

Chapter 1: Probability Theory

Instructor: Nan Liu

Chapter 1 roadmap

Foundations

- 1.1 Set theory
- 1.2 Basic probability theory
- 1.3 Conditional probability and Independence

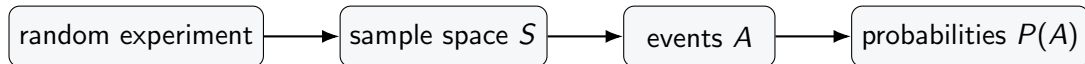
Random variables

- 1.4 Random variables
- 1.5 Distribution functions
- 1.6 Density and mass functions

The organization follows Casella and Berger, *Statistical Inference*, 2nd ed., Chapter 1.

From a random experiment to a probability model

A *random experiment* is a process with well-defined possible outcomes, but the realized outcome is uncertain before the experiment is performed.



Example: Examples of uncertainty

A coin toss

The number of accidents in a month

The lifetime of a component

Set Theory

Sample spaces, events, operations, partitions

Textbook: Section 1.1

1.1 Set Theory

Definition: Sample space

For a random experiment, the *sample space* S is the set of all possible basic outcomes.

Example: Three types of sample spaces

Experiment	Sample space
Roll a die	$S = \{1, 2, 3, 4, 5, 6\}$ (finite)
Number of accidents in a month	$S = \{0, 1, 2, \dots\}$ (countably infinite)
Lifetime of a light bulb	$S = [0, \infty)$ (uncountable)

Definition: Countable

A set is countable if its elements can be placed in one-to-one correspondence with a subset of the integers.

1.1 Set Theory

Definition: Event

An *event* is any subset of the sample space S .

If the realized outcome belongs to $A \subseteq S$, we say that event A occurs.

Example: Two dice

Let $S = \{(i, j) : i, j \in \{1, \dots, 6\}\}$. Then

$$A = \{(i, j) : i + j = 7\}$$

is the event “the sum equals 7.”

1.1 Set Theory

Definition: Empty set

The set containing no outcomes is denoted by $\emptyset$ and represents an impossible event.

Definition: Containment

$$A \subseteq B \iff x \in A \Rightarrow x \in B,$$

Definition: Equality of events

$$A = B \iff A \subseteq B \text{ and } B \subseteq A.$$

Proof strategy for set equality

To prove $A = B$, a reliable method is to prove both inclusions: $A \subseteq B$ and $B \subseteq A$.

1.1 Set Theory

Definition: Union

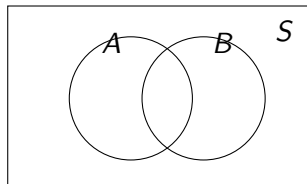
$$A \cup B = \{x : x \in A \text{ or } x \in B\}.$$

Definition: Intersection

$$A \cap B = \{x : x \in A \text{ and } x \in B\}.$$

Definition: Complement

$$A^c = \{x \in S : x \notin A\}.$$



1.1 Set Theory

Example: Event operations

Select one card from a deck and record only its suit. Let

$$S = \{C, D, H, S\}, \quad A = \{C, D\}, \quad B = \{D, H, S\}.$$

Then

$$A \cup B = \{C, D, H, S\} = S,$$

$$A \cap B = \{D\},$$

$$A^c = \{H, S\}.$$

Since $A \cup B = S$, we also have $(A \cup B)^c = \emptyset$.

1.1 Set Theory

Theorem (Algebra of sets)

- (a) *Commutative*: $A \cup B = B \cup A$, $A \cap B = B \cap A$;
- (b) *Associative*: $A \cup (B \cup C) = (A \cup B) \cup C$, $A \cap (B \cap C) = (A \cap B) \cap C$;
- (c) *Distributive*: $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$, $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$;
- (d) *De Morgan laws*: $(A \cup B)^c = A^c \cap B^c$, $(A \cap B)^c = A^c \cup B^c$.

1.1 Set Theory

Theorem (General De Morgan law:)

$$\left(\bigcup_i A_i\right)^c = \bigcap_i A_i^c, \quad \left(\bigcap_i A_i\right)^c = \bigcup_i A_i^c.$$

Proof:

For any outcome x ,

$$\begin{aligned} x \in \left(\bigcup_i A_i\right)^c &\iff x \notin A_i \text{ for every } i \\ &\iff x \in A_i^c \text{ for every } i \\ &\iff x \in \bigcap_i A_i^c. \end{aligned}$$

1.1 Set Theory

Definition: Countable unions

$$\bigcup_{i=1}^{\infty} A_i = \{x : x \in A_i \text{ for at least one } i\},$$

Definition: Countable intersections

$$\bigcap_{i=1}^{\infty} A_i = \{x : x \in A_i \text{ for every } i\}.$$

Example: Nested intervals

Let $A_i = [1/i, 1]$. Then $\bigcup_{i=1}^{\infty} A_i = (0, 1]$, $\bigcap_{i=1}^{\infty} A_i = \{1\}$.

