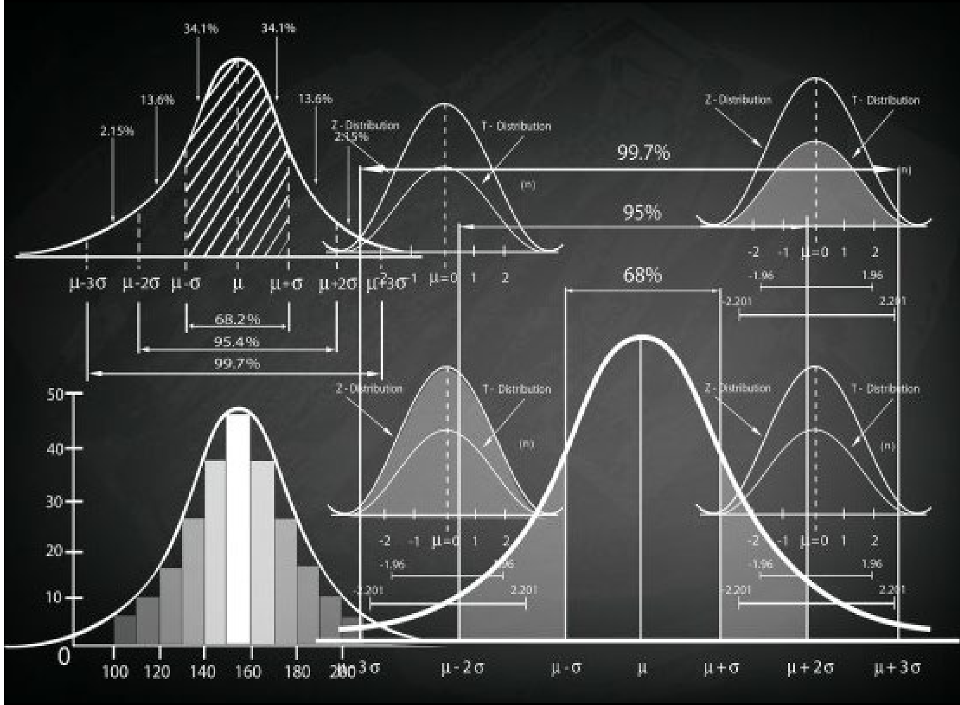
CS1510: Statistics for CS

Lecture 01:
Practicalities & Introduction to Probability Theory

Marijn ten Thij & Tim Dick

Probability Theory
vs
Statistics

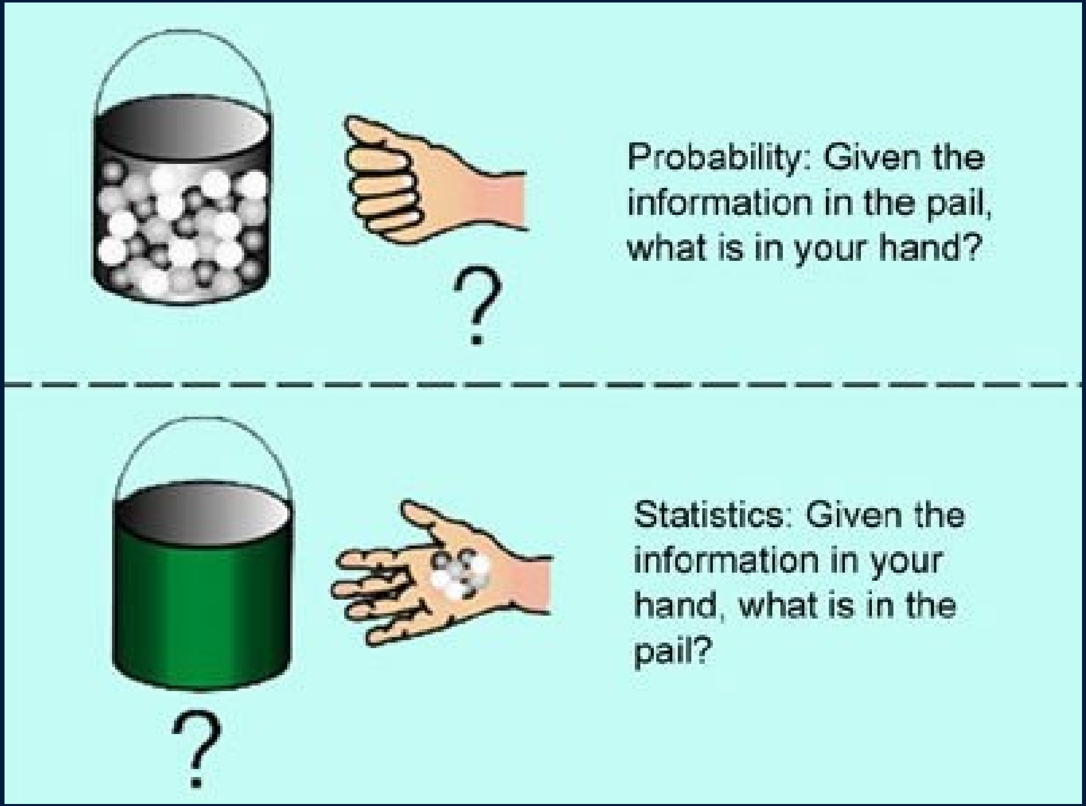
- ***Probability theory:***
  calculate the likelihood of an occurrence of a given event.
- ***Statistics:***
  how to collect, organize, and interpret information



