Advanced statistics and modelling

2025 spring

- **Gergely Palla** / pallag@hal.elte.hu
  Dept. of Biological Physics, Room 3.90.
- **Bendegúz Borkovits** / borbende@phys-gs.elte.hu
  Dept. of Biological Physics, Room 3.72.
- **László Oroszlány** / oroszlanyl@gmail.com
  Dept. of Physics of Complex Systems, Room 5.79
- **Zoltán Kaufmann** / kaufmann@complex.elte.hu
  Dept. of Physics of Complex Systems, Room 5.53
- **Illés J. Farkas** / fij@hal.elte.hu
  Quant AVP at Citi MQA,
  Dept. of Biological Physics, Room 3.90
- **János Báskay** / baskayj@caesar.elte.hu
  Dept. of Biological Physics, Room 3.72.

Grading:

- 1 written test,
- if not satisfied with the offered grade, then an oral exam,

Course material:

- all slides, jupyter notebooks, etc. are going to be uploaded to the ELTE moodle site of the course.

Schedule

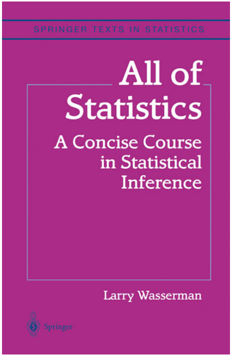
- Sample space, Probability, Variable and Distribution
- Expectation, Inequalities, Convergence
- Models, Inference, Learning
- Empirical PDF, CDF, Smoothing, Binning
- Bootstrap, Maximum Likelihood, Hypothesis testing
- Regression, Inference About Independence
- Extreme Statistics, Posthoc Analysis
- Survival time analysis
- Hierarchical Bayesian models
- Classical chaos, Ljapunov exponents, Frobenius Perron operators
- Random Matrices and Eigenvalue Statistics
- Statistics in fincance
- Stochastic Processes with applications in finance

Schedule

Advanced statistics and modelling 2025 spring		
1. week	Feb. 13. Thursday Gergely Palla	Probability, variable, distribution
2. week	Feb. 20. Thursday Gergely Palla	Expectation, inequalities, convergence
3. week	Feb. 27. Thursday Gergely Palla	Statistical inference, estimating the CDF, PDF, Smoothing
4. week	Mar. 06. Thursday Bendegúz Borkovits	Bootstrap, Maximum Likelihood, Hypothesis testing
5. week	Mar. 13. Thursday Bendegúz Borkovits	Regression, inference about independence
6. week	Mar. 20. Thursday Bendegúz Borkovits	Extreme statistics, posthoc analysis
7. week	Mar. 27. Thursday Gergely Palla	Hierarchical Bayesian modelling
8. week	Apr. 03. Thursday János Báskay	Survival time analysis
9. week	Apr. 10. Thursday Zoltán Kaufmann	Classical chaos, Ljapunov exponent, Frobenius-Perron
10. week	Apr. 17. Thursday	Spring break
11. week	Apr. 24. Thursday László Oroszlány	Random matrices and eigenvalue statistics
12. week	May. 01. Thursday	National Holiday
13. week	May 08. Thursday Illés Farkas	Statistics in finance
14. week	May 15. Thursday	!! FINAL TEST !!

Recommended literature

Larry Wasserman: **All of Statistics**
A Concise Course in Statistical Inference



SAMPLE SPACE, PROBABILITY, VARIABLE AND DISTRIBUTION

SAMPLE SPACE

Sample space and events

Sample space

$\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$ is the set of possible outcomes of an experiment. The points ω_i in Ω are called **sample outcomes** or **realisations**.

Events

A subset $A \subset \Omega$ is called an **event**.

- True event: $A = \Omega$, (always true).
- Null event: $A = \emptyset$, (always false).

Examples

- Coin tossing: $\Omega = \{\omega = (\omega_1, \omega_2, \omega_3, \ldots,), \ \ \omega_i \in \{H, T\}\}$
  Event A can be e.g., that the first head appears on the third toss:
  $A = \{(\omega_1, \omega_2, \omega_3, \ldots,) : \omega_1 = T, \omega_2 = T, \omega_3 = H, \omega_i \in \{H, T\}\}$.
- Let ω be the outcome of a measurement of some physical quantity, e.g., temperature. Then $\Omega = [-\infty, \infty]$.
  Event A can be e.g., that he measurement is between 10 and 20: $A = [10, 20]$.

Sample space and events

Sample space

$\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$ is the set of possible outcomes of an experiment. The points ω_i in Ω are called **sample outcomes** or **realisations**.

Events

A subset $A \subset \Omega$ is called an **event**.

- True event: $A = \Omega$, (always true).
- Null event: $A = \emptyset$, (always false).

Examples

- Coin tossing: $\Omega = \{\omega = (\omega_1, \omega_2, \omega_3, \ldots,), \ \ \omega_i \in \{H, T\}\}$
  Event A can be e.g., that the first head appears on the third toss:
  $A = \{(\omega_1, \omega_2, \omega_3, \ldots,) : \omega_1 = T, \omega_2 = T, \omega_3 = H, \omega_i \in \{H, T\}\}$.
- Let ω be the outcome of a measurement of some physical quantity, e.g., temperature. Then $\Omega = [-\infty, \infty]$.
  Event A can be e.g., that he measurement is between 10 and 20:
  $A = [10, 20]$.

Sample space and events

Sample space

$\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$ is the set of possible outcomes of an experiment. The points ω_i in Ω are called **sample outcomes** or **realisations**.

Events

A subset $A \subset \Omega$ is called an **event**.

- True event: $A = \Omega$, (always true).
- Null event: $A = \varnothing$, (always false).

Examples

- Coin tossing: $\Omega = \{\omega = (\omega_1, \omega_2, \omega_3, \ldots,), \ \ \omega_i \in \{H, T\}\}$
  Event A can be e.g., that the first head appears on the third toss: $A = \{(\omega_1, \omega_2, \omega_3, \ldots,) : \omega_1 = T, \omega_2 = T, \omega_3 = H, \omega_i \in \{H, T\}\}$.
- Let ω be the outcome of a measurement of some physical quantity, e.g., temperature. Then $\Omega = [-\infty, \infty]$.
  Event A can be e.g., that he measurement is between 10 and 20: $A = [10, 20]$.

Sample space and events

Sample space

$\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$ is the set of possible outcomes of an experiment. The points ω_i in Ω are called **sample outcomes** or **realisations**.

Events

A subset $A \subset \Omega$ is called an **event**.

- True event: $A = \Omega$, (always true).
- Null event: $A = \emptyset$, (always false).

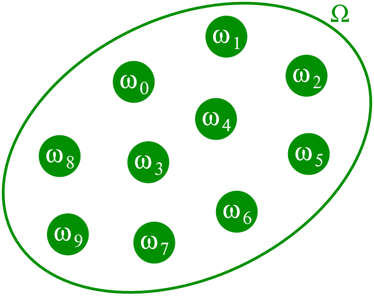
Examples

- Coin tossing: $\Omega = \{\omega = (\omega_1, \omega_2, \omega_3, \ldots,),\ \ \omega_i \in \{H, T\}\}$
  Event A can be e.g., that the first head appears on the third toss:
  $A = \{(\omega_1, \omega_2, \omega_3, \ldots,) : \omega_1 = T, \omega_2 = T, \omega_3 = H, \omega_i \in \{H, T\}\}$.
- Let ω be the outcome of a measurement of some physical quantity, e.g., temperature. Then $\Omega = [-\infty, \infty]$.
  Event A can be e.g., that he measurement is between 10 and 20:
  $A = [10, 20]$.

Operations between events

Operations between events

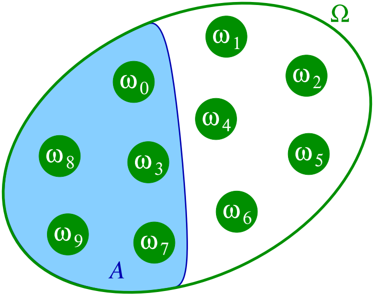
- Union, $A \cup B$: at least one of them is true (OR): $A + B$.
- Intersection, $A \cap B$: both of them are true (AND): AB.
- Complement, A^c: it is false (NEGATION): $\overline{A}$.
- Inclusion: $A \subset C$.
- Set difference, $A \setminus B$: A is true and B is false: $A \setminus B = A \cap \overline{B}$.



Operations between events

Operations between events

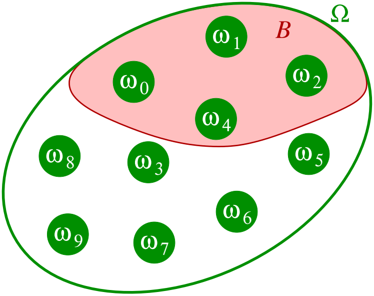
- Union, $A \cup B$: at least one of them is true (OR): $A + B$.
- Intersection, $A \cap B$: both of them are true (AND): AB.
- Complement, A^c: it is false (NEGATION): $\overline{A}$.
- Inclusion: $A \subset C$.
- Set difference, $A \setminus B$: A is true and B is false: $A \setminus B = A \cap \overline{B}$.



Operations between events

Operations between events

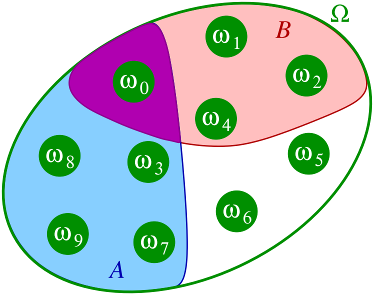
- Union, $A \cup B$: at least one of them is true (OR): $A + B$.
- Intersection, $A \cap B$: both of them are true (AND): AB.
- Complement, A^c: it is false (NEGATION): $\overline{A}$.
- Inclusion: $A \subset C$.
- Set difference, $A \setminus B$: A is true and B is false: $A \setminus B = A \cap \overline{B}$.



Operations between events

Operations between events

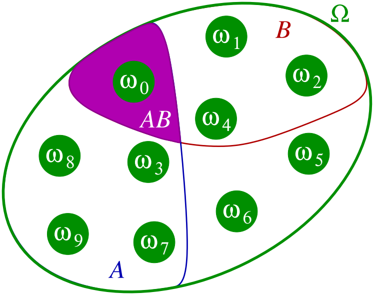
- Union, $A \cup B$: at least one of them is true (OR): $A + B$.
- Intersection, $A \cap B$: both of them are true (AND): AB.
- Complement, A^c: it is false (NEGATION): $\overline{A}$.
- Inclusion: $A \subset C$.
- Set difference, $A \setminus B$: A is true and B is false: $A \setminus B = A \cap \overline{B}$.



Operations between events

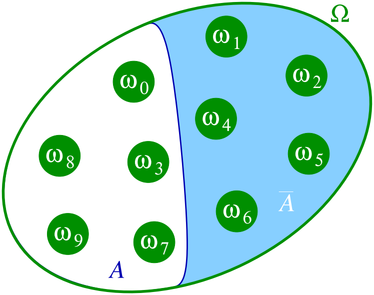
Operations between events

- Union, $A \cup B$: at least one of them is true (OR): $A + B$.
- Intersection, $A \cap B$: both of them are true (AND): AB.
- Complement, A^c: it is false (NEGATION): $\overline{A}$.
- Inclusion: $A \subset C$.
- Set difference, $A \setminus B$: A is true and B is false: $A \setminus B = A \cap \overline{B}$.



