it is called* ***continuous****.* ***Example:*** *The height of students in a classroom can be modeled as a continuous random variable.*

Definition 1.2 (Probability Law)**.** *The* probability law *(or distribution) of a random variable* X *is the probability measure* $\mathbb{P}_X$ *defined on* $\mathbb{R}$ *by*

$$\mathbb{P}_X(B) = \mathbb{P}(X \in B), \quad B \in \mathcal{B}(\mathbb{R}),$$

where $\mathcal{B}(\mathbb{R})$ *denotes the Borel* σ*-algebra of* $\mathbb{R}$*.*

Equivalently:

- For a **discrete** random variable X, the law is described by its **probability mass function (pmf)**:

  $$p(x) = \mathbb{P}(X = x), \quad x \in \mathbb{R}.$$

- For a **continuous** random variable X, the law is described by its **probability density function (pdf)** $f(x)$, satisfying

  $$\mathbb{P}(a \le X \le b) = \int_a^b f(x)\, dx, \quad a < b.$$

Definition 1.3 (Cumulative Distribution Function)**.** *The* cumulative distribution function (cdf) *of* X*, denoted* F*, is defined for all* $x \in \mathbb{R}$ *by*

$$F(x) = \mathbb{P}(X \le x).$$

Thus, $F(x)$ *gives the probability that* X *takes a value less than or equal to* x*.*

Notation: We will write $X \sim F$ to mean that F is the distribution function of X. All probability questions about X can be answered in terms of its cdf F.

Remark:

- If X is discrete with pmf $p(x)$, then
$$F(a) = \sum_{x \le a} p(x).$$
- If X is continuous with pdf $f(x)$, then
$$F(a) = \int_{-\infty}^{a} f(x)\, dx.$$

1.2 Numerical Characteristics

Definition 1.4 (Expectation)**.** *The* expectation *(or mean) of a random variable X is defined by*
$$\mathbb{E}[X] = \sum_x x\, p(x), \quad \textit{(discrete case)},$$
$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f(x)\, dx, \quad \textit{(continuous case)},$$
provided the sum or integral is absolutely convergent.

Remark 1.1. *The expectation represents the "average value" or the "center of gravity" of the distribution. It is also called the* first moment*. Intuitively, it tells us the long-run average outcome if the experiment is repeated many times.*

However, expectation alone is not sufficient to fully describe a distribution. For example, two distributions can have the same expectation: one may be unimodal while another is bimodal. In such cases, we need an additional measure to capture the spread of the distribution—this is given by the variance *or the* standard deviation*.*

Definition 1.5 (Variance and Standard Deviation)**.** *The* variance *of X is*
$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2].$$
An equivalent formula is
$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.$$

The variance σ^2 is related to the second moment and measures the average squared deviation of X from its mean $\mu = \mathbb{E}[X]$. The standard deviation *is defined as*
$$\mathrm{SD}(X) = \sqrt{\mathrm{Var}(X)},$$
and represents the spread of the distribution in the same units as X.

Definition 1.6 (Higher Moments)**.** *The k-th (raw) moment of X is*
$$\mu'_k = \mathbb{E}[X^k].$$
The k-th central moment is
$$\mu_k = \mathbb{E}[(X - \mathbb{E}[X])^k].$$

Remark 1.2. *Taken together, higher-order moments serve as a kind of "fingerprint" of the distribution. Two important shape measures are:*

- ***Skewness:*** $\gamma_1 = \frac{\mu_3}{\sigma^3}$, *which measures the asymmetry of the distribution.*
- ***Kurtosis:*** $\gamma_2 = \frac{\mu_4}{\sigma^4}$, *which measures the concentration of mass and the heaviness of the tails.*