Picture from:
https://online.stanford.edu/courses/epi259-introduction-probability-and-statistics-epidemiology

Maastricht University

Probability Theory vs Statistics



Picture from: https://www.quora.com/What-is-the-difference-between-probability-and-statistics

Main goals of the course

- To have an ***understanding of fundamental concepts*** in probability and statistics
- To be ***familiar*** with the ***most frequently used*** probability distributions/densities and statistical procedures (statistical estimation and hypothesis tests)
- To be able to ***recognize*** several ***probability distributions*** in real life situations to which they typically apply

Intuition often fails us when dealing with probabilities and data!

Picture from: https://medium.com/illumination/understanding-probability-through-intuition-0e99b1c5e700

Course organization

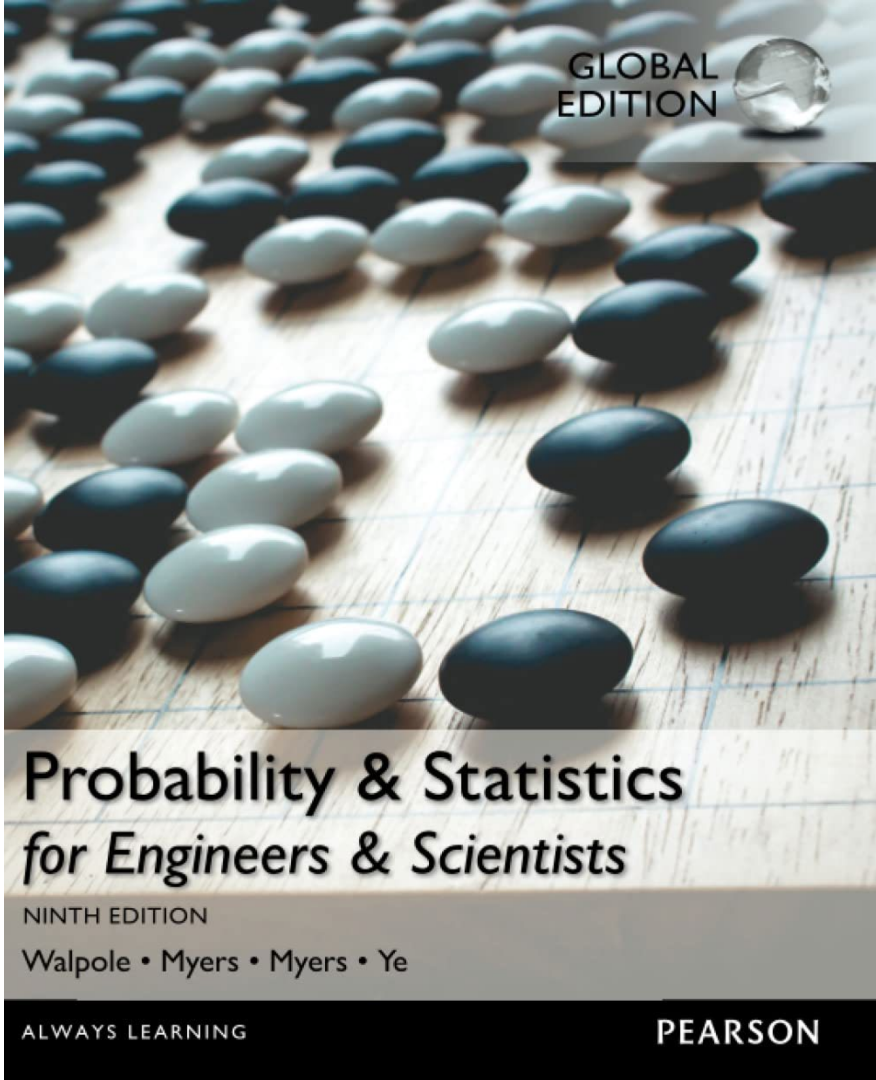
Textbook

Probability and Statistics for Engineers and Scientists

9^{th} *edition*

Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, and Keying E. Ye

Maastricht University



Instruction format

Lectures

Tutorials

Assignment

Organization of Tutorials

- Tutorials take place in two groups

Group	***01***	***02***
Lecturer	Marijn ten Thij	Tim Dick

- Both rooms will cover the exact same content
- Group Assignment will be visible in your schedule
- Exercises are posted on Canvas in advance of tutorials

