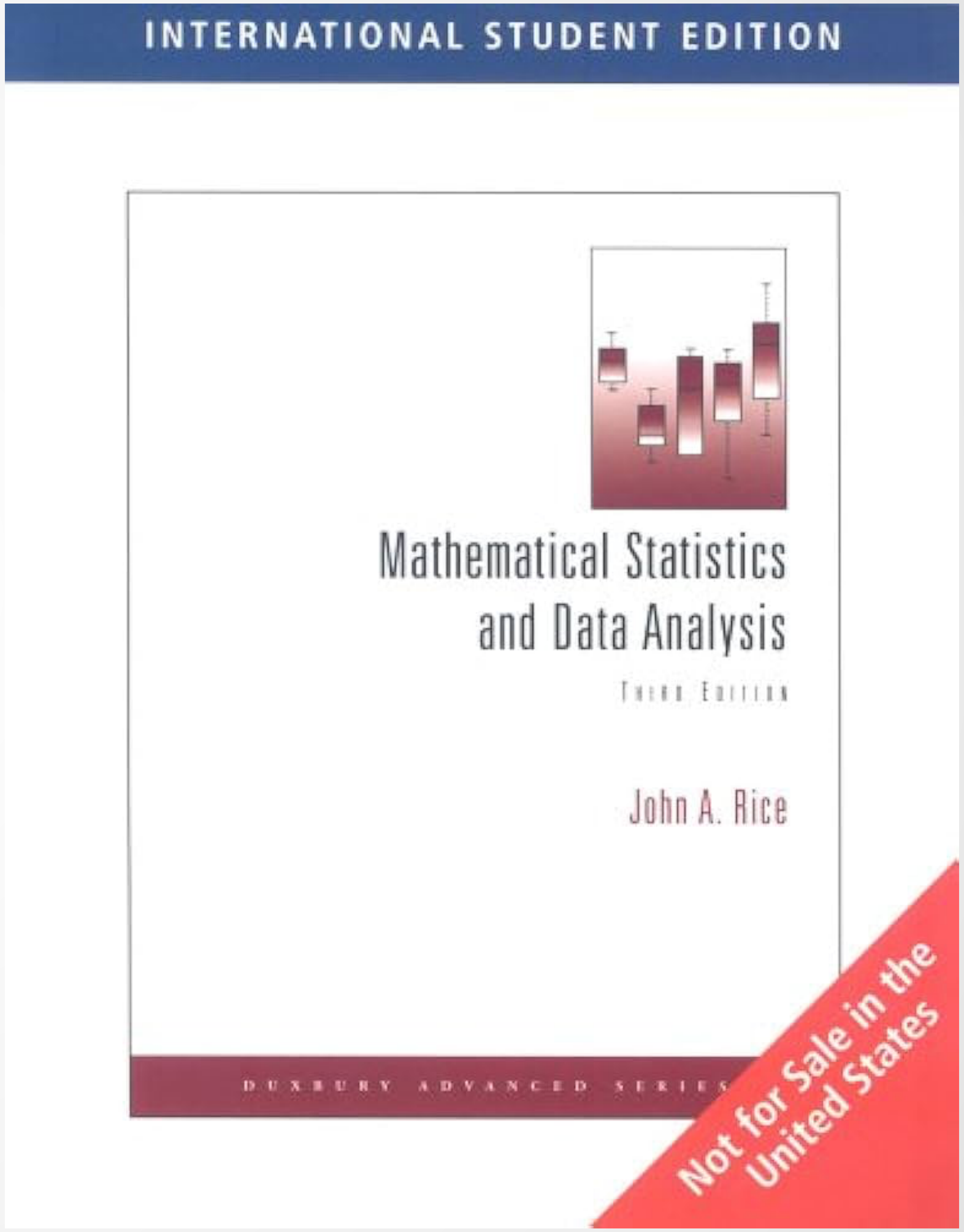
Mathematical Statistics

Chapter 1: Probability

Huei-Wen Teng

Dept Information Management and Finance
National Yang Ming Chiao Tung University

This course is based on



1.1 Introduction

Here is the beef

- Once, a student bravely asked me: Teacher, why do we need to learn mathematical statistics? I really feel that I have learned nothing and it is of no use.
- short term (1-2 Y):
  - Graduate financial engineering entrance exam
  - P-exam for American Actuary Association
- Mid-term (3-5 Y): find a satisfying job
  - Financial engineering, model validation
  - Risk management, Basel III accord
  - Portfolio management, fund management
  - FinTech

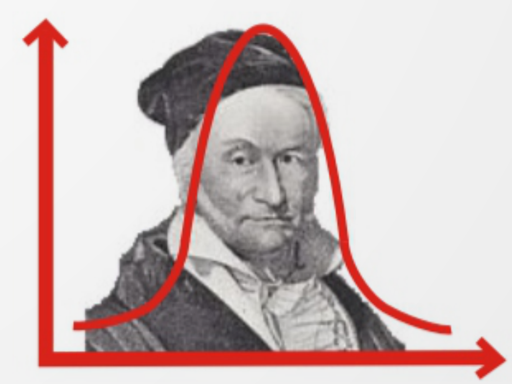
A real-world example in Statistics

- We would like to know if the average height of NYCU male students is 170 cm. Suppose we find 5 male students and record their heights: $185, 168, 160, 172, 170$. How do you answer this question?
- With data, we calculate the sample mean $\overline{X} = 171$ and sample variance $s^2 = 82$.

