# STA 35C: Statistical Data Science III

## Lecture 3: Bayes' Theorem and Random Variables

Dogyoon Song

Spring 2026, UC Davis

# Agenda

Last time:

- Set theory basics
- Probabilistic models
- Conditional probability: definition and multiplication rule

Today:

- Complete probability review: the law of total probability and independence
- Bayes' theorem
- Random variables
  - PMFs, PDFs, and CDFs
  - Expectation and variance
  - Joint, marginal, and conditional distributions
  - Covariance and correlation
  - Sums of random variables

# Recap: Outcomes, events, probability

**Outcome:** a possible result of an experiment (often denoted $\omega$)

- "Head" or "Tail"
- The outcome 6 on a die roll
- Stock price change of $5

**Event:** a subset $A \subseteq \Omega$; the event $A$ occurs if the realized outcome $\omega$ lies in $A$

- $\emptyset, \{6\}, \{1, 2, 3, 4, 5, 6\}$
- For example, if $A = \{1, 2, 3\}$, then $A$ occurs if and only if $\omega \in A$, i.e., iff $\omega = 1, 2$, or 3
    - This does **NOT** mean "$\omega = 1$ and $\omega = 2$ and $\omega = 3$"

**Probability law:** a function $P$ that assigns each event $A$ a number $P(A)$ such that

- $P(A) \geq 0$
- $A \cap B = \emptyset \implies$[1] $P(A \cup B) = P(A) + P(B)$
- $P(\Omega) = 1$

[1] Additivity also applies to countable collections of pairwise disjoint events.

# Recap: Conditional probability

For an event $B$ with $P(B) > 0$, the **conditional probability**[2] of $A$ given $B$ is

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

- Conditioning restricts attention to $B$ and rescales its probability to 1
- **Die roll example:** $A = \{2, 3, 5\}$ (prime), $B = \{2, 4, 6\}$ (even)

$$P(A) = 1/2, \qquad P(A \mid B) = \frac{1/6}{1/2} = 1/3$$

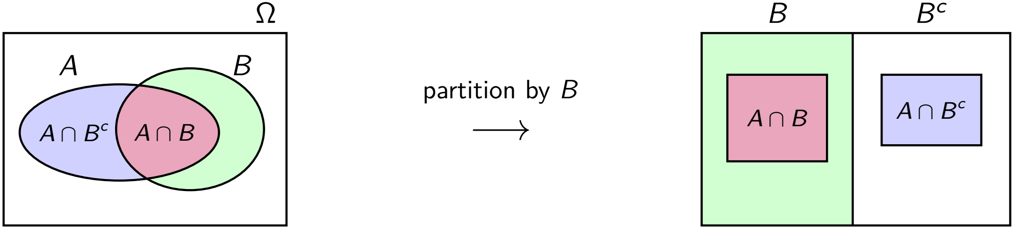
- **Multiplication rule:** $P(A \cap B) = P(B)P(A \mid B)$

[2]For fixed $B$ with $P(B) > 0$, $P(\cdot \mid B)$ satisfies the probability axioms

# Law of total probability

For $0 < P(B) < 1$, the events $B$ and $B^c$ partition $\Omega$. Split $A$ into two disjoint pieces, then use the multiplication rule:

$$\begin{aligned} P(A) &= P(A \cap B) + P(A \cap B^c) \\ &= P(A \mid B)\, P(B) + P(A \mid B^c)\, P(B^c). \end{aligned}$$




- **Same die example:** Among even faces, $1/3$ are prime; among odds, $2/3$ are prime

$$P(A) = \frac{1}{3} \cdot \frac{1}{2} + \frac{2}{3} \cdot \frac{1}{2} = \frac{1}{2}.$$

# Independence

Events $A$ and $B$ are **independent** if $P(A \cap B) = P(A)P(B)$

- Recall that $P(A \cap B) = P(B)P(A \mid B)$
- When $P(B) > 0$, independence is equivalent to $P(A \mid B) = P(A)$
  - Knowing that $B$ occurred does not change the probability that $A$ occurs
- Similarly, when $P(A) > 0$, independence is equivalent to $P(B \mid A) = P(B)$