Organization of Assignment

- Design and perform an experiment in groups
  - min 8 students and max 12 students per group
- The experiment should test a hypothesis
- Each group will act as participants for other groups
- Each group member will write an individual report (max 3 pages) that discusses:
  - The experiment itself
  - The hypothesis that is tested
  - The statistical analysis of the results of the experiment

Desired prior knowledge

Calculus

Discrete Mathematics

Examination: Grading

The following ***mandatory*** components will be examined, and will count towards the final grade of the course:

- Written Exam (WE), graded 0.0-10.0, weight 80%
- Individual Assignment (AS), graded 0.0-10.0 , weight 20%

Assessment component	Form and extent (length)	Grading scheme
Written exam (mandatory)	2 hours	Points per exam question depend on difficulty and expected length.
Individual Assignment (mandatory)	Short report (max. 3 pages) which describes an experiment and the analysis of its outcomes	Rubrics based

The final grade is calculated as:

$$IF\left(AND(WE > 0, AS > 0), MAX\left(1, ROUND(0.8 \cdot WE + 0.2 \cdot AS)\right), "NG"\right)$$

Examination: Resit

Assessment component	Form of the resit
Written exam (mandatory)	Regular resit in the form of a written exam in the ongoing academic year
Individual Assignment (mandatory)	If you have a score between 40% and 60% for the assignment, and you failed the overall course at the regular opportunity, you are eligible for a repair opportunity for the assignment. If you make the repair opportunity in a satisfactory way, you will get a 60% score for the assignment, otherwise the score will stay as it was.

Individual Assignment

- **Report:** Hand-in a single PDF file of at most 3 pages!
  - Additional pages will be disregarded.
- **No submission:** grade 0.0 for assignment
- **Late policy:** **(this grace period lasts 15 minutes)**
  - The assignment will disappear shortly after the deadline.
  - If you submit in this grade period, you lose **25%** of the grade.

    (e.g., score: $8.0 \Rightarrow$ **grade:** 5.5**)**
- **Academic Integrity:**
  - You are expected to perform the statistical analysis of the experimental data on your own, you cannot collaborate on this part.
  - The use of generative AI is **not** allowed.
  - Plagiarism checks **will** be performed!

Exam

- Closed book exam
- **Duration:** 120 minutes, without breaks.
- **Allowed aids:** Pen, calculator from DACS allowed calculator list.
- Mock exam will be posted on CANVAS (time TBD)
  - Multiple Choice: related to discussed concepts
  - Statements: true/false + explanation
  - Open problems: similar to tutorial exercises

Communication

- Lectures
- Tutorials
- Canvas discussion board
- Please, do NOT send us e-mails with questions

Course Overview

Tentative Schedule (Probability Theory)

W	Event	Topic	Day	Date	Time	Room
1	Lecture 1	Intro to Probability and Random Variables (Sections 2.1 – 2.5 and 3.1 – 3.3)	Mon	08/04/2024	13:30h – 15:30h	EPD150 MSM
	Tutorial 1	Intro to Probability Theory (Sections 2.1 – 2.5)	Thu	11/04/2024	08:30h – 10:30h	PHS1 C0.016/C.020
	Tutorial 2	Intro to Random Variables (Sections 3.1 – 3.3)	Fri	12/04/2024	13:30h – 15:30h	PHS1 C0.016/C.020
2	Lecture 2	Conditional Probability, Bayes Rule and Joint Distributions (Sections 2.6, 2.7 and 3.4)	Mon	15/04/2024	13:30h – 15:30h	EPD150 MSM
	Tutorial 3	Conditional Probability and Bayes Rule (Sections 2.6 and 2.7)	Thu	18/04/2024	08:30h – 10:30h	PHS1 C0.016/C.020
	Tutorial 4	Joint Distributions (Section 3.4)	Fri	19/04/2024	13:30h – 15:30h	PHS1 C0.016/C.020

Tentative Schedule (Probability Theory)