Mathematical notations

- Let height follows $X \sim N(\mu, \sigma^2)$. But both μ and σ^2 are unknown.

Carl Friedrich Gauss



Six steps in Hypothesis test

- We set up a hypothesis test: $H_0 : \mu = 170$ vs. $H_a : \mu \neq 170$.
- Set up $\alpha = 0.05$
- The test statistic is $t_{\text{stat}} = \dfrac{\bar{X} - 170}{s/\sqrt{5}} \sim t_4$
- Collect data to calculate the realized statistic: $t^* = 0.25$.
- Find the rejection region $\{t : |t| > 2.776\}$
- Conclusion: Because t^* does not fall in to the rejection region, we do not reject H_0

Conclusion with plain language

- There is no significant evidence that the average height of male student at NYCU is different from 170 cm. Thus, it is likely that the average height of male student at NYCU is 170 cm.

Goals

- We would like to understand, why these six steps work?
- Specifically, why Step 3 is correct?
  - A statistic, a countable noun, is a function of random samples $X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} F(\theta)$
  - Statistics, an uncountable noun, is a set of knowledge that includes many statistics.
  - We would like to know the distribution of a statistic, $G(X_1, \cdots, X_n)$.
  - How? We need knowledge accumulated from Chapters 1 to 6 to understand the distribution or the density function of $G(X_1, \cdots, X_n)$

Extensions to this course

- Financial Econometrics: Regression, endogenous problem, simultaneous equations, panel data
- Time series: portfolio management, risk management
- Mathematical Finance: financial derivatives
- Machine learning: FinTech, AI

Motivations

- Chapters 3 to 6 provide fundamental to prove the step 3 seen in Statistics.
- Applications of probability: bioinformatics, kinetic theory of gases, computer operating systems, queues, electrical devices and communication systems, atmospheric turbulence, operations research, actual science, commercial or military aircraft, finance
- To understand statistics in a deeper level, we start with learning probability.

1.2 Sample Spaces

What is probability theory

- Probability theory is concerned with situations in which the outcomes occur randomly.
- Generically, such situations are called *experiments*, and the set of all possible outcomes is the *sample space* corresponding to an experiment.
- The sample space is denoted by Ω, and an *element* of Ω is denoted by ω.
- This element is also called a *simple event*.

Example

- Driving to work, a computer passes through a sequence of three intersections with traffic lights. At each light, she either stops, s, or continues, c. The sample space is the set of all possible outcomes $\Omega = \{\mathrm{ccc}, \mathrm{ccs}, \mathrm{css}, \mathrm{csc}, \mathrm{sss}, \mathrm{ssc}, \mathrm{scc}, \mathrm{scs}\}$
- The length of time between successive earthquakes is $\Omega = \{t \mid t \geq 0\}$

Events

- We are often interested in particular subsets of Ω, which in probability language are called *events.*
  - The empty set is the set with no element, denoted by $\varnothing$.
  - If $A \cap B = \varnothing$, A and B are said to be disjoint.

Set operations

- Suppose A is the event that the computer stops at the first light and B is the event that she stops that she stops at the third light

$$A = \{\text{sss}, \text{ssc}, \text{scc}, \text{scs}\}$$
$$B = \{\text{sss}, \text{scs}, \text{ccs}, \text{css}\}$$

Events

- The union of two events:

  $C = A \cup B = \{\text{sss}, \text{ssc}, \text{scc}, \text{scs}, \text{ccs}, \text{css}\}$
- The intersection of two events: $C = A \cap B = \{\text{sss}, \text{scs}\}$
- The complement of an event: $A^c = \{\text{ccc}, \text{ccs}, \text{css}, \text{csc}\}$

