MIE286 Lecture Notes

Probability and Statistics

Cheryl Shi

Contents

Chapter 1: Probability Foundations 1

1.1 Sampling Methods 1

1.2 Data, Variables, and Distributions 1

1.3 Graphical Representations of Data 3

Chapter 2: Random Variables and Distributions 4

2.1 Experiments, Sample Spaces, and Events 4

2.2 Event Operations and Probability Rules 5

2.3 Counting Techniques and Equally Likely Outcomes 8

2.4 Conditional Probability and Independence 30

Chapter 3: Statistical Inference 18

3.1 Random Variables and Their Interpretation 18

3.2 Discrete Random Variables 19

3.3 Probability Mass Functions 30

3.3.1 PMFs for Discrete Random Variables 30

3.4 Cumulative Distribution Function (CDF) 23

3.4.1 CDF for Discrete Random Variables 23

3.4.2 CDF for Continuous Random Variables 27

3.5 Continuous Sample Space and Continuous Random Variables 25

3.6 Probability Density Function (PDF) 30

3.7 Joint Probability Distributions 28

3.7.1 Joint Distributions for Discrete Random Variables 28

3.7.2 Joint Distributions for Continuous Random Variables 31

3.8 Marginal Distributions 29
3.9 Conditional Distributions 30
3.10 Independence 38

Chapter 4 40

4.1 Expected Value 40
4.1.1 Expected Value of a Random Variable 40
4.1.2 Expected Value of a Function of a Random Variable 45
4.1.3 Expected Value from a Joint Distribution 45
4.1.4 Properties of Expected Value 47
4.2 Variance 47

Chapter 1

Statistics Definitions

Global definition: Statistics involves collecting, organizing, summarizing, presenting, and analyzing data, as well as making inferences, conclusions, and decisions based on data.

Statistical definition: A statistic is a numerical value calculated from data (e.g. mean, proportion, standard deviation).

$$\begin{tikzcd} \boxed{\text{Population}} \arrow[r, bend left, "\text{Probability (deductive)}"] & \boxed{\text{Sample}} \arrow[l, bend left, "\text{Statistics Inference (inductive)}"] \end{tikzcd}$$

Basic Terminology

Individuals: Objects on which data are collected (people, animals, plots of land, etc.).
Variable: Any characteristic of an individual.
Population: The entire group of individuals of interest.
Sample: A subset of individuals taken from the population.
Statistical Inference: Drawing conclusions about a population based on a sample.

Sampling Methods

Simple Random Sample (SRS):

- Every possible group of size n has an equal chance of being selected.
- Helps avoid bias in sampling.
- Can be selected using random number tables or software.

Stratified Random Sampling:

- The population is divided into homogeneous groups *(individuals are similar with respect to the variable being studied)* called strata.
- A simple random sample is taken from each stratum. *(one subgroup of the population created)*
- Ensures that important subgroups are neither over nor under represented.

Data, Variables, and Distributions

Types of Variables

Categorical Variable: Places individuals into categories (e.g. gender, major). These are qualitative.

Quantitative Variable: Takes numerical values for which arithmetic operations are meaningful.

- Discrete
- Continuous

Distributions

Distribution: Describes what values a variable takes and how often those values occur. When examining a distribution, look for:

- **Shape**
- **Center**
- **Spread**
- **Outliers**

Outlier: An individual value that falls outside the overall pattern of the data.

Describing Distributions with Numbers

Central Tendency: Describes where the data cluster or center.

Central Tendency: Describes where the data cluster or center.

- Mean: average value
- Median: middle value

Mean (Arithmetic Mean):

$$\boxed{\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i}$$

Median:

$$\boxed{\tilde{x} = \begin{cases} x_{\left(\frac{n+1}{2}\right)}, & \text{if } n \text{ is odd} \\ \dfrac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2}, & \text{if } n \text{ is even} \end{cases}}$$

Theorem 1.1

1. The mean is more sensitive to extreme values than the median.
2. Changing a single data value will always change the mean, but may not change the median.
3. If a distribution is exactly symmetric, the mean and median are equal.

Trimmed Mean: The mean computed after removing extreme values.

$$\boxed{\bar{x}_{\text{trim}} = \frac{1}{n-2k} \sum_{i=k+1}^{n-k} x_{(i)}}$$

where k values are removed from both ends of the ordered data. (normally given in question like 10%)

Measures of Spread

Range: Maximum minus minimum. Very sensitive to extreme values.

Sample Variance: Measures the average squared deviation from the mean.

$$\boxed{s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2}$$

Standard Deviation: The square root of the sample variance.

$$\boxed{s = \sqrt{s^2}}$$

Degrees of Freedom: The number of independent pieces of information available to estimate variability. For sample variance: $df = n - 1$.

Graphical Representations of Data

Scatter Plot: Used to display the relationship between two quantitative variables (x, y). A scatter plot helps identify trends, patterns, and associations between variables.

Stem-and-Leaf Plot: An intermediate step between raw data and a frequency table. Preserves the original data values while showing the distribution.

Stem	Leaf
1	2 4 7
2	1 3 5 8
3	0 4 6

Relative Frequency Table: Shows the proportion of observations in each class.

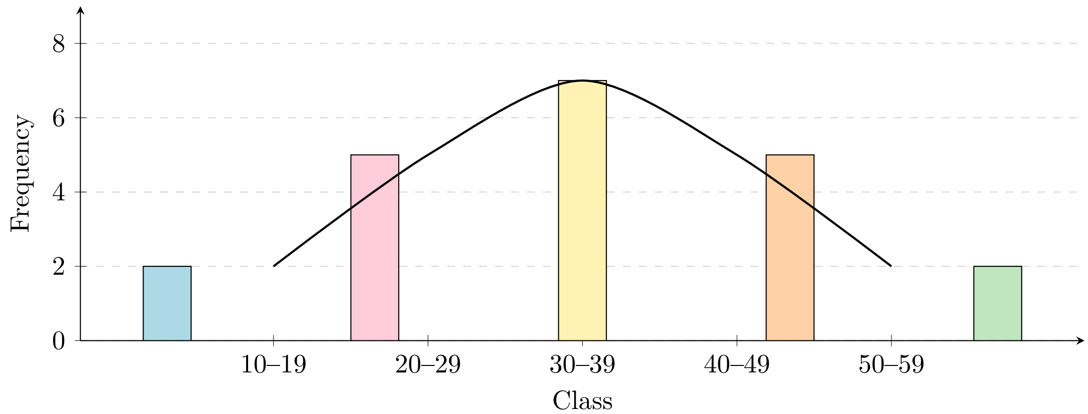
Class Interval	Class Midpoint	Frequency	Relative Frequency
10–19	14.5	3	0.30
20–29	24.5	4	0.40
30–39	34.5	3	0.30

Histogram: A graphical representation of a frequency or relative frequency table using contiguous bars.

When describing the shape of a histogram, we commonly classify it as:

- **Symmetric**
- **Skewed right** (positively skewed)
- **Skewed left** (negatively skewed)

Symmetric distribution example



Chapter 2, Jan 9th

Experiments, Sample Spaces, and Events

Experiment: A process that generates an outcome.

Sample Space (S): The set of all possible outcomes of an experiment.

Example 1:

Select 3 items from a production line. Each item can be classified as either defective (D) or non-defective (N).

$$S = \{DDD, DDN, DND, NDD, DNN, NDN, NND, NNN\}$$

Since each item has 2 possible outcomes,

$$|S| = 2^3 = 8$$

Example 2:

$$S = \{(x, y) \mid x^2 + y^2 \leq 4\}$$

Event (A): A subset of the sample space S.

Examples of events:

$$A = \{DDD, DDN, DND, NDD\}$$

$$B = \{NNN\}$$

$$C = \{(x, y) \mid x^2 + y^2 \leq 4\}$$

Event Operations and Probability Rules

Event Operations:

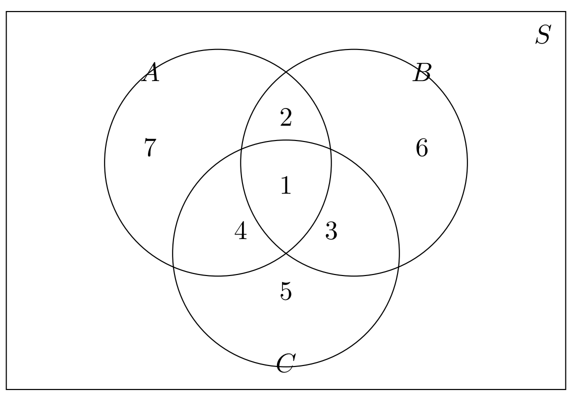
- Complement: A^c (or A')
- Intersection: $A \cap B$
- Union: $A \cup B$
- Null Event: $\varnothing$

If

$$A \cap B = \varnothing,$$

then A and B are mutually exclusive.

Example (Venn Diagram):



$$A = \{DDD, DDN, DND, NDD\}, \quad B = \{NNN\}$$

$$A \cup B = \{DDD, DDN, DND, NDD, NNN\}$$

$$A \cap B = \varnothing$$

Chapter 2: January 12

Review

1. *Experiment:* *A process that generates an outcome.*
2. *Sample Space* (S)*: The set of all possible outcomes of an experiment.*
3. *Event Operations:*
   - *Complement:* A' (A^c)
   - *Intersection:* $A \cap B$
   - *Union:* $A \cup B$
   - *Null Event:* $\varnothing$
4. *If* $A \cap B = \varnothing$*, then* A *and* B *are called mutually exclusive.*

$$(A \cap B)' = A' \cup B'$$

$$(A \cup B)' = A' \cap B'$$

$$A \cap \varnothing = \varnothing$$

$$A \cup \varnothing = A$$

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

Probability

$P(A)$ = probability of event A: the proportion of times the event occurs in infinitely many repetitions of the experiment.

