Teoretične osnove računalništva 2
Teorija informacij

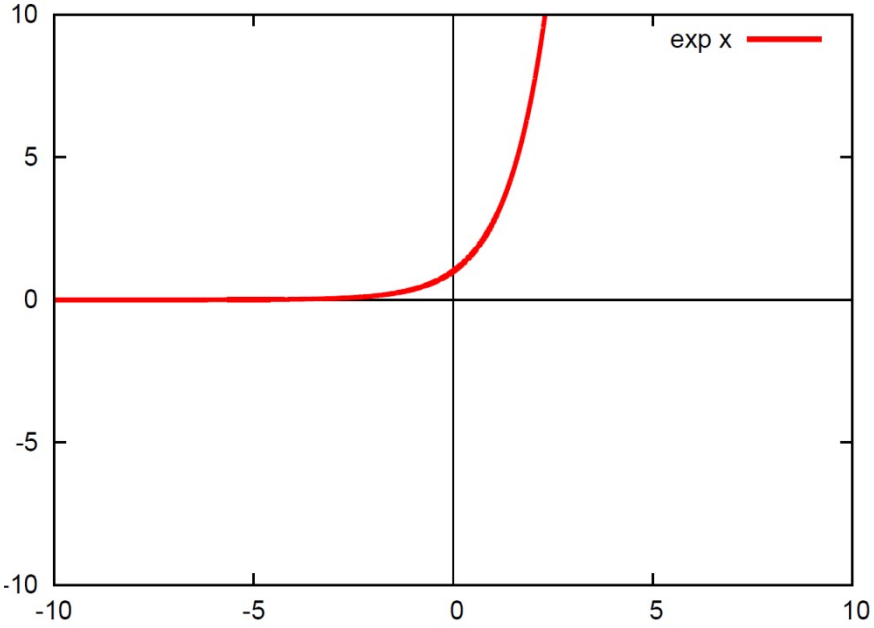
Mathematical background for information theory

Enes Pasalic UP FAMNIT študijsko leto 22/23

Lecture topics

- Mathematical analysis:
  - limits and convergence
  - convexity
- Probability:
  - Basics on probability, random variables, distribution,
  - Chain rules and conditional probabilities,
  - Mathematical expectation,
  - Law of large numbers velikih števil
- Nekaj mat. neenakosti
  - Jensen‘s inequality,
  - Gibbs‘ inequality

Exponential function

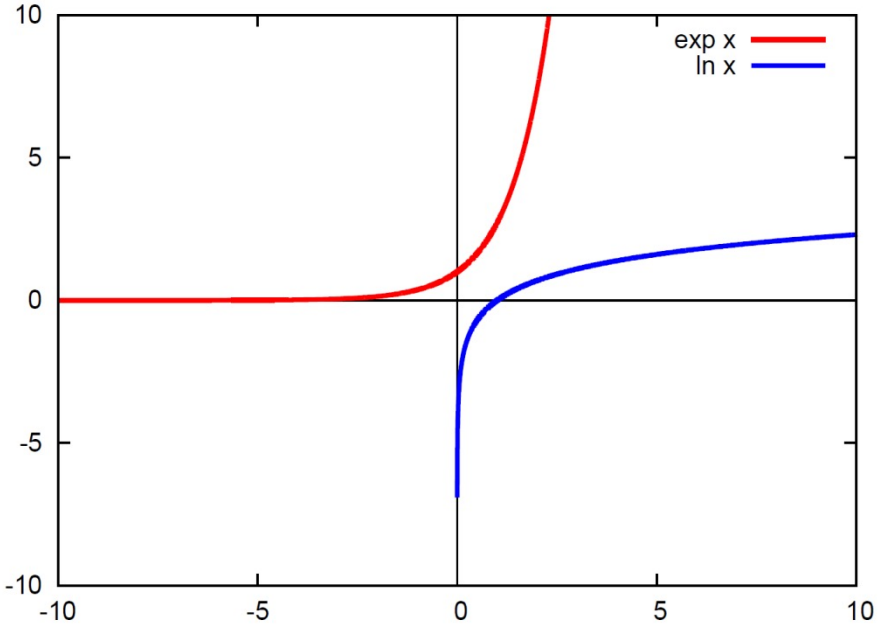


- Exponential function: $\exp : \mathbb{R} \to \mathbb{R}^+, \exp k = e^k = \overbrace{e \times e \times \ldots \times e}^{k}$