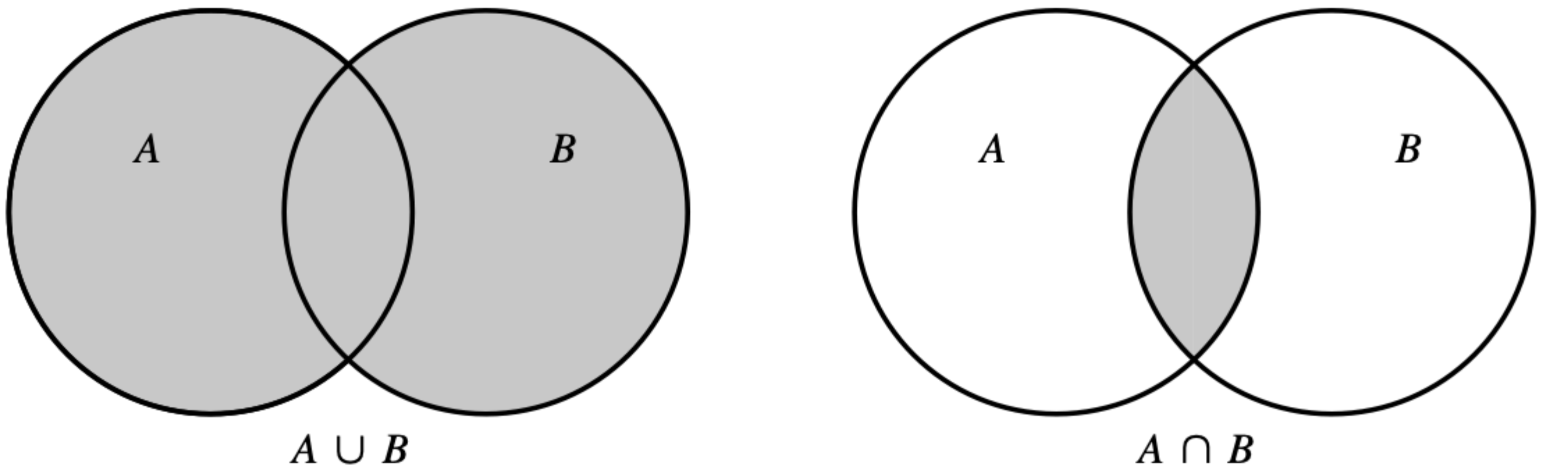


FIGURE 1.1 Venn diagrams of $A \cup B$ and $A \cap B$.

Events

- Commutative laws:
$$A \cup B = B \cup A\,,$$
$$A \cap B = B \cap A\,.$$

- Associative laws:
$$\left(A \cup B\right) \cup C = A \cup \left(B \cup C\right),$$
$$\left(A \cap B\right) \cap C = A \cap \left(B \cap C\right).$$

- Distributive laws:
$$\left(A \cup B\right) \cap C = \left(A \cap C\right) \cup \left(B \cap C\right),$$
$$\left(A \cap B\right) \cup C = \left(A \cup C\right) \cap \left(B \cup C\right).$$

1.3 Probability Measures

An original definition of probability

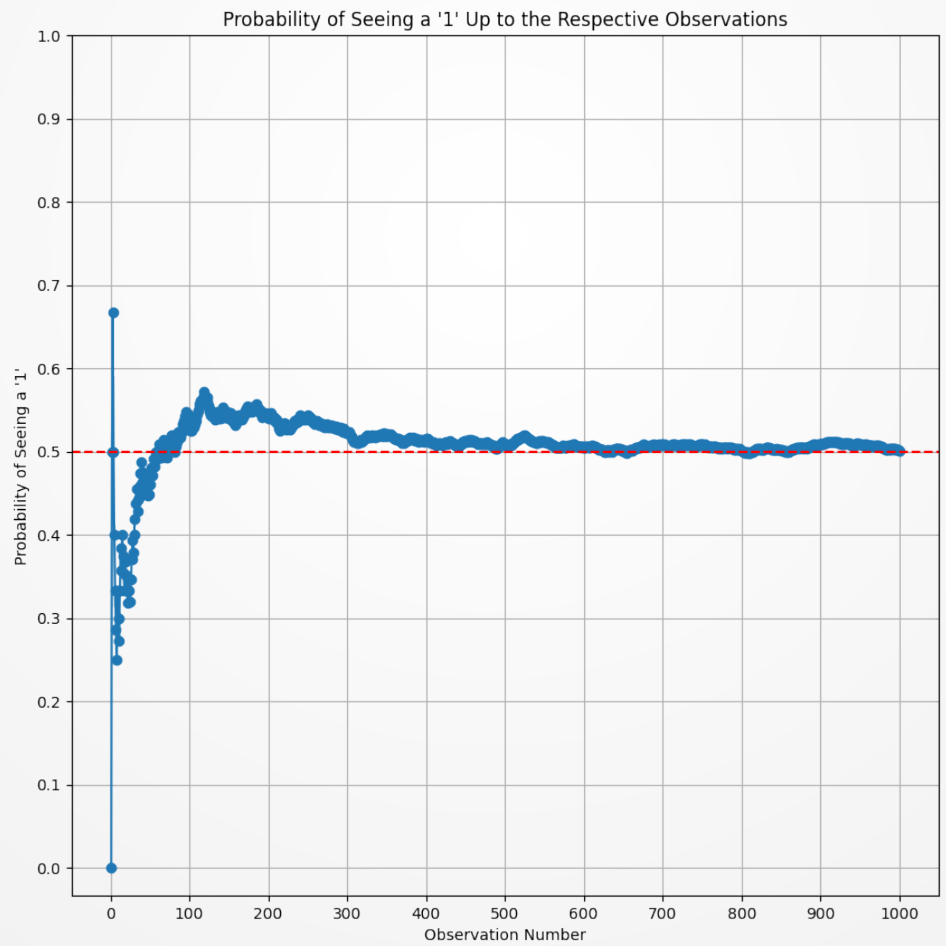
- What is the probability of observing a head (H) or tail (T), when tossing a fair coin?
- The sample space is $\Omega = \{H, T\}$. We are interested in the event $A = \{H\}$.