1.1 Set Theory

Definition: mutually exclusive/pairwise Disjoint

Events $A_1, \dots, A_n$ are *mutually exclusive* (pairwise disjoint) if

$$A_i \cap A_j = \emptyset \quad (i \neq j).$$

Definition: collectively exhaustive

Events $A_1, \dots, A_n$ are *collectively exhaustive* if $\bigcup_{i=1}^n A_i = S$.

Definition: Partition

A collection $A_1, A_2, \dots$ is a partition of S if the sets are pairwise disjoint and collectively exhaustive.

Partitions will be central in the law of total probability and Bayes' rule.

Quiz 1: Sets and events

Quiz:

1. Let $A_i = [0, 1 + 1/i)$ for $i \geq 1$. Find $\bigcap_{i=1}^n A_i$ and $\bigcap_{i=1}^{\infty} A_i$.
2. Prove $(\bigcup_{i=1}^n A_i)^c = \bigcap_{i=1}^n A_i^c$ by the element method.
3. Express “exactly one of A and B occurs” using set.

Basics of Probability Theory

Sigma algebras, probability axioms, counting

Textbook: Section 1.2

1.2 Basics of Probability Theory– Axiomatic Foundations

Definition: σ -algebra

A collection $\mathcal{B}$ of subsets of S is a σ -algebra if

1. $\emptyset \in \mathcal{B}$;
2. $A \in \mathcal{B} \Rightarrow A^c \in \mathcal{B}$;
3. $A_1, A_2, \dots \in \mathcal{B} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{B}$.

By De Morgan's laws, a σ -algebra is also closed under countable intersections.

Example: Examples

If $S = \{1, 2, 3\}$, the power set 2^S is a σ -algebra. For every nonempty S , $\{\emptyset, S\}$ is also a σ -algebra.

f

1.2 Basics of Probability Theory– Axiomatic Foundations

A complete probability model is $(S, \mathcal{B}, P)$, where S is the sample space, $\mathcal{B}$ is the collection of events, and P assigns probabilities to events.

Definition: Probability function

A probability function $P : \mathcal{B} \rightarrow [0, 1]$ satisfies

1. $P(A) \geq 0$ for every $A \in \mathcal{B}$;
2. $P(S) = 1$;
3. if $A_1, A_2, \dots$ are pairwise disjoint, then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

1.2 Basics of Probability Theory– Axiomatic Foundations

Example: Probability axioms versus modeling assumptions

For $S = \{H, T\}$, both

$$P(H) = P(T) = \frac{1}{2}$$

and

$$P(H) = 0.7, \quad P(T) = 0.3$$

satisfy the probability axioms.

Important

The axioms specify how probabilities must behave. They do not determine which probability model is appropriate for a particular experiment.

1.2 Basics of Probability Theory– Axiomatic Foundations

Theorem (Constructing a probability on a countable space)

Let $S = \{s_1, s_2, \dots\}$ be finite or countable. If $p_i \geq 0$ and $\sum_i p_i = 1$, then

$$P(A) = \sum_{i:s_i \in A} p_i$$

defines a probability function on S .

Example: Fair die

For $S = \{1, 2, 3, 4, 5, 6\}$ and $p_i = 1/6$,

$$P(A) = \frac{|A|}{6}$$

for every event $A \subseteq S$.

1.2 Basics of Probability Theory – Calculus of Probabilities

Theorem (Basic consequences of the axioms)

For every event A ,

$$P(\emptyset) = 0, \quad 0 \leq P(A) \leq 1, \quad P(A^c) = 1 - P(A).$$

Proof of the complement rule

Since $A \cap A^c = \emptyset$ and $A \cup A^c = S$,

$$1 = P(S) = P(A) + P(A^c).$$

1.2 Basics of Probability Theory – Calculus of Probabilities

Theorem (Probability calculus for two events)

For any events A and B in $\mathcal{B}$

- a $P(B \cap A^c) = P(B) - P(A \cap B),$
- b $P(A \cup B) = P(A) + P(B) - P(A \cap B),$
- c *If $A \subseteq B$, then $P(A) \leq P(B).$*

Example: Bonferroni lower bound

Because $P(A \cup B) \leq 1,$

$$P(A \cap B) \geq P(A) + P(B) - 1.$$

For $P(A) = P(B) = 0.95$, we must have $P(A \cap B) \geq 0.90.$

1.2 Basics of Probability Theory – Calculus of Probabilities

Theorem (Partitions and Boole's inequality)

If $C_1, C_2, \dots$ form a partition of S , then

$$P(A) = \sum_{i=1}^{\infty} P(A \cap C_i).$$

For arbitrary events $A_1, A_2, \dots$,

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} P(A_i), \quad \text{Boole's inequality.}$$

1.2 Basics of Probability Theory – Calculus of Probabilities

Example: General Bonferroni inequality

Applying Boole's inequality to the complements A_i^c gives

$$P\left(\bigcap_{i=1}^n A_i\right) = 1 - P\left(\bigcup_{i=1}^n A_i^c\right) \geq 1 - \sum_{i=1}^n P(A_i^c),$$

so

$$P\left(\bigcap_{i=1}^n A_i\right) \geq \sum_{i=1}^n P(A_i) - (n - 1).$$

This is general version of Bonferroni inequality.