**Example:** A coin toss and a die roll, with all 12 outcome pairs equally likely

- Knowing the coin shows heads does not change the probability of an even die roll

**Counter-example:** Seeing someone wearing sunglasses and a rainy day

- If we see someone wearing sunglasses, it is less likely to be a rainy day
- **Die example:** Prime and even are not independent: $P(A \mid B) = 1/3 \neq P(A) = 1/2$

**Disjoint vs independent:** If $A \cap B = \emptyset$ and $P(A), P(B) > 0$, then

$$0 = P(A \cap B) < P(A)P(B) \qquad \Longrightarrow \qquad \text{A and B are not independent}$$

# Bayes' theorem

Often, we know $P(B \mid A)$, but want $P(A \mid B)$

- $A$: a hypothesis/model/state of the world
- $B$: observed data or evidence
- Example: $A =$ having a disease, $\quad B =$ positive screening result

After observing $B$, Bayes' theorem updates $P(A)$ to $P(A \mid B)$

- *Prior:* $P(A)$ and $P(A^c)$
- *Likelihood:* $P(B \mid A)$ and $P(B \mid A^c)$
- *Posterior:* $P(A \mid B)$ and $P(A^c \mid B)$

For $0 < P(A) < 1$ and $P(B) > 0$, **Bayes' theorem** states that

$$P(A \mid B) = \frac{P(B \mid A)\,P(A)}{P(B)} \qquad \text{where} \qquad P(B) = P(B \mid A)P(A) + P(B \mid A^c)P(A^c)$$

- *Posterior* $=$ *Likelihood* $\times$ *Prior* / *Evidence*

# Bayes' theorem: Example 1

## Example (Disease screening)

Let $A$ = disease and $B$ = positive screening result

- Suppose $P(A) = 0.01$
- $P(B \mid A) = 0.8$ (true-positive rate)
- $P(B \mid A^c) = 0.1$ (false-positive rate: positive despite not having the disease)

**Q:** If the test is positive, how likely is it that the person actually has the disease?

$$\begin{aligned} P(A \mid B) &= \frac{P(B \mid A)\, P(A)}{P(B \mid A)\, P(A) + P(B \mid A^c)\, P(A^c)} \\ &= \frac{0.8 \times 0.01}{0.8 \times 0.01 + 0.1 \times 0.99} \approx 0.075 \end{aligned}$$

- A positive result raises the probability from 1% to about 7.5%, but it is still far from certain
- When the disease is rare, false positives can outnumber true positives

# Bayes' theorem: Example 2

## Example (Fair or double-headed?)

Let $A$ = the coin is double-headed and $A^c$ = the coin is fair

Let $B$ = the first toss is Head

- Suppose $P(A) = 0.5$
- $P(B \mid A) = 1$ (the double-headed (biased) coin always lands Head)
- $P(B \mid A^c) = 0.5$ (the fair coin lands Head with probability $1/2$)

**Q:** After observing one Head, how likely is it that the coin is double-headed?

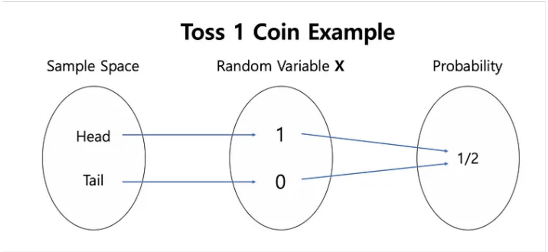
$$\begin{aligned} P(A \mid B) &= \frac{P(B \mid A)P(A)}{P(B \mid A)P(A) + P(B \mid A^c)P(A^c)} \\ &= \frac{1 \times 0.5}{1 \times 0.5 + 0.5 \times 0.5} = \frac{2}{3} \approx 0.667. \end{aligned}$$

- One Head raises the probability from 50% to 66.7%
- Data increase the probability of the model that better explains the observation

# Random variables

A **random variable**[3] $X : \Omega \to \mathbb{R}$ assigns a real number to each outcome $\omega \in \Omega$

- Once the outcome $\omega$ is realized, $X = X(\omega)$ is just a number
- Probabilities on $\Omega$ induce a distribution on the values of $X$



