MS-A0503 First course in probability and statistics

Jukka Kohonen

Department of mathematics and systems analysis
Aalto SCI

Academic year 2022–2023
Period III

Course overview

Lectures: Mon and Fri 10-12 in U2
(except 27.1. in hall B)

- Lecturer: Jukka Kohonen

Weekly exercises 2 x 2h
+ Online STACK exercises

- Head assistant: Markus Hirvensalo
  markus.hirvensalo@aalto.fi
- For practical questions about the exercises, contact the head assistant or your own assistant.

https://mycourses.aalto.fi/course/view.php?id=36271

Passing the course

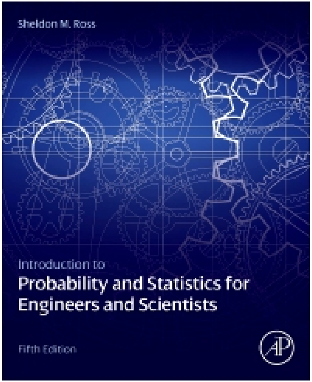
Two alternatives:

(a) Exam 60% + exercises 40%

(b) Exam 100%

- For every student taking the exam, we calculate both grades, and apply the better one. You do not need to ask for this separately.
- In each case, 50% of points is enough to pass the course.
- Exercise points are valid for exams in periods III and IV.

Reading materials



Lecture slides from the course page:
https://mycourses.aalto.fi/course/view.php?id=36271

Course book, available as e-book from Aalto library:
http://www.sciencedirect.com.libproxy.aalto.fi/science/book/9780123948113

Also available but in Finnish: L. Leskelä's lecture notes
http://math.aalto.fi/~lleskela/LectureNotes003.html

Learning outcomes

After passing the course, the student

- can compute probabilities of composite events by applying operations of set theory
- is familiar with the most important discrete and continuous probability distributions and recognizes situations that can modeled with them
- can apply joint distributions to compute statistics of random vectors and recognizes when two random variables are stochastically independent
- knows methods for estimating the parameters of a statistical model
- can compute the posterior distributions and make conclusions based on them
- can explain what can and what cannot be concluded from a p-value of chosen statistical test

Workload

- Participating in lectures 24 h (4 h/week)
- Participating in exercise classes 24 h (4 h/week)
- Weekly independent study 36–72 h (6–12 h/week)
- Participating and preparing for exam 4–40 h

Altogether 88–160 h $\approx$ 5 credits

Independent study is crucial in mathematics. Including:

- reading the material (lecture slides, textbook)
- solving exercises
- also reading the exercise solutions, and thinking about them
- asking questions from the teachers

Lecture plan

L1A Probability: Concept and basic rules
L1B Random variables and distributions

L2A Expected value and transformations
L2B Standard deviation and correlation

L3A Normal approximation
L3B Statistical datasets

L4A Parameter estimation
L4B Confidence intervals

L5A Bayesian inference
L5B Bayesian estimates

L6A Significance tests
L6B Wrap-up

After this course...?

Aalto University

Probability and Statistics Courses 2019–2020

- MSc course
- BSc course
- Prequisite course
- Recommended order

Large random systems
MS-E1602, Next time in spring 2021, 5 cr

Brownian motion and stochastic analysis
MS-E1991, Period V, 5 cr
Lecturer: Kalle Kytölä

Brownian motion is a fundamentally important stochastic process, discovered in the contexts of financial markets and statistical physics. It relates to diverse mathematical topics from partial differential equations to constructive quantum field theory. This course introduces you to key techniques for working with Brownian motion, including stochastic integration, martingales, and Itô's formula.

Multivariate location and scatter
MS-E1992, Period II, 5 cr
Lecturer: Pauliina Ilmonen

When we are dealing with multivariate observations, the very first questions that come to our mind are: Where is the data? How is it scattered? These questions are discussed, different multivariate location and scatter functionals and estimates are presented, and some selected applications are considered. This is an advanced course in statistics for master's students and doctoral students.

Multivariate statistical analysis
MS-E2112, Periods III–IV, 5 cr
Lecturer: Pauliina Ilmonen

You get introduced to common multivariate data analysis techniques. You learn to apply these methods in practice using R programming. Course topics include multivariate location and scatter, principal component analysis, multivariate correspondence analysis, canonical correlation analysis, discriminant analysis, classification, and clustering.

Random graphs and network statistics
MS-E1603, Period I, 5 cr
Lecturer: Lasse Leskelä

You get introduced to the theory of statistical models (uniform random graphs, stochastic block models, graphons) used in predicting and learning structural properties of networks based on incomplete and noisy observations. The course is targeted to students in mathematics, operations research, and computer science.

Probability theory
MS-E1600, Period III, 5 cr
Lecturer: Lasse Leskelä

This course introduces you to the mathematical theory of randomness. You learn to work with probabilistic concepts such as stochastic independence, convergence of random sequences, information contained in a sigma-algebra, characteristic and generating functions of probability measures, laws of large numbers, and central limit theorems.

