HIROSHIMA UNIVERSITY

Fundamental Data Science (30104001)

Lecture 9 — Probability

Jorge N. Tendeiro
Hiroshima University

Lecture schedule

Lecture	Topic
1	Guidance and Introduction
2	Data acquisition and open data, data science ethics
3	Types of data and descriptive statistics
4	Descriptive statistics
5	Visualize data in R
6	Correlation and regression
7	Simple regression analysis in Excel
8	Principal Component Analysis, Cluster Analysis in R
9	Probability
10	Random variables and probability distributions
11	Basic probability distributions
12	Bivariate probability distributions
13	Methods for data collection
14	Point estimate and interval estimation
15	Interval estimation

Descriptive statistics

Data analysis methods

Probability

Inferential statistics

Example of inferential statistics

Inferential statistics:

> *Infer the tendency or characteristics of the entire data (population) based on a part of the data (sample).*

One example is given by the so-called interval estimation:

> *Based on a sample of test scores, we computed the so-called 95% confidence interval; it is found to be equal to $(55, 60)$. We can infer that this interval was computed using a procedure that, in the long run, is correct with 95% probability.*

In inferential statistics, probability is key!

Today

- Probability
- Conditional probability
- Bayes' theorem

Probability

What is probability

Probability:

> *Probability is the branch of mathematics concerning events and numerical descriptions of how likely they are to occur.*

Ref.: Wikipedia

Familiar probabilities (taking values between 0 and 1):

- The probability of getting 1 when throwing a fair die once.
- The probability of a baseball player getting a hit.
- The probability of precipitation.

Probabilities are used to guide rational decision making procedures.

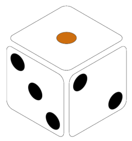
Calculating probabilities

$$\boxed{p = \frac{\text{number of favorable cases}}{\text{total number of cases}}}$$

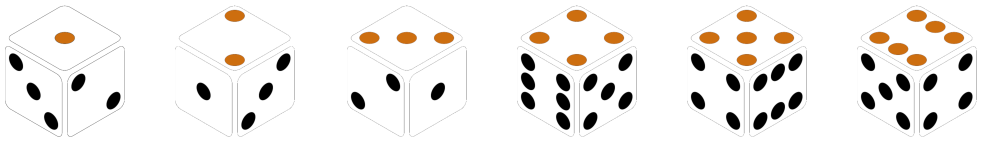
Example:

> *What is the probability of getting 1 when throwing a fair die once?*

Favorable cases:



Possible cases:



Therefore:

$$p = \frac{1}{6}.$$

Trial, sample space

Trial:

> *Any type of experiment involving uncertainty.*

Example:
Tossing a die once.

Sample space (Ω):

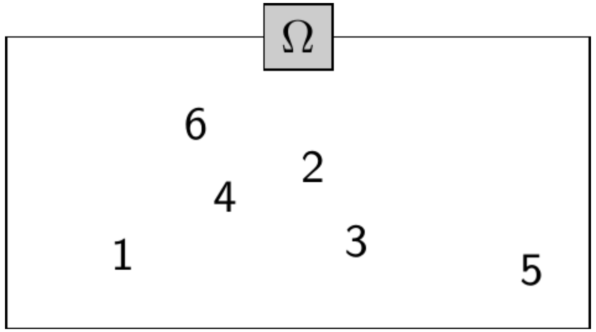
> *Set of all possible outcomes of a trial.*

Example:
The sample space of the trial consisting of tossing a die once is given by

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

or, using words,

$$\Omega = \{\text{getting } 1, \text{getting } 2, \cdots, \text{getting } 6\}$$



Event

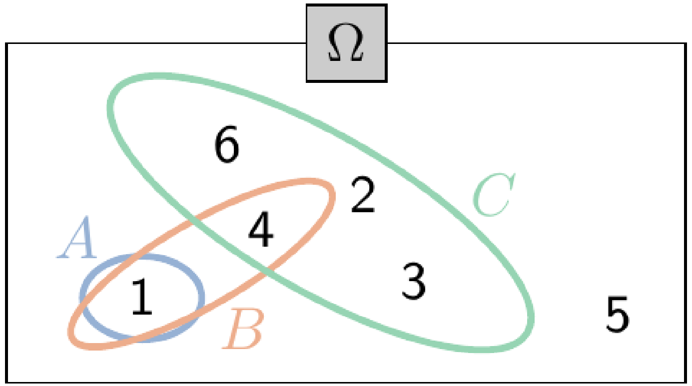
Event:

> *Any subset of the sample space Ω.*

Example:
Consider again the trial of tossing a die once.