$$P(H) = \lim_{n \to \infty} \frac{\#H}{n}$$

  - where n is the number of independent trial and $\#H$ is the number of heads observed in the n trials.
  - We denote H by 1 and T by 0.
- Essentially, probability contains two ideas:
  - Ratio
  - Repeated sampling (a dynamic procedure)

An example

- See the result of randomly selection: [0 1 0 1 0 1 0 0 1 0]
- Ratios of observing 1 from the *I*-th observations:
  [0, 1/2, 1/3, 2/4, 2/5, 3/6, 3/7, 3/8, 4/9, 4/10]



https://colab.research.google.com/drive/1QM4lwy3EDU7HWFIU8O4Kokdzq3HeuhWf#scrollTo=7cMDCen6Yo2R

Probability

A probability measure

- A probability measure on Ω is a function P from subsets of Ω to the real numbers that satisfies the following axioms:
  - $P(\Omega) = 1$;
  - If $A \in \Omega$, then $P(A) \geq 0$;
  - If $A_1, A_2, \cdots,$ are mutually disjoint, then

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

Properties

- $P(A^c) = 1 - P(A)$.
- $P(\varnothing) = 0$.
- If $A \subset B$, then $P(A) \leq P(B)$.
- Addition law: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

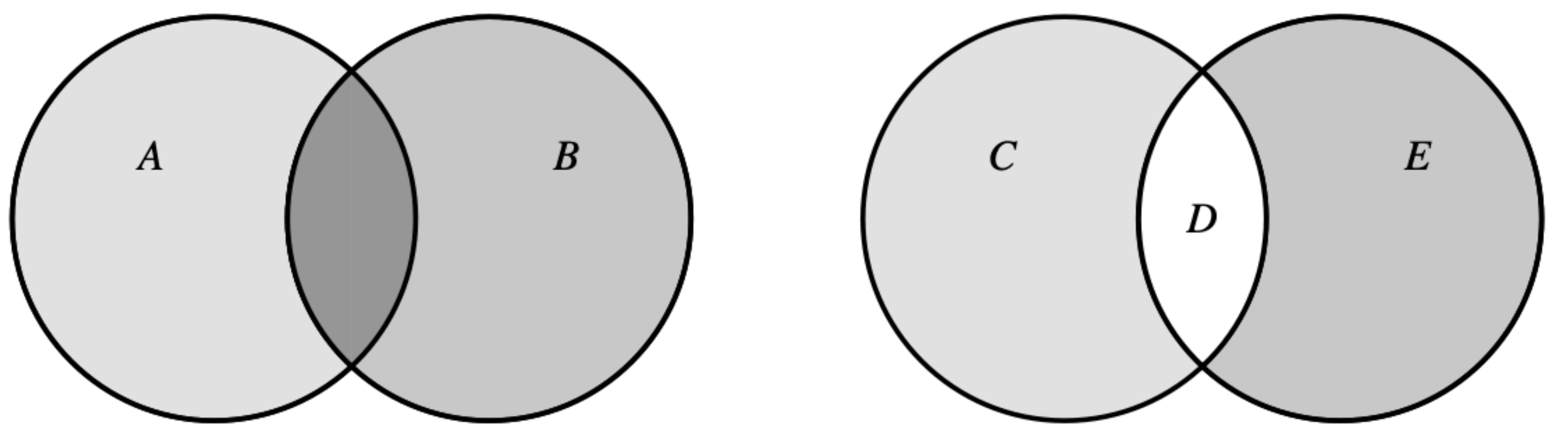


FIGURE 1.2 Venn diagram illustrating the addition law.

1.4 Computing Probabilities: Counting Methods

Counting rules to probabilities

- The element of Ω all have equal probabilities; so if there are N elements in Ω, each of them has probability $1/N$.
- If A can occur if any of n mutually exclusive ways, then $P(A) = n/N$, or

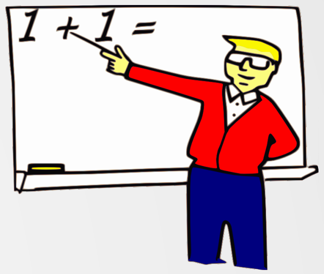
$$P(A) = \frac{\text{number of ways } A \text{ can occur}}{\text{total number of outcomes}}$$

Multiplication principle

- If there are p experiments and the first has n_1 possible outcomes, the second n_2, and the p-th n_p possible outcomes, then there is a total of $n_1 \times n_2 \times \cdots \times n_p$ possible outcomes for the p experiments.

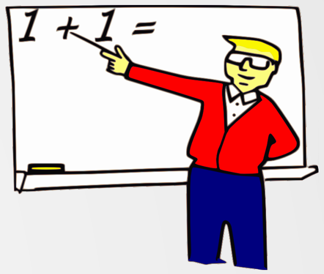
Example 1

- A class has 12 boys and 8 girls. The teacher selects 1 boy and 1 girl to act as representative to the student government.
- Solution
  - She can do this in any $12 \times 8 = 96$ different ways.



