Foundations of Inference: Probability

Lecture 8A handout

Name: _____

Reminder: sample statistics and population parameters

Table 1: Notation for sample statistics and population parameters

Measure	Sample statistic	Population parameter
Mean	$\bar{x}$	μ
Standard deviation	s	σ
Variance	s^2	σ^2
Proportion	p	π

We calculate sample statistics from the data. Population parameters are usually unknown, and inference is about learning them from the sample.

Probability basics

A **random process** has more than one possible outcome, and we cannot predict for certain which one will happen. For example, flipping a coin or rolling a die.

The _____, S, is the set of all possible outcomes of a random process.

- A single roll of a die: $S =$ _____
- Flipping a coin twice: $S =$ _____

An _____ is a subset of the sample space. Roll a die once and let

- A: the roll is even. $A =$ _____
- B: the roll is odd. $B =$ _____
- C: the roll is prime. $C =$ _____

Venn diagram. Write the numbers 1 to 6 in the right regions:

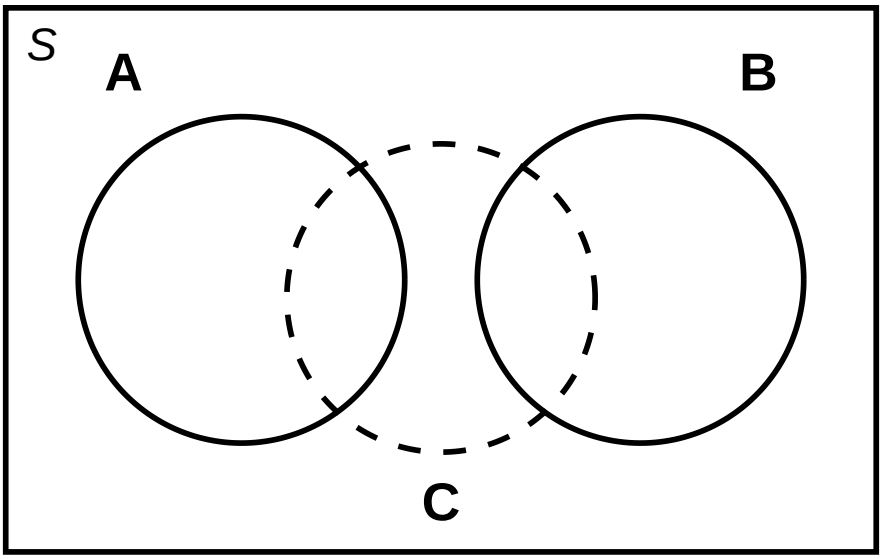


Table 2: Complement, union, and intersection of events for one roll of a die

Name	Notation	Read as	Outcomes
Complement of A	A^c	not A	
Union	$A \cup C$	A or C	
Intersection	$A \cap C$	A and C	

Two events are **mutually exclusive** (disjoint) if they cannot happen at the same time.

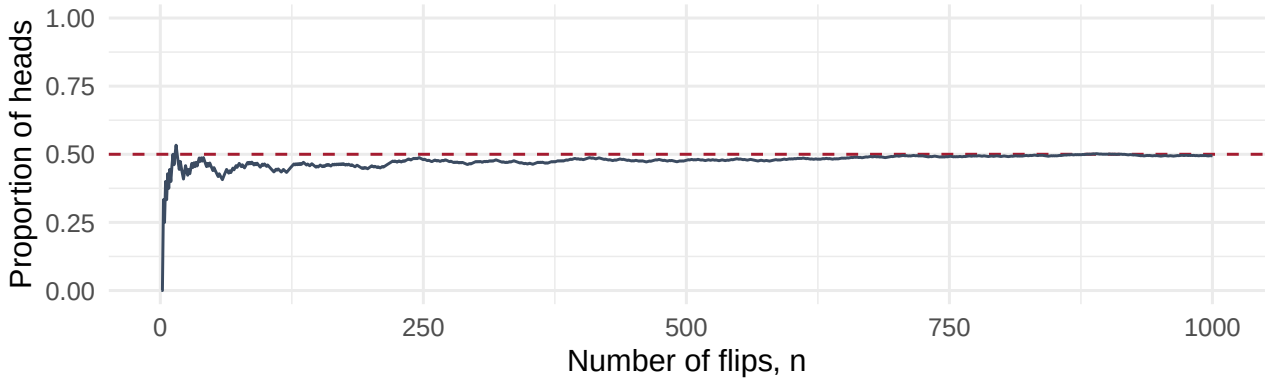
Are A and B mutually exclusive? Explain.