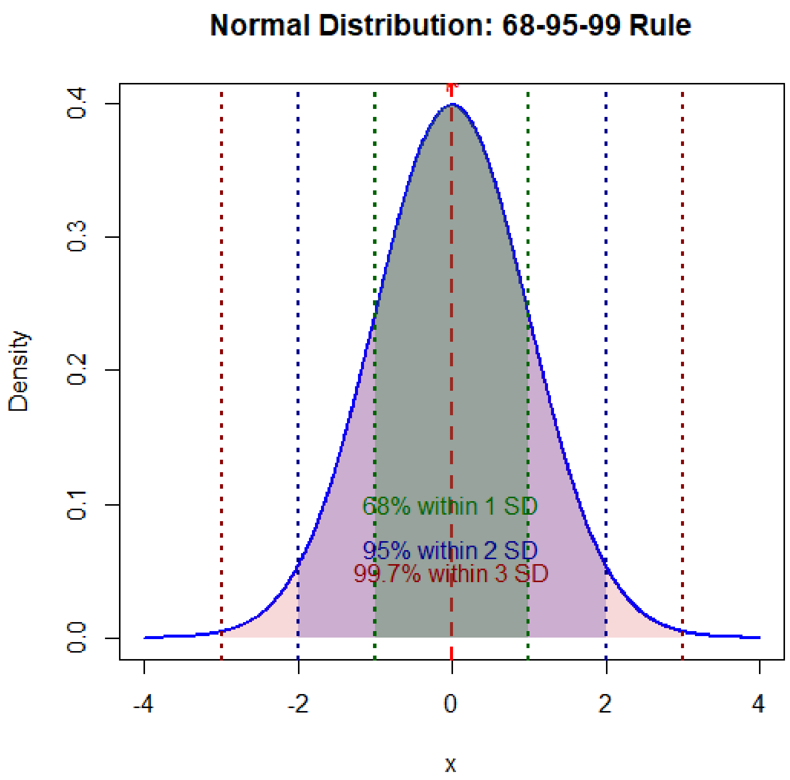

Figure 1.1: Normal distribution curve with mean μ and variance σ^2. Shaded areas (e.g., $\mu \pm \sigma$) correspond to probabilities such as $\mathbb{P}(\mu - \sigma \leq X \leq \mu + \sigma) \approx 0.68$, $\mathbb{P}(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \approx 0.95$, and $\mathbb{P}(\mu - 3\sigma \leq X \leq \mu + 3\sigma) \approx 0.997$.

Remark 1.3 (68–95–99 Rule for the Normal Distribution)**.** *Let* $X \sim \mathcal{N}(\mu, \sigma^2)$*. Then:*

$$\mathbb{P}(\mu - \sigma \leq X \leq \mu + \sigma) \approx 0.68,$$

$$\mathbb{P}(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \approx 0.95,$$

$$\mathbb{P}(\mu - 3\sigma \leq X \leq \mu + 3\sigma) \approx 0.997.$$

Example 1.2. *Suppose the heights of students are Normally distributed with mean* $\mu = 170$ *cm and standard deviation* $\sigma = 10$ *cm.*

- *About 68% of students are between* 160 *cm and* 180 *cm.*
- *About 95% of students are between* 150 *cm and* 190 *cm.*
- *Almost all (99.7%) students are between* 140 *cm and* 200 *cm.*

1.3 Main Probability Distributions

Discrete Distributions

- **Bernoulli distribution:** $X \sim \text{Bernoulli}(p)$, with

$$\mathbb{P}(X = 1) = p, \quad \mathbb{P}(X = 0) = 1 - p.$$

  Expectation: $\mathbb{E}[X] = p$, Variance: $\text{Var}(X) = p(1-p)$. A Bernoulli random variable X can take two values: 0 (event does not happen) or 1 (event happens). Examples: coin flip (Head = 1, Tail = 0), dice roll (six = 1, otherwise = 0).

- **Binomial distribution:** $X \sim \text{Bin}(n, p)$, with

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

  Expectation: $\mathbb{E}[X] = np$, Variance: $\text{Var}(X) = np(1-p)$. If n independent Bernoulli trials are performed $(B_1, \ldots, B_n)$, then $X = B_1 + \cdots + B_n \sim \text{Bin}(n, p)$.

- **Poisson distribution:** $X \sim \text{Poisson}(\lambda)$, with

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

  Expectation: $\mathbb{E}[X] = \lambda$, Variance: $\text{Var}(X) = \lambda$. The Poisson distribution can be obtained as an approximation of the Binomial distribution when n is large and p is small. It is extremely useful for modeling rare events such as accidents, catastrophes, or defective items in a factory.