Theorem 2.1

$$0 \leq P(A) \leq 1$$

$$P(A) + P(A') = 1$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$\begin{aligned} P(A \cup B \cup C) = P(A) &+ P(B) + P(C) \\ &- P(A \cap B) - P(A \cap C) - P(B \cap C) \\ &+ P(A \cap B \cap C) \end{aligned}$$

Mutually Exclusive Events

Definition: If $A_1, A_2, \ldots, A_n$ are mutually exclusive, then

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = P(A_1) + P(A_2) + \cdots + P(A_n)$$

If

$$A_1 \cup A_2 \cup \cdots \cup A_n = S,$$

then $\{A_1, A_2, \ldots, A_n\}$ is a partition of S.

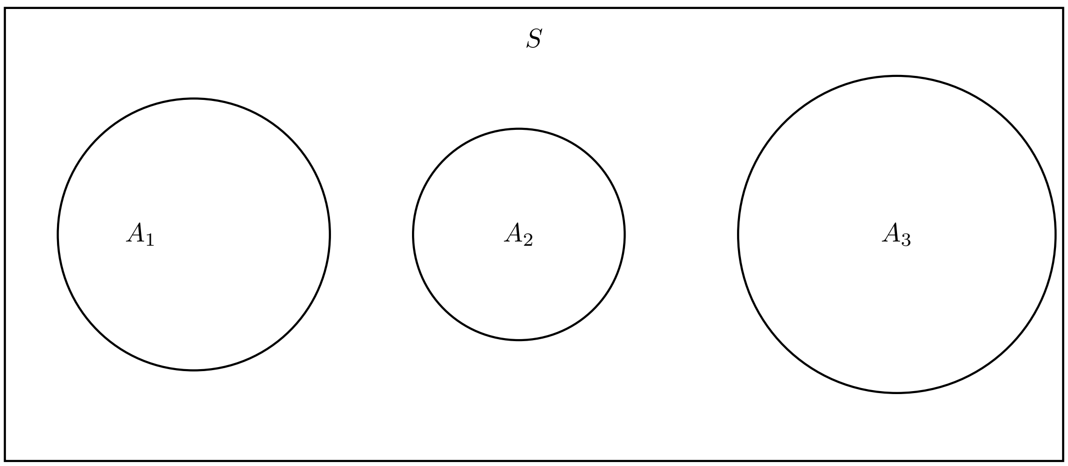


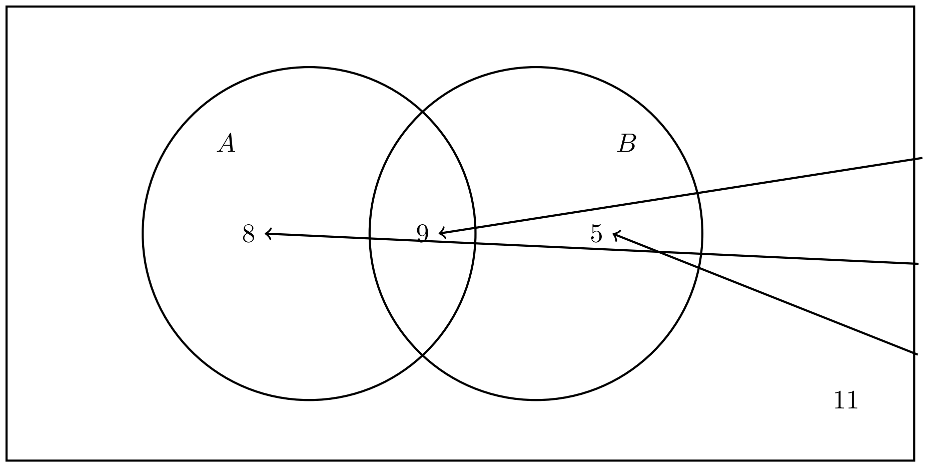
Figure 1: Partition of the sample space S into A_1, A_2, A_3

Example

In a class of 33 students:

- 17 earned an A on the midterm
- 14 earned an A on the final
- 11 earned no A on either exam

Find the probability that a randomly selected student earned A's on both exams.



$$|A \cup B| = 33 - 11 = 22$$

$$|A \cap B| = |A| + |B| - |A \cup B|$$
$$= 17 + 14 - 22 = 9$$

$$|A \setminus B| = |A| - |A \cap B| = 17 - 9 = 8$$

$$|B \setminus A| = |B| - |A \cap B| = 14 - 9 = 5$$

$$P(A \cap B) = \frac{|A \cap B|}{|S|} = \frac{9}{33} \approx 0.273$$

Figure 2: Events A: A on midterm, B: A on final, with region counts and calculations

Counting Techniques and Equally Likely Outcomes

Theorem 2.2 (Equally Likely Outcomes)

If the sample space S has a finite number of outcomes and all outcomes are equally likely, then for any event A,

$$P(A) = \frac{|A|}{|S|}.$$

where

$|A|$ = number of outcomes in event A, $\qquad$ $|S|$ = number of outcomes in the sample space.

Example 1: Poker Hands Basics

A standard deck has:

$$4 \text{ suits} \times 13 \text{ denominations (A,2,3,\ldots,Q,K)} = 52 \text{ cards.}$$

A poker hand consists of 5 cards chosen from 52:

$$|S| = \binom{52}{5} = 2{,}598{,}960.$$

Combinations Reminder

If there are 3 objects $\{A, B, C\}$ and we choose 2:

$$\binom{3}{2} = \frac{3!}{(3-2)!2!}.$$

Order does <u>not</u> matter.

Example 2: Probability of 2 Aces and 1 Jack

A 5-card hand contains:

- exactly 2 aces,
- exactly 1 jack,
- 2 cards that are neither aces nor jacks.

$$P(2 \text{ aces and } 1 \text{ jack}) = \frac{\binom{4}{2}\binom{4}{1}\binom{44}{2}}{\binom{52}{5}}.$$

Example 3: Probability of a Full House

A full house consists of:

- 3 cards of one denomination
- 2 cards of a different denomination

Number of full house hands:

$$\binom{13}{1}\binom{4}{3}\binom{12}{1}\binom{4}{2}.$$

Thus,

$$P(\text{full house}) = \frac{\binom{13}{1}\binom{4}{3}\binom{12}{1}\binom{4}{2}}{\binom{52}{5}}.$$

Example 4: Probability of Four of a Kind

A four of a kind consists of:

- 4 cards of the same denomination
- 1 remaining card of a different denomination

Number of such hands:

$$\binom{13}{1}\binom{4}{4}\binom{48}{1}.$$

Thus,

$$P(\text{four of a kind}) = \frac{\binom{13}{1}\binom{4}{4}\binom{48}{1}}{\binom{52}{5}}.$$

Example 5: Probability of Exactly One Pair

An **excatly** one-pair hand consists of:

- 1 pair
- 3 cards of different denominations, none matching the pair

Number of such hands:

$$\binom{13}{1}\binom{4}{2}\binom{12}{3}\binom{4}{1}^3.$$

Thus,

$$P(\text{exactly one pair}) = \frac{\binom{13}{1}\binom{4}{2}\binom{12}{3}\binom{4}{1}^3}{\binom{52}{5}}.$$

Note*: Counting Patterns

$$\binom{a}{b}$$

Meaning: Choose b different items from a **at once**, order does not matter.

Key features:

- No repeats
- Grouped choice
- Used when items must be distinct

$$\binom{a}{1}^b$$

Meaning: Make b independent choices, each time choosing 1 item from c.

Key features:

- Repeats allowed
- Choices are independent
- Used when selections do not restrict each other

Rule to Remember:

$$\text{Different items, no repeats} \Rightarrow \binom{a}{b}$$

$$\text{Independent choices} \Rightarrow \binom{c}{1}^b$$

Chapter 2 continue, Jan 14

Review

1. *Probability is the proportion of times the event occurs in infinitely many repetitions of the experiment.*

2. $0 \leq P(A) \leq 1$

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

4. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

5. $$\begin{aligned} P(A \cup B \cup C) = &P(A) + P(B) + P(C) \\ &- P(A \cap B) - P(A \cap C) - P(B \cap C) \\ &+ P(A \cap B \cap C) \end{aligned}$$

6. *Permutation: A permutation counts ordered arrangements.*

$$\boxed{{}_nP_r = \frac{n!}{(n-r)!}}$$

Example 1: Two fair dice

A pair of fair dice are rolled. Find the probability that the second die lands on a smaller value than the first.

The outcomes where the second die is smaller than the first are represented below.

First Die (Stem)	Second Die (Leaf)
2	1
3	1 2
4	1 2 3
5	1 2 3 4
6	1 2 3 4 5

There are 15 favorable outcomes and 36 total outcomes.

$$P(\text{second} < \text{first}) = \frac{15}{36} = \frac{5}{12}.$$

Conditional Probability and Independence

Conditional Probability

The conditional probability of an event B given that event A has occurred is the probability that B occurs when it is known that A has occurred.

$$\boxed{P(B \mid A) = \frac{P(A \cap B)}{P(A)}, \quad P(A) > 0}$$

Example 2: Drinking Survey

A survey records the following data:

	D	N	Total
M	19	41	60
F	12	28	40
Total	31	69	100

The symbols used above are defined as follows:

- M: male
- F: female
- D: the individual drinks
- N: the individual does not drink

$$P(D|M) = \frac{19}{60} \qquad P(M|D) = \frac{19}{31}$$

Law of Total Probability

Theorem 2.3 (Law of Total Probability)

If $B_1, B_2, \ldots, B_k$ form a partition of the sample space S with $P(B_i) > 0$ for all i, then for any event A,

$$P(A) = \sum_{i=1}^{k} P(A \mid B_i)\, P(B_i).$$

Example 3: Monty Hall (3 doors)

Car location	Monty opens	Probability	Stay	Switch
Door 1	Door 2	$\frac{1}{6}$	Car	Goat
Door 1	Door 3	$\frac{1}{6}$	Car	Goat
Door 2	Door 3	$\frac{1}{3}$	Goat	Car
Door 3	Door 2	$\frac{1}{3}$	Goat	Car

Staying wins only when the car is behind Door 1, so

$$P(\text{win by staying}) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}.$$

Switching wins when the car is behind Door 2 or Door 3, so

$$P(\text{win by switching}) = \frac{1}{3} + \frac{1}{3} = \frac{2}{3}.$$

Example 4: Birthday Problem

Assume the following:

- Leap years are ignored
- All 365 birthdays are equally likely
- Birthdays of different people are independent

Question: What is the probability that at least two people share the same birthday in a group of n people?