Example 2

- An 8-bit binary word is a sequence of 8 digits, of which each may be either a 0 or a 1.
- Solution
  - There are $2^8 = 256$ such words.

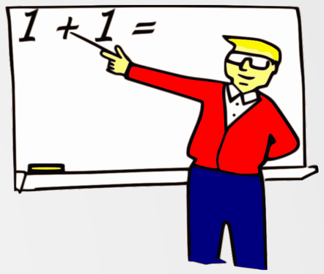


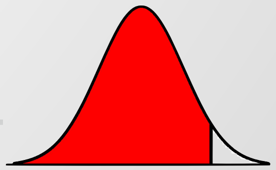
Proposition A

- For a set of size n and a sample of size r,
  - there are n^r different ordered sampled samples with replacement
  - there are $n(n-1)(n-2)\cdots(n-r+1)$ different order samples without replacement.

Corollary A

- The number of ordering n elements is $n(n-1)(n-2)\cdots = n!$.
- Example:
  - How many ways can five children be lined up?
- Solution
  - There are $5!$ ways.





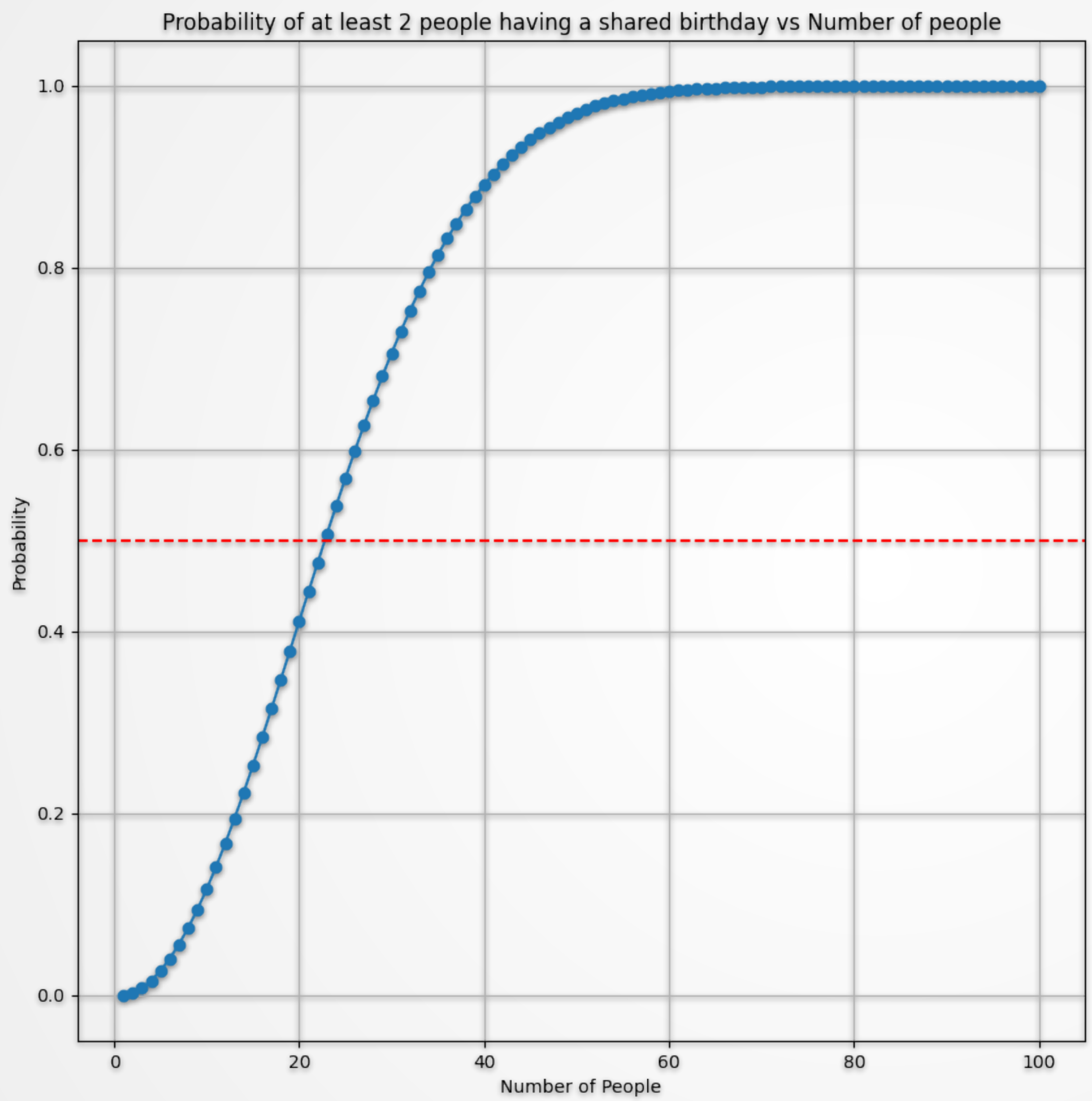
Birthday problem

- Suppose that a room contains n people. What is the probability that at least two of them have a common birthday? Assume that every day of the year is equally likely to be a birthday. Let A denote the event that at least two people have a common birthday. Then,

$$P\left(A^c\right) = \frac{365 \times 364 \times \cdots \times (365 - n + 1)}{365^n}$$

$$P(A) = 1 - P\left(A^c\right) = 1 - \frac{365 \times 364 \times \cdots \times (365 - n + 1)}{365^n}$$