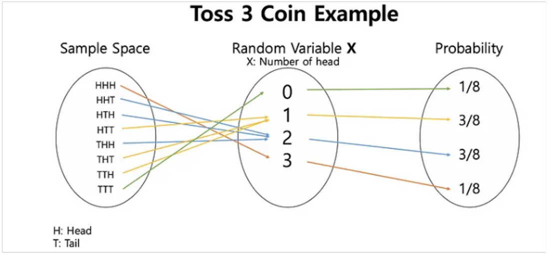




Figure: Random variable example: outcome of tossing a coin[4]

[3] A fully rigorous definition is beyond the scope of STA 35C

[4] Source: https://medium.com/jun94-devpblog/prob-stats-1-random-variable-483c45242b3c

# Distribution of a random variable

How do we describe how probability is distributed over the values of a random variable?

- If $X$ is **discrete**, its **probability mass function** (PMF) is

$$p_X(x) = P(X = x) \qquad \text{where} \qquad p_X(x) \geq 0 \quad \text{and} \quad \sum_x p_X(x) = 1$$

- If $X$ is **continuous**, its **probability density function** (PDF) is a function $f_X \geq 0$ such that

$$P(a \leq X \leq b) = \int_a^b f_X(x)\,dx \qquad \text{and therefore} \qquad \int_{-\infty}^{\infty} f_X(x)\,dx = 1$$

  - If $X$ is continuous, then $P(X = x) = 0$ for every single $x$
  - Thus, $f_X(x)$ is a density, not a point probability

- In both cases, the **cumulative distribution function** (CDF) is

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

# Distribution functions: Examples of discrete RVs

## Example (Biased coin)

Suppose a biased coin lands Head twice as often as Tail, and define

$$X = \mathbf{1}\{\text{Head}\} = \begin{cases} 1 & \text{Head}, \\ 0 & \text{Tail}. \end{cases} \qquad \to \qquad p_X(k) = \begin{cases} 2/3 & k = 1, \\ 1/3 & k = 0, \\ 0 & \text{otherwise} \end{cases}$$

The CDF is

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

**Exercise:** Let $X$ denote the number on the face after rolling a fair die:

$$p_X(k) = \frac{1}{6}, \qquad \forall k \in \{1, 2, 3, 4, 5, 6\}$$

Write down the CDF of $X$ and draw its graph

# Distribution functions: Examples of continuous RVs

## Example (Uniform)

Let $X$ be a random variable uniform on the interval $[0, 2]$

$$f_X(x) = \begin{cases} \frac{1}{2}, & 0 \leq x \leq 2, \\ 0, & \text{otherwise.} \end{cases}$$

For $a, b \in \mathbb{R}$ such that[a] $0 \leq a \leq b \leq 2$

$$P(a \leq X \leq b) = \int_a^b f_X(x)\, dx = \frac{b-a}{2}$$

The CDF

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt = \begin{cases} 0 & x < 0, \\ \frac{x}{2} & 0 \leq x < 2, \\ 1 & x \geq 2. \end{cases}$$

[a] **Question:** What if $a < 0$ or $b > 2$?

# Expectation and variance

The distribution of a random variable $X$ determines probabilities concerning $X$

Describing the full distribution can be complicated

$\rightarrow$ We often summarize it by location and spread

- The **expected value** of a random variable $X$ is its probability-weighted average:
  - Discrete: $\mathbb{E}[X] = \sum_x x \cdot p_X(x)$
  - Continuous: $\mathbb{E}[X] = \int_{-\infty}^{\infty} x \, f_X(x) \, dx$
- The **variance** of a random variable $X$ is the "spread" around the mean:
  - $\mathrm{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]$
  - Alternative formula: $\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$

# Expectation and variance: Example 1

## Example (Tossing a coin with head probability $p$)

The PMF of a Bernoulli random variable $X \sim \mathrm{Bern}(p)$, $0 \leq p \leq 1$, is

$$p_X(x) = p(X = x) = \begin{cases} p & \text{if } x = 1 \text{ (head)}, \\ 1 - p & \text{if } x = 0 \text{ (tail)}, \\ 0 & \text{otherwise.} \end{cases}$$