Prediction and time series analysis
MS-C2128, Period II, 5 cr
Lecturer: Pauliina Ilmonen

This course introduces you to methods and concepts used in analyzing and predicting statistical time series. Course topics include linear models and regression diagnostics, stationary random processes, various ARIMA models, Kalman filters, and an introduction to dynamic regression models. R programming is used in the exercises.

Euklidiset avaruudet
MS-C1540, 5 op

Stochastic processes
MS-C2111, Period II, 5 cr
Lecturer: Lasse Leskelä

This course will get you introduced to stochastic processes, the theory of time-dependent random phenomena. You learn to mathematically model and analyze particle and population flows using Markov processes, unpredictable time instants using Poisson processes, and gambling and investment strategies using martingales.

Statistical inference
MS-C1620, Periods III–IV, 5 cr
Lecturer: Jukka Kohonen

This course is an introduction to statistical analysis and statistical inference, with emphasis on statistical hypothesis testing and prediction methods. Course topics include estimation, simple parametric and nonparametric tests, statistical dependence and correlation, linear regression analysis, and analysis of variance. R programming is used in the exercises.

Todennäköisyyslaskennan ja tilastotieteen peruskurssi
MS-A050X, Periodit I, II, III, IV, 5 op
Luennoitsijat: Jukka Kohonen, Georg Metsalo

Kurssilla opitaan laskemaan todennäköisyyksiä ja odotusarvoja joukkojen, summien ja integraalien avulla. Lisäksi tutustutaan tilastollisiin menetelmiin, joiden avulla voi laatia estimaatteja ja ennusteita sekä analysoida tilastollista merkitsevyyttä havaitun datan ja prioritiedon valossa.

Introduction to R programming
MS-E1994, One-weekend intensive course, 1 cr

This course gets you started with R programming. No prior programming experience is required.

Matriisilaskenta
MS-A000X, 5 op

Differentiaali- ja integraalilaskenta 1
MS-A010X, 5 op

Interested in writing a BSc, MSc, or PhD thesis in stochastics and statistics?
Contact Pauliina Ilmonen, Kalle Kytölä, or Lasse Leskelä

Obtain credits by participating in the MS-E1609 Stochastics and statistics seminar?
Join **stochastics@list.aalto.fi** to stay updated

MS-A0503 First course in probability and statistics

1A Probability: Concept and basic rules

Jukka Kohonen

Deparment of mathematics and systems analysis
Aalto SCI

Academic year 2022–2023
Period III

Contents

Introduction

Random experiments

Basic rules of probability

Conditional probability

Probability and combinatorics

Contents

Introduction

Random experiments

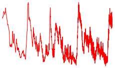
Basic rules of probability

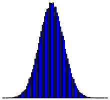
Conditional probability

Probability and combinatorics

Statistics and stochastics

Statistics is a science that aims to develop methods for making informed guesses and decisions based on incomplete and uncertain information.





Stochastics is the field of mathematics concerned with modelling randomness and probability.

Computation and visualization by a computer may be enough to study the properties of a particular data set.

The mathematical models of stochastics are needed whenever one wants to use the data to generalize and predict.

Probability: The concept

Probability is a way of quantifying the belief of something being true or false.

- Tossing a coin gives "heads" with probability $\frac{1}{2}$
- Next Monday in Otaniemi it will rain with probability
  - 14% (says Ilmatieteen laitos)
  - 19% (says Foreca)

"Interpretations" of probability

- Objective (relative *frequency* in the long run)
- Subjective (*degree of belief*, based on some information)

The two interpretations are not in conflict, but support each other. Also, the *mathematical laws* of probability are the same in both cases.

http://www.stat.berkeley.edu/~aldous/Real-World/100.html

Contents

Introduction

Random experiments

Basic rules of probability

Conditional probability

Probability and combinatorics

Random experiment

A random experiment is a process that will result in something occurring (an outcome), but we do not exactly know what.

- Sample space S is the set of all possibile outcomes
- Outcome $=$ an element of the sample space, $s \in S$
- Event $=$ a set of outcomes; a subset of the sample space, $A \subset S$

Terminology

- An event A occurs, if it contains the outcome that occurs
- The full set S is the certain event
- The empty set $\emptyset$ is the impossible event

Example: Rolling a die

- Outcome i = the result of the roll
- Sample space $S = \{1, 2, \dots, 6\}$
  - Events are all subsets of S, for example
    - A = "outcome is even" $= \{2, 4, 6\}$.
    - B = "outcome is bigger than four" $= \{5, 6\}$.

Example: Two rolls of a die

- An outcome is a pair of integers (i, j), where i is the first roll result and j is the second roll result
- Sample space is

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

Events are for example

- $A =$ "the two results are equal" $= \{(1,1),(2,2),(3,3),(4,4),(5,5),(6,6)\}$.
- $B =$ "first roll was one" $= \{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6)\}$.