W	Event	Topic	Day	Date	Time	Room
3	Lecture 3	Properties of Distributions & Probability Distributions (Sections 4.1 – 4.3 and 5.1)	Mon	22/04/2024	13:30h – 15:30h	EPD150 MSM
	Tutorial 5	Properties of Distributions & Probability Distributions (Sections 4.1 – 4.3 and 5.1)	Thu	25/04/2024	08:30h – 10:30h	PHS1 C0.016/C.020
4	Lecture 4	Discrete and Continuous Probability Distributions (Sections 5.1 – 5.5 and 6.1 – 6.6)	Mon	29/04/2024	13:30h – 15:30h	EPD150 MSM
	Tutorial 6	Discrete and Continuous Probability Distributions (Sections 5.1 – 5.5 and 6.1 – 6.6)	Thu	01/05/2024	11:00h – 13:00h	PHS1 C0.016/C.020

Tentative Schedule (Statistics)

W	Event	Topic	Day	Date	Time	Room
5	Lecture 5	Introduction to Statistics and Estimators (Sections 8.1 – 8.3 and 9.1 – 9.3)	Mon	06/05/2024	13:30h – 15:30h	EPD150 MSM
	Tutorial 7	Introduction to Statistics and Estimators (Sections 8.1 – 8.3 and 9.1 – 9.3)	Wed	08/05/2024	11:00h – 13:00h	PHS1 C0.016/C.020
6	Lecture 6	CLT, Interval Estimate and Differences of Means (Sections 8.4 – 8.6 and 9.4, 9.5, 9.8 – 9.10 & 9.12)	Mon	13/05/2024	13:30h – 15:30h	EPD150 MSM
	Tutorial 8	CLT, Interval Estimate and Differences of Means (Sections 8.4 – 8.6 and 9.4, 9.5, 9.8 – 9.10 & 9.12)	Thu	16/05/2024	08:30h – 10:30h	PHS1 C0.016/C.020
7	Lecture 7	One and Two Sample Tests of Hypothesis (Sections 10.1 – 10.5, 10.8, 10.11 – 10.13)	Wed	22/05/2024	11:00h – 13:00h	EPD150 MSM
	Tutorial 9	Tests of Hypothesis for the Mean (Sections 10.1 – 10.5)	Thu	23/05/2024	08:30h – 10:30h	PHS1 C0.016/C.020
	Tutorial 10	Addition statistical tests (Sections 10.8, 10.11 – 10.13)	Fri	24/05/2024	11:00h – 13:00h	PHS1 C0.016/C.020

Tentative Schedule (Practical)

W	Topic	Day	Date	Time	Room
7	Assignment is posted on Canvas	Wed	22/05/2024	13:00h	
	Deadline to submit the experimental design on Canvas	Thu	23/05/2024	18:00h	
8	Question hour	Mon	27/05/2024	13:30h – 15:30h	EPD150 MSM
	Deadline to submit the worked out experiment on Canvas	Mon	27/05/2024	18:00h	
	Deadline to submit the outcomes of the experiment on Canvas	Wed	29/05/2024	18:00h	
	Deadline to submit the ***individual*** report on Canvas	Fri	31/05/2024	18:00h	

Probability Theory Basics

Sections 2.1 and 2.2

Probability Theory Basics

Experiment

An action whose outcome is determined by chance

Example: Rolling a single die

Sample Space

The set of possible outcomes of an experiment (denoted by S)

Example: the number of spots on each side of the die

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

Event

A subset of the sample space (denoted by $E \subseteq S$)

Example: the outcome of the experiment is a multiple of 3.

$$E = \{3, 6\}$$

Probability Theory Basics (example)

Experiment: Draw a card out a full deck of cards

Sample space S: the cards in a full deck of cards.

$$S = \{\clubsuit 2, \clubsuit 3, \dots, \clubsuit A, \diamondsuit 2, \dots, \diamondsuit A, \heartsuit 2, \dots, \heartsuit A, \spadesuit 2, \dots, \spadesuit K, \spadesuit A\}$$

(54 outcomes)

Event E: the card drawn from the deck is a queen.

$$E = \{\clubsuit Q, \diamondsuit Q, \heartsuit Q, \spadesuit Q\}$$

Probability Theory Basics (example)

Experiment: Flipping a coin twice

Sample space S: all possible outcomes (H and T denote heads and tails, respectively).

$$S = \{HH, HT, TH, TT\} \text{ (4 outcomes)}$$

Event E: one flip lands heads and one flip lands tails.

$$E = \{HT, TH\}$$

Probability Theory Basics (example)

Experiment: Rolling a die 3 times

Sample space S: all possible outcomes (respectively).

$$S = \{111, 112, 113, \ldots, 666\} \text{ } (6 \times 6 \times 6 = 216 \text{ outcomes})$$