- *Expectation:*

$$\begin{aligned} \mathbb{E}[X] &= \sum_x x \cdot p_X(x) \\ &= 0 \cdot (1 - p) + 1 \cdot p \\ &= p \end{aligned}$$

- *Variance:*

$$\begin{aligned} \mathrm{Var}(X) &= \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] \\ &= \sum_x (x - p)^2 \cdot p_X(x) \\ &= (-p)^2 \cdot (1 - p) + (1 - p)^2 \cdot p \\ &= p(1 - p) \end{aligned}$$

# Expectation and variance: Example 2

## Example (Uniform distribution)

For $a < b$, the PDF of a uniform random variable on $[a, b]$ is

$$f_X(x) = \begin{cases} \dfrac{1}{b-a}, & a \leq x \leq b, \\ 0, & \text{otherwise.} \end{cases}$$

**Expectation:**

$$\begin{aligned} \mathbb{E}[X] &= \int_a^b x \cdot \frac{1}{b-a}\, dx \\ &= \frac{a+b}{2} \end{aligned}$$

**Variance:**

$$\begin{aligned} \mathbb{E}[X^2] &= \int_a^b x^2 \cdot \frac{1}{b-a}\, dx \\ &= \frac{a^2 + ab + b^2}{3} \\ \mathrm{Var}(X) &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \\ &= \frac{(b-a)^2}{12} \end{aligned}$$

# Some properties of expectation and variance

Let $X$ be a random variable and $a, b \in \mathbb{R}$

**Expectation:**

- $\mathbb{E}[a] = a$

$$\because \; \mathbb{E}[a] = \sum_x a\, p_X(x) = a \underbrace{\sum_x p_X(x)}_{=1} = a$$

- $\mathbb{E}[bX] = b \cdot \mathbb{E}[X]$

$$\because \; \mathbb{E}[bX] = \sum_x (bx)\, p_X(x) = b \sum_x x\, p_X(x) = b\, \mathbb{E}[X]$$

**Variance:**

- $\mathrm{Var}(a) = 0$
- $\mathrm{Var}(bX) = b^2 \cdot \mathrm{Var}(X)$

$$\because \; \mathrm{Var}(bX) = \mathbb{E}[(bX)^2] - \mathbb{E}[bX]^2 = b^2 \cdot \mathbb{E}[X^2] - \left(b \cdot \mathbb{E}[X]\right)^2 = b^2\, \mathrm{Var}(X)$$

# Distribution of multiple random variables

Let $X, Y$ be discrete random variables

**Joint distribution:**

- The joint PMF of $X$ and $Y$ satisfies $p_{X,Y}(x, y) = P(X = x,\ Y = y)$
- The joint CDF of $X$ and $Y$ is defined by $F_{X,Y}(x, y) = P(X \leq x,\ Y \leq y)$

**Marginal distributions:** The marginal PMFs of $X$ and $Y$ are:

$$p_X(x) = \sum_y p_{X,Y}(x, y), \qquad p_Y(y) = \sum_x p_{X,Y}(x, y)$$

**Conditional distribution:** For $p_Y(y) > 0$, the conditional PMF of $X$ given $Y = y$ is

$$p_{X|Y}(x \mid y) = \frac{p_{X,Y}(x, y)}{p_Y(y)} = \frac{p_{X,Y}(x, y)}{\sum_{x'} p_{X,Y}(x', y)}$$

- **Independence:** $X$ and $Y$ are independent if $p_{X,Y}(x, y) = p_X(x)\, p_Y(y)$ for all $x, y$

For continuous RVs, PMFs are replaced by PDFs and sums are replaced by integrals

# Joint, marginal, conditional distributions: Example

## Example (One fair die)

Suppose rolling a fair die and let

$$X = \mathbf{1}\{\text{roll is even}\}, \qquad Y = \mathbf{1}\{\text{roll} > 3\}.$$

$$(X, Y) = \begin{cases} (0,0) & \text{if } \omega \in \{1,3\} \\ (0,1) & \text{if } \omega \in \{5\} \\ (1,0) & \text{if } \omega \in \{2\} \\ (1,1) & \text{if } \omega \in \{4,6\} \end{cases} \implies$$

| | $Y = 0$ | $Y = 1$ | $p_X(x)$ |
|---|---|---|---|
| $X = 0$ | $1/3$ | $1/6$ | $1/2$ |
| $X = 1$ | $1/6$ | $1/3$ | $1/2$ |
| $p_Y(y)$ | $1/2$ | $1/2$ | |