Quiz 2: Probability axioms and calculus

Quiz:

1. If $P(A) = 0.6$, $P(B) = 0.5$, and $P(A \cap B) = 0.3$, find $P(A \cup B)$, $P(A^c)$, and $P(B \cap A^c)$.
2. Suppose $A \subseteq B$. Use the probability axioms to prove $P(A) \leq P(B)$.

1.2 Basics of Probability Theory – Counting

Theorem (Fundamental theorem of counting)

If a procedure consists of k successive tasks, and task i can be performed in n_i ways for each admissible choice made earlier, then the entire procedure can be completed in

$$n_1 n_2 \cdots n_k$$

ways.

Definition: Factorial

For a positive integer n ,

$$n! = n(n-1)(n-2) \cdots 2 \cdot 1, \quad 0! = 1.$$

1.2 Basics of Probability Theory – Counting

Suppose we select r objects from a collection of n distinct objects.

There are two questions to consider:

1. Is sampling *with replacement* or *without replacement*?
2. Does the *order* of the selected objects matter?

Thus, there are four possible cases:

	Without replacement	With replacement
Ordered	1	2
Unordered	3	4

We consider these four cases separately.

1.2 Basics of Probability Theory – Counting

1. Ordered, without replacement

The first object can be selected in n ways.

Since it is not replaced, the second object can be selected in $n - 1$ ways, the third in $n - 2$ ways, and so on.

Therefore, the number of ordered samples of size r is

$$n(n-1) \cdots (n-r+1) = \frac{n!}{(n-r)!}.$$

Example: Three-digit password

A three-digit password is formed using the digits $0, 1, \dots, 9$, with no repeated digits. Since order matters,

$$10 \times 9 \times 8 = \frac{10!}{7!} = 720.$$

1.2 Basics of Probability Theory – Counting

2. Ordered, with replacement

If the selected object is replaced after each draw, then each of the r positions has n possible choices.

Therefore, the number of ordered samples is

$$\underbrace{n \times n \times \cdots \times n}_{r \text{ times}} = n^r.$$

Example: Three-digit password

If repetition of digits is allowed, then each of the three positions has 10 possible choices.

Hence,

$$10^3 = 1,000$$

three-digit passwords are possible.

1.2 Basics of Probability Theory – Counting

3. Unordered, without replacement

From Case 1, there are

$$\frac{n!}{(n-r)!}$$

ordered selections. However, if order does not matter, each group of r selected objects has been counted

$$r!$$

times. Therefore, the number of unordered samples is

$$\frac{n!}{r!(n-r)!}.$$

1.2 Basics of Probability Theory – Counting

Definition: Binomial coefficient

For integers $n \geq r \geq 0$,

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}.$$

$\binom{n}{r}$ counts selections of r distinct objects from n when order does not matter and replacement is not allowed.

Example: Committee

Suppose we select 3 students from a class of 10 students to form a committee. Since the order of the students on the committee does not matter and no student can be selected twice,

$$\binom{10}{3} = \frac{10!}{3!7!} = 120.$$

1.2 Basics of Probability Theory – Counting

4. Unordered, with replacement (not required)

This is the most subtle of the four cases.

We select r objects from n types, repetition is allowed, but the order of selection is not recorded.

Think of placing r markers into n bins. The n bins can be separated by $n - 1$ walls.

Thus, we arrange

$$r + (n - 1) = n + r - 1$$

objects consisting of r identical markers and $n - 1$ identical walls.

Therefore, the number of unordered samples is

$$\boxed{\binom{n + r - 1}{r}} = \binom{n + r - 1}{n - 1}.$$

1.2 Basics of Probability Theory – Counting

Example: Example: Unordered sampling with replacement (not required)

Suppose we choose $r = 3$ numbers from $\{1, 2, 3, 4\}$, with replacement, and order does not matter. For example,

$$\{1, 1, 3\}, \quad \{1, 2, 4\}, \quad \{3, 3, 3\}$$

are possible outcomes. The number of distinct unordered samples is

$$\binom{4 + 3 - 1}{3} = \binom{6}{3} = 20.$$

1.2 Basics of Probability Theory – Counting

Summary: four methods of counting

Suppose we select r objects from n distinct objects.