Rather than computing this directly, we use the complement rule.

$$P(\text{at least one match}) = 1 - P(\text{no match})$$

Probability of no shared birthdays

- Person 1 can have any birthday: probability 1
- Person 2 must avoid that birthday: $\frac{364}{365}$
- Person 3 must avoid the first two birthdays: $\frac{363}{365}$
- $\ldots$
- Person n must avoid the previous $n-1$ birthdays: $\frac{365-(n-1)}{365}$

Therefore,

$$P(\text{no match}) = \frac{365}{365} \cdot \frac{364}{365} \cdot \frac{363}{365} \cdots \frac{365-(n-1)}{365}$$

or equivalently,

$$P(\text{no match}) = \prod_{k=0}^{n-1} \frac{365-k}{365}$$

Final result

$$P(\text{at least one shared birthday}) = 1 - \prod_{k=0}^{n-1} \frac{365-k}{365}$$

Important values

- For $n = 23$: $P(\text{at least one match}) \approx 0.507$
- For $n = 57$: $P(\text{at least one match}) \approx 0.99$

Chapter 2 — Jan 16

Review: Conditional Probability

Conditional Probability:

The probability of event B given that event A has occurred is

$$\boxed{P(B \mid A) = \frac{P(A \cap B)}{P(A)}, \quad P(A) > 0}$$

Read as: the probability of B given A.

Independence of Events

Definition (Independence): Events A and B are independent if and only if

$$\boxed{P(B \mid A) = P(B)}$$

Equivalently,

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

or

$$\boxed{P(A \cap B) = P(A)\,P(B)}$$

Multiple Independent Events

Definition:

If events $A_1, A_2, \ldots, A_k$ are independent, then

$$\boxed{P(A_1 \cap A_2 \cap \cdots \cap A_k) = P(A_1)\,P(A_2) \cdots P(A_k)}$$

Mutual Independence

Mutual Independence : A collection of events $A_1, A_2, \ldots, A_n$ is mutually independent if and only if for *every* subcollection $\{A_{i_1}, \ldots, A_{i_k}\}$,

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

Example (Three Events):

Events A_1, A_2, A_3 are mutually independent if all of the following hold:

$$P(A_1 \cap A_2) = P(A_1)P(A_2)$$

$$P(A_1 \cap A_3) = P(A_1)P(A_3)$$

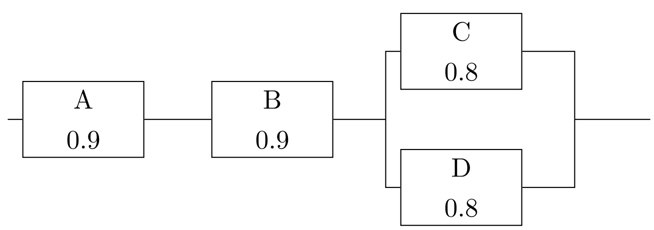
$$P(A_2 \cap A_3) = P(A_2)P(A_3)$$

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3)$$

Note: Mutually exclusive events are dependent. If one event occurs, the other cannot occur.

Example: Component Reliability

An electrical system has four components A, B, C, D. The system works if A and B work and at least one of C or D works. Assume all components are independent.



$$P(A) = 0.9, \quad P(B) = 0.9, \quad P(C) = 0.8, \quad P(D) = 0.8$$

(a) Probability the entire system works

The system works if A and B work and either C or D works.

$$\begin{aligned} P(\text{system works}) &= P(\text{all work}) + P(A, B, C \text{ work}, D \text{ does not}) \\ &\quad + P(A, B, D \text{ work}, C \text{ does not}) \end{aligned}$$

$$= (0.9)(0.9)(0.8)(0.8) + (0.9)(0.9)(0.8)(1-0.8) + (0.9)(0.9)(0.8)(1-0.8)$$

$$\boxed{P(\text{system works}) = 0.7776}$$

(b) Conditional probability

$$P(C^c \mid \text{system works}) = \frac{P(C^c \cap \text{system works})}{P(\text{system works})}$$

$$P(C^c \cap \text{system works}) = (0.9)(0.9)(0.8)(1 - 0.8)$$

$$\boxed{P(C^c \mid \text{system works}) = \frac{(0.9)(0.9)(0.8)(1-0.8)}{0.7776} = 0.16}$$

Theorem of Total Probability

Let $B_1, B_2, \ldots, B_k$ be a partition of the sample space S such that $P(B_i) > 0$ for all i. Then for any event $A \subseteq S$,

$$\boxed{P(A) = \sum_{i=1}^{k} P(A \mid B_i)\, P(B_i) = \sum_{i=1}^{k} P(A \cap B_i)}$$

Theorem: Bayes' Rule (1701–1761)

Let $B_1, B_2, \ldots, B_k$ be a partition of the sample space S such that $P(B_i) > 0$ for $i = 1, \ldots, k$. For any event $A \subseteq S$ with $P(A) > 0$,

$$\boxed{P(B_r \mid A) = \frac{P(B_r \cap A)}{\sum_{i=1}^{k} P(B_i \cap A)} = \frac{P(B_r)\, P(A \mid B_r)}{\sum_{i=1}^{k} P(B_i)\, P(A \mid B_i)}, \quad r = 1, \ldots, k}$$

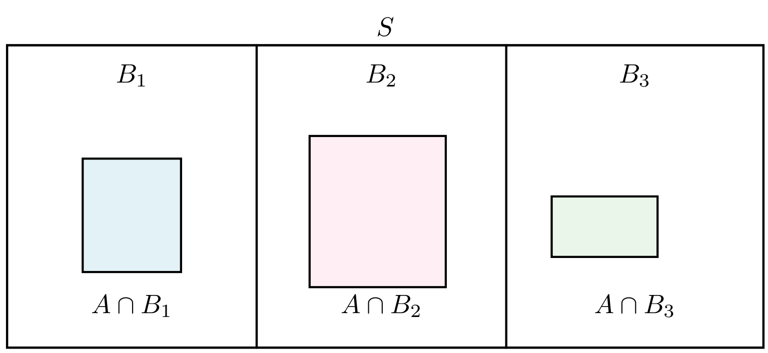


Figure 3: Partition of S into B_1, B_2, B_3 with shaded regions $A \cap B_i$

Example (Medical Test)

The fraction of people in a population who have a certain disease is 0.01.

$$P(D) = 0.01, \quad P(D^c) = 0.99$$

The test characteristics are:

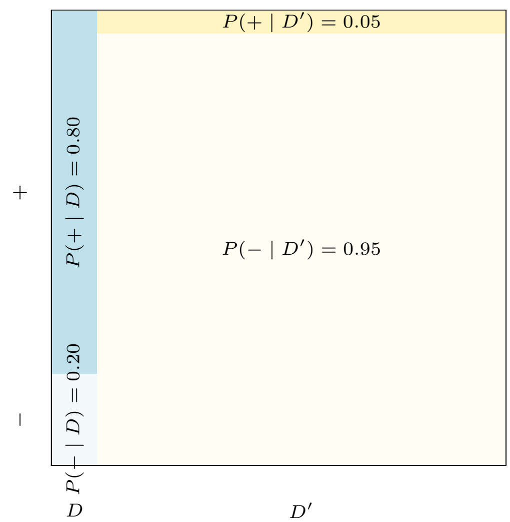
$$P(\text{test says } D \mid D^c) = 0.05 \quad \text{(false positive rate)}$$

$$P(\text{test says } D^c \mid D) = 0.20 \quad \text{(false negative rate)}$$

Thus,

$$P(\text{test says } D \mid D) = 1 - 0.20 = 0.80$$

<u>Note:</u> $1 - P(\text{test says } D^c \mid D)$ is called the <u>sensitivity</u> of the test, and $1 - P(\text{test says } D \mid D^c)$ is called the <u>specificity</u>.



(a) Probability the test says disease

$$P(\text{test says } D) = P(D \cap \text{test says } D) + P(D^c \cap \text{test says } D)$$

$$= P(\text{test says } D \mid D)P(D) + P(\text{test says } D \mid D^c)P(D^c)$$

$$= (0.80)(0.01) + (0.05)(0.99) = 0.0575$$

(b) Probability of disease given positive test

$$P(D \mid \text{test says } D) = \frac{P(D \cap \text{test says } D)}{P(\text{test says } D)}$$

$$= \frac{P(\text{test says } D \mid D)P(D)}{0.0575} = \frac{(0.80)(0.01)}{0.0575}$$

$$\boxed{P(D \mid \text{test says } D) \approx 0.139}$$

(c) Probability of disease given negative test

$$P(D \mid \text{test says } D^c) = \frac{P(D \cap \text{test says } D^c)}{P(\text{test says } D^c)}$$

$$= \frac{P(\text{test says } D^c \mid D)P(D)}{1 - P(\text{test says } D)}$$

$$= \frac{(0.20)(0.01)}{1 - 0.0575}$$

$$\boxed{P(D \mid \text{test says } D^c) \approx 0.00212}$$

Chapter 3 — January 19

Random Variables and Their Interpretation

Definition: A random variable (r.v.) is a rule that assigns a **real number** to each outcome in the sample space.
Alternative definition: A random variable is a function that takes the outcome of an experiment and assigns it a number so that probabilities can be calculated.

Example 1: Three Electronic Components

Each component is classified as either defective (D) or non-defective (N).

$$S = \{NNN, DNN, NDN, NND, DDN, DND, NDD, DDD\}$$

- **Defective (D):** the component does not meet required specifications and fails inspection.
- **Non-defective (N):** the component meets specifications and passes inspection.