Birthday problem



Birthday coincidence

https://colab.research.google.com/drive/1QM4lwy3EDU7HWFIU8O4Kokdzq3HeuhWf#scrollTo=YEGBmNkPWcX_

Proposition B

- Denote

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

- The number of unordered samples of k objects selected from n objects without replacements is $\binom{n}{k}$. The numbers $\binom{n}{k}$, called the binomial coefficients, occur in the expansion

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$$

Proposition B

- In particular,

$$2^n = \sum_{k=0}^{n} \binom{n}{k}$$

- The latter results can be interpreted as the number of subsets of a set of n objects.

Capture/Recapture method

- We assume that there are n animals in the population, of which 10 are tagged. If the 20 animals captured later are taken in such a way that all $\binom{n}{20}$ possible groups are equally likely, the probability that 4 of them are tagged is

$$\frac{\binom{10}{4}\binom{n-10}{16}}{\binom{n}{20}}$$

Likelihood for Example 1

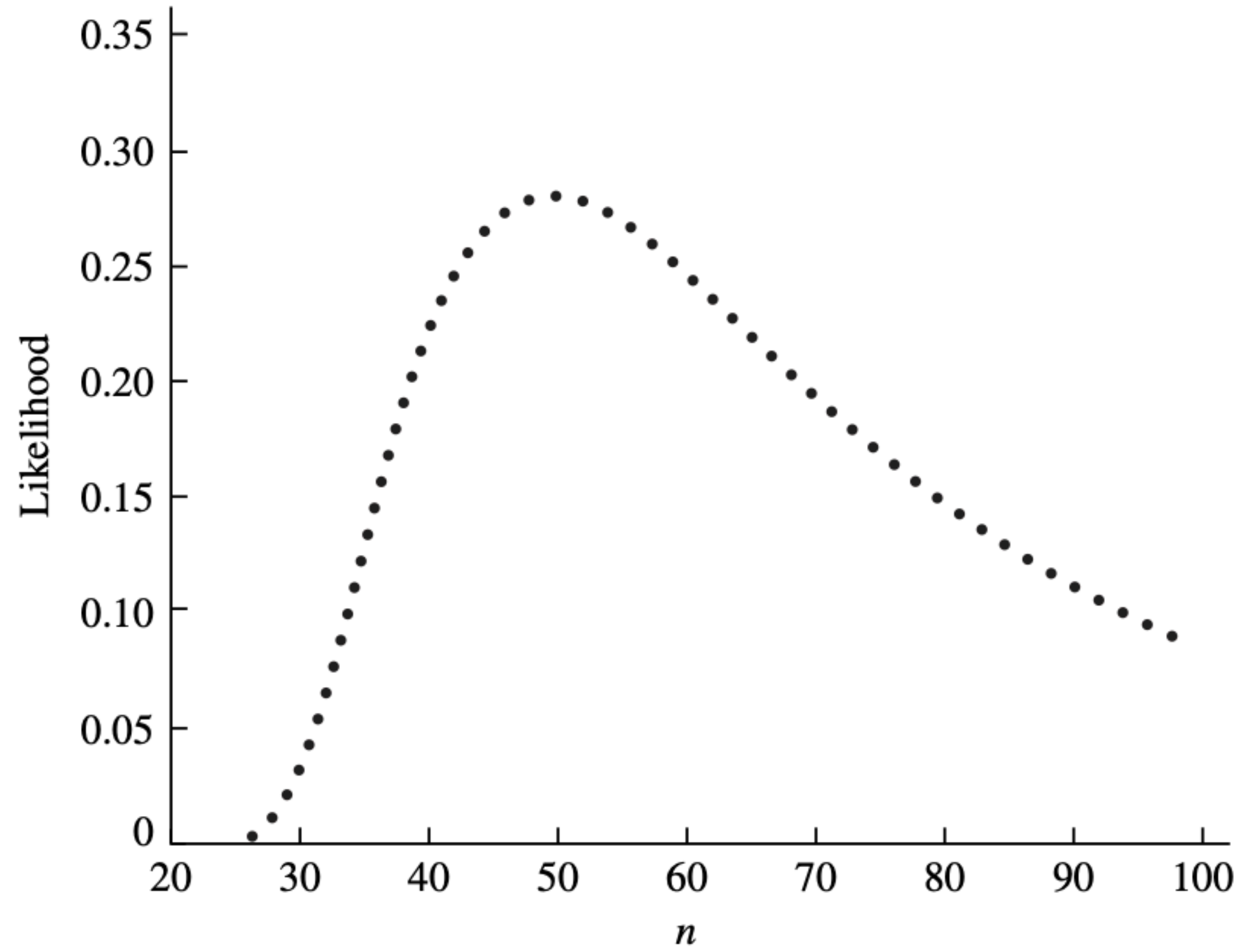


FIGURE **1.3** Likelihood for Example I.