Example: Tomorrow's rainfall in Otaniemi (mm)

- Outcomes are real numbers $x \geq 0$.
- Sample space $S = [0, \infty)$.

Events are e.g.

- $A =$ "rainfall exceeds 10 mm" $= (10, \infty)$
- $B =$ "no rain tomorrow" $= \{0\}$

Combining events

We can combine events into new events by different logical operations.

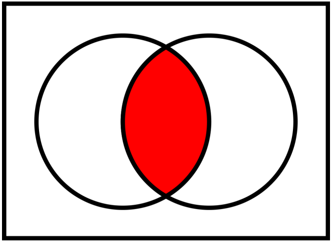
- "A and B occur"
- "A or B occurs"
- "A does not occur"
- "B occurs but A does not"

In probability, it is customary to use the language of *set theory*, because events are *sets* of possible outcomes.

Intersection of events

The intersection of two events, denoted $A \cap B$, contains every outcome that belongs to A and also belongs to B.

$$A \cap B = \{s \in S : s \in A \text{ and } s \in B\}.$$



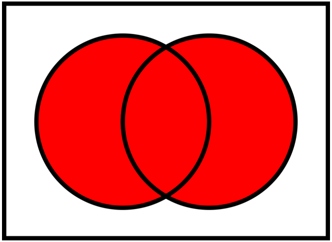
Example (One die)

- $A =$ "Result exceeds 3" $= \{4, 5, 6\}$
- $B =$ "Result is even" $= \{2, 4, 6\}$
- $A \cap B =$ "Result exceeds 3 and is even" $= \{4, 6\}$

Union of events

The union of two events, denoted $A \cup B$, contains every such outcome that belongs to A or to B (or both).

$$A \cup B = \{s \in S : s \in A \text{ or } s \in B\}.$$



Example (Die roll)

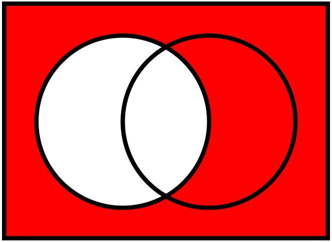
- $A =$ "Result exceeds 3" $= \{4, 5, 6\}$
- $B =$ "Result is even" $= \{2, 4, 6\}$
- $A \cup B =$ "Result exceeds 3 or is even" $= \{2, 4, 5, 6\}$

Note that "or" is usually understood as "inclusive or", that is, it allows the possibility of both happening.

Complement of an event

The complement of an event, denoted A^c, contains exactly those outcomes that are not in A.

$$A^c = \{s \in S : s \notin A\}.$$



Example (Die roll)

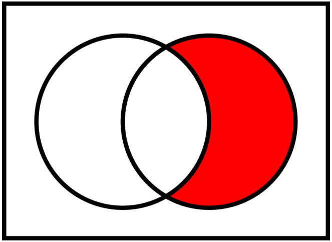
- $A =$ "Result exceeds 3" $= \{4, 5, 6\}$
- $A^c =$ "Result does not exceed 3"
  $=$ "Result at most 3" $= \{1, 2, 3\}$

Careful with inequalities! The complement of "bigger than 3" is not "smaller than 3", but "smaller or equal to 3".

Difference of events

The difference event $B \setminus A = B \cap A^c$ contains the outcomes that are in B, but are not in A.

$$B \setminus A = \{s \in S : s \in B \text{ and } s \notin A\}.$$



Example (Die roll)

- $A =$ "Result exceeds 3" $= \{4, 5, 6\}$
- $B =$ "Result is even $= \{2, 4, 6\}$
- $B \setminus A =$ "Result is even but does not exceed 3" $= \{2\}$

Mutually exclusive events

Two events A and B are mutually exclusive (or disjoint), if they cannot both occur:

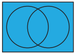
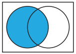
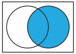
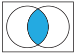
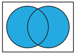
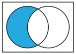
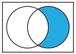
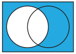
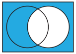
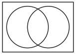
$$A \cap B = \emptyset.$$

Several events $A_1, A_2, \dots$ are mutually exclusive, if every pair of events is mutually exclusive (only one of the events can occur at the same time).

Example (Die roll)

- $A = \{1, 2\}$ and $\{3, 4\}$ are mutually exclusive.
- $A = \{1, 2, 3\}$ and $\{2, 4, 5\}$ are not mutually exclusive.
- $A = \{1, 2\}$, $B = \{3, 4\}$ and $C = \{5, 6\}$ are mutually exclusive.