- **Geometric distribution:** The probability that the first success occurs on the n-th trial is

$$\mathbb{P}(X = n) = (1-p)^{n-1} p, \quad n = 1, 2, \ldots$$

  Example: sequence of Bernoulli trials with success probability p.

Continuous Distributions

- **Uniform distribution:** $X \sim U(a, b)$, with

$$f(x) = \frac{1}{b-a}, \quad a \leq x \leq b.$$

  Expectation: $\mathbb{E}[X] = \frac{a+b}{2}$, Variance: $\text{Var}(X) = \frac{(b-a)^2}{12}$.

- **Exponential distribution:** $X \sim \text{Exp}(\lambda)$, with

$$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0.$$

  Expectation: $\mathbb{E}[X] = \frac{1}{\lambda}$, Variance: $\text{Var}(X) = \frac{1}{\lambda^2}$. Models waiting times, lifetimes, or times between rare events. Memoryless property: $\mathbb{P}(T > t + s \mid T > s) = \mathbb{P}(T > t)$.

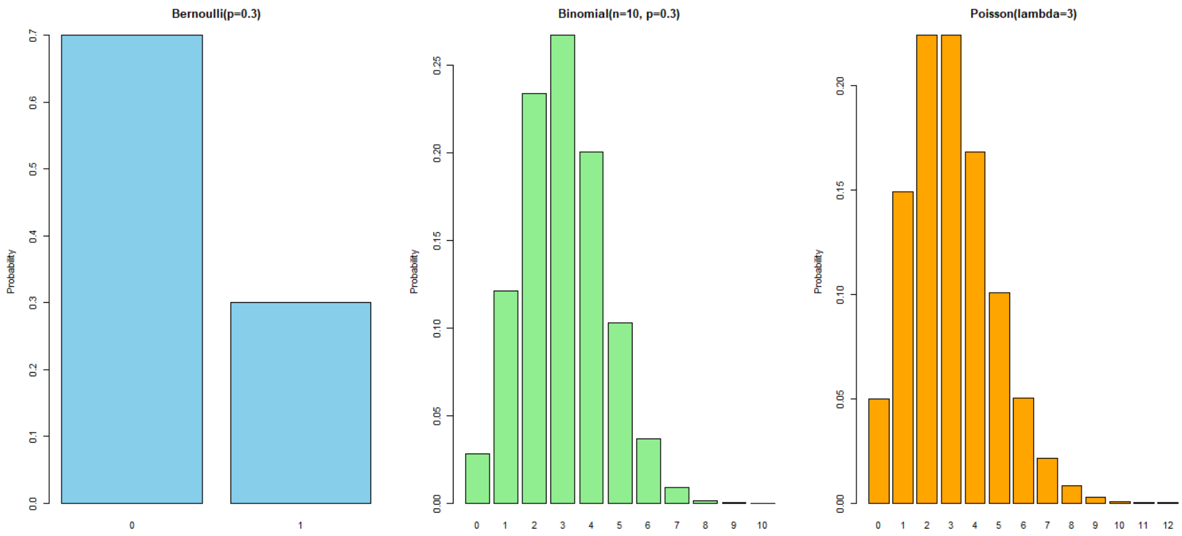

Figure 1.2: Discrete distributions

- **Gamma distribution:** $X \sim \text{Gamma}(r, \lambda)$, with

$$f(x) = \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x}, \quad x > 0,$$

  where $r > 0$ (shape) and $\lambda > 0$ (rate). Expectation: $\mathbb{E}[X] = \frac{r}{\lambda}$, Variance: $\text{Var}(X) = \frac{r}{\lambda^2}$.

- **Normal distribution:** $X \sim \mathcal{N}(\mu, \sigma^2)$, with

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

  Expectation: $\mathbb{E}[X] = \mu$, Variance: $\text{Var}(X) = \sigma^2$.