Are A and C mutually exclusive? Explain.

Your turn. List the outcomes in B^c, $B \cap C$, and $B \cup C$.

Defining probability

Frequentist definition: The probability of an outcome is the proportion of times the outcome is observed when the random process is repeated a very large number of times.



Bayesian definition: The probability of an event is how plausible the event is, based on the information we have, including prior information. When we get new information (data), we update the probability.

Example: what is the probability that a bottle cap lands top side up? Unlike a coin, we do not know the answer ahead of time.

- A **frequentist** flips the bottle cap a very large number of times. The probability is the long-run proportion of flips that land top side up.
- A **Bayesian** starts with a prior belief, for example "probably around 0.5, but I am not sure," and updates it after each flip.

Suppose we flip the bottle cap 3 times and all 3 land top side up.

What would the frequentist estimate be? _____

Would the Bayesian's answer be higher or lower than that? Why?

What do you think happens after 500 flips?

Rules of probability

1. The probability of any event is between 0 and 1: $0 \leq P(A) \leq 1$

2. The probability of the sample space is 1: $P(S) = 1$

Example: when you roll a fair die, $P(S) = P(\{1, 2, 3, 4, 5, 6\}) =$ _____. Something in the sample space always happens.

3. If A and B are mutually exclusive, $P(A \cup B) = P(A) + P(B)$

Example: rolling a fair die, find $P(A \cup B)$, the probability that the roll is even or odd.

Complement rule: $P(A^c) =$ _____

Example: find $P(C^c)$, the probability that the roll is not prime.

Addition rule: for **any** two events, $P(A \cup C) = P(A) + P(C) -$ _____

Example: for a fair die, find $P(A \cup C)$. Why do we subtract?

Independence: two processes are independent if knowing the outcome of one does not help predict the outcome of the other. If A and B are from independent processes,

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

Example: what is the probability of getting two heads in two flips of a fair coin? Check your answer with the sample space.

Your turn. You roll a fair die and flip a fair coin. What is the probability of getting a prime number and heads?

Random variables

A _____ is a numerical outcome of a random process.

Flip a coin twice and let X be the number of heads. X can be _____.

A random variable can be…

- a **random observation**, X: the height of a randomly selected student. Technically, X is an observation in the population. Every observation in a sample is also in the population, so the observations in a sample, $X_1, X_2, \dots, X_n$, are random observations too.
- a **random sample statistic**, $\bar{X}$: the mean height of a random sample of 25 students, or p: the proportion of a random sample of students who drink coffee daily
- a **random population parameter**, μ: the mean height of all students. In Bayesian inference, even a population parameter is treated as a random variable to describe our uncertainty about it.

A **discrete** random variable takes values we can _____.

A **continuous** random variable can take _____.

Your turn. Write the sample space of each random variable. Is it discrete or continuous?

Table 3: Sample space and type of each random variable

Random variable	Sample space	Discrete or continuous?
Number on a die	$\{1, 2, 3, 4, 5, 6\}$	discrete
Number of courses you take this semester	_____	_____
Hours you sleep tonight	_____	_____
Number of texts you get today	_____	_____

Random variable	Sample space	Discrete or continuous?
Time the broken alarm rings (below)	_____	_____

Probability distributions

The **probability distribution** of a random variable tells us all its possible values and how probable each one is.