Operations between events

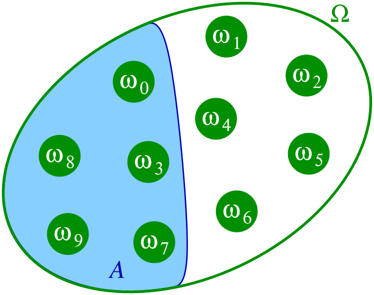
- Union, $A \cup B$: at least one of them is true (OR): $A + B$.
- Intersection, $A \cap B$: both of them are true (AND): AB.
- Complement, A^c: it is false (NEGATION): $\overline{A}$.
- Inclusion: $A \subset C$.
- Set difference, $A \setminus B$: A is true and B is false: $A \setminus B = A \cap \overline{B}$.



Operations between events

Operations between events

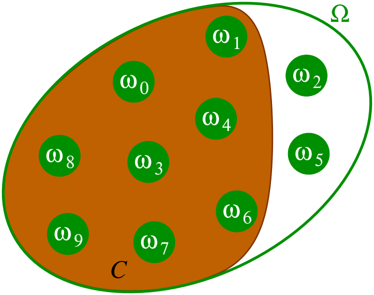
- Union, $A \cup B$: at least one of them is true (OR): $A + B$.
- Intersection, $A \cap B$: both of them are true (AND): AB.
- Complement, A^c: it is false (NEGATION): $\overline{A}$.
- Inclusion: $A \subset C$.
- Set difference, $A \setminus B$: A is true and B is false: $A \setminus B = A \cap \overline{B}$.



Operations between events

Operations between events

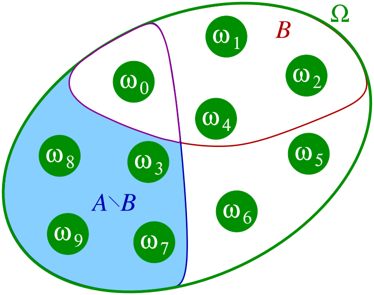
- Union, $A \cup B$: at least one of them is true (OR): $A + B$.
- Intersection, $A \cap B$: both of them are true (AND): AB.
- Complement, A^c: it is false (NEGATION): $\overline{A}$.
- Inclusion: $A \subset C$.
- Set difference, $A \setminus B$: A is true and B is false: $A \setminus B = A \cap \overline{B}$.



Operations between events

Operations between events

- Union, $A \cup B$: at least one of them is true (OR): $A + B$.
- Intersection, $A \cap B$: both of them are true (AND): AB.
- Complement, A^c: it is false (NEGATION): $\overline{A}$.
- Inclusion: $A \subset C$.
- Set difference, $A \setminus B$: A is true and B is false: $A \setminus B = A \cap \overline{B}$.



Operational identities

Operational identities

„AND”

- $A \cap B = B \cap A$
- $A \cap A = A$
- $A \cap (B \cap C) = (A \cap B) \cap C$
- $A \cap \overline{A} = \varnothing$
- $A \cap \varnothing = \varnothing$
- $A \cap \Omega = A$

„OR”

- $A \cup B = B \cup A$
- $A \cup A = A$
- $A \cup (B \cup C) = (A \cup B) \cup C$
- $A \cup \overline{A} = \Omega$
- $A \cup \varnothing = A$
- $A \cup \Omega = \Omega$

- $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
- $A \cup (A \cap B) = A$
- $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
- $\overline{A \cap B} = \overline{A} \cup \overline{B}$
- $\overline{A \cup B} = \overline{A} \cap \overline{B}$

Partition

Mutually exclusive events

A and B are mutually exclusive events if $A \cap B = \varnothing$

Partition

$A_1, A_2, \ldots, A_n$ is a partition of Ω, if for all $k = 1, \ldots, n$

a) $A_k \neq \varnothing$,

b) $A_j \cap A_k = \varnothing$ if $j \neq k$,

c) $A_1 \cup A_2 \cup \ldots \cup A_n = \Omega$.

Partition

Mutually exclusive events

A and B are mutually exclusive events if $A \cap B = \varnothing$

Partition

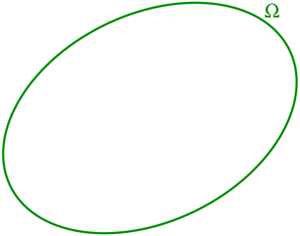
$A_1, A_2, \ldots, A_n$ is a partition of Ω, if for all $k = 1, \ldots, n$

a) $A_k \neq \varnothing$,

b) $A_j \cap A_k = \varnothing$ if $j \neq k$,

c) $A_1 \cup A_2 \cup \ldots \cup A_n = \Omega$.

Illustration:



Partition

Mutually exclusive events

A and B are mutually exclusive events if $A \cap B = \emptyset$

Partition

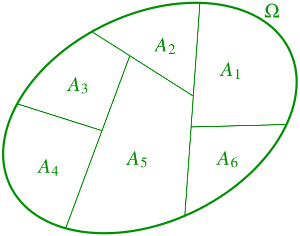
$A_1, A_2, \ldots, A_n$ is a partition of Ω, if for all $k = 1, \ldots, n$

a) $A_k \neq \emptyset$,

b) $A_j \cap A_k = \emptyset$ if $j \neq k$,

c) $A_1 \cup A_2 \cup \ldots \cup A_n = \Omega$.

Illustration:



PROBABILITY

Definition of probability (Kolmogorov)

Given a sample space Ω, a function $P(A)$ over the subsets of Ω is a **probability distribution** or a **probability measure** if

- K1 $0 \leq P(A) \leq 1 \ \forall A \subset \Omega$,
- K2 $P(\Omega) = 1$,
- K3 If $A_1, A_2, \ldots$ are mutually exclusive, then

$$P(\bigcup_{i=1}^{\infty} A_k) = \sum_{i=1}^{\infty} P(A_i).$$

Definition of probability (Kolmogorov)

Given a sample space Ω, a function $P(A)$ over the subsets of Ω is a **probability distribution** or a **probability measure** if

K1 $0 \leq P(A) \leq 1 \ \forall A \subset \Omega$,

K2 $P(\Omega) = 1$,

K3 If $A_1, A_2, \ldots$ are mutually exclusive, then

$$P(\bigcup_{i=1}^{\infty} A_k) = \sum_{i=1}^{\infty} P(A_i).$$

Definition of probability (Kolmogorov)

Given a sample space Ω, a function $P(A)$ over the subsets of Ω is a **probability distribution** or a **probability measure** if

K1 $0 \leq P(A) \leq 1 \ \forall A \subset \Omega$,

K2 $P(\Omega) = 1$,

K3 If $A_1, A_2, \ldots$ are mutually exclusive, then

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

Probability identities

The probability of the false event $\varnothing$ is 0.

$$
\begin{array}{rcl}
A \cup \varnothing = A, & & A \cap \varnothing = \varnothing \\
P(A) = P(A \cup \varnothing) & \underset{K3}{=} & P(A) + P(\varnothing) \\
P(\varnothing) & = & 0
\end{array}
$$

For a partition $A_1, A_2, \ldots, A_n$ the sum of the probabilities is $\sum_i P(A_i) = 1$.

- $A_i \cap A_j = \varnothing$ and $\bigcup_{i=1}^{n} A_i = \Omega$.
- Since $P(\Omega) = 1$, according to (K3) we obtain $\sum_i P(A_i) = 1$.

The probability of the complement $\overline{A}$ is $P(\overline{A}) = 1 - P(A)$.

(A and $\overline{A}$ together form a partition.)

Probability identities

The probability of the false event $\varnothing$ is 0.

$$A \cup \varnothing = A, \qquad A \cap \varnothing = \varnothing$$

$$\begin{aligned} P(A) = P(A \cup \varnothing) &\underset{K3}{=} P(A) + P(\varnothing) \\ P(\varnothing) &= 0 \end{aligned}$$

For a partition $A_1, A_2, \ldots, A_n$ the sum of the probabilities is $\sum_i P(A_i) = 1$.

- $A_i \cap A_j = \varnothing$ and $\bigcup_{i=1}^{n} A_i = \Omega$.
- Since $P(\Omega) = 1$, according to (K3) we obtain $\sum_i P(A_i) = 1$.

The probability of the complement $\overline{A}$ is $P(\overline{A}) = 1 - P(A)$.

(A and $\overline{A}$ together form a partition.)

Probability identities

Probability of a subset

If A is included in B, i.e., $A \subset B$, then $P(B \setminus A) = P(B) - P(A)$.

- Since $A \subset B$, we can write B as $B = A \cup (B \setminus A)$.
- Also $A \cap (B \setminus A) = \varnothing$, thus according to (K3) $P(B) = P(A) + P(B \setminus A)$.

Probability of the union

For any events A and B the probability of their union is $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

- Since $A \cup B = A \cup (B \setminus A \cap B)$ and $A \cap (B \setminus A \cap B) = \varnothing$, according to (K3) $P(A \cup B) = P(A) + P(B \setminus A \cap B)$.
- Since $A \cap B \subset B$, according to the previous result $P(B \setminus A \cap B) = P(B) - P(A \cap B)$, which substituted back yields the end result.

Probability identities

Probability of a subset

If A is included in B, i.e., $A \subset B$, then $P(B \setminus A) = P(B) - P(A)$.

- Since $A \subset B$, we can write B as $B = A \cup (B \setminus A)$.
- Also $A \cap (B \setminus A) = \emptyset$, thus according to (K3) $P(B) = P(A) + P(B \setminus A)$.

Probability of the union

For any events A and B the probability of their union is $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

- Since $A \cup B = A \cup (B \setminus A \cap B)$ and $A \cap (B \setminus A \cap B) = \emptyset$, according to (K3) $P(A \cup B) = P(A) + P(B \setminus A \cap B)$.
- Since $A \cap B \subset B$, according to the previous result $P(B \setminus A \cap B) = P(B) - P(A \cap B)$, which substituted back yields the end result.

Probability identities

Probability of a subset

If A is included in B, i.e., $A \subset B$, then $P(B \setminus A) = P(B) - P(A)$.

- Since $A \subset B$, we can write B as $B = A \cup (B \setminus A)$.
- Also $A \cap (B \setminus A) = \varnothing$, thus according to (K3) $P(B) = P(A) + P(B \setminus A)$.

Probability of the union

For any events A and B the probability of their union is $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

- Since $A \cup B = A \cup (B \setminus A \cap B)$ and $A \cap (B \setminus A \cap B) = \varnothing$, according to (K3) $P(A \cup B) = P(A) + P(B \setminus A \cap B)$.
- Since $A \cap B \subset B$, according to the previous result $P(B \setminus A \cap B) = P(B) - P(A \cap B)$, which substituted back yields the end result.

Probability identities

Probability of a subset

If A is included in B, i.e., $A \subset B$, then $P(B \setminus A) = P(B) - P(A)$.