Define the random variable

$$X = \text{number of defective components.}$$

Then:

$$\begin{aligned} X &= 0 \quad \text{for } \{NNN\} \\ X &= 1 \quad \text{for } \{DNN, NDN, NND\} \\ X &= 2 \quad \text{for } \{DDN, DND, NDD\} \\ X &= 3 \quad \text{for } \{DDD\} \end{aligned}$$

Thus, the possible values of X are:

$$\{0, 1, 2, 3\}.$$

Example 2: One Component (Dummy Variable)

$$S = \{D, N\}$$

Define the random variable

$$X = \begin{cases} 1, & \text{if the component is defective (D)} \\ 0, & \text{if the component is non-defective (N)} \end{cases}$$

This is called a **dummy variable** because the outcome is categorical, but is encoded numerically.
A <u>dummy variable</u> is a special type of random variable that assigns numerical labels to categorical outcomes, where the numbers have no quantitative meaning beyond identification.

Discrete Random Variables

Definition: A random variable is called <u>discrete</u> if its set of possible values is **countable** (finite or countably infinite).

Example 3: Sampling Until First Defective

Components are tested one(independently) at a time until the first defective component is observed.

$$S = \{D,\ ND,\ NND,\ NNND, \ldots\}$$

Define

$$X = \text{number of components tested until the first defective.}$$

Then:

$$\begin{aligned} X &= 1 \quad \text{for } \{D\} \\ X &= 2 \quad \text{for } \{ND\} \\ X &= 3 \quad \text{for } \{NND\} \\ &\vdots \end{aligned}$$

Hence,

$$X = 1, 2, 3, \ldots$$

Since the possible values can be listed, X is a discrete random variable.

Non-discrete version of the same experiment:

Define

$$Y = \text{time (in seconds) until the first defective component is observed.}$$

Since Y can take any real value in $[0, \infty)$, it cannot be listed and is therefore a continuous (non-discrete) random variable.

Discrete vs. Continuous Random Variables

Discrete Random Variable	**Continuous Random Variable**
Counts things	Measures things
Possible values are countable	Possible values fill an interval
$P(X = x)$ can be > 0	$P(X = x) = 0$ for all x
Uses a probability mass function (PMF)	Uses a probability density function (PDF)

Probability Mass Function (PMF)

Definition:

Let X be a discrete random variable. The probability mass function (PMF) of X, denoted $f(x)$, is defined by:

$$\boxed{\begin{array}{ll} 1) & f(x) \geq 0 \quad \text{for all } x \\ 2) & \displaystyle\sum_x f(x) = 1 \end{array}}$$

Note:

- Capital X: random variable
- Lowercase x: a specific value

Bernoulli and Binomial Random Variables

I. Bernoulli Random Variable (Single Trial)

A Bernoulli random variable: X models a single experiment with only two possible outcomes: success or failure.

$$X = \begin{cases} 1, & \text{success} \\ 0, & \text{failure} \end{cases}$$

If $p = P(X = 1)$, then the PMF is

x	0	1
$P(X = x)$	$1 - p$	p

Here, p is the probability of success (e.g. observing a defective component).

Binomial Random Variable (Multiple Bernoulli Trials)

The binomial random variable extends the Bernoulli case to multiple independent trials.

Definition:

A random variable X is called a binomial random variable if it represents the number of successes in n independent Bernoulli trials, each with success probability p.

$$X = \text{number of successes in } n \text{ trials}$$

In this case,

$$X \sim \text{Bin}(n, p)$$

and the probability mass function is

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

Conditions for a Binomial Model

A binomial model applies only if:

- each trial has exactly two outcomes (success or failure),
- the probability of success p is the same for every trial,
- the trials are independent,
- the number of trials n is fixed.

Example: Three Components Tested

Assume each component is defective with probability

$$p = 0.1, \quad n = 3.$$

where p = P(a single component is defective), n = number of trials

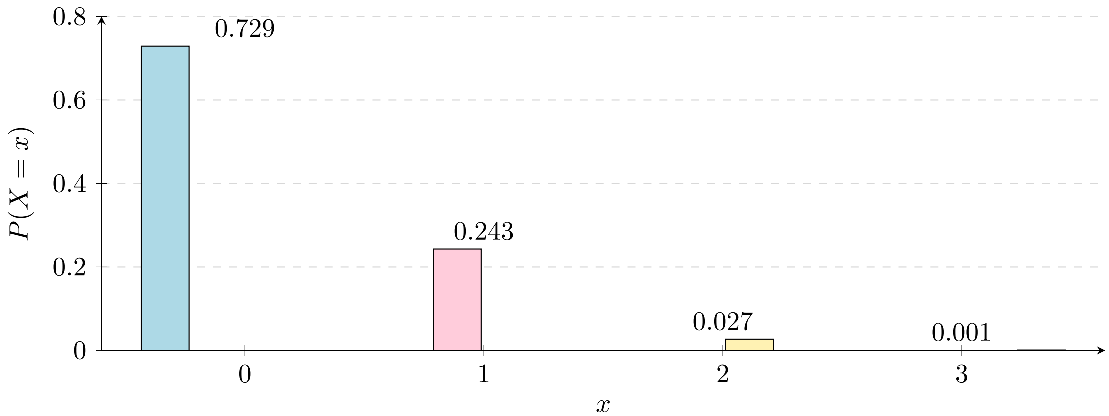
Let

$$X = \text{number of defective components.}$$

x	$P(X = x)$
0	$\binom{3}{0}(0.9)^3 = 0.729$
1	$\binom{3}{1}(0.1)(0.9)^2 = 0.243$
2	$\binom{3}{2}(0.1)^2(0.9) = 0.027$
3	$\binom{3}{3}(0.1)^3 = 0.001$

$$0.729 + 0.243 + 0.027 + 0.001 = 1.$$

PMF of X



Geometric Random Variable

Example: Sampling Until First Defective

Components are sampled one at a time until the first defective component is observed. Assume the probability that a component is defective is

$$p = 0.1.$$

Define the random variable

$$X = \text{number of samples collected until the first defective.}$$

x	$P(X = x) = f(x)$
1	0.1
2	$0.9(0.1)$
3	$0.9^2(0.1)$
$\vdots$	$\vdots$

Geometric Random Variable

Definition:

A random variable X is called a geometric random variable if it represents the number of trials needed to obtain the first success in a sequence of independent Bernoulli trials with success probability p.

$$\boxed{P(X = x) = (1-p)^{x-1}p, \quad x = 1, 2, 3, \ldots}$$

In this example,

$$P(X = x) = 0.9^{x-1}(0.1).$$

Verification That Probabilities Sum to 1

$$\sum_{x=1}^{\infty} 0.9^{x-1}(0.1) = 0.1 \sum_{x=0}^{\infty} 0.9^{x} = 0.1\left(\frac{1}{1-0.9}\right) = 1.$$

Cumulative Distribution Function (CDF)

Definition:

The cumulative distribution function (CDF) of a discrete random variable X with PMF $f(x)$ is defined as

$$\boxed{F(x) = P(X \le x) = \sum_{t \le x} f(t), \quad -\infty < x < \infty.}$$

Example: Binomial Distribution (Three Components)

Let X be the number of defective components when three components are tested, with

$$P(X=0) = 0.729, \quad P(X=1) = 0.243, \quad P(X=2) = 0.027, \quad P(X=3) = 0.001.$$

$$\begin{aligned}
F(0) &= P(X \le 0) = 0.729 \\
F(1) &= P(X \le 1) = 0.729 + 0.243 = 0.972 \\
F(2) &= P(X \le 2) = 0.729 + 0.243 + 0.027 = 0.999 \\
F(3) &= P(X \le 3) = 0.729 + 0.243 + 0.027 + 0.001 = 1
\end{aligned}$$

Thus, the CDF can be written as

$$F(x) = \begin{cases} 0, & x < 0 \\ 0.729, & 0 \le x < 1 \\ 0.972, & 1 \le x < 2 \\ 0.999, & 2 \le x < 3 \\ 1, & x \ge 3 \end{cases}$$

Properties of the CDF

- $F(x)$ is monotone non-decreasing.

- If $x < y$, then $F(x) \le F(y)$.
- $0 \le F(x) \le 1$.

Note: A function is **monotone** non-decreasing if its value never decreases as the input increases.

Using the CDF to Compute Probabilities

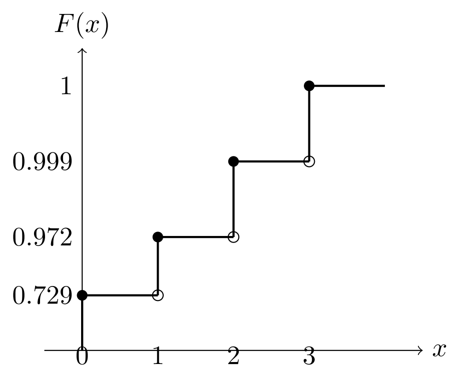
For $a < b$,

$$P(a < X \le b) = F(b) - F(a).$$

Example:

$$P(0 < X \le 2) = F(2) - F(0) = 0.999 - 0.729 = 0.27.$$

CDF Histogram (Step Function)



Note: For a discrete random variable, the PMF is drawn as a bar chart since it shows probabilities at individual points, while the CDF is drawn as a step function since it represents cumulative probability and is monotone non-decreasing.

Chapter 3 — January 21

Review

1. *Random Variable (RV), X*: A random variable assigns a real number to each outcome.
2. *Discrete Random Variable*: If X is discrete,
   - $P(X = x) = f(x)$
   - $f(x)$ is the *probability mass function (PMF)*
   - $f(x) \ge 0$
   - $\sum_x f(x) = 1$
3. *Cumulative Distribution Function (CDF), $F(x)$*:

- $F(x) = P(X \leq x)$
- If X is discrete:

$$F(x) = \sum_{t \leq x} f(t), \quad -\infty < x < \infty$$

Continuous Sample Space and Continuous Random Variables

If the sample space contains an infinite number of outcomes equal to the number of points on a line segment, it is called a continuous sample space.