	Without replacement	With replacement
Ordered	$\frac{n!}{(n-r)!}$	n^r
Unordered	$\binom{n}{r}$	$\binom{n+r-1}{r}$

Before counting, always ask

1. Is replacement allowed?
2. Does order matter?

Quiz 3: Counting

Quiz:

1. From 8 students, how many ways can we choose (a) a 3-person committee; (b) a president, vice-president, and secretary?
2. Write the formula for the probability that at least two people in a group of k share a birthday under the 365-day model.
3. From 10 women and 8 men, 5 students are chosen at random. Find the probability that exactly 2 men are chosen.

1.2 Basics of Probability Theory – Enumerating Outcomes

The counting techniques of the previous section are especially useful when the sample space S is finite and all outcomes in S are equally likely. Suppose

$$S = \{s_1, \dots, s_N\},$$

and

$$P(\{s_i\}) = \frac{1}{N}, \quad i = 1, \dots, N.$$

Then, for any event A ,

$$P(A) = \sum_{s_i \in A} P(\{s_i\}) = \frac{\#\{\text{elements in } A\}}{\#\{\text{elements in } S\}} = \frac{|A|}{|S|}.$$

1.2 Basics of Probability Theory – Enumerating Outcomes

Example: Birthday problem

Assume 365 equally likely birthdays and ignore leap day. For k people,

$$P(\text{all birthdays different}) = \frac{365 \cdot 364 \cdots (365 - k + 1)}{365^k}.$$

$$P(\text{at least one shared birthday}) = 1 - \frac{365!}{(365 - k)! 365^k}.$$

For $k = 23$, this probability is approximately 0.507.

Example: Selecting students from two groups

A class has 15 boys and 30 girls. Ten students are selected uniformly from the 45 students. The probability of selecting exactly 3 boys is

$$P(\text{exactly 3 boys}) = \frac{\binom{15}{3} \binom{30}{7}}{\binom{45}{10}}.$$

1.2 Basics of Probability Theory – Enumerating Outcomes

Theorem (Counting arrangements with repeated values)

Suppose k positions contain m distinct values, where value j appears k_j times and

$$k_1 + \cdots + k_m = k.$$

Then the number of distinct ordered arrangements is

$$\frac{k!}{k_1! k_2! \cdots k_m!}$$

1.2 Basics of Probability Theory – Enumerating Outcomes

Example: Example: Assigning employees to shifts

A small company has 9 employees. Each day,

- 4 employees work during the day,
- 3 employees work at night,
- 2 employees do not work.

How many different assignments of the 9 employees are possible?

$$\frac{9!}{4!3!2!} = \binom{9}{4, 3, 2} = 1260.$$

Quiz 4: Enumerating outcomes

Quiz:

1. Five cards are drawn from a standard deck. Find the probability of exactly two aces.
2. Two numbers are drawn independently from $\{1, 2, 3, 4\}$ with replacement. What is the probability that the unordered result is $\{1, 3\}$?

Conditional Probability and Independence

Updating information, Bayes' rule, independence

Textbook: Section 1.3

1.3-1 Conditional Probability

Definition: Conditional probability

For events A and B with $P(B) > 0$,

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

Conditioning on B means that the probability model is recalibrated after learning that B occurred.

Theorem (Multiplication rule)

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

1.3-1 Conditional Probability

For events $A_1, \dots, A_n$, repeated use of the multiplication rule gives the chain rule

$$P\left(\bigcap_{i=1}^n A_i\right) = P(A_1) \prod_{i=2}^n P\left(A_i \middle| \bigcap_{j=1}^{i-1} A_j\right),$$

whenever the conditional probabilities are defined.

Example: Four aces by sequential conditioning

$$P(\text{first four cards are aces}) = \frac{4}{52} \frac{3}{51} \frac{2}{50} \frac{1}{49}.$$

1.3-1 Conditional Probability

Example: Sampling without replacement

A box contains r red balls and b blue balls. Draw two balls without replacement. Let

$$A = \{\text{first ball is red}\}, \quad B = \{\text{second ball is blue}\}.$$

Then

$$P(A) = \frac{r}{r+b}, \quad P(B | A) = \frac{b}{r+b-1},$$

so

$$P(A \cap B) = \frac{r}{r+b} \frac{b}{r+b-1}.$$

1.3-1 Conditional Probability

Theorem (Law of total probability)

If $A_1, A_2, \dots$ form a partition of S , then for any event B ,

$$P(B) = \sum_i P(B \mid A_i)P(A_i).$$

Example: Three groups

If $P(B_i) = 1/3$ and $P(A \mid B_i) = i/6$ for $i = 1, 2, 3$, then

$$P(A) = \frac{1}{3} \left(\frac{1}{6} + \frac{2}{6} + \frac{3}{6} \right) = \frac{1}{3}.$$

1.3-1 Conditional Probability

Theorem (Bayes' rule)

If $A_1, A_2, \dots$ form a partition and $P(B) > 0$, then

$$P(A_i | B) = \frac{P(B | A_i)P(A_i)}{\sum_j P(B | A_j)P(A_j)}.$$

Bayes' rule reverses the direction of conditioning.