- **Chi-square distribution:** $X \sim \chi^2_k$, with

$$f(x) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{\frac{k}{2}-1} e^{-x/2}, \quad x > 0,$$

  where k is the degrees of freedom. Expectation: $\mathbb{E}[X] = k$, Variance: $\text{Var}(X) = 2k$.

- **Beta distribution:** $X \sim \text{Beta}(\alpha, \beta)$, with

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \quad 0 < x < 1,$$

  where $\alpha, \beta > 0$. Expectation: $\mathbb{E}[X] = \frac{\alpha}{\alpha+\beta}$, Variance: $\text{Var}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$. Useful for modeling proportions and probabilities.

- **Cauchy distribution:** $X \sim \text{Cauchy}(x_0, \gamma)$, with

$$f(x) = \frac{1}{\pi\gamma} \frac{1}{1 + \left(\frac{x-x_0}{\gamma}\right)^2}, \quad -\infty < x < \infty,$$

where x_0 is the location and $\gamma > 0$ the scale. Heavy-tailed distribution: expectation and variance are *undefined*. Appears in physics (resonance) and as the ratio of two standard normals.

- **Lognormal distribution:** $X \sim \text{Lognormal}(\mu, \sigma^2)$, with

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right), \quad x > 0.$$

Expectation: $\mathbb{E}[X] = e^{\mu + \frac{\sigma^2}{2}}$, Variance: $\text{Var}(X) = \left(e^{\sigma^2} - 1\right)e^{2\mu+\sigma^2}$. Models skewed positive data such as incomes, stock prices, and survival times.