Here are various possible events:

- Getting 1: $A = \{1\}$
- Getting 1 or 4: $B = \{1, 4\}$
- Not getting 1 or 5: $C = \{2, 3, 4, 6\}$
- Getting an integer larger than 0: $D = \Omega$ (i.e., the entire sample space).
- Getting 2.5: $E = \{\} = \varnothing$ (i.e., the empty set).



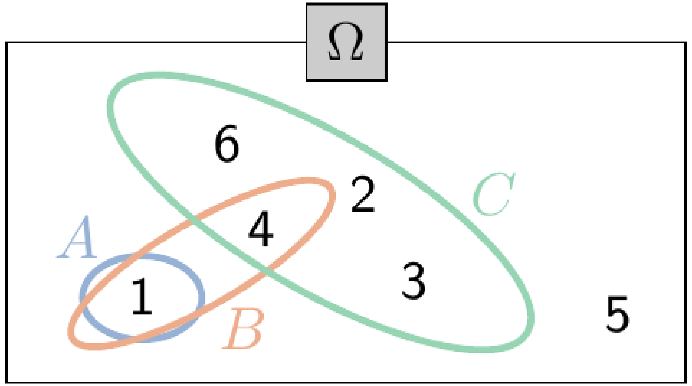
Number of cases

Number of cases:

> *The number of cases of an event.*

Example:

- $A = \{1\} \longrightarrow \#(A) = 1$
- $B = \{1, 4\} \longrightarrow \#(B) = 2$
- $C = \{2, 3, 4, 6\} \longrightarrow \#(C) = 4$
- $D = \Omega \longrightarrow \#(D) = 6$
- $E = \{\} = \varnothing \longrightarrow \#(E) = 0$



(Classical) probability

The (*classical*, or *frequentist*) probability that an event A occurs is defined as

$$\boxed{P(A) = \frac{\text{number of cases that } A \text{ occurs}}{\text{total number of cases}} = \frac{\#(A)}{\#(\Omega)}}$$

Examples:

- Probability of getting 1 when throwing a fair die once:
  - Event: $A = \{1\}$.
  - Sample space $= \Omega = \{1, 2, 3, 4, 5, 6\}$.

$$\longrightarrow \quad P(A) = \frac{\#(A)}{\#(\Omega)} = \frac{1}{6}$$

- Probability of getting heads when throwing a fair coin once:
  - Event: $A = \{\text{heads}\}$.
  - Sample space $= \Omega = \{\text{heads}, \text{tails}\}$.

$$\longrightarrow \quad P(A) = \frac{\#(A)}{\#(\Omega)} = \frac{1}{2}$$

(Classical) probability — Example

This is said on a TV show:

> *"This supplement helps a lot to lose your weight.*
> *In fact, all 5 people out of 5 have lost weight!"*

A TV viewer may think:

> *If it really works, I want to buy this supplement!*

We may conceive of using the following rule to guide which decision to make:

> Assuming that the supplement does *not* work, if the probability of all 5 people losing weight is
>
> - Less than $.01$ (i.e., 1%) $\Longrightarrow$ I will buy the supplement.
> - Larger than $.01$ (1%) $\Longrightarrow$ I will not buy the supplement.

(Classical) probability — Example

So, let's assume that the supplement does *not* work.
In other words, let's assume that all 5 people lost their weight due to chance alone.