Event E: three times the same number, i.e.,

$$E = \{111, 222, 333, 444, 555, 666\}$$

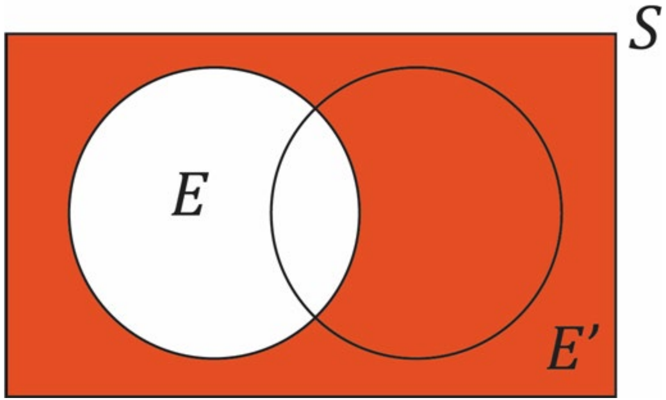
Events are sets, as a consequence we also have

Complement of an Event E (denoted by E', or E^C)

Example: Rolling a single die

$$S = \{1,2,3,4,5,6\},$$
$$E = \{1,3,5\}$$
$$E' = \{2,4,6\}$$

Figure: E' highlighted



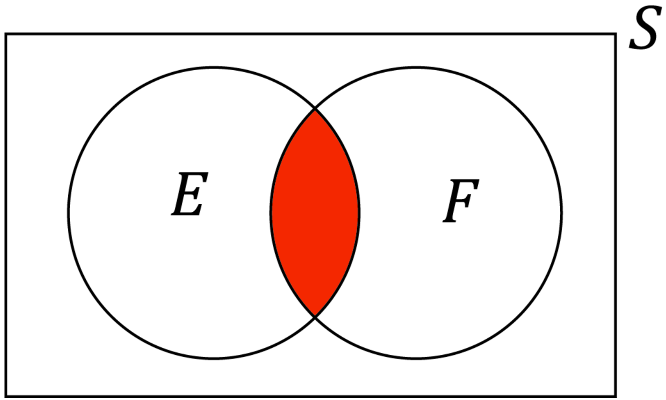
Events are sets, as a consequence we also have

Intersection of Events E and F (denoted by $E \cap F$)

Example: Rolling a single die

$$S = \{1,2,3,4,5,6\},$$
$$E = \{1,3,5\}, F = \{3,6\}$$
$$E \cap F = \{3\}$$

Figure: $E \cap F$ highlighted



Events are sets, as a consequence we also have

Mutually Exclusive Events E and F: $E \cap F = \emptyset$

Example: Rolling a single die

$$S = \{1,2,3,4,5,6\},$$
$$E = \{1,3,5\}, F = \{4\}$$
$$E \cap F = \emptyset$$

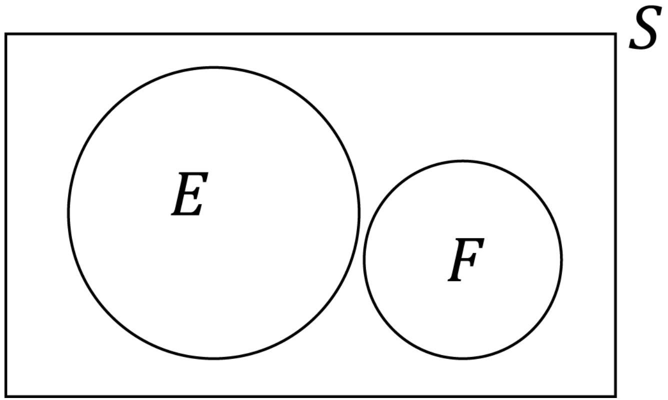


Figure: $E \cap F$ highlighted

Events are sets, as a consequence we also have

Union of Events E and F (denoted by $E \cup F$)

Example: Rolling a single die

$$S = \{1,2,3,4,5,6\},$$
$$E = \{1,3,5\}, F = \{3,6\}$$
$$E \cup F = \{1,3,5,6\}$$

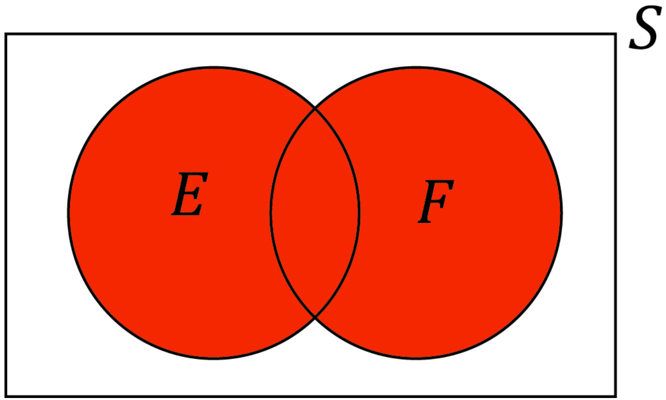


Figure: $E \cup F$ highlighted

Combinatorics

Section 2.3

Calculating the number of possibilities

When you draw k times from a collection of size n, the number of possible outcomes, given the restrictions imposed on the drawing, can be calculated using:

	With replacement	Without replacement
Ordered sampling	n^k	$\frac{n!}{(n-k)!}$ (i)
Unordered sampling	$\binom{n+k-1}{k}$ (ii)	$\binom{n}{k}$ (iii)

Ordered sampling without replacement

Intuition:

$$\underbrace{n \times (n-1) \times (n-2) \times \cdots \times (n-k+1)}_{k \text{ factors}} = \boxed{\frac{n!}{(n-k)!}}$$

Unordered sampling without replacement

Intuition:

$$\underbrace{\overbrace{\times \quad \times \quad \times}^{k \text{ items}}}_{} \qquad \underbrace{\ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots}_{n \text{ positions}} \quad = \quad \boxed{\binom{n}{k}}$$

Unordered sampling with replacement

Intuition:

$$\overbrace{\times \quad \times \quad \times}^{k \text{ items}}$$

$$\underbrace{| \quad | \quad | \quad | \quad | \quad | \quad | \quad | \quad | \quad |}_{n \text{ positions} = n-1 \text{ borders between positions}} = \boxed{\binom{n+k-1}{k}}$$

Probability Theory Basics (continued)

Sections 2.4 and 2.5

Probability Theory Basics

Probability measure

A function that maps each possible event E to a number in $[0, 1]$.

Properties of probability measures:

1. $P(S) = 1$
2. $P(\emptyset) = 0$
3. $0 \leq P(E) \leq 1 \forall E \subseteq S$

Probability of an event

The probability of an event E is the sum of the probabilities of the elements in E (denoted by $P(E)$)

Probability measures (example)

A fair die

$$S = \{1,2,3,4,5,6\}, P(\{1\}) = P(\{2\}) = \cdots = P(\{6\}) = \frac{1}{6}$$

An unfair die

$$S = \{1,2,3,4,5,6\}, P(\{1\}) = \cdots = P(\{5\}) = \frac{1}{10}, P(\{6\}) = \frac{1}{2}$$

5 apples, 4 pears

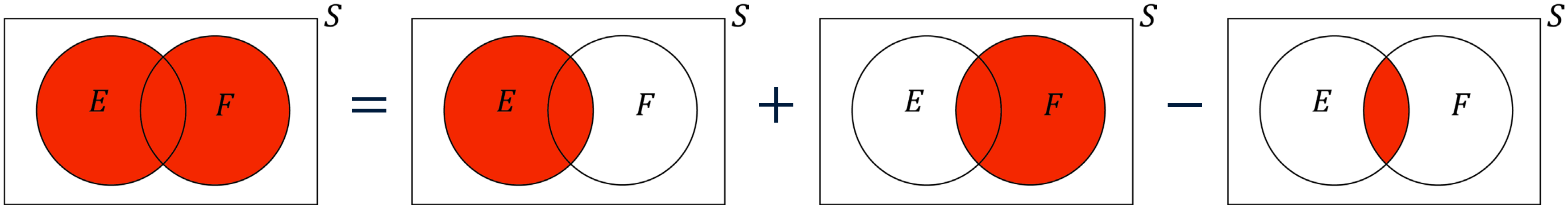
$$S = \{A_1, A_2, A_3, A_4, A_5, P_1, P_2, P_3, P_4\}$$

$$P(\{A_1\}) = \cdots = P(\{P_4\}) = \frac{1}{9}, P(\{apple\}) = \frac{5}{9}, P(\{pear\}) = \frac{4}{9}$$

Probability Theory Basics

Rules of calculating probabilities

- $P(E) = \frac{|E|}{|S|}$, if each element in S has equal probability, where $|\cdot|$ denotes the cardinality of the set.
- Given events E and F: $P(E \cup F) = P(E) + P(F) - P(E \cap F)$
- For disjoint events E and F: $P(E \cup F) = P(E) + P(F)$