Combining events — Summary

Name	Notation	Definition	Venn diagram	Interpretation
Sample space	S	$\{x \in S : x \in S\}$		Certain event
Event	A	$\{x \in S : x \in A\}$		A occurs
Event	B	$\{x \in S : x \in B\}$		B occurs
Intersection	$A \cap B$	$\{x \in S : x \in A \text{ and } x \in B\}$		A and B occur
Union	$A \cup B$	$\{x \in S : x \in A \text{ or } x \in B\}$		A or B occurs (or both)
Difference	$A \setminus B$	$\{x \in S : x \in A \text{ and } x \notin B\}$		A occurs but B does not
Difference	$B \setminus A$	$\{x \in S : x \in B \text{ and } x \notin A\}$		B occurs but A does not
Complement	A^c	$\{x \in S : x \notin A\}$		A does not occur
Complement	B^c	$\{x \in S : x \notin B\}$		B does not occur
Empty set	$\emptyset$	$\{x \in S : x \notin S\}$		Impossible event

Contents

Introduction

Random experiments

Basic rules of probability

Conditional probability

Probability and combinatorics

The axioms

A probability on the sample space S is a *function* from events to numbers.
The probability of event A is denoted $\mathbb{P}(A)$.
We have some basic requirements for the function.

(i) The whole sample space S has probability $\mathbb{P}(S) = 1$.

(ii) For every event A we have $0 \leq \mathbb{P}(A) \leq 1$.

(iii) Additivity: For any collection of *mutually exclusive* events $A_1, A_2, \ldots$ we have

$$\mathbb{P}(A_1 \cup A_2 \cup \cdots) \ = \ \mathbb{P}(A_1) + \mathbb{P}(A_2) + \cdots$$

Other rules of calculating with probability can be deduced from these "axioms".

Further rules

- General addition rule:

$$\mathbb{P}(A \cup B) \ = \ \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

- Additivity of two *mutually exclusive* events:

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B), \quad \text{kun } A \cap B = \emptyset.$$

- Probability of complement and difference:

$$\mathbb{P}(A^c) = 1 - \mathbb{P}(A),$$
$$\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

- Monotonicity:

$$\mathbb{P}(A) \leq \mathbb{P}(B), \quad \text{kun } A \subset B.$$

These can be deduced from the axioms.

Contents

Introduction

Random experiments

Basic rules of probability

Conditional probability

Probability and combinatorics

Conditional probability

If A and B are two events, we define the conditional probability of A given that B occurs, by the formula

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}, \quad \mathbb{P}(B) \neq 0.$$

- Read as "probability of A given B", or "P of A given B"
- Interpretation: This is the probability of A occurring if B occurs.
- Note that $\mathbb{P}(A|B)$ is not the same as $\mathbb{P}(A \cap B)$.
- Also $\mathbb{P}(B|A)$ is different.
- If $\mathbb{P}(B) = 0$, we leave $\mathbb{P}(A|B)$ undefined.

General product rule

From the definition of conditional probability, we can simply deduce the general product rule.

Rule

If $\mathbb{P}(A) \neq 0$, *then*

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

Interpretation

Probability of the event "both A and B occur" is obtained by multiplying the probability of A with the conditional probability of B.

Product rule for several events (chain rule)

Rule

If $\mathbb{P}(A_1 \cap \cdots \cap A_{k-1}) \neq 0$, *then*

$$\begin{aligned}&\mathbb{P}(A_1 \cap \cdots \cap A_k)\\ &\qquad = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1 \cap A_2)\cdots\mathbb{P}(A_k|A_1 \cap \cdots \cap A_{k-1}).\end{aligned}$$

Interpretation

The probability for the event "all of $A_1, \ldots, A_k$ ocrrus" is obtained by multiplying together:

- probability of A_1,
- then conditional probability of A_2 given A_1,
- then conditional probability of A_3 given A_1 and A_2,
- ...
- conditional probability of A_k given $A_1, A_2, \ldots, A_{k-1}$.

Product rule — Example

From a well-shuffled deck (of 52 cards) we deal three cards. What is the probability that all three are spades?

- A_i = "ith card is spade"
- $A = A_1 \cap A_2 \cap A_3$

Apply the chain rule on three events.

$$\mathbb{P}(A) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1 \cap A_2) = \frac{13}{52} \cdot \frac{12}{51} \cdot \frac{11}{50} \approx 0.013.$$

There is another method that involves combinatorics (we'll learn about this later).

Stochastic dependence and independence

Two events A and B are independent, if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

If this does not hold, we say the events are (stochastically) dependent.

Several events $\{A_i, i \in I\}$ are independent, if

$$\mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_k}) \;=\; \mathbb{P}(A_{i_1}) \cdots \mathbb{P}(A_{i_k})$$

for all $i_1, i_2, \ldots, i_k \in I$.

Example

Some situations where independence is intuitively clear.

- Physically separate tosses of a coin (or a die).
- Sampling *with replacement*. Pick a lottery ticket from a box, *place it back in the box* and shuffle, then pick again a lottery ticket.

Independence and conditional probability

Fact

If $\mathbb{P}(A) \neq 0$ *and* $\mathbb{P}(B) \neq 0$*, then these three conditions are equivalent:*

- *A and B are independent.*
- $\mathbb{P}(A|B) = \mathbb{P}(A)$.
- $\mathbb{P}(B|A) = \mathbb{P}(B)$.