$$\exp x \cdot \exp y = \exp(x + y)$$

Odvod: $\frac{d \exp x}{dx} = \exp x$

Logarithm



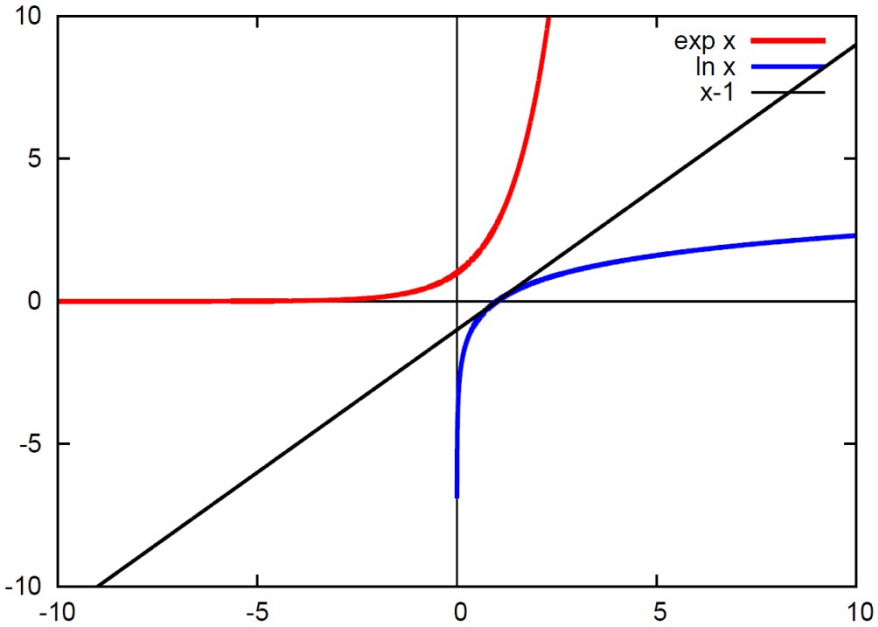
- Natural logarithm: $\ln : \mathbb{R}^+ \to \mathbb{R}, \ln \exp x = x$

$$\log_a a^x = x$$

- Connection for basis:

$$\log_a x = \frac{1}{\ln a} \ln x$$

Properties of logarithm



- Some properties:

$$\ln xy = \ln x + \ln y \qquad \ln x^r = r \ln x \qquad \ln \frac{1}{x} = -\ln x \qquad \ln \frac{x}{y} = \ln x - \ln y$$

- Inequality: $\ln x \leq x - 1$, equality exactly when $x = 1$
  Not valid if $\log_a x \quad a \neq e$
- Derivation:

$$\frac{d \ln x}{dx} = \frac{1}{x}$$

Limits and convergence

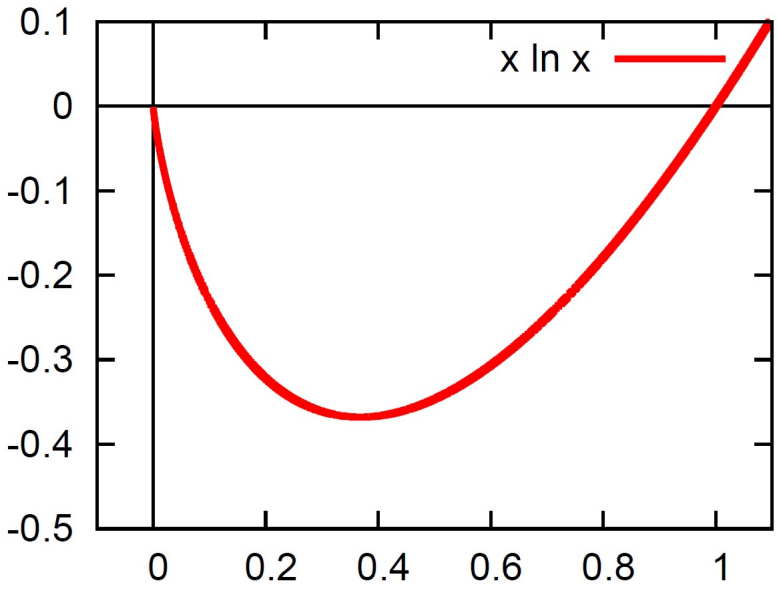
- Zaporedje števil $x_i : i \in \mathbb{N}$ konvergira k limiti L, $\lim_{i\to\infty} x_i = L$, natanko tedaj, ko za poljuben $\varepsilon > 0$ obstaja število $N \in \mathbb{N}$, tako da velja