1.3-1 Conditional Probability

Example: Auto-insurance risk classification

An insurance company classifies its customers as high risk (H), medium risk (M), or low risk (L). Suppose

$$P(H) = 0.25, \quad P(M) = 0.25, \quad P(L) = 0.50,$$

and the probabilities of having at least one accident during the year are

$$P(A | H) = 0.25, \quad P(A | M) = 0.16, \quad P(A | L) = 0.10.$$

If a customer has had an accident, what is the probability that the customer is high risk?

By Bayes' formula,

$$\begin{aligned} P(H | A) &= \frac{P(A | H)P(H)}{P(A | H)P(H) + P(A | M)P(M) + P(A | L)P(L)} \\ &= \frac{0.25(0.25)}{0.25(0.25) + 0.16(0.25) + 0.10(0.50)} \approx 0.410. \end{aligned}$$

An observed accident raises the probability of the high-risk type from 0.25 to about 0.41.

1.3-1 Conditional Probability

Example: Color blindness

In a population, the ratio of the number of men to the number of women is r . Thus,

$$P(M) = \frac{r}{1+r}, \quad P(W) = \frac{1}{1+r}.$$

Suppose the incidence of color blindness is

$$P(D | M) = p, \quad P(D | W) = p^2.$$

If a randomly selected person is color blind, what is the probability that the person is a man?

By Bayes' formula,

$$\begin{aligned} P(M | D) &= \frac{P(D | M)P(M)}{P(D | M)P(M) + P(D | W)P(W)} \\ &= \frac{p \frac{r}{1+r}}{p \frac{r}{1+r} + p^2 \frac{1}{1+r}} = \frac{r}{r+p}. \end{aligned}$$

1.3-1 Conditional Probability

Example: TV game

There are three curtains A , B , and C . A big prize is hidden behind one of them, with

$$P(A) = P(B) = P(C) = \frac{1}{3}.$$

You choose curtain A .

The host knows where the prize is and opens one of the other curtains, say B , showing that it is empty. He then offers you the opportunity to switch from A to C .

Should you stay with A or switch to C ?

1.3-1 Conditional Probability

Let B^* denote the event that the host opens curtain B and shows that it is empty.
The host's behavior implies

$$P(B^* | A) = \frac{1}{2}, \quad P(B^* | B) = 0, \quad P(B^* | C) = 1.$$

Hence,

$$P(A | B^*) = \frac{\frac{1}{2} \cdot \frac{1}{3}}{\frac{1}{2} \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}} = \frac{1}{3}.$$

Similarly,

$$P(C | B^*) = \frac{2}{3}.$$

Conclusion

Switching is better:

$$P(C | B^*) = \frac{2}{3} > \frac{1}{3} = P(A | B^*).$$

1.3-2: Independence

Definition: Statistical independence

Events A and B are independent if

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

If $P(B) > 0$, this is equivalent to

$$P(A \mid B) = P(A).$$

Disjointness is different

If A and B are disjoint with positive probability, then $P(A \cap B) = 0$ but $P(A)P(B) > 0$. Therefore they cannot be independent.

1.3-2: Independence

Example: At least one six

Roll a fair die four times independently. Then

$$P(\text{at least one six}) = 1 - P(\text{no six}) = 1 - \left(\frac{5}{6}\right)^4.$$

More generally, if $A_1, \dots, A_n$ are independent,

$$P\left(\bigcup_{i=1}^n A_i\right) = 1 - \prod_{i=1}^n \{1 - P(A_i)\}.$$

If all $P(A_i) = p$, this becomes $1 - (1 - p)^n$.

1.3-2: Independence

Theorem (Independence is preserved by complements)

If A and B are independent, then

$$(A, B^c), \quad (A^c, B), \quad (A^c, B^c)$$

are also independent pairs.