Interpretation

If $\mathbb{P}(A|B) \neq \mathbb{P}(A)$, then knowing whether B occurs or not affects the probability of A occurring (i.e. either makes it *more probable* or *less probable*).

Example: Two events when dealing one card

A random card is dealt from a shuffled deck.

- A = "the card is a spade"
- B = "the card is an ace"

Are A and B dependent or independent?

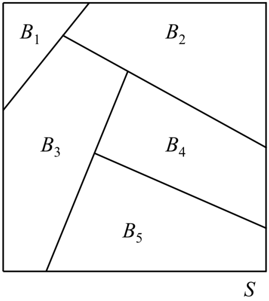
Let us calculate whether $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

- $\mathbb{P}(A) = \frac{13}{52} = \frac{1}{4}$.
- $\mathbb{P}(B) = \frac{4}{52} = \frac{1}{13}$.
- $\mathbb{P}(A \cap B) = \mathbb{P}(\text{"ace of spades"}) = \frac{1}{52}$.

Because $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$, we see that A and B are *independent* events.

Law of total probability

If we divide the sample space S into mutually exclusive events $B_1, \ldots, B_n$ whose union is the whole S, we have a partition of S.



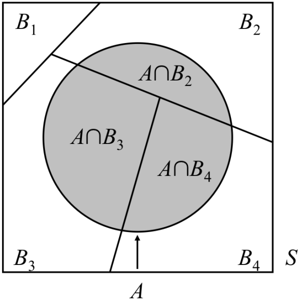
Rule

If $B_1, \ldots, B_n$ are a partition of the sample space and $\mathbb{P}(B_i) \neq 0$ for all i, then

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(B_i)\mathbb{P}(A|B_i).$$

Proof.

Events $C_i = A \cap B_i$ are mutually exclusive and their union is A.



Applying both additivity, and the product rule $\mathbb{P}(A \cap B_i) = \mathbb{P}(B_i)\mathbb{P}(A|B_i)$, we have

$$\mathbb{P}(A) = \mathbb{P}\left(\bigcup_{i=1}^{n} C_i\right) = \sum_{i=1}^{n} \mathbb{P}(C_i) = \sum_{i=1}^{n} \mathbb{P}(A \cap B_i)$$
$$= \sum_{i=1}^{n} \mathbb{P}(B_i)\mathbb{P}(A|B_i).$$

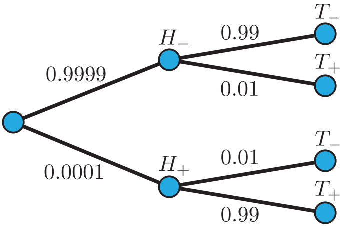
□

Example. A rare disease

In a population, 1/10000 of the people carry a certain disease. There is a test for the disease, but it makes *false positive* and *false negative* results, both with a probability 1%. What is the probability that a random person shows a positive ("diseased") test result?

$$\begin{aligned} H_- &= \text{"not diseased"} \qquad & T_- &= \text{"test is negative"} \\ H_+ &= \text{"diseased"} & T_+ &= \text{"test is positive"} \end{aligned}$$

Law of total probability $\Longrightarrow$
$$\begin{aligned} \mathbb{P}(T_+) &= \mathbb{P}(H_-)\mathbb{P}(T_+ \mid H_-) + \mathbb{P}(H_+)\mathbb{P}(T_+ \mid H_+) \\ &= 0.9999 \cdot 0.01 + 0.0001 \cdot 0.99 \\ &= 0.010098. \end{aligned}$$



Bayes' rule

Question: How are $\mathbb{P}(A|B)$ and $\mathbb{P}(B|A)$ related to each other?

Rule (Bayes' rule)

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

Proof.

Apply the definition of the conditional probability twice.

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

$\square$

Example. A rare disease

In a population, 1/10000 of the people carry a certain disease. There is a test for the disease, but it makes *false positive* and *false negative* results, both with a probability 1%. If we test a random person and the test is positive, what is the probability that the person indeed has the disease?

$$\begin{aligned} H_- &= \text{"not diseased"} & T_- &= \text{"test is negative"} \\ H_+ &= \text{"diseased"} & T_+ &= \text{"test is positive"} \end{aligned}$$

Previously we found $\mathbb{P}(T_+) = 0.010098$. Bayes' rule $\Longrightarrow$

$$\mathbb{P}(H_+ \mid T_+) \;=\; \frac{\mathbb{P}(H_+)\mathbb{P}(T_+ \mid H_+)}{\mathbb{P}(T_+)} \;=\; \frac{0.0001 \cdot 0.99}{0.010098} \;\approx\; 0.0098.$$

Is there something odd here?

:

- 99% of all test results are correct
- Over 99% of *positive* results are wrong!