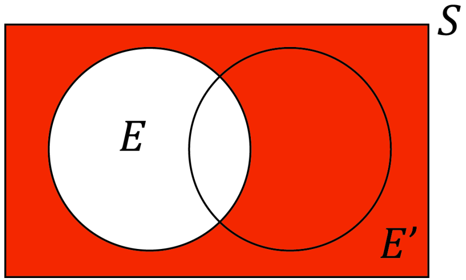


Probability Theory Basics

Rules of calculating probabilities

- $P(E) = \frac{|E|}{|S|}$, if each element in S has equal probability, where $|\cdot|$ denotes the cardinality of the set.
- Given events E and F: $P(E \cup F) = P(E) + P(F) - P(E \cap F)$
- For disjoint events E and F: $P(E \cup F) = P(E) + P(F)$
- For any event E: $P(E) = 1 - P(E')$

Proof: $1 = P(S) = P(E \cup E') = P(E) + P(E')$



Calculating Probabilities (example)

Example: Rolling a fair die

$S = \{1,2,3,4,5,6\}$, events: $A = \{1\}, B = \{2,4,6\}, C = \{1,2,3\}$

- $P(A) = \frac{|A|}{|S|} = \frac{1}{6}$
- $P(B \cup C) = P(B) + P(C) - P(B \cap C) = \frac{|B|}{|S|} + \frac{|C|}{|S|} - \frac{|B \cap C|}{|S|} = \frac{5}{6}$

  $= \frac{|B \cup C|}{|S|} = \frac{5}{6}$
- $P(A \cup B) = P(A) + P(B) = \frac{|A|}{|S|} + \frac{|B|}{|S|} = \frac{2}{3}$
- $P(C') = 1 - P(C) = 1 - \frac{|C|}{|S|} = 1 - \frac{3}{6} = \frac{1}{2}$

Introduction to Random Variables

Sections 3.1 – 3.3

Introduction to Random Variables

Random Variable (RV)

X is a random variable for the sample space S if it assigns a real number to each element of S, $X\colon\ S \to \mathbb{R}$

Example 1

$$S = \{1,2,3,4\}$$

X: square: $\forall s \in S \colon X(s) = s^2$

$$Y\colon \begin{cases} 2s & \text{if } s \text{ is odd} \\ \frac{s}{2} & \text{if } s \text{ is even} \end{cases}$$

$$Y(1) = 2, Y(2) = 1, Y(3) = 6, Y(4) = 2$$

Introduction to Random Variables

Random Variable (RV)

X is a random variable for the sample space S if it assigns a real number to each element of S, $X\colon\ S \to \mathbb{R}$

Example 2

Rolling a die until a 6 comes up $S = \{6, N6, NN6, NNN6, \dots\}$
where N denotes an outcome of 1,2,3,4 or 5
X: number of rolls required

$$X(6) = 1, X(N6) = 2, X(NN6) = 3, \dots$$

Introduction to Random Variables

Discrete vs Continuous Random Variables

X is a discrete if S is finite or countable, otherwise it is continuous

Example

$S = [1,2]$ (notice: S is uncountable)

$X(s) = \frac{1}{s} \forall s \in S \Rightarrow X \in \left[\frac{1}{2}, 1\right]$ is continuous

$Y(s) = \lfloor 2s \rfloor \forall s \in S \Rightarrow Y(s) = \begin{cases} 2 & if\ s \in \left[\frac{1,3}{2}\right) \\ 3 & if\ s \in \left[\frac{1,3}{2}\right) \\ 4 & if\ s = 2 \end{cases} \Rightarrow Y$ is discrete

Discrete Random Variables

Probability distribution

The probability distribution of a discrete random variable (RV) is defined as $f(x) = P(X = x)$

Example: Rolling a die until a 6 comes up

$S = \{6, N6, NN6, NNN6, \dots\}$

X: number of rolls required, i.e. $X(6) = 1, X(N6) = 2, \dots$

$$f(1) = \frac{1}{6}$$

$$f(2) = \frac{5}{6} \cdot \frac{1}{6}$$

Discrete Random Variables

Probability distribution

The probability distribution of a discrete random variable (RV) is defined as $f(x) = P(X = x)$

Example: Rolling a die

$S = \{1,2,3,4,5,6\}$, $\mathrm{X}(s) = \left\lfloor \frac{s}{2} \right\rfloor$, where s is the outcome of the roll

$$f(0) = P(X = 0) = P(\{1\}) = \frac{1}{6}$$

$$f(1) = P(X = 1) = P(\{2,3\}) = \frac{1}{3}, f(2) = \frac{1}{3}, f(3) = \frac{1}{6}$$

Discrete Random Variables

Cumulative distribution (denoted by $F(x)$)

$$F(x) = P(X \leq x) = \sum_{y \leq x} f(y)$$

Example: Rolling a die

$S = \{1,2,3,4,5,6\}$, $X(s) = \left\lfloor \frac{s}{2} \right\rfloor$, where s is the outcome of the roll

x	0	1	2	3
$f(x) = P(X = x)$	1/6	1/3	1/3	1/6
$F(x) = P(X \leq x)$	1/6	1/2	5/6	1

Exercise

A shipment of 20 similar laptops to a retail outlet contains 3 that are defective. If a school makes a random purchase of 2 of these laptops.