Proposition C

- The number of ways that n objects can be grouped into r classes with n_i in the i-th class, $i = 1, \cdots, r$, and $\sum_{i=1}^{r} n_i = n$ is

$$\binom{n}{n_1 \, n_2 \cdots \, n_r} = \frac{n!}{n_1! \, n_2! \, \cdots \, n_r!}$$

Multinomial coefficients

- The numbers $\binom{n}{n_1\, n_2\, \cdots\, n_r}$ are called multinomial coefficients.

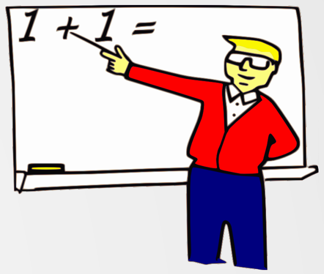
  They occur in the expansion

$$(x_1 + x_2 + \cdots + x_r)^n = \sum \binom{n}{n_1\, n_2\, \cdots\, n_r} x_1^{n_1} x_2^{n_2} \cdots x_r^{n_r}$$

where the sum is over all non-negative integers $n_1, n_2, \cdots, n_r$ such that $n_1 + \cdots + n_r = n$.

Example

- A committee of seven members is to be divided into three subcommittees of size three, two, and two.
- Solution
  - This can be done in



$$\binom{7}{3\;2\;2} = \frac{7!}{3!2!2!} = 210$$

ways.

1.5 Conditional Probability

Conditional probability

- Let A and B be two events with $P(B) \neq 0$.
- The conditional probability of A given B is defined to be

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

Multiplication law

- Let A and B be events and assume that $P(B) \neq 0$. Then.

$$P\left(A \cap B\right) = P\left(A \mid B\right) P(B)$$

Law of total probability

- Let $B_1, B_2, \cdots, B_n$ be such that $\bigcup_{i=1}^{n} B_i = \Omega$ and $B_1 \cap B_j = \varnothing$ for $i \neq j$, with $P(B_i) > 0$ for all i. Then for any event A,

$$P(A) = \sum_{i=1}^{n} P\left(A \mid B_i\right) P(B_i)$$

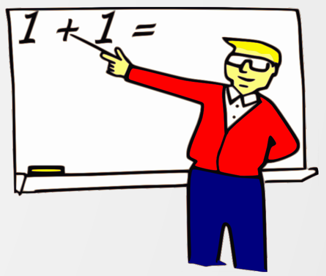
Proof

$$
\begin{aligned}
P(A) &= P\left(A \cap \Omega\right) \\
&= P\left(A \cap \left(\bigcup_{i=1}^{n} B_i\right)\right) \\
&= P\left(\bigcup_{i=1}^{n} (A \cap B_i)\right) \\
&= \sum_{i=1}^{n} P\left(A \cap B_i\right) \\
&= P\left(A \mid B_i\right) P(B_i) \,.
\end{aligned}
$$

Example 1

- An urn contains three red balls and one blue ball. Two balls are selected without replacement. What is the probability that they are both red?
- Solution
  - Let R_i denote the events that a red ball is draw on the i-th trial. Then

$$P(R_1 \cap R_2) = P(R_1)P(R_2 | R_1) = \frac{3}{4}\frac{2}{3} = \frac{1}{2}$$



Example 2

- An urn contains three red balls and one blue ball. Two balls are selected without replacement. What is the probability that a red ball is selected on the second draw?
- Solution

$$P(R_2) = P\left(R_2 \mid R_1\right)P(R_1) + P\left(R_2 \mid B_1\right)P(B_1) = \frac{2}{3}\frac{3}{4} + \frac{1}{1}\frac{1}{4} = \frac{3}{4}$$

Bayes' rule

- Let A and $B_1, \cdots, B_n$ be events where the B_i are disjoint, $\bigcup_{i=1}^{n} B_i = \Omega$, and $P(B_i) > 0$ for all i. Then

$$P\left(B_j \mid A\right) = \frac{P\left(A \mid B_j\right) P(B_j)}{\sum_{i=1}^{n} P\left(A \mid B_i\right) P(B_i)}$$

Proof

- Because $P\left(B_j \mid A\right) = \dfrac{P\left(A \cap B_j\right)}{P(A)}$, we have

  $$P\left(A \cap B_j\right) = P\left(A \mid B_j\right)P(B_j)\,.$$

- Second, we have

$$\begin{aligned} P(A) &= P\left(A \cap \Omega\right) \\ &= P\left(A \cap \left(\bigcup_{i=1}^{n} B_i\right)\right) \\ &= P\left(\bigcup_{i=1}^{n} (A \cap B_i)\right) \\ &= \sum_{i=1}^{n} P\left(A \cap B_i\right) \\ &= P\left(A \mid B_i\right)P(B_i)\,. \end{aligned}$$