- Since $A \subset B$, we can write B as $B = A \cup (B \setminus A)$.
- Also $A \cap (B \setminus A) = \varnothing$, thus according to (K3) $P(B) = P(A) + P(B \setminus A)$.

Probability of the union

For any events A and B the probability of their union is $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

- Since $A \cup B = A \cup (B \setminus A \cap B)$ and $A \cap (B \setminus A \cap B) = \varnothing$, according to (K3) $P(A \cup B) = P(A) + P(B \setminus A \cap B)$.
- Since $A \cap B \subset B$, according to the previous result $P(B \setminus A \cap B) = P(B) - P(A \cap B)$, which substituted back yields the end result.

Probability identities

Probability of a subset

If A is included in B, i.e., $A \subset B$, then $P(B \setminus A) = P(B) - P(A)$.

- Since $A \subset B$, we can write B as $B = A \cup (B \setminus A)$.
- Also $A \cap (B \setminus A) = \varnothing$, thus according to (K3) $P(B) = P(A) + P(B \setminus A)$.

Probability of the union

For any events A and B the probability of their union is $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

- Since $A \cup B = A \cup (B \setminus A \cap B)$ and $A \cap (B \setminus A \cap B) = \varnothing$, according to (K3) $P(A \cup B) = P(A) + P(B \setminus A \cap B)$.
- Since $A \cap B \subset B$, according to the previous result $P(B \setminus A \cap B) = P(B) - P(A \cap B)$, which substituted back yields the end result.

Probability identities

Probability of a subset

If A is included in B, i.e., $A \subset B$, then $P(B \setminus A) = P(B) - P(A)$.

- Since $A \subset B$, we can write B as $B = A \cup (B \setminus A)$.
- Also $A \cap (B \setminus A) = \varnothing$, thus according to (K3) $P(B) = P(A) + P(B \setminus A)$.

Probability of the union

For any events A and B the probability of their union is $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

- Since $A \cup B = A \cup (B \setminus A \cap B)$ and $A \cap (B \setminus A \cap B) = \varnothing$, according to (K3) $P(A \cup B) = P(A) + P(B \setminus A \cap B)$.
- Since $A \cap B \subset B$, according to the previous result $P(B \setminus A \cap B) = P(B) - P(A \cap B)$, which substituted back yields the end result.

Probability identities

Limit theorems

L I. If $A_1, A_2, \ldots$ is an infinite series of events where $A_n \supset A_{n+1}$ for all n, and $\bigcap_{i=1}^{\infty} A_i = A$, then $\lim_{n\to\infty} P(A_n) = P(A)$.

L II. If $A_1, A_2, \ldots$ is an infinite series of events where $A_n \subset A_{n+1}$ for all n, and $\bigcup_{i=1}^{\infty} A_i = A$, then $\lim_{n\to\infty} P(A_n) = P(A)$.

Conditional probability

Conditional probability

The conditional probability of event A given event B is defined as

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

(where we assumed that $P(B) > 0$).

Properties:

- $0 \leq P(A \mid B) \leq 1$,
- $P(B \mid B) = 1$,
- If $A_1, A_2, \ldots,$ are pairwise disjoint, then $P(\bigcup_i A_i \mid B) = \sum_i P(A_i \mid B)$.
- The $P(A \mid B)$ function over the subsets A in Ω is also a probability distribution.
- $P(A \mid B)P(B) = P(A \cap B) = P(B \mid A)P(A)$.

Conditional probability

Conditional probability

The conditional probability of event A given event B is defined as

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

(where we assumed that $P(B) > 0$).

Properties:

- $0 \leq P(A \mid B) \leq 1$,
- $P(B \mid B) = 1$,
- If $A_1, A_2, \ldots,$ are pairwise disjoint, then $P(\bigcup_i A_i \mid B) = \sum_i P(A_i \mid B)$.
- The $P(A \mid B)$ function over the subsets A in Ω is also a probability distribution.
- $P(A \mid B)P(B) = P(A \cap B) = P(B \mid A)P(A)$.

Conditional probability

Conditional probability

The conditional probability of event A given event B is defined as

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

(where we assumed that $P(B) > 0$).

Properties:

- $0 \leq P(A \mid B) \leq 1$,
- $P(B \mid B) = 1$,
- If $A_1, A_2, \ldots,$ are pairwise disjoint, then $P(\bigcup_i A_i \mid B) = \sum_i P(A_i \mid B)$.
- The $P(A \mid B)$ function over the subsets A in Ω is also a probability distribution.
- $P(A \mid B)P(B) = P(A \cap B) = P(B \mid A)P(A)$.

Conditional probability

Conditional probability

The conditional probability of event A given event B is defined as

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

(where we assumed that $P(B) > 0$).

Properties:

- $0 \leq P(A \mid B) \leq 1$,
- $P(B \mid B) = 1$,
- If $A_1, A_2, \ldots,$ are pairwise disjoint, then $P(\bigcup_i A_i \mid B) = \sum_i P(A_i \mid B)$.
- The $P(A \mid B)$ function over the subsets A in Ω is also a probability distribution.
- $P(A \mid B)P(B) = P(A \cap B) = P(B \mid A)P(A)$.

Conditional probability

Conditional probability

The conditional probability of event A given event B is defined as

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

(where we assumed that $P(B) > 0$).

Properties:

- $0 \leq P(A \mid B) \leq 1$,
- $P(B \mid B) = 1$,
- If $A_1, A_2, \ldots,$ are pairwise disjoint, then $P(\bigcup_i A_i \mid B) = \sum_i P(A_i \mid B)$.
- The $P(A \mid B)$ function over the subsets A in Ω is also a probability distribution.
- $P(A \mid B)P(B) = P(A \cap B) = P(B \mid A)P(A)$.

Conditional probability

Conditional probability

The conditional probability of event A given event B is defined as

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

(where we assumed that $P(B) > 0$).

Properties:

- $0 \leq P(A \mid B) \leq 1$,
- $P(B \mid B) = 1$,
- If $A_1, A_2, \ldots,$ are pairwise disjoint, then $P(\bigcup_i A_i \mid B) = \sum_i P(A_i \mid B)$.
- The $P(A \mid B)$ function over the subsets A in Ω is also a probability distribution.
- $P(A \mid B)P(B) = P(A \cap B) = P(B \mid A)P(A)$.

Multiplication rule

Multiplication rule

For any events $A_1, A_2, \ldots, A_n$

$$\begin{aligned} P(A_1 \cap A_2 \cap \cdots \cap A_n) &= P(A_1 \mid A_2 \cap \cdots \cap A_n) P(A_2 \mid A_3 \cap \cdots \cap A_n) \cdots \\ & \cdots P(A_{n-1} \mid A_n) P(A_n) = \\ & P(A_n) \prod_{i=1}^{n-1} P(A_i \mid A_{i+1} \cap \cdots \cap A_n) \end{aligned}$$

Proof: by applying the multiplication rule for two events recursively, e.g.,

$$P(A_1 \cap A_2 \cap A_3) = P(A_1 \mid A_2 \cap A_3) P(A_2 \cap A_3) = P(A_1 \mid A_2 \cap A_3) P(A_2 \mid A_3) P(A_3),$$

etc.

Multiplication rule

Multiplication rule

For any events $A_1, A_2, \ldots, A_n$

$$\begin{aligned} P(A_1 \cap A_2 \cap \cdots \cap A_n) &= P(A_1 \mid A_2 \cap \cdots \cap A_n) P(A_2 \mid A_3 \cap \cdots \cap A_n) \cdots \\ &\cdots P(A_{n-1} \mid A_n) P(A_n) = \\ &P(A_n) \prod_{i=1}^{n-1} P(A_i \mid A_{i+1} \cap \cdots \cap A_n) \end{aligned}$$

Proof: by applying the multiplication rule for two events recursively, e.g.,

$$P(A_1 \cap A_2 \cap A_3) = P(A_1 \mid A_2 \cap A_3) P(A_2 \cap A_3) = P(A_1 \mid A_2 \cap A_3) P(A_2 \mid A_3) P(A_3),$$

etc.

Law of total probability

Law of total probability

If $A_1, A_2, \ldots, A_n$ provide a partition and $P(A_i) > 0\ \forall i$, then for any event B

$$P(B) = \sum_{i=1}^{n} P(B \mid A_i) P(A_i).$$

- Since the A_i are pairwise disjoint, so are $B \cap A_i$.
- Based on $\bigcup_i A_i = \Omega$, the union of the events $B \cap A_i$ is

$$(B \cap A_1) \cup (B \cap A_2) \cup \cdots \cup (B \cap A_n) = B \cap (A_1 \cup A_2 \cup \cdots \cup A_n) = B \cap \Omega = B.$$

- According to (K3) $P\left(\bigcup_i (B \cap A_i)\right) = \sum_i P(B \cap A_i)$.
- Multiplication rule for two events $P(B \cap A_i) = P(B \mid A_i) P(A_i)$.

$$\to P(B) = P\left(\bigcup_{i=1}^{n} B \cap A_i\right) = \sum_{i=1}^{n} P(B \cap A_i) = \sum_{i=1}^{n} P(B \mid A_i) P(A_i)$$

Law of total probability

Law of total probability

If $A_1, A_2, \ldots, A_n$ provide a partition and $P(A_i) > 0 \ \forall i$, then for any event B

$$P(B) = \sum_{i=1}^{n} P(B \mid A_i)P(A_i).$$

- Since the A_i are pairwise disjoint, so are $B \cap A_i$.
- Based on $\bigcup_i A_i = \Omega$, the union of the events $B \cap A_i$ is

$$(B \cap A_1) \cup (B \cap A_2) \cup \cdots \cup (B \cap A_n) = B \cap (A_1 \cup A_2 \cup \cdots \cup A_n) = B \cap \Omega = B.$$

- According to (K3) $P\left(\bigcup_i (B \cap A_i)\right) = \sum_i P(B \cap A_i)$.
- Multiplication rule for two events $P(B \cap A_i) = P(B \mid A_i)P(A_i)$.

$$\rightarrow P(B) = P\left(\bigcup_{i=1}^{n} B \cap A_i\right) = \sum_{i=1}^{n} P(B \cap A_i) = \sum_{i=1}^{n} P(B \mid A_i)P(A_i)$$

Law of total probability

Law of total probability

If $A_1, A_2, \dots, A_n$ provide a partition and $P(A_i) > 0\ \forall i$, then for any event B

$$P(B) = \sum_{i=1}^{n} P(B \mid A_i) P(A_i).$$

- Since the A_i are pairwise disjoint, so are $B \cap A_i$.
- Based on $\bigcup_i A_i = \Omega$, the union of the events $B \cap A_i$ is

$$(B \cap A_1) \cup (B \cap A_2) \cup \cdots \cup (B \cap A_n) = B \cap (A_1 \cup A_2 \cup \cdots \cup A_n) = B \cap \Omega = B.$$

- According to (K3) $P\left(\bigcup_i (B \cap A_i)\right) = \sum_i P(B \cap A_i)$.
- Multiplication rule for two events $P(B \cap A_i) = P(B \mid A_i) P(A_i)$.