A continuous random variable has

$$\boxed{P(X = x) = 0 \quad \text{for all } x,}$$

so probabilities are computed over intervals instead of single values.

***Alternative definition**: A continuous sample space contains infinitely many outcomes, like the points on a line segment. For a continuous random variable, the probability of taking any exact value is zero, i.e. $P(X = x) = 0$ for all x. Therefore, probabilities are computed over intervals rather than at single points.

Probability Density Function (PDF)

A function $f(x)$ is called a probability density function (PDF) of a continuous random variable X, defined over $\mathbb{R}$, if:

1. $f(x) \geq 0$ for all $x \in \mathbb{R}$
2. $\displaystyle\int_{-\infty}^{\infty} f(x)\, dx = 1$
3. For any $a < b$,

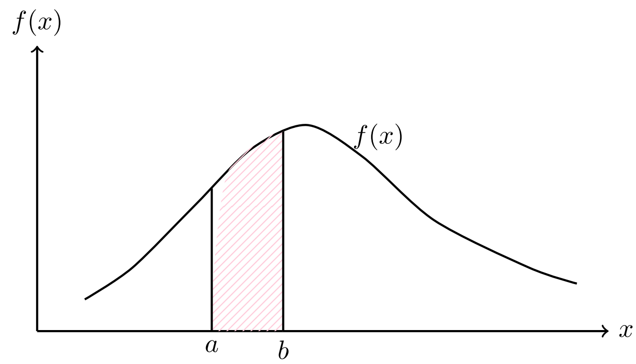
$$P(a < X < b) = \int_a^b f(x)\, dx$$

For continuous random variables,

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

$$\boxed{P(X = a) = 0}$$

because a single point has zero area under the probability density function.



$$\int_a^b f(x)\,dx = P(a < X < b)$$

Example: Uniform Distribution

Let the probability density function be

$$f(x) = \begin{cases} c, & 5 < x < 10, \\ 0, & \text{otherwise.} \end{cases}$$

1) Determine the value of c

Since $f(x)$ is a probability density function, the total area under the curve must equal 1:

$$\int_{-\infty}^{\infty} f(x)\,dx = 1.$$

Because $f(x) = 0$ outside the interval $(5, 10)$,

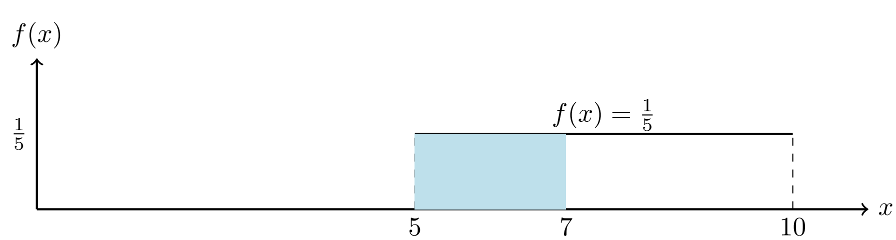
$$\int_5^{10} c\,dx = 1.$$

Evaluating the integral,

$$c(10-5) = 1 \quad \Rightarrow \quad c = \frac{1}{5}.$$

Thus,

$$f(x) = \begin{cases} \frac{1}{5}, & 5 < x < 10, \\ 0, & \text{otherwise.} \end{cases}$$



2): Compute $P(X < 7)$

Formula used:

$$P(a < X < b) = \int_a^b f(x)\, dx.$$

Applying this formula,

$$P(X < 7) = \int_5^7 \frac{1}{5}\, dx = \frac{1}{5}(7 - 5) = \frac{2}{5}.$$

CDF for Continuous Random Variables

Def: Let X be a continuous random variable with pdf: probability density function $f(x)$. The cumulative distribution functio
is

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)\, dt.$$

Consequently:

$$\boxed{f(x) = \frac{d}{dx} F(x), \qquad P(a < X < b) = F(b) - F(a).}$$

Example

Let X be the time until a chemical reaction is complete (in msec). Suppose the CDF is

$$F(x) = \begin{cases} 0, & x < 0, \\ 1 - e^{-0.01x}, & x \ge 0. \end{cases}$$

(a) Find the pdf.

Use $f(x) = \dfrac{d}{dx} F(x)$.
For $x < 0$, $F(x) = 0$, so

$$f(x) = 0.$$

For $x \ge 0$,

$$f(x) = \frac{d}{dx}\left(1 - e^{-0.01x}\right) = 0.01e^{-0.01x}.$$

Therefore,

$$f(x) = \begin{cases} 0, & x < 0, \\ 0.01e^{-0.01x}, & x \ge 0. \end{cases}$$

(b) Find $P(X < 200)$**.**

Use the CDF directly:

$$P(X < 200) = F(200) = 1 - e^{-0.01(200)} = 1 - e^{-2} \approx 0.8647.$$

(c) Check if this is a valid CDF.

Theorem: Properties of a Cumulative Distribution Function

A function $F(x)$ is a valid cumulative distribution function (CDF) if and only if:

- $0 \le F(x) \le 1$ for all x
- $F(x)$ is monotone non-decreasing, i.e.

$$x \le y \implies F(x) \le F(y)$$

- $\lim\limits_{x \to -\infty} F(x) = 0$
- $\lim\limits_{x \to \infty} F(x) = 1$

For this $F(x)$:

$$\lim_{x \to -\infty} F(x) = 0, \qquad \lim_{x \to \infty} F(x) = 1,$$

and for $x \ge 0$, $1 - e^{-0.01x}$ increases as x increases, so $F(x)$ is monotone non-decreasing.

Thus, $F(x)$ is a valid CDF.

Joint Probability Distributions (Discrete)

Def: The function $f(x, y)$ is a joint probability mass function (PMF) of discrete random variables X and Y if:

1. $f(x, y) \ge 0$ for all (x, y)
2. $\sum\limits_{x} \sum\limits_{y} f(x, y) = 1$
3. $P(X = x, Y = y) = f(x, y)$

That is, $f(x, y)$ gives the probability that the two random variables X and Y take the values x and y *simultaneously*.

For any region A in the xy-plane,

$$\boxed{P\big((X, Y) \in A\big) = \sum_{(x,y) \in A} f(x, y)}$$

Example: Pen Refills

Two refills are selected at random and without replacement from a box containing:

- 3 blue refills
- 2 red refills
- 3 green refills

Define the random variables

$$X = \text{number of blue refills selected}, \qquad Y = \text{number of red refills selected}.$$

The total number of possible selections is

$$\binom{8}{2}.$$

Joint PMF Table

For each pair (x, y),

$$f(x,y) = \frac{\text{number of favorable outcomes}}{\binom{8}{2}}.$$

$X \backslash Y$	0	1	2	Row Total
0	$\frac{\binom{3}{2}}{\binom{8}{2}}$	$\frac{\binom{2}{1}\binom{3}{1}}{\binom{8}{2}}$	$\frac{\binom{2}{2}}{\binom{8}{2}}$	$\frac{\binom{5}{2}}{\binom{8}{2}}$
1	$\frac{\binom{3}{1}\binom{3}{1}}{\binom{8}{2}}$	$\frac{\binom{3}{1}\binom{2}{1}}{\binom{8}{2}}$	0	$\frac{15}{\binom{8}{2}}$
2	$\frac{\binom{3}{2}}{\binom{8}{2}}$	0	0	$\frac{3}{\binom{8}{2}}$
Column Total	$\frac{15}{\binom{8}{2}}$	$\frac{12}{\binom{8}{2}}$	$\frac{1}{\binom{8}{2}}$	1

Marginal Distributions

Def: Let $f(x, y)$ be the joint PMF of X and Y.

The marginal PMF of X is obtained by summing over all values of Y:

$$g(x) = \sum_y f(x,y).$$

The marginal PMF of Y is obtained by summing over all values of X:

$$h(y) = \sum_x f(x,y).$$

Marginal example:

$$P(X=1) = \sum_y P(X=1, Y=y).$$

Conditional Distributions (Discrete)

Def: The conditional PMF of Y given $X = x$ is

$$f_{Y|X}(y|x) = \frac{f(x,y)}{g(x)}, \quad g(x) > 0.$$

Similarly, the conditional PMF of X given $Y = y$ is

$$f_{X|Y}(x|y) = \frac{f(x,y)}{h(y)}, \quad h(y) > 0.$$

Example: Conditional Probabilities

1. $$P(Y = 1 \mid X = 1) = \frac{P(X = 1, Y = 1)}{P(X = 1)} = \frac{3/28}{15/28} = \frac{3}{15}.$$

2. $$P(X = 0 \mid Y = 1) = \frac{P(X = 0, Y = 1)}{P(Y = 1)} = \frac{9/28}{15/28} = \frac{9}{15}.$$

3. $$P(Y = 2 \mid X = 1) = 0.$$

Check:

$$\sum_y P(Y = y \mid X = 1) = 1.$$

Review: Chapter 3 — Random Variables

1. *Random Variable* A random variable is a function that maps outcomes of an experiment to real numbers.
   - Domain: sample space outcomes
   - Range: real numbers
   - Can be **discrete** or **continuous**

2. *Probability Mass Function (PMF)* The PMF of a discrete random variable X is

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

   - Only for discrete random variables
   - $f_X(x) \geq 0$
   - $\sum_x f_X(x) = 1$

3. *Probability Density Function (PDF)* For a continuous random variable X, probability is defined by

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

- Only for continuous random variables
- Area under the curve gives probability
- $P(X = x) = 0$

4. *Cumulative Distribution Function (CDF)* The CDF is defined as

$$F_X(x) = P(X \leq x)$$

- Discrete: $F(x) = \sum_{t \leq x} f(t)$
- Continuous: $F(x) = \int_{-\infty}^{x} f(t)\, dt$
- $0 \leq F(x) \leq 1$, non-decreasing

5. *Joint Distribution* The joint distribution describes probabilities involving two random variables X and Y.