Summary of the rules of probability

Addition rules (two versions)

$$\begin{aligned}\mathbb{P}(A \cup B) &= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) \\ &= \mathbb{P}(A) + \mathbb{P}(B) \quad \text{(if } A \text{ and } B \text{ mutually exclusive)}\end{aligned}$$

Product rule (two versions)

$$\begin{aligned}\mathbb{P}(A \cap B) &= \mathbb{P}(A)\mathbb{P}(B|A) \\ &= \mathbb{P}(A)\mathbb{P}(B) \quad \text{(if } A \text{ and } B \text{ independent)}\end{aligned}$$

Law of total probability

$$\mathbb{P}(A) = \sum_i \mathbb{P}(B_i)\mathbb{P}(A|B_i) \quad \text{(if the } B_i\text{'s are a partition of } S\text{)}$$

Bayes' rule

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

Contents

Introduction

Random experiments

Basic rules of probability

Conditional probability

Probability and combinatorics

Probability and combinatorics

If we have a finite sample space S that has n equally probable outcomes, then each outcome has probability $1/n$.

Then if an event A contains k outcomes, by additivity its probability is

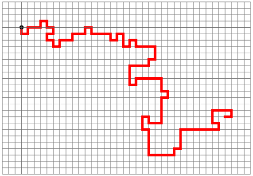
$$\mathbb{P}(A) = \frac{k}{n} = \frac{\#A}{\#S} = \frac{\text{the number of outcomes in } A}{\text{the number of outcomes in } S}.$$

A sample space whose outcomes are equally probable is called *symmetric*. It seems that probability in such spaces is trivially easy.

However ...

- It is difficult to "count" the elements of a large set one by one.
- Sometimes you can "calculate" the number of elements with more effient methods.
- Combinatorics is a field of mathematics that provides tools for this.

Sometimes combinatorics is difficult



Example (Difficult combinatorial problem)

A random walk in a square grid: Each step goes randomly in one of the four cardinal directions, except that it never goes back to where it has been. If the random walk is performed for 10^8 steps, what is the probability that it ends at a distance 10^6 from where it started?

Basic principles of combinatorics

Often, all elements of some set X can be produced by making *several choices consecutively*.

Sum rule.

1. Find how many choices there are for the first step.
2. In each case, find how many choices there are for the second step. *Be careful.* What you did in the first step may (or may not) affect how many choices you have now!
3. Continue until done (you have constructed an element of X).
4. Finally, *add up* the choices.

Product rule. If the first step has n choices, and each leads to m choices in the second step, then there are on $m + m + \ldots + m = n \cdot m$ possible results.

Ross calls this the *basic rule of counting*. You can obviously generalize it for more than two steps.

Typical combinatorial tasks

- Counting (ordered) sequences made of given elements,
  - if the same element can be used again
  - if the same element cannot be used again
- Counting the ways of ordering a given set of elements.
- Counting (unordered) subsets of a given set.

All of these can be solved by the product rule. The results are well-known formulas such as *power*, *factorial*, and *binomial coefficient*.

Advice: Learn both the basic formulas (for the most common situations) and the underlying principle of counting (for more difficult situations).

Number of sequences (repetition allowed)

How many four-digit PIN codes can you make of the ten digits?

Let us do it in four steps.

1. Choose the 1st digit for the PIN; 10 choices
2. Choose the 2nd digit for the PIN; 10 choices
3. Choose the 3rd digit for the PIN; 10 choices
4. Choose the 4th digit for the PIN; 10 choices

$\implies$ In total $10^4 = 10000$ choices

0000, 0001, 0002, 0003, 0004, 0005, 0006, 0007, 0008, 0009, 0010, 0011, 0012, 0013, 0014, 0015, 0016, 0017, 0018, 0019, 0020, 0021, 0022, 0023, 0024, 0025, 0026, 0027, 0028, 0029, 0030, 0031, 0032, 0033, 0034, 0035, 0036, 0037, 0038, 0039, 0040, 0041, 0042, 0043, 0044, 0045, 0046, 0047, 0048, 0049, 0050, 0051, 0052, 0053, 0054, 0055, 0056, 0057, 0058, 0059, 0060, 0061, 0062, 0063, 0064, 0065, 0066, 0067, 0068, 0069, 0070, 0071, 0072, 0073, 0074, 0075, 0076, 0077, 0078, 0079, 0080, 0081, 0082, 0083, 0084, 0085, 0086, 0087, 0088, 0089, 0090, 0091, 0092, 0093, 0094, 0095, 0096, 0097, 0098, 0099, 0100, 0101, 0102, 0103, 0104, 0105, 0106, 0107, 0108, 0109, 0110, 0111, 0112, 0113, 0114, 0115, 0116, 0117, 0118, 0119, 0120, 0121, 0122, 0123, 0124, 0125, 0126, 0127, 0128, 0129, 0130, 0131, 0132, 0133, 0134, 0135, 0136, 0137, 0138, 0139, 0140, 0141, 0142, 0143, 0144, 0145, 0146, 0147, 0148, 0149, 0150, 0151, 0152, 0153, 0154, 0155, 0156, 0157, 0158, 0159, 0160, 0161, 0162, 0163, 0164, 0165, 0166, 0167, 0168, 0169, 0170, 0171, 0172, 0173, 0174, 0175, 0176, 0177, 0178, 0179, 0180, 0181, 0182, 0183, 0184, 0185, 0186, 0187, 0188, 0189, 0190, 0191, 0192, 0193, 0194, 0195, 0196, 0197, . . . , 9983, 9984, 9985, 9986, 9987, 9988, 9989, 9990, 9991, 9992, 9993, 9994, 9995, 9996, 9997, 9998, 9999