If we abide to the following rules:

- Each person either lost or did not lose their weight.
- Losing weight is independent between persons,

then we can compute the probability of observing 5-in-5 persons losing weight:

- Total number of cases ($\mathrm{P}_i = i$th person):

$$\underbrace{2}_{\mathrm{P}_1} \times \underbrace{2}_{\mathrm{P}_2} \times \underbrace{2}_{\mathrm{P}_3} \times \underbrace{2}_{\mathrm{P}_4} \times \underbrace{2}_{\mathrm{P}_5} = 32 \text{ cases}$$

- Number of cases where all 5 people lost weight:

$$(\underbrace{\text{lose}}_{\mathrm{P}_1}, \underbrace{\text{lose}}_{\mathrm{P}_2}, \underbrace{\text{lose}}_{\mathrm{P}_3}, \underbrace{\text{lose}}_{\mathrm{P}_4}, \underbrace{\text{lose}}_{\mathrm{P}_5}) \longrightarrow 1 \text{ case}$$

Hence, the probability of observing all 5 persons losing weight is

$$p = \frac{1}{32} = .03125 \simeq 3\%$$

.

(Classical) probability — Example

Recall our 'decision rule':

> Assuming that the supplement does *not* work, if the probability of all 5 people losing weight is
>
> - Less than $.01$ (i.e., 1%) $\implies$ I will buy the supplement.
> - Larger than $.01$ (1%) $\implies$ I will not buy the supplement.

Since $p = .03 > .01$, we decide not to buy the supplement.

The reasoning is this:

> *Assuming that the supplement does not work, the probability of all 5 people losing weight is still to high to be ignored (at least larger than 1%).*

Careful

This does **not** mean that we can conclude that the supplement does not work!
Instead, the observed data are *too likely* if we assume that the supplement does not work!

(Classical) probability — Example

Now suppose that the following had been said instead on the TV show:

> *"This supplement helps a lot to lose your weight.*
> *In fact, all 10 people out of 10 have lost weight!"*

Under the same assumptions as before, now the probability of observing all 10 persons losing weight is

$$p = \frac{\text{number of favorable cases}}{\text{total number of cases}} = \frac{1}{\underbrace{2 \times \cdots \times 2}_{10 \text{ times}}} = \frac{1}{1024} \simeq .001$$

(Classical) probability — Example

Based on our 'decision rule',

> Assuming that the supplement does *not* work, if the probability of all 10 people losing weight is
>
> - Less than $.01$ (i.e., 1%) $\Longrightarrow$ I will buy the supplement.
> - Larger than $.01$ (1%) $\Longrightarrow$ I will not buy the supplement.

we now decide to buy the supplement since $p = .001 < .01$.

The reasoning is this:

> *Assuming that the supplement does not work, the probability of all 10 people losing weight is too small (at least smaller than 1%).*

Careful

This does **not** mean that we can conclude that the supplement does work!
Instead, the observed data are *too unlikely* if we assume that the supplement does not work!

(Classical) probability — One caveat

$$\boxed{p = \frac{\text{number of favorable cases}}{\text{total number of cases}}}$$

This way of computing probabilities predicates on one crucial assumption:

> *All cases are equally likely.*

Example:

Let's consider an uneven die.
The sample space is still the same as before:

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$

But for an event such as $A = \{6\}$ we now actually have that

$$p(A) \neq \frac{1}{6}$$

Conditional probability

Conditional probability

Conditional probabilities allow incorporating additional information into the computation.

Here is an example.

> *Suppose you throw a fair die, once, but you did not see the result.*
> *What is the probability of getting 3 or lower?*

Denoting this event by A, the probability is given by

$$p(A) = \frac{\text{number of favorable cases}}{\text{total number of cases}} = \frac{\#\{1,2,3\}}{\#\{1,2,3,4,5,6\}} = \frac{3}{6} = \frac{1}{2}.$$

Conditional probability

Now suppose you received some extra information:

> A person seeing you throwing a die said, "it was an odd number".