- Discrete: $f_{X,Y}(x, y) = P(X = x, Y = y)$
- Continuous: joint PDF $f_{X,Y}(x, y)$

6. *Marginal Distribution* A marginal distribution is obtained by eliminating the other variable.

- $f_X(x) = \sum_y f_{X,Y}(x, y)$ or $f_X(x) = \int f_{X,Y}(x, y)\, dy$
- $f_Y(y) = \sum_x f_{X,Y}(x, y)$ or $f_Y(y) = \int f_{X,Y}(x, y)\, dx$

Joint Probability Density Function (Continuous)

Def: A function $f(x, y)$ is a joint probability density function (PDF) of continuous random variables X and Y if:

1. $f(x, y) \geq 0$ for all (x, y)
2. $$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dx\, dy = 1$$
3. For any region A in the xy-plane,

$$P\big((X, Y) \in A\big) = \iint_A f(x, y)\, dx\, dy$$

Geometric interpretation: The joint PDF is a surface above the xy-plane. Probabilities correspond to the **volume under the surface** over a specified region.

Example 1:

$$f(x, y) = \begin{cases} \dfrac{12}{7}(x^2 + xy), & 0 \leq x \leq 1,\ 0 \leq y \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

(a) Verify it is a valid joint PDF:

$$\int_0^1 \int_0^1 \frac{12}{7}(x^2 + xy)\, dx\, dy = 1$$

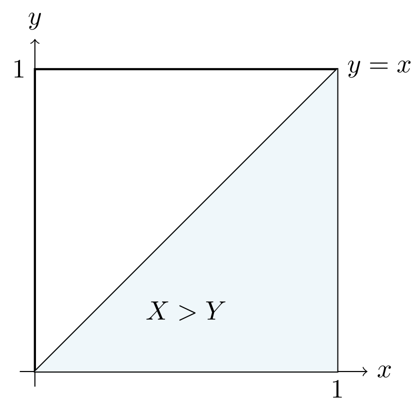
(b) Find $P(0 < X < 0.2,\ 0 < Y < 1)$**:**

$$P = \int_0^1 \int_0^{0.2} \frac{12}{7}(x^2 + xy)\, dx\, dy$$

(c) Find $P(X > Y)$**:**

The region $X > Y$ corresponds to the area below the line $y = x$ in the unit square $0 \leq x \leq 1,\ 0 \leq y \leq 1$.

$$P(X > Y) = \int_0^1 \int_0^x \frac{12}{7}(x^2 + xy)\, dy\, dx = \frac{9}{14}$$



(d) Find $P(X = Y)$**:**

$$P(X = Y) = \int_0^1 \int_y^y \frac{12}{7}(x^2 + xy)\, dx\, dy = 0$$

(Probability along a line is zero for continuous random variables.)

Example:

Let the joint PDF be

$$f(x, y) = \begin{cases} e^{-y}, & 0 < x < y < \infty \\ 0, & \text{otherwise} \end{cases}$$

We want to compute

$$P(X + Y \geq 1).$$

The support of the joint PDF is the region $0 < x < y$, which lies above the line $y = x$ in the first quadrant.

The boundary of the event $X + Y \geq 1$ is the line $x + y = 1$.

Rather than integrating over the unbounded region $X+Y \geq 1$, we compute the complement:

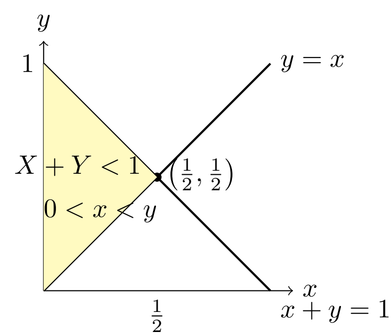
$$P(X+Y \geq 1) = 1 - P(X+Y < 1).$$

The region $X+Y<1$ that lies within the support is bounded by:

$$0 \leq x \leq \tfrac{1}{2}, \qquad x \leq y \leq 1-x.$$

Therefore,

$$P(X+Y \geq 1) = 1 - \int_0^{1/2} \int_x^{1-x} e^{-y}\, dy\, dx = 2e^{-1/2} - e^{-1}.$$



Marginal and Conditional PDFs

Def: The marginal PDFs of X and Y are defined as

$$g_X(x) = \int_{-\infty}^{\infty} f(x,y)\, dy, \qquad h_Y(y) = \int_{-\infty}^{\infty} f(x,y)\, dx$$

Example (continued):

$$f(x,y) = \begin{cases} \dfrac{12}{7}(x^2+xy), & 0 \leq x \leq 1,\ 0 \leq y \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

Marginal PDF of X:

$$g_X(x) = \int_0^1 \frac{12}{7}(x^2+xy)\, dy = \frac{12}{7}\left(x^2 + \frac{x}{2}\right), \qquad 0 \leq x \leq 1$$

Marginal PDF of Y:

$$h_Y(y) = \int_0^1 \frac{12}{7}(x^2+xy)\, dx = \frac{12}{7}\left(\frac{1}{3} + \frac{y}{2}\right), \qquad 0 \leq y \leq 1$$

Conditional PDF of Y given $X=x$:

$$f(y|x) = \frac{f(x,y)}{g(x)} = \frac{\frac{12}{7}(x^2+xy)}{\frac{12}{7}\left(x^2+\frac{x}{2}\right)}, \qquad 0 \le y \le 1,\ 0 < x \le 1$$

Why the bounds are $0 < x \le 1$ and not $0 \le x \le 1$:

The marginal PDF

$$g(x) = \frac{12}{7}\left(x^2 + \frac{x}{2}\right)$$

satisfies $g(0) = 0$.

Since the conditional PDF is defined as

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

it is **undefined at $x = 0$** due to division by zero.

Therefore, the conditional density is only defined for values of x such that

$$g(x) > 0 \ \Rightarrow\ 0 < x \le 1.$$

Key takeaway: The bounds of a conditional PDF exclude points where the conditioning density is zero.

Statistical Independence

Def: Random variables X and Y (discrete or continuous) are statistically independent if and only if

$$f(x,y) = g(x)\,h(y) \quad \text{for all } (x,y) \text{ in their range}$$

Consequences:

- $f(x|y) = g(x)$
- $f(y|x) = h(y)$

Recall Example: Pen Refills

Two refills are selected at random and without replacement from a box containing:

- 3 blue refills
- 2 red refills
- 3 green refills

Define the random variables

$$X = \text{number of blue refills selected}, \qquad Y = \text{number of red refills selected}.$$

Total outcomes:

$$\binom{8}{2} = 28.$$

Joint PMF Table

For each pair (x, y),

$$f(x, y) = \frac{\text{number of favorable outcomes}}{28}.$$

$X \backslash Y$	0	1	2	$g(x) = P(X = x)$
0	$\frac{3}{28}$	$\frac{3}{14}$	$\frac{1}{28}$	$\frac{5}{14}$
1	$\frac{9}{28}$	$\frac{3}{14}$	0	$\frac{15}{28}$
2	$\frac{3}{28}$	0	0	$\frac{3}{28}$
$h(y) = P(Y = y)$	$\frac{15}{28}$	$\frac{3}{7}$	$\frac{1}{28}$	1

Marginals: $g(x) = P(X = x)$ is the marginal PMF of X and is given by the **row totals**. $h(y) = P(Y = y)$ is the marginal PMF of Y and is given by the **column totals**.

$$g(0) = \frac{5}{14}, \quad g(1) = \frac{15}{28}, \quad g(2) = \frac{3}{28}$$
$$h(0) = \frac{15}{28}, \quad h(1) = \frac{3}{7}, \quad h(2) = \frac{1}{28}$$

Independence check: If X and Y were statistically independent, then $f(x, y) = g(x)h(y)$ for all (x, y). Check $(x, y) = (0, 1)$:

$$f(0, 1) = \frac{3}{14}, \qquad g(0)h(1) = \left(\frac{5}{14}\right)\left(\frac{3}{7}\right) = \frac{15}{98}$$

$$\frac{3}{14} \neq \frac{15}{98} \qquad \Rightarrow \qquad X \text{ and } Y \text{ are not independent.}$$

Example: Independence via Factorization (Discrete Case)

Let X and Y be discrete random variables whose values in the nonnegative integers.

Suppose the joint PMF is

$$f(x, y) = \frac{1}{x!\, y!} \lambda^x \mu^y e^{-(\lambda+\mu)}, \qquad x, y = 0, 1, 2, \ldots$$

Factorization:

We can write

$$f(x, y) = \left(\frac{\lambda^x e^{-\lambda}}{x!}\right)\left(\frac{\mu^y e^{-\mu}}{y!}\right) = g(x)\, h(y)$$

Marginal PMF of X:

$$g(x) = \sum_{y=0}^{\infty} f(x,y) = \sum_{y=0}^{\infty} \frac{1}{x!\,y!}\,\lambda^x \mu^y\, e^{-(\lambda+\mu)}$$

Factor out terms that do not depend on y:

$$g(x) = \frac{1}{x!}\,\lambda^x e^{-(\lambda+\mu)} \sum_{y=0}^{\infty} \frac{\mu^y}{y!}$$

Using $\displaystyle\sum_{y=0}^{\infty} \frac{\mu^y}{y!} = e^{\mu}$:

$$g(x) = \frac{1}{x!}\,\lambda^x e^{-(\lambda+\mu)} e^{\mu} = \frac{1}{x!}\,\lambda^x e^{-\lambda}, \qquad x = 0, 1, 2, \ldots$$

Marginal PMF of Y:

$$h(y) = \sum_{x=0}^{\infty} f(x,y) = \sum_{x=0}^{\infty} \frac{1}{x!\,y!}\,\lambda^x \mu^y\, e^{-(\lambda+\mu)}$$

Factor out terms that do not depend on x:

$$h(y) = \frac{1}{y!}\,\mu^y e^{-(\lambda+\mu)} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!}$$

Using $\displaystyle\sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{\lambda}$:

$$h(y) = \frac{1}{y!}\,\mu^y e^{-(\lambda+\mu)} e^{\lambda} = \frac{1}{y!}\,\mu^y e^{-\mu}, \qquad y = 0, 1, 2, \ldots$$

Conclusion:

Since the joint PMF can be written as

$$f(x,y) = g(x)\,h(y)$$

for all x, y, the random variables X and Y are statistically independent. ✓

Important notes:

- Factorization of the joint PMF is **sufficient** to prove independence.
- The constant terms (such as $e^{-\lambda}$, $e^{-\mu}$) must be included to obtain the **correct marginals**.
- Independence requires that every combination of values with positive marginal probability also has positive joint probability.
  This means that if $X = x$ is possible and $Y = y$ is possible, then the pair $(X = x, Y = y)$ must also be possible.

$$\{(x,y) : f(x,y) > 0\} = \{x : g(x) > 0\} \times \{y : h(y) > 0\}.$$

This means that every value of X with positive marginal probability can occur with every value of Y with positive marginal probability.