Number of sequences (repetition not allowed)

There are 15 teams in a hockey series, namely (HPK, IFK, ILV, JUK, JYP, KAL, KÄR, KOO, LUK, PEL, SAI, SPO, TAP, TPS, ÄSS). How many different gold-silver-bronze sequences are possible?

Let us do it in three steps.

1. For gold, pick any team: 15 choices
2. For silver, pick any other team: 14 choices
3. For bronze, pick any team not yet picked: 13 choices

$\Longrightarrow$ In total $15 \times 14 \times 13 = 2730$ choices

(HPK,IFK,ILV),	(HPK,IFK,JUK),	(HPK,IFK,JYP),	(HPK,IFK,KAL),	(HPK,IFK,KÄR),	(HPK,IFK,KOO),	(HPK,IFK,LUK),
(HPK,IFK,PEL),	(HPK,IFK,SAI),	(HPK,IFK,SPO),	(HPK,IFK,TAP),	(HPK,IFK,TPS),	(HPK,IFK,ÄSS),	(HPK,ILV,IFK),
(HPK,ILV,JUK),	(HPK,ILV,JYP),	(HPK,ILV,KAL),	(HPK,ILV,KÄR),	(HPK,ILV,KOO),	(HPK,ILV,LUK),	(HPK,ILV,PEL),
(HPK,ILV,SAI),	(HPK,ILV,SPO),	(HPK,ILV,TAP),	(HPK,ILV,TPS),	(HPK,ILV,ÄSS),	(HPK,JUK,IFK),	(HPK,JUK,ILV),
(HPK,JUK,JYP),	(HPK,JUK,KAL),	(HPK,JUK,KÄR),	(HPK,JUK,KOO),	(HPK,JUK,LUK),	(HPK,JUK,PEL),	(HPK,JUK,SAI),
(HPK,JUK,SPO),	(HPK,JUK,TAP),	(HPK,JUK,TPS),	(HPK,JUK,ÄSS),	(HPK,JYP,IFK),	(HPK,JYP,ILV),	(HPK,JYP,JUK),
(HPK,JYP,KAL),	(HPK,JYP,KÄR),	(HPK,JYP,KOO),	(HPK,JYP,LUK),	(HPK,JYP,PEL),	(HPK,JYP,SAI),	(HPK,JYP,SPO),
(HPK,JYP,TAP),	(HPK,JYP,TPS),	(HPK,JYP,ÄSS),	(HPK,KAL,IFK),	(HPK,KAL,ILV),	(HPK,KAL,JUK),	(HPK,KAL,JYP),
(HPK,KAL,KÄR),	(HPK,KAL,KOO),	(HPK,KAL,LUK),	(HPK,KAL,PEL),	(HPK,KAL,SAI),	(HPK,KAL,SPO),	(HPK,KAL,TAP),
(HPK,KAL,TPS),	(HPK,KAL,ÄSS),	(HPK,KÄR,IFK),	(HPK,KÄR,ILV),	(HPK,KÄR,JUK),	(HPK,KÄR,JYP),	(HPK,KÄR,KAL),
(HPK,KÄR,KOO),	(HPK,KÄR,LUK),	(HPK,KÄR,PEL),	(HPK,KÄR,SAI),	(HPK,KÄR,SPO),	(HPK,KÄR,TAP),	(HPK,KÄR,TPS),
(HPK,KÄR,ÄSS),	(HPK,KOO,IFK),	(HPK,KOO,ILV),	(HPK,KOO,JUK),	(HPK,KOO,JYP),	(HPK,KOO,KAL),	(HPK,KOO,KÄR),
(HPK,KOO,LUK),	(HPK,KOO,PEL),	(HPK,KOO,SAI),	(HPK,KOO,SPO),	(HPK,KOO,TAP),	(HPK,KOO,TPS),	(HPK,KOO,ÄSS),
(HPK,LUK,IFK),	(HPK,LUK,ILV),	(HPK,LUK,JUK),	(HPK,LUK,JYP),	(HPK,LUK,KAL),	(HPK,LUK,KÄR),	(HPK,LUK,KOO),
(HPK,LUK,PEL),	(HPK,LUK,SAI),	(HPK,LUK,SPO),	(HPK,LUK,TAP),	(HPK,LUK,TPS),	(HPK,LUK,ÄSS),	(HPK,PEL,IFK),
(HPK,PEL,ILV),	(HPK,PEL,JUK),	(HPK,PEL,JYP),	(HPK,PEL,KAL),	(HPK,PEL,KÄR),	(HPK,PEL,KOO),	(HPK,PEL,LUK),
(HPK,PEL,SAI),	(HPK,PEL,SPO),	(HPK,PEL,TAP),	(HPK,PEL,TPS),	(HPK,PEL,ÄSS),	(HPK,SAI,IFK),	...