Now the probability becomes

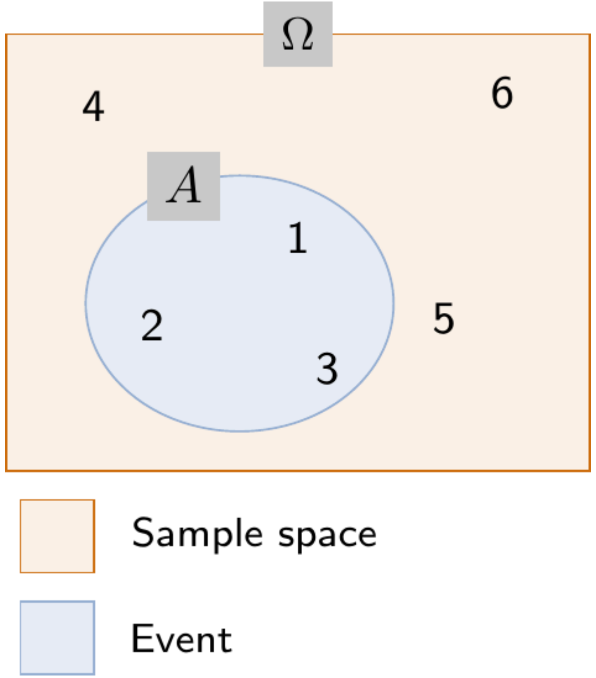
$$p(A) = \frac{\text{number of favorable cases}}{\text{total number of cases}} = \frac{\#\{1,3\}}{\#\{1,3,5\}} = \frac{2}{3}.$$

This is a so-called conditional probability.

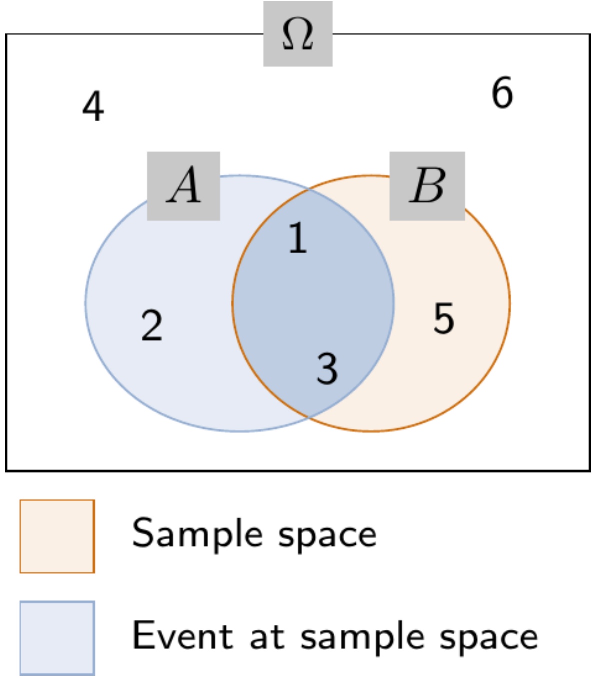
Conditional probability

Using additional information to compute probabilities has the effect of changing the sample space.

Without other information:



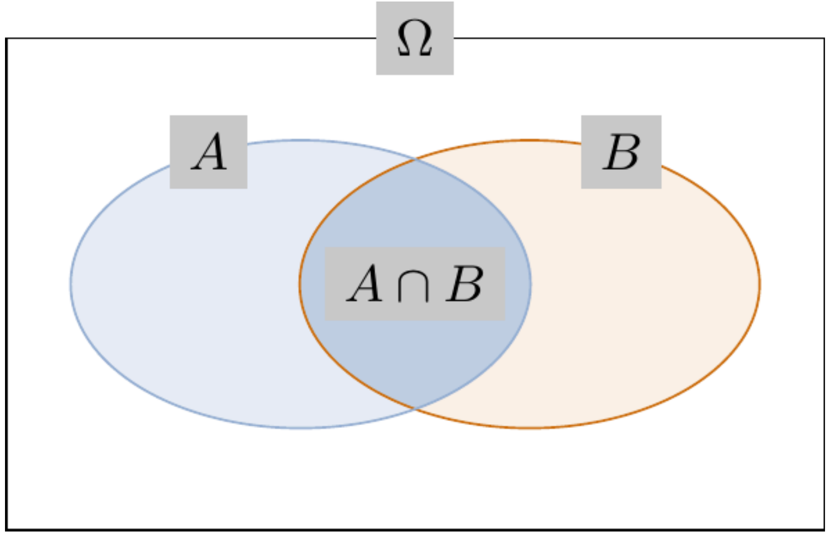
With additional information ("The outcome is an odd number"):



Conditional probability — Definition

Consider two events A and B such that $P(B) > 0$.
The conditional probability of A occuring given that B occurred is

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



Conditional probability — Example 1

> *My friends, who are a couple, have two children.*
> *What is the probability that the two children are both boys?*

Assume that $P(\text{boy}) = P(\text{girl}) = .5$.