a) Find the probability distribution for the number of defective laptops

b) Find the cumulative distribution function $F(x)$

Solution

$X = $ # defective laptops among the purchased laptops.

$$P(X = x) = \frac{\binom{17}{2-x} \cdot \binom{3}{x}}{\binom{20}{2}}$$

x	0	1	2
$f(x) = P(X = x)$	0.716	0.268	0.016
$F(x) = P(X \leq x)$	0.716	0.984	1

Continuous Random Variables

Example:

$S = [0,1]$, $X(s) = s$, and suppose that all outcomes are "equally likely".

Q: What is $P\left(X = \frac{1}{2}\right)$?

Q: What is $P\left(X \leq \frac{1}{2}\right)$? $\frac{1}{2}$

Cumulative distribution (denoted by $F(x)$)

$F(x) = P(X \leq x)$ (is well defined this way)

Continuous Random Variables

Probability density function (denoted by $f(x)$)

$$f(x) = F'(x)$$

Example

$$F(x) = \begin{cases} 0 & if\ x \leq 0, \\ x & if\ 0 \leq x \leq 1, \\ 1 & if\ x \geq 1 \end{cases}$$

$$f(x) = \begin{cases} 0 & if\ x \leq 0, \\ 1 & if\ 0 \leq x \leq 1, \\ 0 & if\ x \geq 1 \end{cases}$$

Continuous Random Variables

Properties of $F(x)$ and $f(x)$

- $0 \leq F(x) \leq 1$ for all x
- $F(x)$ is non-decreasing, and therefore $f(x) \geq 0$
- $\lim_{x \to \infty} F(x) = 1$ and $\lim_{x \to -\infty} F(x) = 0$
- $\lim_{x \to \infty} f(x) = 0$ and $\lim_{x \to -\infty} f(x) = 0$
- $\int_{-\infty}^{\infty} f(x)dx = 1$
- $F(x) = \int_{-\infty}^{x} f(t)dt$
- $P(a \leq X \leq b) = F(b) - F(a) = \int_{a}^{b} f(x)dx = 1$

Continuous Random Variables

Example: a) Show that f is indeed a density function

X is an RV with density $f(x) = \begin{cases} \frac{7-2x}{2}, & 2 < x < 3 \\ 0, & \text{elsewhere} \end{cases}$

a) Solution: to show that $\int_{-\infty}^{\infty} f(x)dx = 1$

$$\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{2} 0\, dx + \int_{2}^{3} \frac{7-2x}{2} dx + \int_{3}^{\infty} 0\, dx = \left[\frac{7}{2}x - \frac{1}{2}x^2\right]_{x=2}^{3} = 1$$

Continuous Random Variables

Example: b) Find $F(x)$

X is an RV with density $f(x) = \begin{cases} \frac{7-2x}{2}, & 2 < x < 3 \\ 0, & \text{elsewhere} \end{cases}$

b) Solution: $F(x) = \int_{-\infty}^{x} f(t)dt$, so

$$F(x) = \begin{cases} \displaystyle\int_{-\infty}^{x} 0\,dt = 0 \quad if\ x < 2, \\ \displaystyle\int_{2}^{x} \frac{7-2x}{2}\,dt = \left[\frac{7}{2}t - \frac{1}{2}t^2\right]_{x=2}^{3} = -\frac{1}{2}x^2 + \frac{7}{2}x - 5 \quad if\ 2 \leq x \leq 3, \\ \displaystyle\int_{-\infty}^{x} f(t)dt = \int_{2}^{3} f(t)dt = 1 \quad if\ x > 3 \end{cases}$$

Continuous Random Variables

Example: c) Find $P\left(1\frac{1}{2} \leq X \leq 2\frac{1}{2}\right)$

X is an RV with density $f(x) = \begin{cases} \frac{7-2x}{2}, & 2 < x < 3 \\ 0, & \text{elsewhere} \end{cases}$

c) Solution:

$$P\left(1\frac{1}{2} \leq X \leq 2\frac{1}{2}\right) = F\left(2\frac{1}{2}\right) - F\left(1\frac{1}{2}\right) = -\frac{1}{2}\left(2\frac{1}{2}\right)^2 + \frac{7}{2}\left(2\frac{1}{2}\right) - 5 - 0$$

$$= -\frac{25}{8} + \frac{35}{4} - 5 = \frac{-25 + 70 - 40}{8} = \frac{5}{8}$$