So the conditional probability

$$p_{X|Y}(1 \mid 1) = \frac{p_{X,Y}(1,1)}{p_Y(1)} = \frac{1/3}{1/2} = \frac{2}{3}$$

Also, $X$ and $Y$ are not independent, since

$$p_{X,Y}(1,1) = \frac{1}{3} \neq p_X(1)p_Y(1) = \frac{1}{4}$$

# Covariance and correlation

The **covariance** between $X$ and $Y$ measures how they vary together:

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

- $\mathrm{Cov}(X, Y) > 0$: large values of $X$ tend to occur with large values of $Y$
- $\mathrm{Cov}(X, Y) < 0$: large values of $X$ tend to occur with small values of $Y$
- If $X$ and $Y$ are independent, then $\mathrm{Cov}(X, Y) = 0$; the converse is false in general

Correlation is the normalized version of covariance:

$$\rho(X, Y) := \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}} \in [-1, 1] \qquad \text{(when } \mathrm{Var}(X), \mathrm{Var}(Y) > 0\text{).}$$

## Example (One fair die)

| | $Y = 0$ | $Y = 1$ |
|---|---|---|
| $X = 0$ | $1/3$ | $1/6$ |
| $X = 1$ | $1/6$ | $1/3$ |

- $\mathrm{Cov}(X, Y) = \frac{1}{12}$
- $\rho(X, Y) = \frac{1}{3}$

# Sum of random variables: Expectation and variance are easy

Sums arise naturally when we combine several sources of randomness:

- Total number of heads in $n$ tosses: $S_n = X_1 + \cdots + X_n$
- Total score, total revenue, total waiting time
- Total measurement error from several noisy components

For $Z = X + Y$, the expectation and variance are often easy to compute:

- $\mathbb{E}[Z] = \mathbb{E}[X] + \mathbb{E}[Y]$

$$\because \ \mathbb{E}[X+Y] = \sum_{x,y}(x+y)\, p_{X,Y}(x,y) = \sum_{x} x\, p_X(x) + \sum_{y} y\, p_Y(y)$$
$$= \mathbb{E}[X] + \mathbb{E}[Y]$$

- $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X,Y)$
  - *Exercise:* Verify this using properties of expectation
  - If $X$ and $Y$ are independent, then $\mathrm{Cov}(X,Y) = 0$, so $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$

# Sum of random variables: Distributions can be harder

Even when $\mathbb{E}[X+Y]$ and $\mathrm{Var}(X+Y)$ are easy, the full distribution of $X+Y$ may not be

- We usually need the *joint distribution* of $(X, Y)$, not just the separate marginals

### Example (Same marginals, different joint)

Suppose $X_1$ and $X_2$ each take values 0 or 1 with probability $1/2$

- If $X_1$ and $X_2$ are independent, then

$$P(X_1 + X_2 = k) = \begin{cases} 1/4 & k = 0, \\ 1/2 & k = 1, \\ 1/4 & k = 2 \end{cases}$$

- If $X_1 = X_2$ always, then

$$P(X_1 + X_2 = k) = \begin{cases} 1/2 & k = 0, \\ 1/2 & k = 2 \end{cases}$$

# Wrap-up

- Independence means that learning one event occurred does not change the probability of the other (when conditioning is defined)
- Bayes' theorem updates prior beliefs to posterior beliefs based on observed data
- A random variable is a numerical function $X : \Omega \to \mathbb{R}$
  - Probabilities on $\Omega$ induce a distribution on the values of $X$
- PMFs, PDFs, and CDFs describe distributions
- Expectation and variance summarize center and spread, and facilitate computation
- Joint, marginal, and conditional distributions describe multiple random variables
  - Covariance and correlation summarize linear association
- For sums of random variables, expectations and variances are often easy to compute