Let's compute the required probability under each of the following scenarios:

1. Without any extra information.
2. Knowing that *at least one child is a boy*.
3. Knowing that *the older child is a boy*.

Conditional probability — Example 1

> *My friends, who are a couple, have two children.*
> *What is the probability that the two children are both boys?*

1. Without any extra information.

Denote event $A = \{\text{Both children are boys}\}$.

$$P(A) = \frac{\#\{(\text{boy}, \text{boy})\}}{\#\{(\text{boy}, \text{boy}), (\text{boy}, \text{girl}), (\text{girl}, \text{boy}), (\text{girl}, \text{girl})\}} = \frac{1}{4}$$

Conditional probability — Example 1

> *My friends, who are a couple, have two children.*
> *What is the probability that the two children are both boys?*

2. Knowing that *at least one child is a boy*.

Denote event $B = \{\text{At least one child is a boy}\}$.

$$P(A|B) = \frac{\#(A \cap B)}{\#(B)} = \frac{\#\{(\text{boy}, \text{boy})\}}{\#\{(\text{boy}, \text{boy}), (\text{boy}, \text{girl}), (\text{girl}, \text{boy})\}} = \frac{1}{3}$$

Conditional probability — Example 1

> *My friends, who are a couple, have two children.*
> *What is the probability that the two children are both boys?*

3. Knowing that *the older child is a boy*.

Denote event $C = \{\text{The older child is a boy}\}$.

$$P(A|C) = \frac{\#(A \cap C)}{\#(C)} = \frac{\#\{(\text{boy}, \text{boy})\}}{\#\{(\text{boy}, \text{boy}), (\text{girl}, \text{boy})\}} = \frac{1}{2}$$

Conditional probability — Example 2

> Mr. A took a detection test for a particular disease.
> The test result was positive.
> *What is the probability that Mr. A actually has the disease?*

Use the following information:

- $P(\text{disease}) = .0001$:
  0.01% *of the population is affected by this disease.*
- $P(\text{positive}|\text{sick}) = .99$.
  The probability of the test correctly detecting the disease is 99%.
  This is known as the test's sensitivity.
- $P(\text{negative}|\text{not sick}) = .99$.
  The probability of the test correctly detecting that a person is not sick is 99%.
  This is known as the test's specificity.

Conditional probability — Example 2

Define the events:

- $A = \{ \text{Is sick} \}$
- $B = \{ \text{Positive result} \}$

We want to compute $P(A|B)$.
The easiest way to approach this problem is to use counts to compute $P(A|B)$.

For $N = 1,000,000$ people.

	Test positive	Test negative	TOTAL
Sick	99 $P(\text{positive}|\text{sick}) \times N_{\text{sick}}$	1 $P(\text{negative}|\text{sick}) \times N_{\text{sick}}$	100 $N_{\text{sick}} = P(\text{sick}) \times N$
Not Sick	9,999 $P(\text{positive}|\text{not sick}) \times N_{\text{not sick}}$	989,901 $P(\text{negative}|\text{not sick}) \times N_{\text{not sick}}$	999,900 $N_{\text{not sick}} = P(\text{not sick}) \times N$
TOTAL	10,098 $P(\text{positive}) \times N$	989,902 $P(\text{negative}) \times N$	1,000,000 N

$$P(A|B) = \frac{\#(A \cap B)}{\#(B)} = \frac{99}{10,098} = .0098 \simeq 1\%$$

Bayes theorem
(Application of conditional probability)