For example,

$$P(A \cap B^c) = P(A) - P(A \cap B) = P(A)\{1 - P(B)\} = P(A)P(B^c).$$

1.3-2: Independence

Pairwise independence is not enough for three or more events. For three events we may have

$$P(A \cap B) = P(A)P(B), \quad P(A \cap C) = P(A)P(C), \quad P(B \cap C) = P(B)P(C),$$

but still

$$P(A \cap B \cap C) \neq P(A)P(B)P(C).$$

Definition: Mutual independence

$A_1, \dots, A_n$ are mutually independent if for every subcollection $A_{i_1}, \dots, A_{i_k}$,

$$P\left(\bigcap_{j=1}^k A_{i_j}\right) = \prod_{j=1}^k P(A_{i_j}).$$

1.3-2: Independence

Example: Pairwise but not mutually independent

Let $S = \{a, b, c, d\}$ with four equally likely outcomes, and define

$$A_1 = \{a, b\}, \quad A_2 = \{b, c\}, \quad A_3 = \{a, c\}.$$

Then

$$P(A_1) = P(A_2) = P(A_3) = \frac{1}{2},$$

and every pair has intersection probability $1/4$. Thus every pair is independent. However,

$$A_1 \cap A_2 \cap A_3 = \emptyset, \quad 0 \neq \frac{1}{8} = P(A_1)P(A_2)P(A_3).$$

Quiz 5: Conditional probability and independence

Quiz:

1. A box has 4 red and 6 blue balls. Two are drawn without replacement. Find the probability that the first is red and the second is blue.
2. In the insurance example, calculate $P(A)$ and then $P(H | A)$.
3. On $S = \{a, b, c, d\}$ with equal probabilities, verify that $A_1 = \{a, b\}$, $A_2 = \{b, c\}$, $A_3 = \{a, c\}$ are pairwise but not mutually independent.

Random Variables

Mapping outcomes into numerical summaries

Textbook: Section 1.4

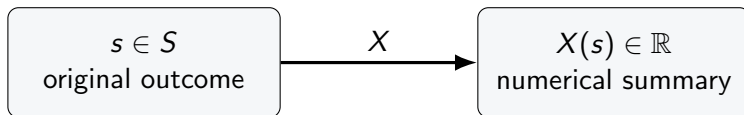
1.4 Random Variables

Definition: Random variable

A random variable X is a function from the sample space S into the real numbers:

$$X : S \longrightarrow \mathbb{R}.$$

The set of possible values of X is its range, often denoted by $\mathcal{X}$.



1.4 Random Variables

Examples of random variables

Experiment	Random variable
Toss two dice	$X =$ sum of the two faces
Toss a coin 25 times	$X =$ number of heads
Measure crop response	$X =$ yield per acre
Sample 50 voters	$X =$ number supporting a proposal

A random variable can reduce a complicated outcome to the numerical feature relevant for analysis.

1.4 Random Variables

If $X : S \rightarrow \mathcal{X}$, the probability on S induces probabilities on values of X :

$$P(X = x) = P(\{s \in S : X(s) = x\}).$$

More generally, for $A \subseteq \mathcal{X}$,

$$P(X \in A) = P(\{s \in S : X(s) \in A\}).$$

Notation

Uppercase letters X, Y, Z denote random variables; lowercase letters x, y, z denote possible or realized values.

1.4 Random Variables

Example: Three coin tosses

Toss a fair coin three times and let X = number of heads.

$$S = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}.$$

Then

x	0	1	2	3
$P(X = x)$	1/8	3/8	3/8	1/8

$$P(0 \leq X \leq 1) = \frac{4}{8}, \quad P(0 \leq X < 1) = P(X = 0) = \frac{1}{8}.$$

Discrete inequalities matter

For a discrete random variable, changing $\leq$ to $<$ may change the event and its probability.

Distribution Functions

CDFs, discrete/continuous variables, identical distributions

Textbook: Section 1.5

1.5 Distribution Functions

Definition: Cumulative distribution function

The cumulative distribution function of X is

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

The cdf is defined for every real x , not only for values in the range of X .

For any $a < b$,

$$P(a < X \leq b) = F_X(b) - F_X(a).$$

1.5 Distribution Functions

Example: CDF for three coin tosses

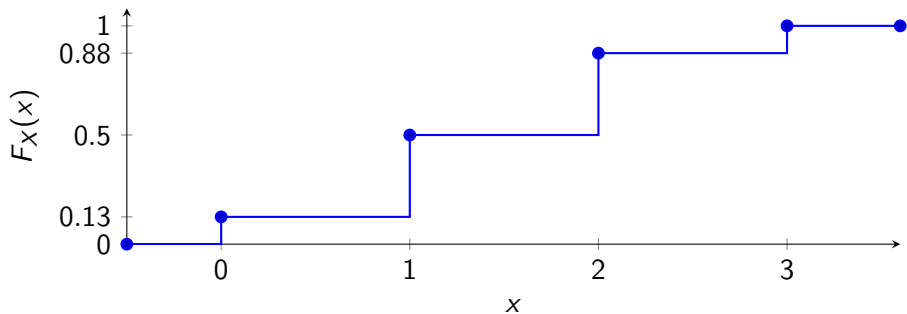
For X = number of heads in three fair tosses,

$$F_X(x) = \begin{cases} 0, & x < 0, \\ 1/8, & 0 \leq x < 1, \\ 1/2, & 1 \leq x < 2, \\ 7/8, & 2 \leq x < 3, \\ 1, & x \geq 3. \end{cases}$$

At a point x_0 , the jump size is

$$F_X(x_0) - F_X(x_0^-) = P(X = x_0).$$

1.5 Distribution Functions

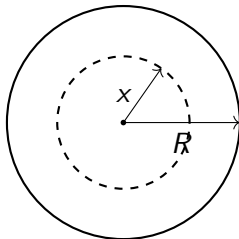


The cdf takes the value at the top of each jump: cdfs are right-continuous.