Jan 30

Joint distribution

Describes probabilities involving two random variables X and Y.

- Discrete:
$$f(X,Y) = P(X = x, Y = y)$$
- Continuous: joint PDF $f(X,Y)$

Marginal distribution

Obtained by eliminating the other variable.

-
$$f(X) = \sum_y f(X,Y), \qquad f(Y) = \sum_x f(X,Y)$$
-
$$f(X) = \int_{-\infty}^{\infty} f(X,Y)\,dy, \qquad f(Y) = \int_{-\infty}^{\infty} f(X,Y)\,dx$$

Joint PDF validity

A function $f(X,Y)$ is a valid joint PDF if:

- $f(X,Y) \geq 0$
-
$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(X,Y)\,dx\,dy = 1$$

Example (valid joint PDF):

$$f(X,Y) = \begin{cases} \frac{12}{7}(x^2 + xy), & 0 \leq x \leq 1,\ 0 \leq y \leq 1, \\ 0, & \text{otherwise} \end{cases}$$

Geometric interpretation

Probabilities correspond to the **volume under the surface** $f(X,Y)$ over a region.

Conditional PDF

Distribution of one variable given the other.

- $$f(Y \mid X) = \frac{f(X,Y)}{f(X)}, \quad f(X) > 0$$

Statistical Independence

Definition: Random variables X and Y are statistically independent if

$$f(X,Y) = f(X)f(Y)$$

This means knowing the value of one variable gives no information about the other.

Support and Independence

Support: The support of a joint distribution is the set

$$\{(x,y) : f(X,Y) > 0\}$$

Key idea: If X and Y are independent, the support must factor as

$$\{x : f(X) > 0\} \times \{y : f(Y) > 0\}$$

That is, every allowed value of X can occur with every allowed value of Y.

Example: Non-Independent Random Variables

$$f(X,Y) = \begin{cases} 4(x+y^2), & xy > 0,\ x+y \le 1, \\ 0, & \text{otherwise} \end{cases}$$

The condition $x + y \le 1$ links x and y, so the support does not factor.

$$\Rightarrow X \text{ and } Y \text{ are not independent}$$

Why this implies dependence (key intuition):

Pick a value of X that is allowed:

$$x = 0.8 \quad (\text{positive and } < 1)$$

Pick a value of Y that is allowed:

$$y = 0.8 \quad (\text{positive and } < 1)$$

Individually, both values are valid. But together:

$$x + y = 0.8 + 0.8 = 1.6 > 1$$

This violates the condition $x + y \le 1$, so the pair $(0.8, 0.8)$ is impossible.

Conclusion: Knowing the value of X restricts which values Y can take. Therefore, X and Y are not independent.

Independent Random Variables

If random variables are independent, joint probabilities factor.

Example: Let X_1, X_2, X_3 be independent with

$$f(X) = e^{-x}, \quad x > 0$$

Then

$$f(X_1, X_2, X_3) = e^{-x_1} e^{-x_2} e^{-x_3}$$

$$P(X_1 < 2,\ 1 < X_2 < 3,\ X_3 > 2) = (1 - e^{-2})(e^{-1} - e^{-3})e^{-2}$$

Explanation (why this works):

Since X_1, X_2, X_3 are independent, the joint probability factors:

$$P(X_1 < 2,\ 1 < X_2 < 3,\ X_3 > 2) = P(X_1 < 2)\, P(1 < X_2 < 3)\, P(X_3 > 2)$$

For an exponential random variable with

$$f(X) = e^{-x}, \quad x > 0,$$

we have:

$$P(X < a) = 1 - e^{-a}, \qquad P(X > a) = e^{-a}$$

Thus,

$$P(X_1 < 2) = 1 - e^{-2}$$

$$P(1 < X_2 < 3) = P(X_2 < 3) - P(X_2 < 1) = e^{-1} - e^{-3}$$

$$P(X_3 > 2) = e^{-2}$$

Multiplying gives:

$$(1 - e^{-2})(e^{-1} - e^{-3})e^{-2}$$

Mutual Independence and Modeling Assumptions

Mutual independence: Random variables $X_1, \ldots, X_n$ are mutually independent if

$$f(X_1, \ldots, X_n) = \prod_{i=1}^{n} f(X_i)$$

Pairwise independence does **not** imply mutual independence.

Independent selection: Selections are independent if:

- each selection is random,
- distributions are identical,
- outcomes do not affect future selections.

Sampling without replacement generally produces dependent variables.

Chapter 4

Motivation

Example: If you roll a fair die repeatedly, what *average value* do you expect in the long run?

This motivates the idea of expectation: a theoretical long-run average.

Expected Value

Expected Value of a Random Variable

Def: For a random variable X, the expectation (or expected value or mean) is the long-run average value of X.

Discrete random variable:

$$\boxed{\mu = E(X) = \sum_{x} x\, f(X)}$$

Continuous random variable:

$$\boxed{\mu = E(X) = \int_{-\infty}^{\infty} x\, f(X)\, dx}$$

Interpretation: Expectation is a weighted average, where values of X are weighted by how likely they are.

Example: Fair Die

Let X be the outcome when a fair die is rolled.

$$E(X) = \sum_{x=1}^{6} x \cdot \frac{1}{6}$$

$$E(X) = \frac{1+2+3+4+5+6}{6} = 3.5$$

Even though 3.5 is not a possible outcome, it represents the long-run average.

Example: Number of Messages per Hour

Let X be the number of messages sent per hour, with PMF:

x	10	11	12	13	14	15
$f(X)$	0.08	0.15	0.30	0.20	0.20	0.07

Check:

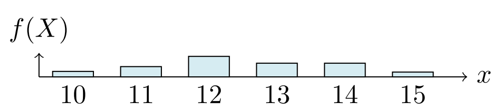
$$\sum f(X) = 1$$

Expected value:

$$\begin{aligned} E(X) &= 10(0.08) + 11(0.15) + 12(0.30) + 13(0.20) + 14(0.20) + 15(0.07) \\ &= 12.5 \end{aligned}$$

This means that over many hours, the average number of messages per hour is about 12.5.

PMF Visualization



Example: Deal or No Deal

Consider a game with two possible outcomes:

- \$1 with probability 1/2
- \$10,000 with probability 1/2

Expected value:

$$E(X) = \frac{1}{2}(1) + \frac{1}{2}(10{,}000) = 5000.5$$

Key idea: The expected value is the average payout in the long run, not the most likely outcome.

Example: Continuous RV (Device Lifetime)

Let X be a random variable that denotes the lifetime (in hours) of a certain device, with PDF

$$f(X) = \begin{cases} \dfrac{20000}{x^3}, & x > 100, \\ 0, & \text{otherwise.} \end{cases}$$

Check (valid PDF):

$$\int_{-\infty}^{\infty} f(X)\,dx = \int_{100}^{\infty} \frac{20000}{x^3}\,dx = 20000\left[\frac{-1}{2x^2}\right]_{100}^{\infty} = 1$$

Expected lifetime:

$$E(X) = \int_{100}^{\infty} x\frac{20000}{x^3}\,dx = \int_{100}^{\infty} \frac{20000}{x^2}\,dx = 20000\left[\frac{-1}{x}\right]_{100}^{\infty} = 200$$

So we expect this type of device to last on average $\boxed{200}$ hours.

Example: Discrete RV and Transformation

Let X be a discrete random variable with PMF:

x	-1	0	1	2
$f(X)$	0.3	0.2	0.3	0.2

Define a new random variable as a transformation of X:

$$g(X) = X^2$$

Possible values of $g(X)$:

$$g(-1) = 1, \quad g(0) = 0, \quad g(1) = 1, \quad g(2) = 4$$

So $g(X)$ can take values $\{0, 1, 4\}$.

Distribution of $g(X)$:

$$P(g(X) = 0) = P(X = 0) = 0.2$$

$$P(g(X) = 1) = P(X = -1) + P(X = 1) = 0.3 + 0.3 = 0.6$$

$$P(g(X) = 4) = P(X = 2) = 0.2$$

$g(X)$	0	1	4
$P(g(X))$	0.2	0.6	0.2

This is called a transformation of a random variable.

Expected value of the transformed RV:

$$\boxed{E(g(X)) = E(X^2) = \sum_x x^2 f(X) = \sum_x g(x)\,f(X)}$$

Numerically:

$$E(X^2) = 0^2(0.2) + (-1)^2(0.3) + (1)^2(0.3) + (2)^2(0.2) = 0 + 0.3 + 0.3 + 0.8 = 1.4$$

Expected Value of a Function of a RV

Let X be a random variable with distribution $f(X)$. The expected value of the random variable $g(X)$ is

$$\boxed{E(g(X)) = \begin{cases} \sum_x g(x)\,f(X), & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} g(x)\,f(X)\,dx, & \text{if } X \text{ is continuous} \end{cases}}$$

Example: Chip Game (Expected Winnings)

A bowl contains 5 chips:

- 3 chips are worth \$1 each
- 2 chips are worth \$4 each

A player draws 2 chips at random (without replacement) and is paid the sum.