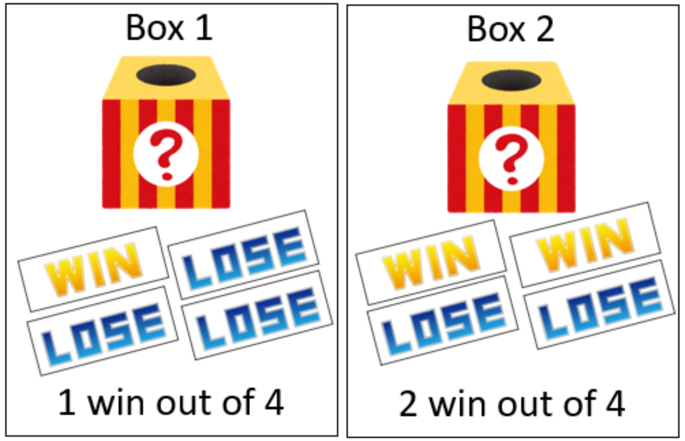
Which box?

> *There are two boxes in a shop which look the same, but one has more winning lotteries than the other box.*
> *A worker brought you one of the boxes and you picked a winning lottery.*
> *Which box is more likely that you picked a lottery from?*

- Without using information about having picked a winning lottery, we would have

$$P(\text{Box } 1) = P(\text{Box } 2) = \frac{1}{2}.$$

- However, we should use the extra information (a winning lottery was picked).
  Since Box 2 has more winning lotteries, perhaps we did use Box 2...



Let's use the extra information to compute $P(\text{Box } 1)$ and $P(\text{Box } 2)$.

Bayes theorem

Consider:

- Event $A =$ outcome
  (e.g., picking a winning lottery).
- Event $B =$ cause
  (e.g., having used Box 1).

Bayes theorem is given as follows:

$$\boxed{P(B|A) = \frac{P(B)P(A|B)}{P(A)}}$$

Here:

- $P(B|A) =$ probability of cause B given that outcome A happened.
- $P(A|B) =$ probability of outcome A given cause B.

Bayes theorem allows reversing conditional probabilities:

> *We can calculate the probability of the cause based on the outcome!*

Which box?

- Event $A =$ picking a winning lottery (outcome)
- Event $B =$ having used Box i (cause)

The probability that you picked a lottery from Box 1 given that you won the lottery is:

$$P(B|A) = \frac{P(B)P(A|B)}{P(A)} = \frac{(^1/_2)(^1/_4)}{^3/_8} = \frac{1}{3}.$$

The probability that you picked a lottery from Box 2 given that you won the lottery is:

$$P(B|A) = \frac{P(B)P(A|B)}{P(A)} = \frac{(^1/_2)(^1/_2)}{^3/_8} = \frac{2}{3}.$$

Conclusion:
Conditional on having picked a winning lottery (A), the probability of having used Box 2 (B) is larger.

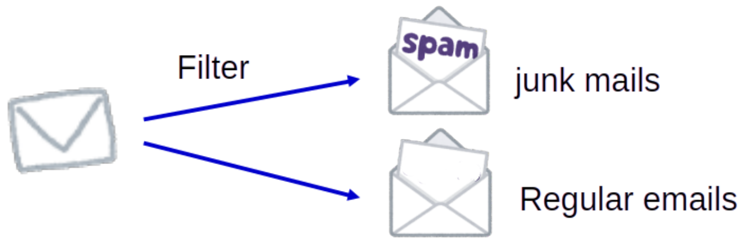
Bayes theorem — Practical examples

- Automatic sorting of junk e-mail.

Example:
Finding the probability that an email is junk, given that it contains the expression "completely free" in its message.

$P(B|A)$, with:

- $A =$ expression in the message
- $B =$ email is junk



- DNA parentage testing.

Example:
Finding the probability that someone is a father, given the child's genes.

$P(B|A)$, with:

- $A =$ child's genes
- $B =$ being father

Summary

- Concept of probability

In statistics, probability is a key concept!

- Conditional probability

An event's probability changes with additional information!

- Bayes theorem

We can reverse conditional probabilities and calculate the probability of a cause based on the outcome!