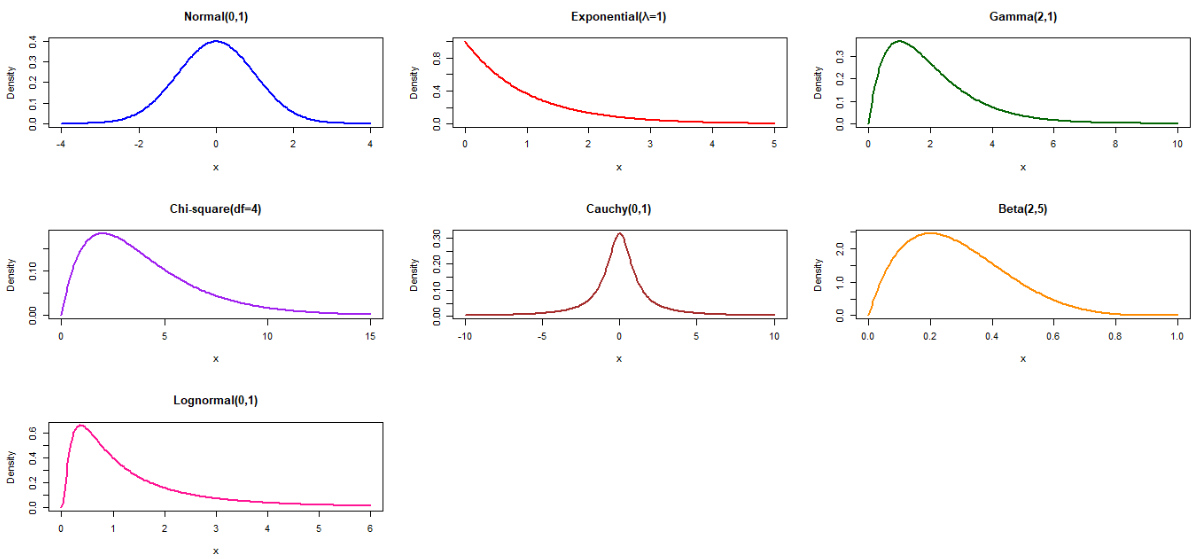

Figure 1.3: Discrete distributions

Distribution	**PMF** $p(x)$	**Support**	$\mathbb{E}[X]$	$\text{Var}(X)$	**Applications**
Bernoulli (p)	$\begin{cases} p & x=1 \\ 1-p & x=0 \end{cases}$	$\{0,1\}$	p	$p(1-p)$	Single trial: coin cess/failure events
Binomial (n,p)	$\binom{n}{k}p^k(1-p)^{n-k}$	$k=0,1,\dots,n$	np	$np(1-p)$	Number of successe dependent trials
Geometric (p)	$(1-p)^{k-1}p$	$k=1,2,\dots$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	Number of trials u success
Poisson (λ)	$e^{-\lambda}\frac{\lambda^k}{k!}$	$k=0,1,2,\dots$	λ	λ	Counts of rare even dents, defects, arri
Negative Binomial (r,p)	$\binom{k+r-1}{k}(1-p)^k p^r$	$k=0,1,2,\dots$	$\frac{r(1-p)}{p}$	$\frac{r(1-p)}{p^2}$	Number of failures successes
Discrete Uniform (a,b)	$\frac{1}{b-a+1}$	$a, a+1, \dots, b$	$\frac{a+b}{2}$	$\frac{(b-a+1)^2-1}{12}$	Random integers, c

Table 1.1: Summary of common discrete distributions.

Distribution	**PDF** $f(x)$	**Support**	$\mathbb{E}[X]$	$\mathrm{Var}(X)$	**Applications**
Uniform (a, b)	$\frac{1}{b-a}$	$a \le x \le b$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	Equal probabilities, number generation
Exponential (λ)	$\lambda e^{-\lambda x}$	$x \ge 0$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	Waiting times, l Poisson processes
Gamma (r, λ)	$\frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x}$	$x > 0$	$\frac{r}{\lambda}$	$\frac{r}{\lambda^2}$	Waiting time for r-t Bayesian stats
Normal (μ, σ^2)	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	$x \in \mathbb{R}$	μ	σ^2	Heights, measuren rors, CLT limit law
Chi-square (k)	$\frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2-1} e^{-x/2}$	$x > 0$	k	$2k$	Hypothesis testing, inference
Beta (α, β)	$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$	$0 < x < 1$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	Modeling prop Bayesian priors
Cauchy (x_0, γ)	$\frac{1}{\pi\gamma} \frac{1}{1+(\frac{x-x_0}{\gamma})^2}$	$x \in \mathbb{R}$	–	–	Physics (resonance) tailed phenomena
Lognormal (μ, σ^2)	$\frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}}$	$x > 0$	$e^{\mu+\frac{\sigma^2}{2}}$	$(e^{\sigma^2}-1)e^{2\mu+\sigma^2}$	Incomes, stock pri vival times

Table 1.2: Summary of common continuous distributions.

1.3.1 Operations on Random Variables

Proposition 1.1 (Linear Transformations)**.** *If* $Y = aX + b$*, then*

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

Proposition 1.2 (Independent Random Variables)**.** *If X and Y are independent, then*

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

Example 1.3. *If $X \sim Bin(n_1, p)$ and $Y \sim Bin(n_2, p)$ are independent, then*

$$X + Y \sim Bin(n_1 + n_2, p).$$

Example 1.4. *If $X \sim Poisson(\lambda_1)$ and $Y \sim Poisson(\lambda_2)$ are independent, then*

$$X + Y \sim Poisson(\lambda_1 + \lambda_2).$$

Functions of a Random Variable

Let X be a random variable with pdf f_X and cdf $F_X(x) = P(X \le x)$, and let $Y = g(X)$.

If g is a strictly monotone differentiable function, then the pdf of Y is given by

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

Example 1.5 (Celsius to Fahrenheit). *Suppose the temperature in Celsius follows* $X \sim \mathcal{N}(\mu, \sigma^2)$*, and we define*

$$Y = g(X) = aX + b, \quad (a = 1.8,\ b = 32).$$

Then

$$Y \sim \mathcal{N}(a\mu + b,\ a^2\sigma^2).$$

Remark 1.4 (Standardization). *If* $X \sim \mathcal{N}(\mu, \sigma^2)$*, then*

$$Z = \frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1).$$

This allows us to compute probabilities using the standard normal cdf Φ*:*

$$P(a \le X \le b) = \Phi\left(\tfrac{b-\mu}{\sigma}\right) - \Phi\left(\tfrac{a-\mu}{\sigma}\right).$$