1.5 Distribution Functions

Example: Distance from the center of a target

A shot lands uniformly by area on a circular target of radius R . Let X be the distance from the hit point to the center.



For $0 \leq x \leq R$,

$$F_X(x) = \frac{\pi x^2}{\pi R^2} = \frac{x^2}{R^2}.$$

Thus

$$F_X(x) = \begin{cases} 0, & x < 0, \\ x^2/R^2, & 0 \leq x \leq R, \\ 1, & x > R. \end{cases}$$

1.5 Distribution Functions

Theorem (Characterization of a cdf)

A function F is a cdf if and only if

1. $\lim_{x \rightarrow -\infty} F(x) = 0$ and $\lim_{x \rightarrow \infty} F(x) = 1$;
2. F is nondecreasing;
3. F is right-continuous.

Jump probability

$$P(X = x_0) = F_X(x_0) - F_X(x_0^-), \quad F_X(x_0^-) = \lim_{x \uparrow x_0} F_X(x).$$

1.5 Distribution Functions

For any random variable,

$$P(a < X \leq b) = F_X(b) - F_X(a),$$

and

$$P(a \leq X \leq b) = F_X(b) - F_X(a^-).$$

If the cdf is continuous at a , then $F_X(a^-) = F_X(a)$.

Why right-continuity?

As $x \downarrow x_0$, the events $\{X \leq x\}$ decrease to $\{X \leq x_0\}$, so probability continuity implies

$$\lim_{x \downarrow x_0} F_X(x) = F_X(x_0).$$

1.5 Distribution Functions

Example: Geometric cdf

Let X be the number of independent Bernoulli trials required for the first success, with success probability p . Then

$$P(X = x) = (1 - p)^{x-1}p, \quad x = 1, 2, \dots$$

and for positive integers x ,

$$F_X(x) = \sum_{i=1}^x (1 - p)^{i-1}p = 1 - (1 - p)^x.$$

Between consecutive integers, the cdf is flat.

1.5 Distribution Functions

Example: A continuous cdf

Consider

$$F(x) = \frac{1}{1 + e^{-x}}.$$

Then

$$\lim_{x \rightarrow -\infty} F(x) = 0, \quad \lim_{x \rightarrow \infty} F(x) = 1,$$

and

$$F'(x) = \frac{e^{-x}}{(1 + e^{-x})^2} > 0.$$

Hence F is continuous and increasing and satisfies the cdf conditions.

1.5 Distribution Functions

Definition: Identically distributed random variables

Random variables X and Y are identically distributed if

$$P(X \in A) = P(Y \in A)$$

for every admissible set A .

Theorem (CDF characterization of identical distributions)

X and Y are identically distributed if and only if

$$F_X(x) = F_Y(x)$$

for every real x .

1.5 Distribution Functions

Example: Heads and tails

For three fair tosses, let

X = number of heads, Y = number of tails.

Then X and Y have the same cdf and hence the same distribution, but they are not equal outcome by outcome. In fact,

$$X + Y = 3.$$

Important

Identically distributed does not mean that two random variables are the same function on the sample space, nor does it imply independence.

Quiz 7: Distribution functions

Quiz:

1. A shot is uniform by area on a target of radius $R = 2$. If X is distance from the center, find $F_X(1)$ and $P(1 < X \leq 1.5)$.
2. Is $F(x) = 0$ for $x < 0$, $F(x) = x/2$ for $0 \leq x < 2$, and $F(x) = 1$ for $x \geq 2$ a valid cdf? Explain.
3. If $F(2^-) = 0.65$ and $F(2) = 0.80$, find $P(X = 2)$.

Density and Mass Functions

PMFs, PDFs, normalization, interval probabilities

Textbook: Section 1.6

1.6 Mass Functions and Density Functions

Definition: Probability mass function

For a discrete random variable X , the probability mass function is

$$f_X(x) = P(X = x).$$

It satisfies

$$f_X(x) \geq 0, \quad \sum_x f_X(x) = 1.$$

The support is

$$\text{supp}(X) = \{x : P(X = x) > 0\}.$$

Moreover,

$$F_X(x) = \sum_{y \leq x} f_X(y), \quad f_X(x) = F_X(x) - F_X(x^-).$$

1.6 Mass Functions and Density Functions

Example: Geometric pmf

For the geometric waiting-time random variable,

$$f_X(x) = \begin{cases} (1-p)^{x-1}p, & x = 1, 2, \dots, \\ 0, & \text{otherwise.} \end{cases}$$

For integers $1 \leq a \leq b$,

$$P(a \leq X \leq b) = \sum_{k=a}^b (1-p)^{k-1}p.$$

The pmf value at x is the jump size of the cdf at x .