$$\rightarrow P(B) = P\left(\bigcup_{i=1}^{n} B \cap A_i\right) = \sum_{i=1}^{n} P(B \cap A_i) = \sum_{i=1}^{n} P(B \mid A_i) P(A_i)$$

Law of total probability

Law of total probability

If $A_1, A_2, \ldots, A_n$ provide a partition and $P(A_i) > 0\ \forall i$, then for any event B

$$P(B) = \sum_{i=1}^{n} P(B \mid A_i) P(A_i).$$

- Since the A_i are pairwise disjoint, so are $B \cap A_i$.
- Based on $\bigcup_i A_i = \Omega$, the union of the events $B \cap A_i$ is

$$(B \cap A_1) \cup (B \cap A_2) \cup \cdots \cup (B \cap A_n) = B \cap (A_1 \cup A_2 \cup \cdots \cup A_n) = B \cap \Omega = B.$$

- According to (K3) $P\left(\bigcup_i (B \cap A_i)\right) = \sum_i P(B \cap A_i)$.
- Multiplication rule for two events $P(B \cap A_i) = P(B \mid A_i) P(A_i)$.

$$\rightarrow P(B) = P\left(\bigcup_{i=1}^{n} B \cap A_i\right) = \sum_{i=1}^{n} P(B \cap A_i) = \sum_{i=1}^{n} P(B \mid A_i) P(A_i)$$

Law of total probability

Law of total probability

If $A_1, A_2, \ldots, A_n$ provide a partition and $P(A_i) > 0\ \forall i$, then for any event B

$$P(B) = \sum_{i=1}^{n} P(B \mid A_i) P(A_i).$$

- Since the A_i are pairwise disjoint, so are $B \cap A_i$.
- Based on $\bigcup_i A_i = \Omega$, the union of the events $B \cap A_i$ is

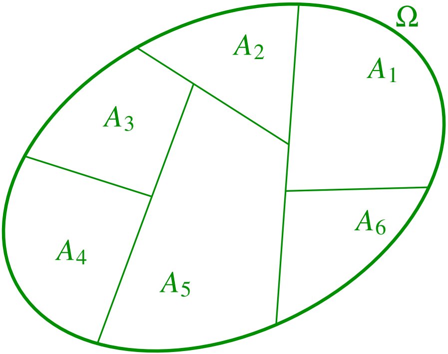
$$(B \cap A_1) \cup (B \cap A_2) \cup \cdots \cup (B \cap A_n) = B \cap (A_1 \cup A_2 \cup \cdots \cup A_n) = B \cap \Omega = B.$$

- According to (K3) $P\left(\bigcup_i (B \cap A_i)\right) = \sum_i P(B \cap A_i)$.
- Multiplication rule for two events $P(B \cap A_i) = P(B \mid A_i) P(A_i)$.

$$\rightarrow P(B) = P\left(\bigcup_{i=1}^{n} B \cap A_i\right) = \sum_{i=1}^{n} P(B \cap A_i) = \sum_{i=1}^{n} P(B \mid A_i) P(A_i)$$

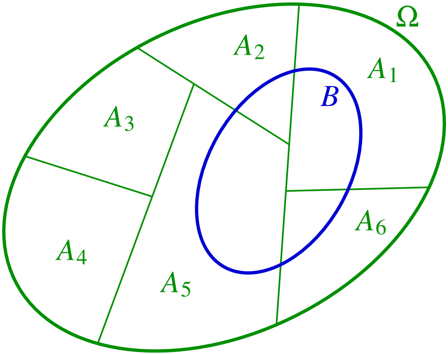
Law of total probability

Illustration:



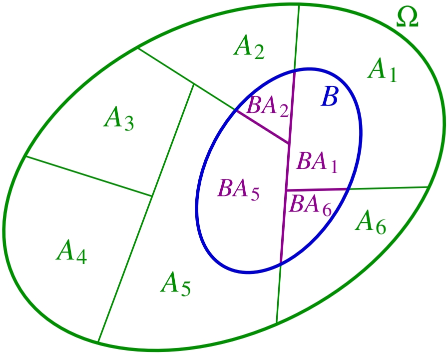
Law of total probability

Illustration:



Law of total probability

Illustration:



Bayes' theorem

Bayes' theorem

Let $A_1, A_2, \ldots, A_n$ be a partition where $P(A_i) > 0\ \forall i$. If $P(B) > 0$ for any event B, then

$$P(A_k \mid B) = \frac{P(B \mid A_k)P(A_k)}{\sum_{i=1}^{n} P(B \mid A_i)P(A_i)}.$$

Proof:

- According to the definition of the conditional probability $P(A_k \mid B)P(B) = P(B \mid A_k)P(A_k)$.
- We divide by $P(B)$ and use the law of total probability,

$$P(A_k \mid B) = \frac{P(B \mid A_k)P(A_k)}{P(B)} = \frac{P(B \mid A_k)P(A_k)}{\sum_{i=1}^{n} P(B \mid A_i)P(A_i)}.$$

Bayes' theorem

Bayes' theorem

Let $A_1, A_2, \ldots, A_n$ be a partition where $P(A_i) > 0\ \forall i$. If $P(B) > 0$ for any event B, then

$$P(A_k \mid B) = \frac{P(B \mid A_k)P(A_k)}{\sum_{i=1}^{n} P(B \mid A_i)P(A_i)}.$$

Proof:

- According to the definition of the conditional probability $P(A_k \mid B)P(B) = P(B \mid A_k)P(A_k)$.
- We divide by $P(B)$ and use the law of total probability,

$$P(A_k \mid B) = \frac{P(B \mid A_k)P(A_k)}{P(B)} = \frac{P(B \mid A_k)P(A_k)}{\sum_{i=1}^{n} P(B \mid A_i)P(A_i)}.$$

Bayes' theorem

Bayes' theorem

Let $A_1, A_2, \ldots, A_n$ be a partition where $P(A_i) > 0\ \forall i$. If $P(B) > 0$ for any event B, then

$$P(A_k \mid B) = \frac{P(B \mid A_k)P(A_k)}{\sum_{i=1}^{n} P(B \mid A_i)P(A_i)}.$$

Proof:

- According to the definition of the conditional probability $P(A_k \mid B)P(B) = P(B \mid A_k)P(A_k)$.
- We divide by $P(B)$ and use the law of total probability,

$$P(A_k \mid B) = \frac{P(B \mid A_k)P(A_k)}{P(B)} = \frac{P(B \mid A_k)P(A_k)}{\sum_{i=1}^{n} P(B \mid A_i)P(A_i)}.$$

Independence of events

Independent events

Events A and B are independent if

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

- According to the definition the conditional probability of A given B in case of independence is

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

→ This can also be the alternative definition of independence.

Independence of events

Independent events

Events A and B are independent if

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

- According to the definition the conditional probability of A given B in case of independence is

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

→ This can also be the alternative definition of independence.

RANDOM VARIABLE, CDF and PDF

Random variable

Random variable

A **random variable** is usually a mapping $X : \Omega \longrightarrow \mathbb{R}$ assigning a real number $X(\omega)$ to each event. However in general a random variable can map also as

$$X(\omega) : \Omega \longrightarrow \begin{cases} \mathbb{N} \\ \mathbb{R} \\ \mathbb{C} \\ \mathbb{R}^n \end{cases}$$

Cumulative distribution function

Cumulative distribution function (CDF)

The **cumulative distribution function** of a random variable X denoted by $F_X(x)$ is defined as

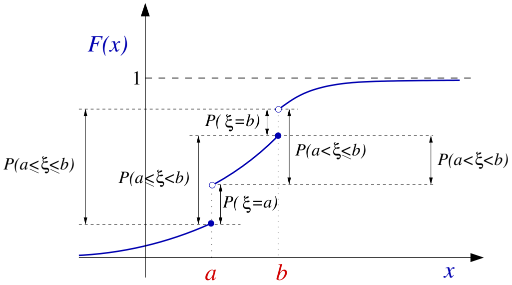
$$F_X(x) := P(X < x) = P\left(\{\omega \in \Omega\} : X(\omega) < x\right)$$

Properties:

- If $x_1 < x_2$ then $F(x_1) \leq F(x_2)$,
  (because event $X < x_1$ is a subset of $X < x_2$).
- $\lim\limits_{x \to -\infty} F(x) = 0$ and $\lim\limits_{x \to \infty} F(x) = 1$.
  (These are the consequences of L I. and L II. limit theorems).
- $F(x)$ is continuous from the left:
  ha $x_1 < x_2 \cdots < x_i < \cdots$ and $\lim\limits_{n \to \infty} x_n = x$ then $\lim\limits_{n \to \infty} F(x_n) = F(x)$.
  (This comes from the limit theorem L II.)

Cumulative distribution function

Illustration:



Point mass function

Point mass function

- A random variable is **discrete** if its values $X(\omega) = x$ can take up finite or countable many different values.

Point mass function

Point mass function

- A random variable is **discrete** if its values $X(\omega) = x$ can take up finite or countable many different values.
- Discrete random variables can be also characterised by their **point mass function** $f_X(x)$. By denoting the i-th discrete value $X(\omega)$ can take as x_i, the $f_X(x)$ can be defined as

$$f_X(x_i) := P(X(\omega) = x_i) = P(\{\omega_j\} : X(\omega_i) = x_i).$$

Point mass function

Point mass function

- A random variable is **discrete** if its values $X(\omega) = x$ can take up finite or countable many different values.
- Discrete random variables can be also characterised by their **point mass function** $f_X(x)$. By denoting the i-th discrete value $X(\omega)$ can take as x_i, the $f_X(x)$ can be defined as

$$f_X(x_i) := P(X(\omega) = x_i) = P(\{\omega_j\} : X(\omega_i) = x_i).$$

- The relation between the CDF and the point mass function can be written as

$$F(x) = \sum_{i:\, x_i < x} f_X(x_i) = \sum_{i:\, x_i < x} P(\{\omega_j\} : X(\omega_i) = x_i).$$

Probability density function

Probability density function

- A random variable is **continuous**, if there is a $\rho(x) \geq 0$ function fulfilling

$$F(b) - F(a) = P(a \leq X < b) = P(a < X < b) = \int_a^b \rho(x)dx.$$

Probability density function

Probability density function

- A random variable is **continuous**, if there is a $\rho(x) \geq 0$ function fulfilling

$$F(b) - F(a) = P(a \leq X < b) = P(a < X < b) = \int_a^b \rho(x)dx.$$

- Using $F(-\infty) = 0$ we can express the CDF as

$$F(x) = \int_{-\infty}^{x} \rho(x')dx'.$$

Probability density function

Probability density function

- A random variable is **continuous**, if there is a $\rho(x) \geq 0$ function fulfilling

$$F(b) - F(a) = P(a \leq X < b) = P(a < X < b) = \int_a^b \rho(x)dx.$$

- Using $F(-\infty) = 0$ we can express the CDF as

$$F(x) = \int_{-\infty}^{x} \rho(x')dx'.$$

- The function $\rho(x)$ is called the **probability density function**.