1.6 Independence

Proof

- Because $P\left(B_j \mid A\right) = \dfrac{P\left(A \cap B_j\right)}{P(A)}$, we have

$$P\left(A \cap B_j\right) = P\left(A \mid B_j\right)P(B_j)\,.$$

- Second, we have

$$\begin{aligned} P(A) &= P\left(A \cap \Omega\right) \\ &= P\left(A \cap \left(\bigcup_{i=1}^{n} B_i\right)\right) \\ &= P\left(\bigcup_{i=1}^{n} (A \cap B_i)\right) \\ &= \sum_{i=1}^{n} P\left(A \cap B_i\right) \\ &= P\left(A \mid B_i\right)P(B_i)\,. \end{aligned}$$

Intuitions for independent events

- Intuitively, we would say that two events, A and B, are independent, if knowing that one had occurred gave us no information about whether the other had occurred; that is,

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

and

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

Definition of independent events

- Now, if

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

then

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

An easier definition of independent events for checking

- A and B are said to be independent events if

$$P\left(A \cap B\right) = P(A)P(B)$$

- $A_1, A_2, \cdots, A_n$ are said to be *mutually independent*, if for any sub collection, $A_{i_1}, \cdots, A_{i_m}$,

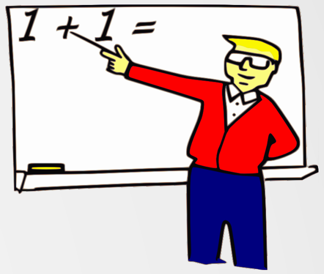
$$P\left(A_{i_1} \cap \cdots \cap A_{i_m}\right) = P\left(A_{i_1}\right) \cdots P\left(A_{i_m}\right)$$

- Note: pairwise independence does not imply mutual independence.

Example

- A fair coin is tossed twice. Let A denote the event of heads on the first toss, B the event of heads on the second toss, and C the event that exactly one head is throw.
  - A and B are independent:
  - A and C are independent:
  - A, B, and C are not mutually independent:

Definition of independent events



Solution

$$
\begin{aligned}
P(A) = P(B) = P(C) &= 1/2, \\
P(A \cap B) = P(A \cap C) = P(B \cap C) &= 1/4, \\
P(A \cap B \cap C) &= 0.
\end{aligned}
$$

Example

- Consider a circuit with three relays. Let A_i denote the event that the i -th relay works, and assume that $P(A_i) = p$ and that the relays are mutually independent. If F denotes the event that current flows through the circuit. Find $P(F)$.

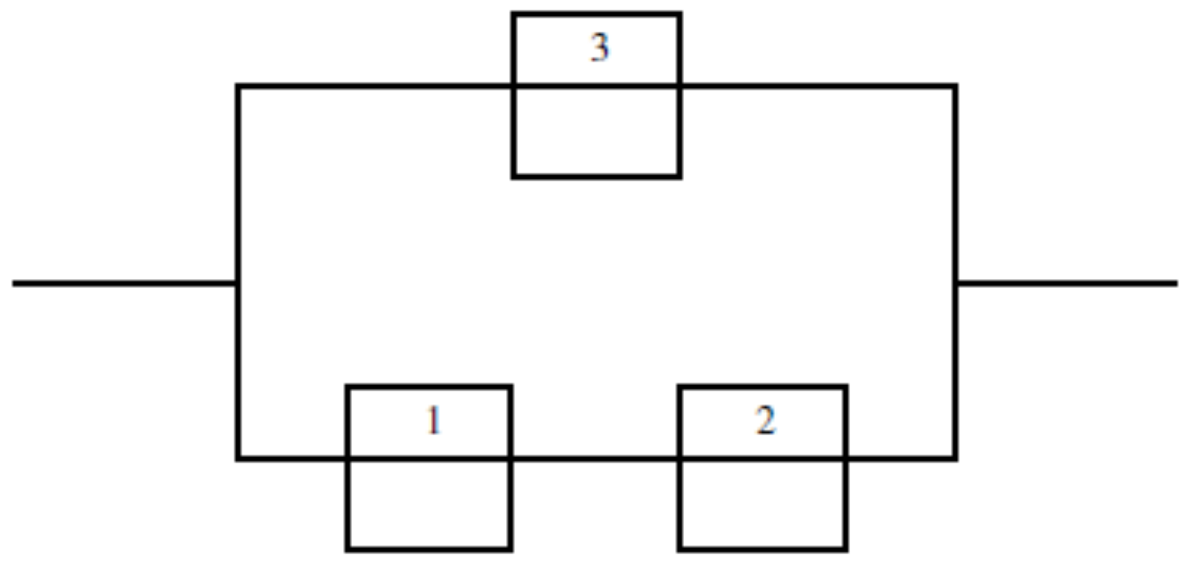
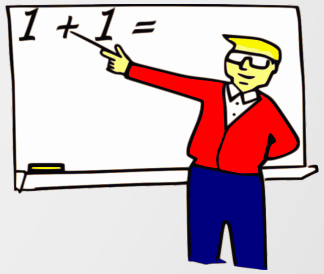


FIGURE 1.4 Circuit with three relays.



- Sol. $P(F) = P(A_3) + P\left(A_1 \cap A_2\right) - P\left(A_1 \cap A_2 \cap A_3\right)$.