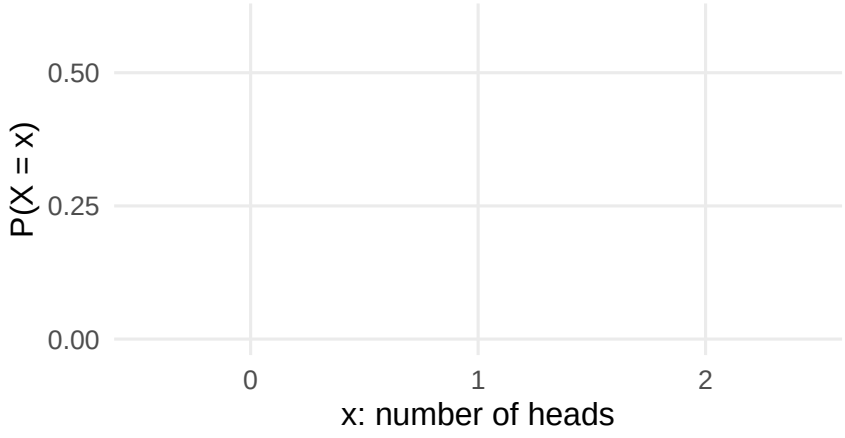
Discrete. We can list $P(X = x)$ for every value. Complete the distribution of X, the number of heads in two flips of a fair coin:

Table 4: Probability distribution of X, the number of heads in two flips of a fair coin

x	0	1	2
$P(X = x)$	_____	_____	_____

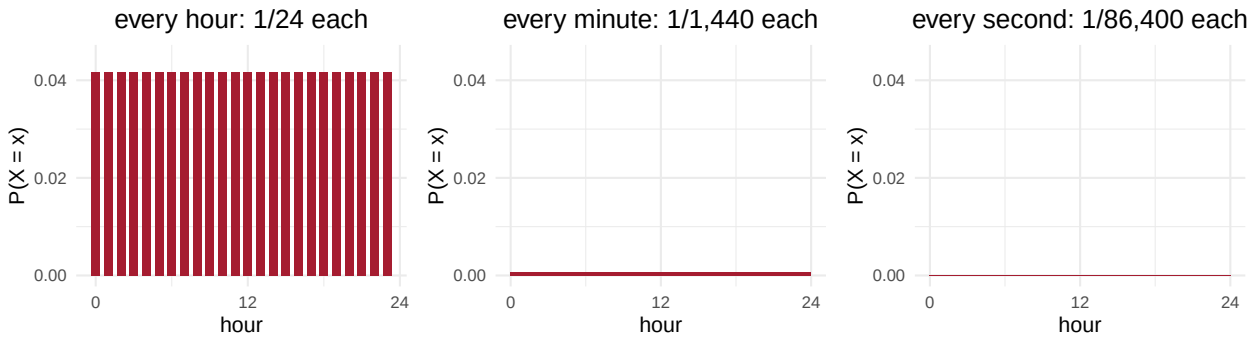
These probabilities add up to _____.

Draw the distribution as a bar plot:



From discrete to continuous. A phone has a broken alarm. It rings once a day, at a completely random time between 0 and 24 hours (midnight to midnight). Let X be the time it rings.

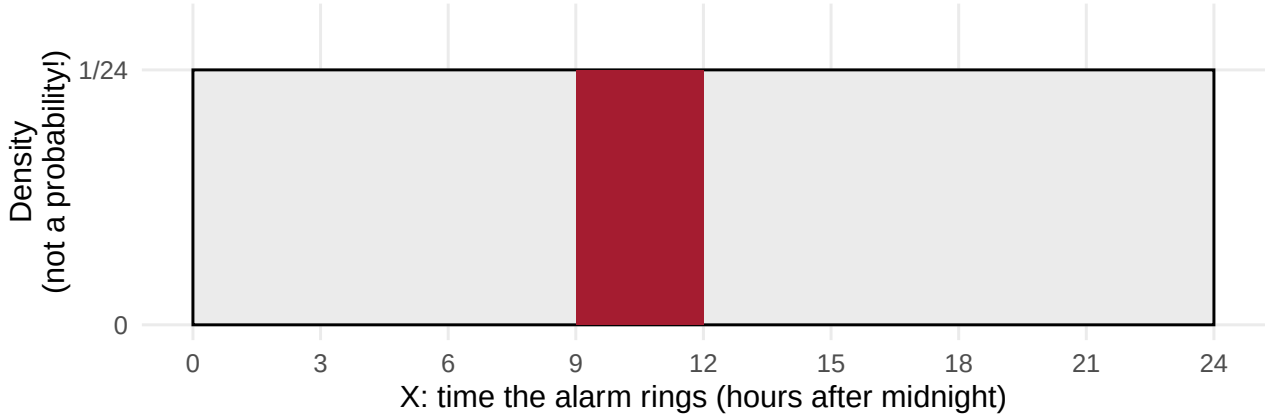
If the alarm could only ring at the start of an hour, there would be 24 equally likely times. At the start of a minute, 1,440. At the start of a second, 86,400:



But time is **continuous**: the alarm could ring at any instant, so there are infinitely many possible times. Then

$P(X = 9) =$ _____

For a continuous random variable, we cannot calculate the probability of a single value. Instead, we calculate the probability of an _____, and the probability is the _____. The total area under the curve is 1.



The height of the curve is _____ because the total area, 24 × height, must be 1.

Switch your mindset!

Table 5: Discrete vs. continuous random variables: what the y-axis shows and where probability is

Type of random variable	y-axis shows	Probability is
Discrete random variable	_____	_____
Continuous random variable	_____	_____

$P(9 < X < 12)$ is the probability that the alarm rings between 9 am and noon:

Your turn. Find the probability that the alarm rings

a. while you sleep, between midnight and 8 am: _____

b. during this lecture (75 minutes): _____

c. at exactly noon: _____

Conditional probability

> "Conditioning is the soul of statistics."
>
> Joe Blitzstein, Harvard Statistics (Stat 110)

Conditional probability answers the question: how should we update what we believe when we learn something new? It is at the heart of statistical inference. Thomas Bayes's famous essay on it was published in 1763, after his death, and it is the root of Bayesian inference, which we will learn in a few weeks.

$P(A \mid B)$ is the probability of A **given that** B has happened. We _____ the sample space to B.

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

Die. Given that the roll is even, what is the probability that it is prime? Write it in notation and calculate it.

Data: GSS 2018. The General Social Survey has surveyed U.S. residents since 1972.

Table 6: Belief in life after death by whether the person took a college science course, GSS 2018 (counts)

College science course	Believes in life after death	Does not believe	Total
Took a college science course	375	75	450
Did not take one	485	115	600
Total	860	190	1050

Let B: a randomly selected person in this sample believes in life after death, and C: they took a college science course.

$P(B \mid C) =$ _____ $P(C \mid B) =$ _____

The order matters! $P(\text{has a dog} \mid \text{likes dogs}) \neq P(\text{likes dogs} \mid \text{has a dog})$. Which one is larger? Why?

Writing conditional probabilities

$$P(\text{what we want the probability of} \mid \text{what we know or assume})$$

The part after | is **given**: we treat it as known (fixed).

Let μ be the mean hours of sleep of **all** Harvard students and $\bar{X}$ the mean of a random **sample** of 30 Harvard students. Write in notation:

a. The probability that the sample mean is at most 7.5 hours, if the population mean is 8 hours.

b. The probability that the population mean is more than 8 hours, given that the sample mean is 7.5 hours.

c. The probability of getting 10 heads in 10 flips, if the coin is fair.

d. The probability that the coin is fair, given that we got 10 heads in 10 flips.

In (a), what is treated as fixed? _____ In (b)? _____

Exercises: if you finish early

1. Roll a fair die. Find $P(A \mid C)$, $P(C \mid B)$, and $P(A \mid B)$.

2. Roll a die four times. What is the probability of getting at least one 6? (Hint: complement and independence.)

3. Using the GSS table, find $P(B^c \mid C)$ and $P(B \mid C) + P(B^c \mid C)$. Why do you get this sum?

4. The broken alarm rings at a random time between 0 and 24 hours. Find $P(X > 20)$ and $P(6 < X < 18)$.

5. Let μ be the mean GPA of all Harvard students and $\bar{X}$ the mean GPA of a random sample of 50 students. Write in words: $P(\bar{X} > 3.8 \mid \mu = 3.7)$ and $P(\mu > 3.8 \mid \bar{X} = 3.75)$.

6. π is the proportion of all Harvard students who drink coffee daily, and p is the proportion in a random sample of 100. Write in notation: the probability that more than half of all students drink coffee daily, given that the sample proportion is 0.6.