Probability density function

Properties of the PDF:

Probability density function

Properties of the PDF:

- It is normalised:

$$\int_{-\infty}^{\infty} \rho(x)dx = F(\infty) - F(-\infty) = 1.$$

Probability density function

Properties of the PDF:

- It is normalised:

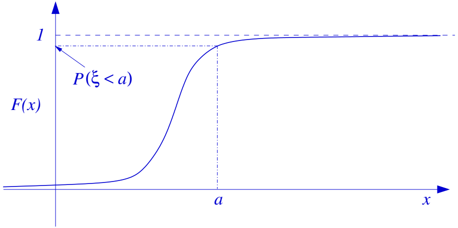
$$\int_{-\infty}^{\infty} \rho(x)dx = F(\infty) - F(-\infty) = 1.$$

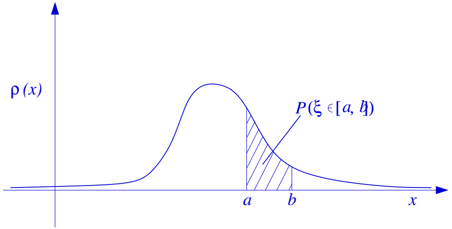
- The probability that $X \in [a, b]$ can be expressed as

$$P(X \in [a, b]) = \int_{a}^{b} \rho(x)dx.$$

F(x) and $\rho(x)$

Illustration





Important distributions

Discrete uniform distribution

X has a uniform distribution on $\{x_1, x_2, \ldots, x_n\}$ if its point mass function is given by

$$f_X(x_i) = \frac{1}{n}.$$

Important distributions

Discrete variables

Discrete uniform distribution

X has a uniform distribution on $\{x_1, x_2, \ldots, x_n\}$ if its point mass function is given by

$$f_X(x_i) = \frac{1}{n}.$$

Bernoulli distribution

Let X represent an experiment with possibly two different outcomes (e.g., coin flip), where $P(X = 1) = p$ and $P(X = 0) = 1 - p$. Then X has a **Bernoulli distribution**, with a point mass function for $x \in 0, 1$ written as

$$f_X(x) = p^x(1-p)^{1-x} = \begin{cases} p, & \text{if } x = 1, \\ 1-p, & \text{if } x = 0. \end{cases}$$

This is usually denoted as $X \sim$Bernoulli(p).

Binomial and Geometric distributions

- Assume a coin flip with probability p for heads and probability $q = 1 - p$ for tails, repeated N times independently.

Binomial and Geometric distributions

- Assume a coin flip with probability p for heads and probability $q = 1 - p$ for tails, repeated N times independently.
- Let X count to the number of heads. X has a **Binomial distribution** with a point mass function

$$f_X(x = k) = P(X = k) = \binom{N}{k} p^k (1-p)^{N-k} = \binom{N}{k} p^k q^{N-k}.$$

This is usually denoted as $X \sim \text{Binomial}(N, p)$.

Binomial and Geometric distributions

- Assume a coin flip with probability p for heads and probability $q = 1 - p$ for tails, repeated N times independently.
- Let X count to the number of heads. X has a **Binomial distribution** with a point mass function

$$f_X(x = k) = P(X = k) = \binom{N}{k} p^k (1 - p)^{N-k} = \binom{N}{k} p^k q^{N-k}.$$

  This is usually denoted as $X \sim \text{Binomial}(N, p)$.
- Let Y count the number of flips needed until the first heads. Y has a **Geometrical distribution** with a point mass function

$$f_Y(y = k) = P(Y = k) = (1 - p)^{k-1} p.$$

  This is usually denoted as $Y \sim \text{Geom}(p)$.

Poisson distribution

X taking up non-negative integer values has a Poisson distribution if its point mass function can be written as

$$f_X(x=k) = P(X=k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$

This is usually denoted as $X \sim$Poisson(λ).

Simeon Poisson

If we take a binomial distribution $X \sim \mathrm{Binom}(N,p)$ in the following limit:

$$\lim N = \infty, \quad \lim p = 0 \quad \lim_{\substack{N\to\infty \\ p\to 0}} pN = \lambda,$$

then its point mass function is converging to a Poisson distribution

$$f_X(x=k) = P(X=k) = \binom{N}{k} p^k (1-p)^{N-k} \xRightarrow{N\to\infty} \frac{(Np)^k}{k!} e^{-Np} = \frac{\lambda^k}{k!} e^{-\lambda}.$$

Poisson distribution

X taking up non-negative integer values has a Poisson distribution if its point mass function can be written as

$$f_X(x=k) = P(X=k) = \frac{\lambda^k}{k!}e^{-\lambda}.$$

Simeon Poisson

This is usually denoted as $X \sim$Poisson(λ).

If we take a binomial distribution $X \sim \mathrm{Binom}(N,p)$ in the following limit:

$$\lim N = \infty, \quad \lim p = 0 \quad \lim_{\substack{N\to\infty \\ p\to 0}} pN = \lambda,$$

then its point mass function is converging to a Poisson distribution

$$f_X(x=k) = P(X=k) = \binom{N}{k} p^k (1-p)^{N-k} \xrightarrow{N\to\infty} \frac{(Np)^k}{k!}e^{-Np} = \frac{\lambda^k}{k!}e^{-\lambda}.$$

Continuous uniform distribution

The continuous random variable $X(\omega) \in [x_1, x_2]$ has a uniform distribution if

$$\rho_X(x) = \begin{cases} \frac{1}{x_2 - x_1} & \text{if } x_1 \leq x \leq x_2 \\ 0 & \text{otherwise} \end{cases} \qquad F_X(x) = \begin{cases} 0 & \text{if } x < x_1 \\ \frac{x - x_1}{x_2 - x_1} & \text{if } x_1 \leq x \leq x_2 \\ 1 & \text{if } x > x_2 \end{cases}$$

Continuous uniform distribution

The continuous random variable $X(\omega) \in [x_1, x_2]$ has a uniform distribution if

$$\rho_X(x) = \begin{cases} \frac{1}{x_2 - x_1} & \text{if } x_1 \le x \le x_2 \\ 0 & \text{otherwise} \end{cases} \qquad F_X(x) = \begin{cases} 0 & \text{if } x < x_1 \\ \frac{x - x_1}{x_2 - x_1} & \text{if } x_1 \le x \le x_2 \\ 1 & \text{if } x > x_2 \end{cases}$$

Exponential distribution

A continuous random variable X over the non-negative real numbers has an **Exponential distribution** if

$$F_X(x) = \begin{cases} 1 - e^{-\frac{x}{\lambda}} & \text{if } x \ge 0 \\ 0 & \text{if } x < 0 \end{cases} \qquad \rho_X(x) = \begin{cases} \frac{1}{\lambda} e^{-\frac{x}{\lambda}} & \text{if } x \ge 0 \\ 0 & \text{if } x < 0 \end{cases}$$

This is usually denoted as $X \sim \text{Exp}(\lambda)$.

Gamma distribution

- A continuous random variable X over the non-negative real numbers has a **Gamma distribution**, usually denoted as $X \sim$Gamma(q, λ) if

$$\rho_X(x) = \frac{1}{\lambda^q \Gamma(q)} x^{q-1} e^{-\frac{x}{\lambda}},$$

where the Gamma function $\Gamma(z)$ is defined as

$$\Gamma(z) = \int_0^\infty x^{z-1} e^{-x} dx.$$

Gamma distribution

- A continuous random variable X over the non-negative real numbers has a **Gamma distribution**, usually denoted as X ~Gamma(q, λ) if

$$\rho_X(x) = \frac{1}{\lambda^q \Gamma(q)} x^{q-1} e^{-\frac{x}{\lambda}},$$

where the Gamma function $\Gamma(z)$ is defined as

$$\Gamma(z) = \int_0^\infty x^{z-1} e^{-x} dx.$$

- Connections with the exponential distribution:
  - The exponential distribution corresponds to the special case of Gamma$(q = 1, \lambda)$.

Important distributions

Continous variables

Gamma distribution

- A continuous random variable X over the non-negative real numbers has a **Gamma distribution**, usually denoted as X ~Gamma(q, λ) if

$$\rho_X(x) = \frac{1}{\lambda^q \Gamma(q)} x^{q-1} e^{-\frac{x}{\lambda}},$$

where the Gamma function $\Gamma(z)$ is defined as

$$\Gamma(z) = \int_0^\infty x^{z-1} e^{-x} dx.$$

- Connections with the exponential distribution:
  - The exponential distribution corresponds to the special case of Gamma$(q = 1, \lambda)$.
  - The sum of n independent random variables X_i ~Exp(λ) has a Gamma distribution Gamma(n, λ).

Gamma distribution

- A continuous random variable X over the non-negative real numbers has a **Gamma distribution**, usually denoted as X ~Gamma(q, λ) if

$$\rho_X(x) = \frac{1}{\lambda^q \Gamma(q)} x^{q-1} e^{-\frac{x}{\lambda}},$$

where the Gamma function $\Gamma(z)$ is defined as

$$\Gamma(z) = \int_0^\infty x^{z-1} e^{-x} dx.$$

- Connections with the exponential distribution:
  - The exponential distribution corresponds to the special case of Gamma$(q = 1, \lambda)$.
  - The sum of n independent random variables X_i ~Exp(λ) has a Gamma distribution Gamma(n, λ).
  - → Thus, also the sum of independent X_i ~Gamma(q_i, λ) has a distribution $\sum_{i=1}^n X_i$ ~Gamma$(\sum_{i=1}^n q_i, \lambda)$.

Normal distribution

A continuous random variable X has a **Normal** (Gaussian) distribution (usually denoted by $X \sim \mathrm{N}(\mu, \sigma)$) if

$$\rho_X(x) = \mathcal{N}(\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad F_X(x) = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{x-\mu}{\sqrt{2}\sigma}\right)\right],$$

A standard normal distribution is corresponding to $\mathrm{N}(\mu = 0, \sigma = 1)$.

Important distributions

Continuous variables

Normal distribution

A continuous random variable X has a **Normal** (Gaussian) distribution (usually denoted by $X \sim \mathrm{N}(\mu, \sigma)$) if

$$\rho_X(x) = \mathcal{N}(\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad F_X(x) = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{x-\mu}{\sqrt{2}\sigma}\right)\right],$$

A standard normal distribution is corresponding to $\mathrm{N}(\mu = 0, \sigma = 1)$.