Number of sequences — Summary

Fact

If you have n different elements and arrange them into k-element sequences, you get

- n^k *sequences if repetition is allowed (kth power of n)*
- $n(n-1)\cdots(n-k+1)$ *sequences if repetition is not allowed (kth falling factorial of n)*

Example (PIN codes)

Sequences of 4 digits out of 10: there are 10^4 sequences, because the same digits can be used again.

Example (Gold-silver-bronze sequences)

Sequences of 3 teams out of 15: there are $15 \times 14 \times 13 = 2730$ sequences, because the same team cannot be used again in the same sequence.

Number of orders

Given the *set* of 15 teams in the hockey series, in how many ways can you arrange *all* of them into an ordered sequence?

We already know how to do this. We are forming 15-element sequences from a 15-element set without repetition. Thus we have $n = k = 15$, and we get $15 \times 14 \times \cdots \times 2 \times 1 \approx 1.31 \times 10^{12}$ sequences.

Fact

n distinct elements can be ordered in $n! = n(n-1)\cdots 1$ ways. (This is called the factorial of n.)

Number of subsets

From a hockey team of 20 players, how many different 5-player units can you form?

If we want *ordered* sequences, we get
$20 \times 19 \times 18 \times 17 \times 16 = 1\ 860\ 480$ of them.

But suppose we only care about *which* five players were selected.
$\Longrightarrow$ Each five-player *set* corresponds to $5! = 120$ five-player *sequences*

$\Longrightarrow$ the number of five-player sets is

$$\frac{20 \times 19 \times 18 \times 17 \times 16}{5!} = \frac{1\ 860\ 480}{120} = 15\ 504$$

Number of subsets — Generalization

Fact

From an n-element set, the number of different k-element subsets is

$$\frac{n(n-1)\cdots(n-k+1)}{k(k-1)\cdots 1} = \frac{n!}{k!(n-k)!} =: \binom{n}{k}.$$

This is called the binomial coefficient (read: "n choose k").

Example: Finnish lottery

There are 40 balls labeled with integers $1, 2, \ldots, 40$. You name a subset of 7 balls. Then 7 balls are picked randomly. What is the probability that you named exactly the correct subset ("7 correct")?

- The sample set is

$$S = \text{"7-ball subsets of } \{1, \ldots, 40\}\text{"}$$

  and its size is $\#S = \binom{40}{7} = 18\ 643\ 560$.

- The event

$$A = \text{"you have all 7 correct"}$$

  contains exactly one outcome, so $\#A = 1$.

- By the symmetry of the process that picks the balls, all outcomes are equally probable, so

$$\mathbb{P}(A) = \frac{\#A}{\#S} = \frac{1}{18\ 643\ 560}$$

Example. Board of executives

A five-member board of executives is being formed. There are 6 male and 10 female applicants. If the board is selected at random, what is the probability that 3 men and 2 women are selected?

The sample space is $S =$ "five-person subsets of the 16 applicants"

$$\#S = \binom{16}{5}.$$

Now find how many outcomes (boards) correspond to the event $A =$ "3 men and 2 women are selected". We could list all such boards like this:

1. Choose some 3 out of the 6 male applicants: $\binom{6}{3}$ choices
2. Choose some 2 out of the 10 female applicants: $\binom{10}{2}$ choices

$$\#A = \binom{6}{3}\binom{10}{2}$$

The probability is

$$\mathbb{P}(A) = \frac{\#A}{\#S} = \frac{\binom{6}{3}\binom{10}{2}}{\binom{16}{5}} = \frac{900}{4368} \approx 20.6\%.$$

Example. Three of a kind in poker

What is the probability of *three of a kind* when five cards are dealt? (Hand of five cards, containing *three* cards of the same value, one of another value, and one of yet another value). Let us form all such five-card hands.

1. Choose the value a of the three cards: 13 choices
2. Choose the other two values as an unordered subset, *out of the 12 values different from a*: $\binom{12}{2} = 66$ choices
3. For value a, there are four cards available; choose three of them: $\binom{4}{3} = 4$ choices
4. Consider the two other values. For the smaller of them, call it b, there are four cards available; pick one: 4 choices
5. For the bigger value, there are 4 cards available: $\binom{4}{1} = 4$ choices

$$\Longrightarrow \mathbb{P}(\text{“three of a kind”}) = \frac{13\binom{12}{2}\binom{4}{3}\binom{4}{1}\binom{4}{1}}{\binom{52}{5}} = \frac{54\ 912}{2\ 598\ 960} \approx 2.1\%.$$

On the next lecture 1B, we talk about random variables and their distributions. . .