Let X be the number of \$1 chips drawn. Then

$$X \in \{0, 1, 2\}.$$

PMF of X: (hypergeometric)

$$f(X) = \begin{cases} \dfrac{\binom{3}{x}\binom{2}{2-x}}{\binom{5}{2}}, & x = 0, 1, 2, \\ 0, & \text{otherwise.} \end{cases}$$

Define payout as a function of X: If you draw x one-dollar chips, then you draw $2 - x$ four-dollar chips, so the payout is

$$g(x) = 1(x) + 4(2 - x) = 8 - 3x.$$

So the payout random variable is $g(X) = 8 - 3X$.

Expected payout:

$$E(g(X)) = \sum_{x=0}^{2} g(x)\,f(X)$$

Compute $f(X)$ values:

$$f(0) = \frac{\binom{3}{0}\binom{2}{2}}{\binom{5}{2}} = \frac{1}{10}, \quad f(1) = \frac{\binom{3}{1}\binom{2}{1}}{\binom{5}{2}} = \frac{6}{10}, \quad f(2) = \frac{\binom{3}{2}\binom{2}{0}}{\binom{5}{2}} = \frac{3}{10}.$$

Then

$$E(g(X)) = \sum_{x=0}^{2}(8-3x)\,f(X) = (8)\left(\frac{1}{10}\right) + (5)\left(\frac{6}{10}\right) + (2)\left(\frac{3}{10}\right) = 4.4$$

Decision: If it costs \$4.75 to play, your expected profit is

$$E(\text{profit}) = E(g(X)) - 4.75 = 4.4 - 4.75 = -0.35$$

So in the long run, you lose about $\boxed{\$0.35}$ per game on average, so you should <u>not</u> play.

Notation: $g(X)$ vs. $g(x)$

- X is a <u>random variable</u>; x is a specific value it can take.
- $g(X)$ is a <u>random variable</u>.
- $g(x)$ is a <u>number</u>.

Key rule:

$$\boxed{E(g(X)) = \sum_x g(x)\,f(X)}$$

Expectation averages the numerical values $g(x)$, weighted by their probabilities.

Exam rule:

$$\boxed{g(X) \text{ is random variable, } g(x) \text{ is a number.}}$$

February 2

Chapter 4 Review: Expectation

1. <u>*Expected value*</u>, $\mu = E(X)$
   - **Discrete:**
     $$E(X) = \sum_x x\,f(X)$$
   - **Continuous:**
     $$E(X) = \int_{-\infty}^{\infty} x\,f(X)\,dx$$
2. <u>*Expected value of a function*</u>, $E[g(X)]$
   - $g(X)$ is a function of X

- **Discrete:**

$$E[g(X)] = \sum_x g(x)\, f(X)$$

- **Continuous:**

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

3. **Example (continuous RV):**

$$f(X) = 4x^3, \quad 0 < x < 1$$

- $$E(X) = \int_0^1 x \cdot 4x^3\, dx$$

- $$E(X^2) = \int_0^1 x^2 \cdot 4x^3\, dx$$

Expected Value of a Function of a Random Variable

1. *Expected value of a function: expected average value in the long wrong*, $E[g(X)]$

- $g(X)$ is a function of X
- **Discrete:**

$$E[g(X)] = \sum_x g(x)\, f(X)$$

- **Continuous:**

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

2. **Example (continuous RV):**

$$f(X) = 4x^3, \quad 0 < x < 1$$

- $$E(X) = \int_0^1 x \cdot 4x^3\, dx$$

- $$E(X^2) = \int_0^1 x^2 \cdot 4x^3\, dx$$

Expected Value from a Joint Distribution

General formula

$$\boxed{E[g(X,Y)] = \int\int g(x,y)\, f(X,Y)\, dx\, dy}$$

Computing $E(X)$ from a joint PDF

There are two equivalent methods to compute $E(X)$.

Method 1: Using the marginal distribution

First find the marginal of X:

$$f(X) = \int_0^1 x(1+3y^2)\,dy$$

$$= x\int_0^1 (1+3y^2)\,dy$$

$$= x\left[y + y^3\right]_0^1 = x(1+1) = 2x$$

Now compute the expected value:

$$E(X) = \int_0^2 x f(X)\,dx = \int_0^2 x(2x)\,dx$$

$$= 2\int_0^2 x^2\,dx = 2\left[\frac{x^3}{3}\right]_0^2 = \frac{16}{3}$$

Method 2: Direct double integral

Apply the general formula with $g(X,Y) = X$:

$$E(X) = \int_0^1 \int_0^2 x\, f(X,Y)\,dx\,dy$$

$$= \int_0^1 \int_0^2 x \cdot x(1+3y^2)\,dx\,dy$$

$$= \int_0^1 (1+3y^2)\left(\int_0^2 x^2\,dx\right)dy$$

$$= \int_0^1 (1+3y^2)\left[\frac{x^3}{3}\right]_0^2 dy = \int_0^1 (1+3y^2)\frac{8}{3}\,dy$$

$$= \frac{8}{3}\int_0^1 (1+3y^2)\,dy = \frac{8}{3}\cdot 2 = \frac{16}{3}$$

Both methods give the same result:

$$\boxed{E(X) = \frac{16}{3}}$$

Properties of Expected Value

Properties of Expectation

Let X, Y be random variables and let $a, b \in \mathbb{R}$ be constants.

1. **Linearity (scaling and shifting):**

$$\boxed{E(aX+b) = aE(X) + b}$$

Continuous case:

$$E(aX+b) = \int_{-\infty}^{\infty} (ax+b)\, f(X)\, dx$$
$$= a\int_{-\infty}^{\infty} x f(X)\, dx + b\int_{-\infty}^{\infty} f(X)\, dx$$

Note that

$$\int_{-\infty}^{\infty} f(X)\, dx = 1$$

because $f(X)$ is a probability density function and the total probability over its entire support must equal 1.

Therefore,

$$E(aX+b) = aE(X) + b$$

2. **Expectation of a constant:**

$$\boxed{E(a) = a}$$

3. **Additivity:**

$$\boxed{E[g(X) \pm h(X)] = E[g(X)] \pm E[h(X)]}$$

4. **Non-negativity:**

$$\boxed{X \geq 0 \;\Rightarrow\; E(X) \geq 0}$$

5. **Zero expectation:**

$$\boxed{X \geq 0 \text{ and } E(X) = 0 \iff P(X=0) = 1}$$

Variance

$E(X)$ is a measure of the **center** of a distribution.

The **variance**, denoted $\sigma^2 = \mathrm{Var}(X)$, measures how closely the distribution is concentrated around the mean μ.

Definition:

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

$\sigma = \sqrt{\sigma^2}$ is called the **standard deviation**.

Theorem (Variance Formula)

$$\mathrm{Var}(X) = \sigma^2 = E(X^2) - \mu^2$$

Proof:

$$\begin{aligned} E[(X-\mu)^2] &= E[X^2 - 2\mu X + \mu^2] \\ &= E(X^2) - 2\mu E(X) + \mu^2 \\ &= E(X^2) - \mu^2 \end{aligned}$$

Example

Weekly demand for a drink (in thousand liters) is a continuous random variable X with PDF

$$f(X) = 2(x-1), \quad 1 < x < 2$$

Mean:

$$\mu = E(X) = \int_1^2 x \cdot 2(x-1)\, dx = \frac{5}{3}$$

Second moment:

$$E(X^2) = \int_1^2 x^2 \cdot 2(x-1)\, dx = \frac{17}{6}$$

Variance:

$$\boxed{\sigma^2 = \mathrm{Var}(X) = \frac{17}{6} - \left(\frac{5}{3}\right)^2 = \frac{1}{18}}$$

Variance of a function of a random variable

Let X be a random variable and let $g(X)$ be a function of X.

Mean of $g(X)$:

$$\mu_{g(X)} = E[g(X)]$$

Variance of $g(X)$:

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

Properties of Variance

Let X, Y be random variables and let $a, b \in \mathbb{R}$.

1. $$\boxed{\mathrm{Var}(aX + b) = a^2 \mathrm{Var}(X)}$$

Proof:

$$\begin{aligned}
\mathrm{Var}(aX + b) &= E[(aX + b)^2] - (E[aX + b])^2 \\
&= E[a^2X^2 + 2abX + b^2] - (aE[X] + b)^2 \\
&= a^2E[X^2] + 2abE[X] + b^2 - a^2E[X]^2 - 2abE[X] - b^2 \\
&= a^2\big(E[X^2] - E[X]^2\big) = a^2\mathrm{Var}(X)
\end{aligned}$$

2. $$\boxed{\mathrm{Var}(X) \geq 0}$$

3. $$\boxed{\mathrm{Var}(a) = 0} \quad \text{for any constant } a$$

4. $$\boxed{\mathrm{Var}(X) = 0 \iff X \text{ is a constant}}$$

Theorem (Variance of a Linear Combination)

Let X, Y be random variables with joint distribution $f(X, Y)$. Then

$$\mathrm{Var}(aX + bY) = a^2\mathrm{Var}(X) + b^2\mathrm{Var}(Y) + 2ab\,\mathrm{Cov}(X, Y)$$

where the **covariance** is defined as

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

Interpretation of covariance

- $\mathrm{Cov}(X, Y) > 0$: X and Y tend to increase together
- $\mathrm{Cov}(X, Y) < 0$: one increases while the other decreases
- $\mathrm{Cov}(X, Y) = 0$: no linear relationship

Example (population intuition):

If X is height and Y is weight in a population,

$$(X - E[X])(Y - E[Y]) > 0$$

for tall–heavy and short–light individuals, so $\mathrm{Cov}(X, Y) > 0$.