χ^2 distribution

- A continuous random variable X over the non-negative real numbers has a χ^2 distribution with n degrees of freedom (usually denoted by $X \sim \chi^2_n$) if

$$\rho_X(x) = \frac{1}{\Gamma\left(\frac{n}{2}\right) 2^{\frac{n}{2}}} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}$$

- Connection with the Normal distribution:
  If $X_1, X_2, \ldots, X_n$ are independent standard Normal random variables, then $\sum_{i=1}^n Z_i^2 \sim \chi^2_n$.

t distribution and Cauchy distribution

- A continuous random variable X over the non-negative real numbers has a t **distribution** (also called as Student's t distribution) with n degrees of freedom (usually denoted by $X \sim t_n$) if

$$\rho_X(x) = \frac{1}{\sqrt{n\pi}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \frac{1}{\left(1 + \frac{x^2}{n}\right)^{\frac{n+1}{2}}}.$$

t distribution and Cauchy distribution

- A continuous random variable X over the non-negative real numbers has a t **distribution** (also called as Student's t distribution) with n degrees of freedom (usually denoted by $X \sim t_n$) if

$$\rho_X(x) = \frac{1}{\sqrt{n\pi}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \frac{1}{\left(1 + \frac{x^2}{n}\right)^{\frac{n+1}{2}}}.$$

- Connection with the Normal distribution:
  If $X_1, X_2, \ldots, X_n$ and Y are independent standard Normal random variables, then the variable

$$Z := \frac{\sqrt{n}Y}{\sqrt{X_1^2 + X_2^2 + \cdots + X_n^2}} \quad \text{has a } t \text{ ditribution, } Z \sim t_n$$

t distribution and Cauchy distribution

- A continuous random variable X over the non-negative real numbers has a t **distribution** (also called as Student's t distribution) with n degrees of freedom (usually denoted by $X \sim t_n$) if

$$\rho_X(x) = \frac{1}{\sqrt{n\pi}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \frac{1}{\left(1 + \frac{x^2}{n}\right)^{\frac{n+1}{2}}}.$$

- Connection with the Normal distribution:
  If $X_1, X_2, \ldots, X_n$ and Y are independent standard Normal random variables, then the variable

$$Z := \frac{\sqrt{n}Y}{\sqrt{X_1^2 + X_2^2 + \cdots + X_n^2}} \quad \text{has a } t \text{ ditribution, } Z \sim t_n$$

- The $n = 1$ special case of the t distribution corresponds to the **Cauchy distribution**, where

$$\rho_X(x) = \frac{1}{\pi} \frac{1}{1 + x^2}.$$

Bivariate distribution

Bivariate distribution

Cumulative distribution function

The joint CDF of random variables X and Y is defined as

$$F_{X,Y}(x, y) := P(X < x, Y < y).$$

Properties:

- Monotonous, non-decreasing function of its variables.
- $\lim_{x \to -\infty} F(x, y) = \lim_{y \to -\infty} F(x, y) = 0$ and $\lim_{x,y \to \infty} F(x, y) = 1.$

Bivariate distribution

Cumulative distribution function

The joint CDF of random variables X and Y is defined as

$$F_{X,Y}(x,y) := P(X < x, Y < y).$$

Properties:

- Monotonous, non-decreasing function of its variables.
- $\lim_{x\to-\infty} F(x,y) = \lim_{y\to-\infty} F(x,y) = 0$ and $\lim_{x,y\to\infty} F(x,y) = 1$.

Bivariate distribution

Cumulative distribution function

The joint CDF of random variables X and Y is defined as

$$F_{X,Y}(x,y) := P(X < x, Y < y).$$

Properties:

- Monotonous, non-decreasing function of its variables.
- $\lim_{x\to-\infty} F(x,y) = \lim_{y\to-\infty} F(x,y) = 0$ and $\lim_{x,y\to\infty} F(x,y) = 1$.

Bivariate distribution

Cumulative distribution function

The joint CDF of random variables X and Y is defined as

$$F_{X,Y}(x,y) := P(X < x, Y < y).$$

Properties:

- Monotonous, non-decreasing function of its variables.
- $\lim_{x\to-\infty} F(x,y) = \lim_{y\to-\infty} F(x,y) = 0$ and $\lim_{x,y\to\infty} F(x,y) = 1$.

Joint mass function

For discrete random variables X and Y the joint mass function is defined as

$$f_{X,Y}(x,y) = P(X = x, Y = y).$$

Bivariate distribution

Joint probability density function

For continuous random variables X and Y the joint probability density function is given by the function $\rho(x,y)$ connected to the joint CDF $F(x,y)$ as

$$F(x,y) = \int_{-\infty}^{x} dx' \int_{-\infty}^{y} dy' \rho(x',y') \qquad \rho(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y}$$

Properties:

- Normalised:

$$\int_{-\infty}^{\infty} dx' \int_{-\infty}^{\infty} dy' \rho(x',y') = 1.$$

- The probability that X and Y fall into given $[x_1,x_2]$ and $[y_1,y_2]$ intervals can be written as

$$P(x_1 < X < x_2, y_1 < Y < y_2) = \int_{x_1}^{x_2} dx \int_{y_1}^{y_2} dy \rho(x,y).$$

Bivariate distribution

Joint probability density function

For continuous random variables X and Y the joint probability density function is given by the function $\rho(x,y)$ connected to the joint CDF $F(x,y)$ as

$$F(x,y) = \int_{-\infty}^{x} dx' \int_{-\infty}^{y} dy' \rho(x',y') \qquad \rho(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y}$$

Properties:

- Normalised:

$$\int_{-\infty}^{\infty} dx' \int_{-\infty}^{\infty} dy' \rho(x',y') = 1.$$

- The probability that X and Y fall into given $[x_1, x_2]$ and $[y_1, y_2]$ intervals can be written as

$$P(x_1 < X < x_2, y_1 < Y < y_2) = \int_{x_1}^{x_2} dx \int_{y_1}^{y_2} dy \rho(x,y).$$

Bivariate distribution

Joint probability density function

For continuous random variables X and Y the joint probability density function is given by the function $\rho(x,y)$ connected to the joint CDF $F(x,y)$ as

$$F(x,y) = \int_{-\infty}^{x} dx' \int_{-\infty}^{y} dy' \rho(x',y') \qquad \rho(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y}$$

Properties:

- Normalised:

$$\int_{-\infty}^{\infty} dx' \int_{-\infty}^{\infty} dy' \rho(x',y') = 1.$$

- The probability that X and Y fall into given $[x_1, x_2]$ and $[y_1, y_2]$ intervals can be written as

$$P(x_1 < X < x_2, y_1 < Y < y_2) = \int_{x_1}^{x_2} dx \int_{y_1}^{y_2} dy \rho(x,y).$$

Marginal distributions

Marginal CDF and probability densities

The joint distribution of X and Y uniquely determines the distribution of either X or Y, and these are called the marginals of the joint CDF:

$$F_X(x) = P(X < x, Y < \infty) = F_{X,Y}(x, \infty) = \int_{-\infty}^{x} dx' \int_{-\infty}^{\infty} dy' \rho(x', y'),$$

$$F_Y(y) = P(X < \infty, Y < y) = F_{X,Y}(\infty, y) = \int_{-\infty}^{\infty} dx' \int_{-\infty}^{y} dy' \rho(x', y').$$

The marginal probability densities:

$$\rho_X(x) = \int_{-\infty}^{\infty} dy \rho(x, y), \qquad \rho_Y(y) = \int_{\infty}^{\infty} dx \rho(x, y).$$

Marginal distributions

Marginal CDF and probability densities

The joint distribution of X and Y uniquely determines the distribution of either X or Y, and these are called the marginals of the joint CDF:

$$F_X(x) = P(X < x, Y < \infty) = F_{X,Y}(x, \infty) = \int_{-\infty}^{x} dx' \int_{-\infty}^{\infty} dy' \rho(x', y'),$$

$$F_Y(y) = P(X < \infty, Y < y) = F_{X,Y}(\infty, y) = \int_{-\infty}^{\infty} dx' \int_{-\infty}^{y} dy' \rho(x', y').$$

The marginal probability densities:

$$\rho_X(x) = \int_{-\infty}^{\infty} dy \rho(x, y), \qquad \rho_Y(y) = \int_{\infty}^{\infty} dx \rho(x, y).$$

Independent random variables

Independence of random variables

The random variables X and Y are independent if for any $x_1 \leq x_2$ and $y_1 \leq y_2$

$$P(x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2) = P(x_1 \leq X \leq x_2)P(y_1 \leq Y \leq y_2).$$

- For discrete variables this means that for any x and y

$$P(X = x, Y = y) = P(X = x)P(Y = y).$$

- For continuous variables this means that for any x and y

$$F_{X,Y}(x, y) = F_X(x)F_Y(y), \qquad p(x, y) = p_X(x)p_Y(y).$$

Independent random variables

Independence of random variables

The random variables X and Y are independent if for any $x_1 \leq x_2$ and $y_1 \leq y_2$

$$P(x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2) = P(x_1 \leq X \leq x_2)P(y_1 \leq Y \leq y_2).$$

- For discrete variables this means that for any x and y

$$P(X = x, Y = y) = P(X = x)P(Y = y).$$

- For continuous variables this means that for any x and y

$$F_{X,Y}(x, y) = F_X(x)F_Y(y), \qquad \rho(x, y) = \rho_X(x)\rho_Y(y).$$

Independent random variables

Independence of random variables

The random variables X and Y are independent if for any $x_1 \leq x_2$ and $y_1 \leq y_2$

$$P(x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2) = P(x_1 \leq X \leq x_2)P(y_1 \leq Y \leq y_2).$$

- For discrete variables this means that for any x and y

$$P(X = x, Y = y) = P(X = x)P(Y = y).$$

- For continuous variables this means that for any x and y

$$F_{X,Y}(x, y) = F_X(x)F_Y(y), \qquad \rho(x, y) = \rho_X(x)\rho_Y(y).$$

Conditional distributions

Conditional distribution

- The conditional CDF of a random variable X given that the variable Y takes the value $Y = y$ can be defined as

$$F_X(x \mid Y = y) := \lim_{\Delta y \to 0} P(X < x \mid y \le Y < y + \Delta y).$$

- The joint CDF of X and Y fully determines this distribution:

$$\begin{aligned} F_X(x \mid y) &= \lim_{\Delta y \to 0} P(X < x \mid y \le Y < y + \Delta y) = \\ & \lim_{\Delta y \to 0} \frac{P(X < x, y \le Y < y + \Delta y)}{P(y \le Y < y + \Delta y)} = \\ & \lim_{\Delta y \to 0} \frac{F(x, y + \Delta y) - F(x, y)}{F_Y(y + \Delta y) - F_Y(y)} = \\ & \lim_{\Delta y \to 0} \frac{\frac{F(x,y+\Delta y) - F(x,y)}{\Delta y}}{\frac{F_Y(y+\Delta y) - F_Y(y)}{\Delta y}} = \frac{\frac{\partial}{\partial y} F(x, y)}{\rho_Y(y)} \end{aligned}$$

Conditional distributions

Conditional distribution

- The conditional CDF of a random variable X given that the variable Y takes the value $Y = y$ can be defined as

$$F_X(x \mid Y = y) := \lim_{\Delta y \to 0} P(X < x \mid y \leq Y < y + \Delta y).$$

- The joint CDF of X and Y fully determines this distribution:

$$\begin{aligned} F_X(x \mid y) &= \lim_{\Delta y \to 0} P(X < x \mid y \leq Y < y + \Delta y) = \\ & \lim_{\Delta y \to 0} \frac{P(X < x, y \leq Y < y + \Delta y)}{P(y \leq Y < y + \Delta y)} = \\ & \lim_{\Delta y \to 0} \frac{F(x, y + \Delta y) - F(x, y)}{F_Y(y + \Delta y) - F_Y(y)} = \\ & \lim_{\Delta y \to 0} \frac{\frac{F(x,y+\Delta y) - F(x,y)}{\Delta y}}{\frac{F_Y(y+\Delta y) - F_Y(y)}{\Delta y}} = \frac{\frac{\partial}{\partial y} F(x, y)}{\rho_Y(y)} \end{aligned}$$