1.6 Mass Functions and Density Functions

Definition: Probability density function

A function f_X is a probability density function for a continuous random variable X when

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

for every x .

When F_X is differentiable,

$$f_X(x) = F'_X(x).$$

For a continuous random variable,

$$P(X = x) = 0$$

for every x .

1.6 Mass Functions and Density Functions

For a continuous random variable,

$$P(a < X \leq b) = F_X(b) - F_X(a) = \int_a^b f_X(x) dx.$$

Since point probabilities are zero,

$$P(a < X < b) = P(a \leq X \leq b) = P(a < X \leq b).$$

Important

A density value $f_X(x)$ is not a point probability. Probability is obtained by integrating the density over a set.

1.6 Mass Functions and Density Functions

Example: Uniform density on an interval

If

$$f_X(x) = \frac{1}{b-a}, \quad a < x < b,$$

and $f_X(x) = 0$ otherwise, then for $a \leq c < d \leq b$,

$$P(c < X < d) = \int_c^d \frac{1}{b-a} dx = \frac{d-c}{b-a}.$$

Probability is proportional to interval length because the density is constant.

1.6 Mass Functions and Density Functions

Theorem (When is a function a pmf or pdf?)

A function f is a valid pmf or pdf if and only if

- 1. $f(x) \geq 0$ for all x ;*
- 2. the total mass equals one:*

$$\sum_x f(x) = 1 \quad (\text{pmf}), \quad \int_{-\infty}^{\infty} f(x) dx = 1 \quad (\text{pdf}).$$

1.6 Mass Functions and Density Functions

Suppose $h(x) \geq 0$ and

$$0 < K = \int_{-\infty}^{\infty} h(x) dx < \infty.$$

Then

$$f(x) = \frac{h(x)}{K}$$

is a valid pdf.

Normalization principle

A nonnegative function specifies the shape of a density; the normalizing constant makes the total probability equal to one.

1.6 Mass Functions and Density Functions

Example: Find a normalizing constant

Suppose

$$f(x) = cx^2, \quad 0 < x < 1,$$

and $f(x) = 0$ otherwise. Then

$$1 = \int_0^1 cx^2 dx = \frac{c}{3}, \quad c = 3.$$

Hence

$$P(X > 1/2) = \int_{1/2}^1 3x^2 dx = 1 - \frac{1}{8} = \frac{7}{8}.$$

1.6 Mass Functions and Density Functions

Example: Logistic density

From

$$F_X(x) = \frac{1}{1 + e^{-x}},$$

we obtain

$$f_X(x) = F'_X(x) = \frac{e^{-x}}{(1 + e^{-x})^2}.$$

Thus

$$P(a < X < b) = \int_a^b f_X(x) dx = F_X(b) - F_X(a).$$

The cdf and pdf contain the same distributional information whenever the density representation exists.

1.6 Mass Functions and Density Functions

	Discrete	Continuous
Point probability	$P(X = x) = f_X(x)$	$P(X = x) = 0$
CDF	$F_X(x) = \sum_{t \leq x} f_X(t)$	$F_X(x) = \int_{-\infty}^x f_X(t) dt$
Interval probability	Sum pmf values	Integrate pdf

Quiz 8: PMFs and PDFs

Quiz:

1. Let $f(x) = c(1 - x)$ for $0 < x < 1$ and 0 otherwise. Find c and $P(X < 1/2)$.
2. A discrete variable has pmf $f(k) = c/2^k$, $k = 1, 2, \dots$. Find c and $P(X \geq 3)$.
3. If X has density $f(x) = 1/4$ on $(0, 4)$, find $P(1 < X < 3)$ and $P(X = 2)$.

Chapter 1: definition map

- Sample space
- Event
- Union, intersection, complement
- Disjoint / pairwise disjoint
- Partition
- Sigma algebra
- Probability function
- Factorial
- Binomial coefficient
- Conditional probability
- Statistical independence
- Mutual independence
- Random variable
- CDF
- Continuous / discrete RV
- Identically distributed
- PMF
- PDF

Problem-solving checklist

1. Define the sample space and events clearly.
2. Ask whether outcomes are equally likely; if yes, consider counting.
3. If information has been revealed, condition on it.
4. Check whether independence is justified rather than assumed.
5. For a random variable, choose the most convenient representation: cdf, pmf, or pdf.
6. Verify probabilities are nonnegative and total to one.
7. Interpret the numerical answer in the original experiment.