Conditional distribution

Conditional probability density

- By taking the partial derivative of $F_X(x \mid y)$-t with respect to x we obtain the conditional probability density of X as

$$\rho_X(x \mid y) = \frac{\partial}{\partial x} F_X(x \mid y) = \frac{\partial}{\partial x} \frac{\frac{\partial}{\partial y} F(x, y)}{\rho_Y(y)} = \frac{\rho(x, y)}{\rho_Y(y)}.$$

Similarly, the conditional CDF and density of Y given that $X = x$ can be written as

$$F_Y(y \mid X = x) = \frac{\frac{\partial}{\partial x} F(x, y)}{\rho_X(x)}, \qquad \rho_Y(y \mid x) = \frac{\rho(x, y)}{\rho_X(x)}.$$

Conditional distribution

Conditional probability density

- By taking the partial derivative of $F_X(x \mid y)$-t with respect to x we obtain the conditional probability density of X as

$$\rho_X(x \mid y) = \frac{\partial}{\partial x} F_X(x \mid y) = \frac{\partial}{\partial x} \frac{\frac{\partial}{\partial y} F(x,y)}{\rho_Y(y)} = \frac{\rho(x,y)}{\rho_Y(y)}.$$

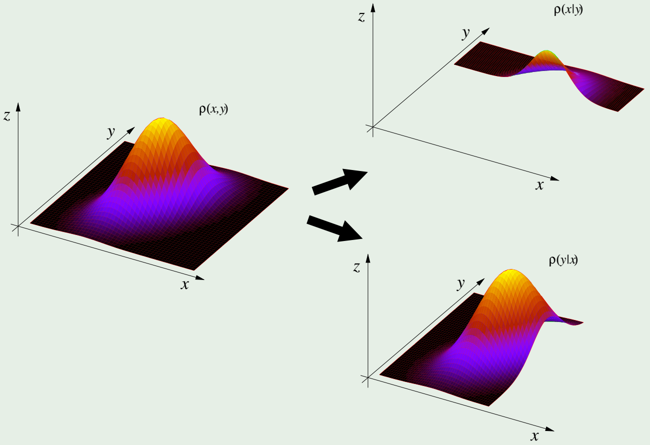
Similarly, the conditional CDF and density of Y given that $X = x$ can be written as

$$F_Y(y \mid X = x) = \frac{\frac{\partial}{\partial x} F(x,y)}{\rho_X(x)}, \qquad \rho_Y(y \mid x) = \frac{\rho(x,y)}{\rho_X(x)}.$$

Conditional probability

Illustration

Marginal density distribution	$\leftrightarrow$	projection
Conditional density distribution	$\leftrightarrow$	intersection



Conditional density and marginal density

- Since

$$\rho(x,y) = \rho_X(x \mid y)\rho_Y(y) = \rho_Y(y \mid x)\rho_X(x),$$

the marginals can be expressed with the help of the conditional densities as

$$\rho_X(x) = \int_{-\infty}^{\infty} dy \rho(x,y) = \int_{-\infty}^{\infty} dy \rho_X(x \mid y)\rho_Y(y),$$

$$\rho_Y(y) = \int_{-\infty}^{\infty} dx \rho(x,y) = \int_{-\infty}^{\infty} dx \rho_Y(y \mid x)\rho_X(x).$$

(These are analogous to the law of total probability).

- An identity analogous to Bayes' theorem:

$$\rho_X(x \mid y) = \frac{\rho(x,y)}{\rho_Y(y)} = \frac{\rho_Y(y \mid x)\rho_X(x)}{\int_{-\infty}^{\infty} dx \rho_Y(y \mid x)\rho_X(x)}$$

Conditional density and marginal density

- Since
$$\rho(x,y) = \rho_X(x \mid y)\rho_Y(y) = \rho_Y(y \mid x)\rho_X(x),$$
the marginals can be expressed with the help of the conditional densities as
$$\rho_X(x) = \int_{-\infty}^{\infty} dy \rho(x,y) = \int_{-\infty}^{\infty} dy \rho_X(x \mid y)\rho_Y(y),$$
$$\rho_Y(y) = \int_{-\infty}^{\infty} dx \rho(x,y) = \int_{-\infty}^{\infty} dx \rho_Y(y \mid x)\rho_X(x).$$
(These are analogous to the law of total probability).
- An identity analogous to Bayes' theorem:
$$\rho_X(x \mid y) = \frac{\rho(x,y)}{\rho_Y(y)} = \frac{\rho_Y(y \mid x)\rho_X(x)}{\int_{-\infty}^{\infty} dx \rho_Y(y \mid x)\rho_X(x)}$$

Multivariate distributions

Multivariate CDF

The joint CDF of random variables $X_1, X_2, \ldots, X_n$ is defined similarly to the bivariate case as

$$F(x_1, x_2, \ldots, x_n) := P(X_1 < x_1, X_1 < x_2, \ldots, X_n < x_n).$$

Properties:

- It is a monotonous non-decreasing function of its variables.
- $\forall i \in [1, n] \quad \lim_{x_i \to -\infty} F(x_1, \ldots, x_i, \ldots, x_n) = 0$ and $\lim_{x_1, \ldots, x_n \to \infty} F(x_1, x_2, \ldots, x_n) = 1.$

Multivariate distributions

Multivariate CDF

The joint CDF of random variables $X_1, X_2, \ldots, X_n$ is defined similarly to the bivariate case as

$$F(x_1, x_2, \ldots, x_n) := P(X_1 < x_1, X_1 < x_2, \ldots, X_n < x_n).$$

Properties:

- It is a monotonous non-decreasing function of its variables.
- $\forall i \in [1, n] \lim_{x_i \to -\infty} F(x_1, \ldots, x_i, \ldots, x_n) = 0$ and $\lim_{x_1, \ldots, x_n \to \infty} F(x_1, x_2, \ldots, x_n) = 1.$

Multivariate distributions

Multivariate CDF

The joint CDF of random variables $X_1, X_2, \ldots, X_n$ is defined similarly to the bivariate case as

$$F(x_1, x_2, \ldots, x_n) := P(X_1 < x_1, X_1 < x_2, \ldots, X_n < x_n).$$

Properties:

- It is a monotonous non-decreasing function of its variables.
- $\forall i \in [1, n] \quad \lim_{x_i \to -\infty} F(x_1, \ldots, x_i, \ldots, x_n) = 0$ and $\lim_{x_1, \ldots, x_n \to \infty} F(x_1, x_2, \ldots, x_n) = 1.$

Multivariate distributions

Multivariate PDF

The joint PDF of continuous random variables $X_1, X_2, \ldots, X_n$ denoted by $\rho(x_1, x_2, \ldots, x_n)$ fulfils

$$F(x_1, x_2, \ldots, x_n) = \int_{-\infty}^{x_1} dx_1' \int_{-\infty}^{x_2} dx_2' \cdots \int_{-\infty}^{x_n} dx_n' \rho(x_1', x_2', \ldots, x_n')$$

$$\rho(x_1, x_2, \ldots, x_n) = \frac{\partial^n F(x_1, x_2, \ldots, x_n)}{\partial x_1 \partial x_2 \cdots \partial x_n}$$

Properties:

- Normalised

$$\int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dx_2 \cdots \int_{-\infty}^{\infty} dx_n \rho(x_1, x_2, \ldots, x_n) = 1.$$

- The probability that $(X_1, X_2, \ldots, X_n)$ falls in some region E is

$$P\left((X_1, X_2, \ldots, X_n) \in E\right) = \iint \cdots \int_E dx_1 dx_2 \ldots dx_n \rho(x_1, x_2, \ldots, x_n).$$

Multivariate distributions

Multivariate PDF

The joint PDF of continuous random variables $X_1, X_2, \ldots, X_n$ denoted by $\rho(x_1, x_2, \ldots, x_n)$ fulfils

$$F(x_1, x_2, \ldots, x_n) = \int_{-\infty}^{x_1} dx_1' \int_{-\infty}^{x_2} dx_2' \cdots \int_{-\infty}^{x_n} dx_n' \rho(x_1', x_2', \ldots, x_n')$$

$$\rho(x_1, x_2, \ldots, x_n) = \frac{\partial^n F(x_1, x_2, \ldots, x_n)}{\partial x_1 \partial x_2 \cdots \partial x_n}$$

Properties:

- Normalised

$$\int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dx_2 \cdots \int_{-\infty}^{\infty} dx_n \rho(x_1, x_2, \ldots, x_n) = 1.$$

- The probability that $(X_1, X_2, \ldots, X_n)$ falls in some region E is

$$P\left((X_1, X_2, \ldots, X_n) \in E\right) = \iint \cdots \int_E dx_1 dx_2 \ldots dx_n \rho(x_1, x_2, \ldots, x_n).$$

IID variables

IID variables

Independent and Identically distributed variables

- The independence of random variables $X_1, X_2, \ldots, X_n$ is defined similarly to the bivariate case, i.e., they are independent if for any $A_1, A_2, \ldots, A_n$ where $A_i \subset \mathbb{R}$,

$$P(X_1 \in A_1, X_2 \in A_2, \ldots, X_n \in A_n) = \prod_{i=1}^{n} P(X_i \in A_i).$$

IID variables

Independent and Identically distributed variables

- The independence of random variables $X_1, X_2, \dots, X_n$ is defined similarly to the bivariate case, i.e., they are independent if for any $A_1, A_2, \dots, A_n$ where $A_i \subset \mathbb{R}$,

$$P(X_1 \in A_1, X_2 \in A_2, \dots, X_n \in A_n) = \prod_{i=1}^{n} P(X_i \in A_i).$$

- This is equivalent to

$$\rho(x_1, x_2, \dots, x_n) = \rho_{X_1}(x_1)\rho_{X_2}(x_2)\cdots\rho_{X_n}(x_n)$$

IID variables

Independent and Identically distributed variables

- The independence of random variables $X_1, X_2, \ldots, X_n$ is defined similarly to the bivariate case, i.e., they are independent if for any $A_1, A_2, \ldots, A_n$ where $A_i \subset \mathbb{R}$,

$$P(X_1 \in A_1, X_2 \in A_2, \ldots, X_n \in A_n) = \prod_{i=1}^{n} P(X_i \in A_i).$$

- This is equivalent to

$$\rho(x_1, x_2, \ldots, x_n) = \rho_{X_1}(x_1)\rho_{X_2}(x_2)\cdots\rho_{X_n}(x_n)$$

- If in addition all X_i have the same marginal distribution, then we call them independent and identically distributed (IID). This means that they correspond to independent draws from the same distribution.

Important multivariate distributions

Multinomial distribution

- This is the **multivariate** version of the **binomial distribution**.

  Let us consider N draws with replacement from an urn with balls of q different colours, where the probability of drawing colour i is p_i. Let $X = (X_1, X_2, \ldots, X_q)$ count the number of drawn balls with the different colours. The probability mass function can be written as

$$f_X(k_1, k_2, \ldots, k_q) = \frac{N!}{k_1!k_2!\cdots k_q!} p_1^{k_1} p_2^{k_2} \cdots p_q^{k_q}.$$

- The marginal distributions of $X = (X_1, X_2, \ldots, X_q)$ ~Multinomial(N, p) where $p = (p_1, p_2, \ldots, p_q)$ are given by X_i ~Binom(N, p_i).

Important multivariate distributions

Multinomial distribution

- This is the **multivariate** version of the **binomial distribution**.

  Let us consider N draws with replacement from an urn with balls of q different colours, where the probability of drawing colour i is p_i. Let $X = (X_1, X_2, \ldots, X_q)$ count the number of drawn balls with the different colours. The probability mass function can be written as

$$f_X(k_1, k_2, \ldots, k_q) = \frac{N!}{k_1!k_2!\cdots k_q!} p_1^{k_1} p_2^{k_2} \cdots p_q^{k_q}.$$

- The marginal distributions of $X = (X_1, X_2, \ldots, X_q) \sim$Multinomial$(N, p)$ where $p = (p_1, p_2, \ldots, p_q)$ are given by $X_i \sim$Binom(N, p_i).

Important multivariate distributions

Multivariate normal distribution

The **multivariate Normal distribution** is parametrised by the vector of expected values $\mu = (\mu_1, \mu_2, \ldots, \mu_n)$ and the covariance matrix Σ. A random vector $X = (X_1, X_2, \ldots, X_n)$ has multivariate Normal distribution if the joint density distribution can be written as

$$\rho(x_1, x_2, \ldots, x_n) = \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu})[\Sigma]^{-1}(\vec{x}-\vec{\mu})} = \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} e^{-\frac{1}{2}\sum_{ij}\left[\Sigma^{-1}\right]_{ij}(x_i-\mu_i)(x_j-\mu_j)}$$

Transformation of random variables

Transformation of random variables

Transformation of variables

- If $Y = f(X)$, then how can we formulate the PDF of Y based on $\rho_X(x)$?

$$y = f(x) \;\rightarrow\; dy = f'(x)dx$$

$$\rho_Y(y)\,|dy| = \rho_X(x)\,|dx|$$

$$\rho_Y(y) = \rho_X(x)\frac{|dx|}{|dy|} = \rho_X(f^{-1}(y))\frac{1}{|f'(f^{-1}(y))|}$$

$$\rho_Y(y) = \sum_n \rho_X(f_n^{-1}(y))\frac{1}{|f'(f_n^{-1}(y))|}$$

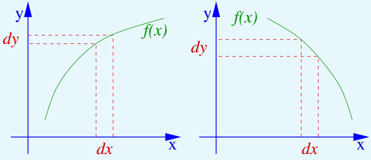
Transformation of random variables

Transformation of variables

- If $Y = f(X)$, then how can we formulate the PDF of Y based on $\rho_X(x)$?

$$y = f(x) \;\rightarrow\; dy = f'(x)dx$$

$$\rho_Y(y)\,|dy| = \rho_X(x)\,|dx|$$



$$\rho_Y(y) = \rho_X(x)\frac{|dx|}{|dy|} = \rho_X(f^{-1}(y))\frac{1}{|f'(f^{-1}(y))|}$$

$$\rho_Y(y) = \sum_n \rho_X(f_n^{-1}(y))\frac{1}{|f'(f_n^{-1}(y))|}$$

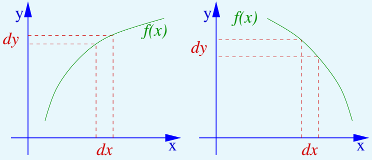
Transformation of random variables

Transformation of variables

- If $Y = f(X)$, then how can we formulate the PDF of Y based on $\rho_X(x)$?

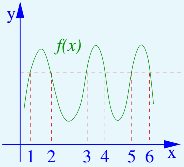
$$y = f(x) \;\rightarrow\; dy = f'(x)dx$$

$$\rho_Y(y)\,|dy| = \rho_X(x)\,|dx|$$



$$\rho_Y(y) = \rho_X(x)\frac{|dx|}{|dy|} = \rho_X(f^{-1}(y))\frac{1}{|f'(f^{-1}(y))|}$$

$$\rho_Y(y) = \sum_n \rho_X(f_n^{-1}(y))\frac{1}{|f'(f_n^{-1}(y))|}$$



Transformations of random variables

Transformation of multivariate distributions

How to obtain the distribution of $\vec{Y} = (Y_1, Y_2, \ldots, Y_q) = \vec{f}(\vec{X}) = \vec{f}(X_1, X_2, \ldots, X_k)$ based on e.g., $\rho_{\vec{X}}(x_1, x_2, \ldots, x_k)$?

The general recipe:

- For each $\vec{y}$ find the set

$$A_{\vec{y}} = \{\vec{x}: \ f_1(\vec{x}) < y_1, f_2(\vec{x}) < y_2, \ldots, f_n(\vec{x}) < y_n\}.$$

- Based on that, the CDF can be written as

$$F_{\vec{Y}}(\vec{y}) = P(\vec{Y} < \vec{y}) = P(\vec{f}(\vec{x}) < \vec{y}) = P\left(\{\vec{x} : \vec{f}(\vec{x}) < \vec{y}\}\right) = \int\int \cdots \int_{A_{\vec{y}}} \rho_{\vec{X}}(x_1, x_2, \ldots, x_k) dx_1 dx_2 \ldots dx_n.$$

- The PDF is

$$\rho_{\vec{Y}}(\vec{y}) = F'_{\vec{Y}}(\vec{y}).$$

Transformations of random variables

Transformation of multivariate distributions

How to obtain the distribution of $\vec{Y} = (Y_1, Y_2, \ldots, Y_q) = \vec{f}(\vec{X}) = \vec{f}(X_1, X_2, \ldots, X_k)$ based on e.g., $\rho_{\vec{X}}(x_1, x_2, \ldots, x_k)$?

The general recipe:

- For each $\vec{y}$ find the set

$$A_{\vec{y}} = \{\vec{x} : \; f_1(\vec{x}) < y_1, f_2(\vec{x}) < y_2, \ldots, f_n(\vec{x}) < y_n\}.$$

- Based on that, the CDF can be written as

$$F_{\vec{Y}}(\vec{y}) = P(\vec{Y} < \vec{y}) = P(\vec{f}(\vec{x}) < \vec{y}) = P\left(\{\vec{x} : \vec{f}(\vec{x}) < \vec{y}\}\right) = \int\int \cdots \int_{A_{\vec{y}}} \rho_{\vec{X}}(x_1, x_2, \ldots, x_k) dx_1 dx_2 \ldots dx_n.$$

- The PDF is

$$\rho_{\vec{Y}}(\vec{y}) = F'_{\vec{Y}}(\vec{y}).$$

Transformations of random variables

Transformation of multivariate distributions

How to obtain the distribution of $\vec{Y} = (Y_1, Y_2, \ldots, Y_q) = \vec{f}(\vec{X}) = \vec{f}(X_1, X_2, \ldots, X_k)$ based on e.g., $\rho_{\vec{X}}(x_1, x_2, \ldots, x_k)$?

The general recipe:

- For each $\vec{y}$ find the set

$$A_{\vec{y}} = \{\vec{x} : \; f_1(\vec{x}) < y_1, f_2(\vec{x}) < y_2, \ldots, f_n(\vec{x}) < y_n\}.$$

- Based on that, the CDF can be written as

$$F_{\vec{Y}}(\vec{y}) = P(\vec{Y} < \vec{y}) = P(\vec{f}(\vec{x}) < \vec{y}) = P\left(\{\vec{x} : \vec{f}(\vec{x}) < \vec{y}\}\right) = \int\int \cdots \int_{A_{\vec{y}}} \rho_{\vec{X}}(x_1, x_2, \ldots, x_k) dx_1 dx_2 \ldots dx_n.$$

- The PDF is

$$\rho_{\vec{Y}}(\vec{y}) = F'_{\vec{Y}}(\vec{y}).$$

Transformation of multivariate distributions

How to obtain the distribution of $\vec{Y} = (Y_1, Y_2, \ldots, Y_q) = \vec{f}(\vec{X}) = \vec{f}(X_1, X_2, \ldots, X_k)$ based on e.g., $\rho_{\vec{X}}(x_1, x_2, \ldots, x_k)$?

The general recipe:

- For each $\vec{y}$ find the set

$$A_{\vec{y}} = \{\vec{x} : \; f_1(\vec{x}) < y_1, f_2(\vec{x}) < y_2, \ldots, f_n(\vec{x}) < y_n\}.$$

- Based on that, the CDF can be written as

$$F_{\vec{Y}}(\vec{y}) = P(\vec{Y} < \vec{y}) = P(\vec{f}(\vec{x}) < \vec{y}) = P\left(\{\vec{x} : \vec{f}(\vec{x}) < \vec{y}\}\right) = \int \int \cdots \int_{A_{\vec{y}}} \rho_{\vec{X}}(x_1, x_2, \ldots, x_k) dx_1 dx_2 \ldots dx_n.$$

- The PDF is

$$\rho_{\vec{Y}}(\vec{y}) = F'_{\vec{Y}}(\vec{y}).$$

Transformations of random variables

- In the special case of $\vec{Y} = \vec{f}(\vec{X})$ where both Y and X have dimension n and f is invertible:

$$\vec{y} = \vec{f}(\vec{x}), \quad \rightarrow \quad d\vec{y} = D\vec{f}(\vec{x})d\vec{x}, \quad \rho_Y(\vec{y})\,|d\vec{y}| = \rho_X(\vec{x})\,|d\vec{x}|,$$

$$\rho_Y(\vec{y}) = \rho_X(\vec{x})\frac{|d\vec{x}|}{|d\vec{y}|} = \rho_X(\vec{f}^{-1}(\vec{y}))\left|\det \boldsymbol{J}\left[\vec{f}^{-1}(\vec{y})\right]\right|,$$

where the determinant of the Jacobi matrix is

$$\det \boldsymbol{J}\left[\vec{g}(\vec{x})\right] = \det\begin{pmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} & \cdots & \frac{\partial g_1}{\partial x_n} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} & \cdots & \frac{\partial g_2}{\partial x_n} \\ \vdots & & \ddots & \vdots \\ \frac{\partial g_n}{\partial x_1} & \frac{\partial g_n}{\partial x_2} & \cdots & \frac{\partial g_n}{\partial x_n} \end{pmatrix}$$

Transformations of random variables

- What if $\vec{X} = (X_1, X_2, \ldots, X_n)$, but $\vec{Y}$ and also $\vec{f}$ are m dimensional?

Transformations of random variables

- What if $\vec{X} = (X_1, X_2, ..., X_n)$, but $\vec{Y}$ and also $\vec{f}$ are m dimensional?

→ We can write an integral form with the help of the Dirac-delta :

$$\begin{aligned} \rho_{\vec{Y}}(\vec{y}) &= \iint \cdots \int \delta(\vec{y} - \vec{f}(\vec{x}))\rho_{\vec{X}}(\vec{x})dx^n = \\ & \iint \cdots \int \delta(\vec{y} - \vec{f}(\vec{x}))\rho_{\vec{X}}(x_1, x_2, ..., x_n)dx^n. \end{aligned}$$