$$|x_i - L| < \varepsilon \text{ za vse } i \geq N.$$

- Funkcija $f(x)$ kovergira k limiti L, ko gre x proti c, $\lim_{x\to c} f(x) = L$, natanko tedaj, ko za poljuben $\varepsilon > 0$ obstaja $\delta > 0$, tako da velja

$$|f(x) - L| < \varepsilon \text{ za vse } x : c - \delta < x < c + \delta.$$

Example: Limes of xlnx



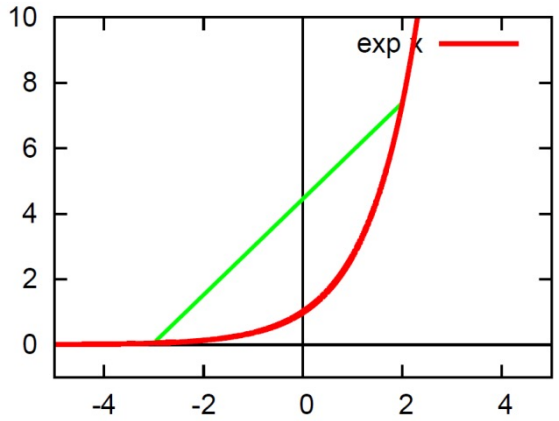
$$\lim_{x \to 0^+} x \ln x = 0$$

L'Hospital's rule

Convexity

- Funkcija $f : \mathcal{X} \to \mathbb{R}$ je **konveksna**, natanko tedaj ko za poljuben $x, y \in \mathcal{X}$ in poljuben $t \in [0, 1]$ velja

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y).$$

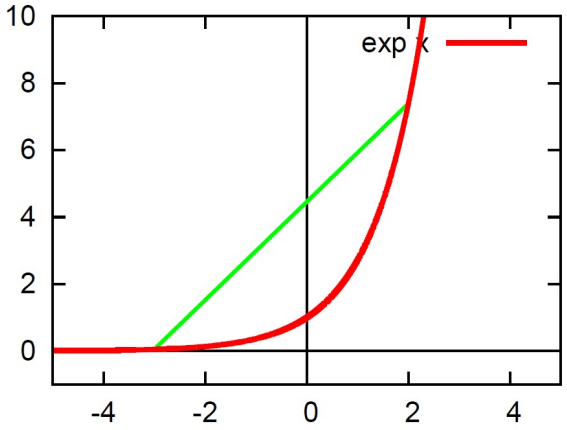


- Funkcija f je **strogo konveksna** natanko tedaj, ko velja stroga neenakost ($<$ namesto $\leq$).
- Funkcija f je (strogo) **konkavna** natanko tedaj ko velja gornja neenakost za $-f$.

Convexity and derivatives

Izrek

Če obstaja drugi odvod funkcije f, f'', in je f'' nenegativen (≥ 0) za vse x, potem je f konveksna. Če je f'' pozitiven (> 0) za vse x je f strogo konveksna.



Primer

$f'(x) = \frac{d \exp(x)}{dx} = \exp(x) \Rightarrow f''(x) = \exp(x) > 0$. Sledi exp je strogo konveksna funkcija.

Probability

- Events and probability spaces
- Basic properties for probability,
- Bayes‘ rules,
- Random variables
- Distribution for random variables,
- Joint and conditional distribution,
- Mathematical expectation,
- Law of large numbers

Probability spaces and outcomes

Probability space $(\Omega, \mathcal{F}, P)$ is defined by:

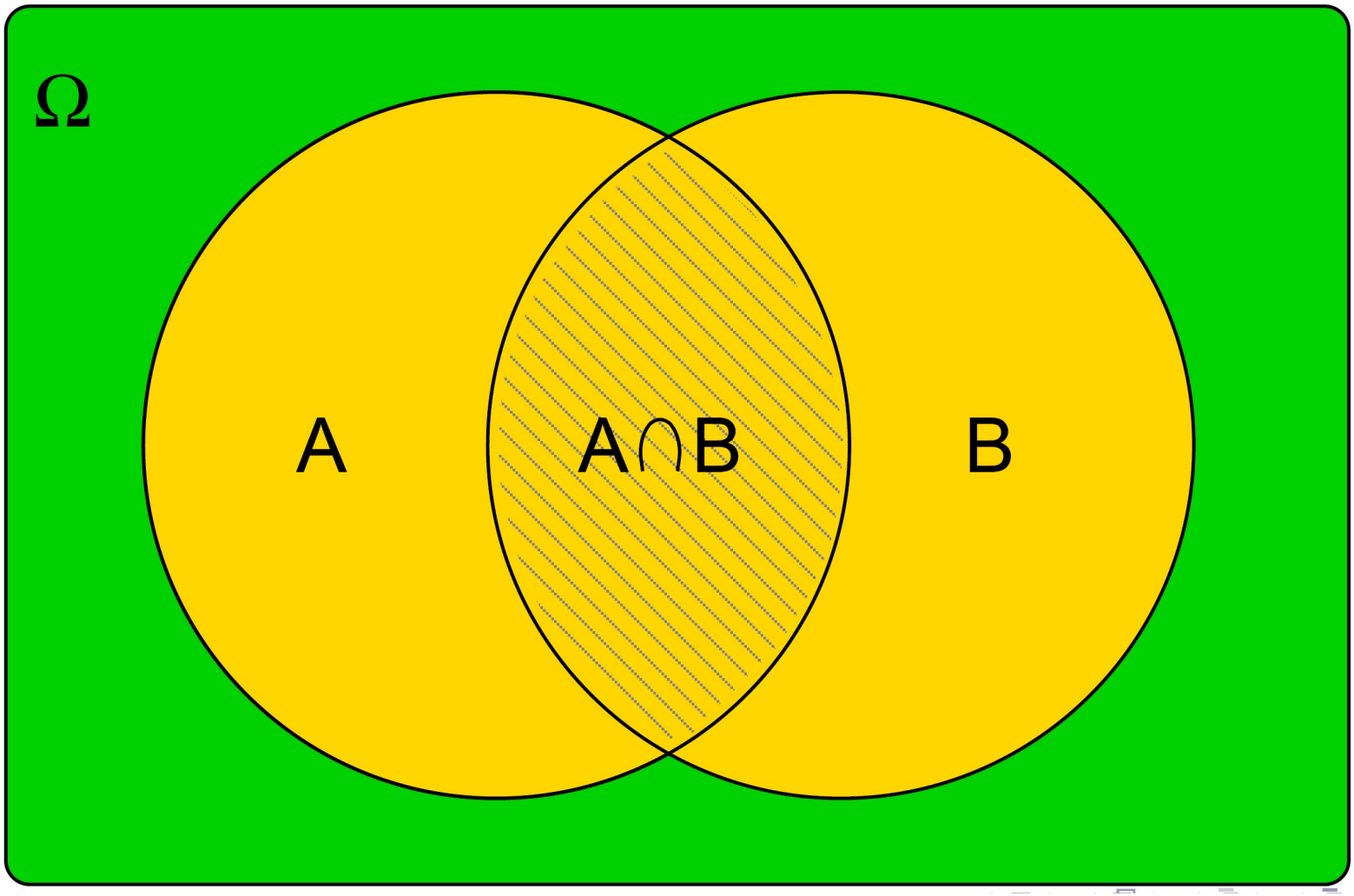
- **Probability space** Ω with elements ω,
- sigma algebra which is a collection of subsets of Ω, with elements E called **events**,
- a measure P, determining the **probability of event**, $P : \mathcal{F} \to [0, 1]$.

P follows the **probability axioms**: $P(E) \geq 0$ for all $E \in \mathcal{F}$, $P(\Omega) = 1$, and $P(E_1 \cup E_2 \cup \ldots) = \sum_i P(E_i)$ when E_i are disjoint events.

From these axioms one can derive e.g.

$$P(A \cup B) = P(A) + P(B) - P(A \cap B),\ P(\Omega / E) = 1 - P(E) \ldots$$

Venn diagram



EXAMPLE *Illustrating the Addition Rule*

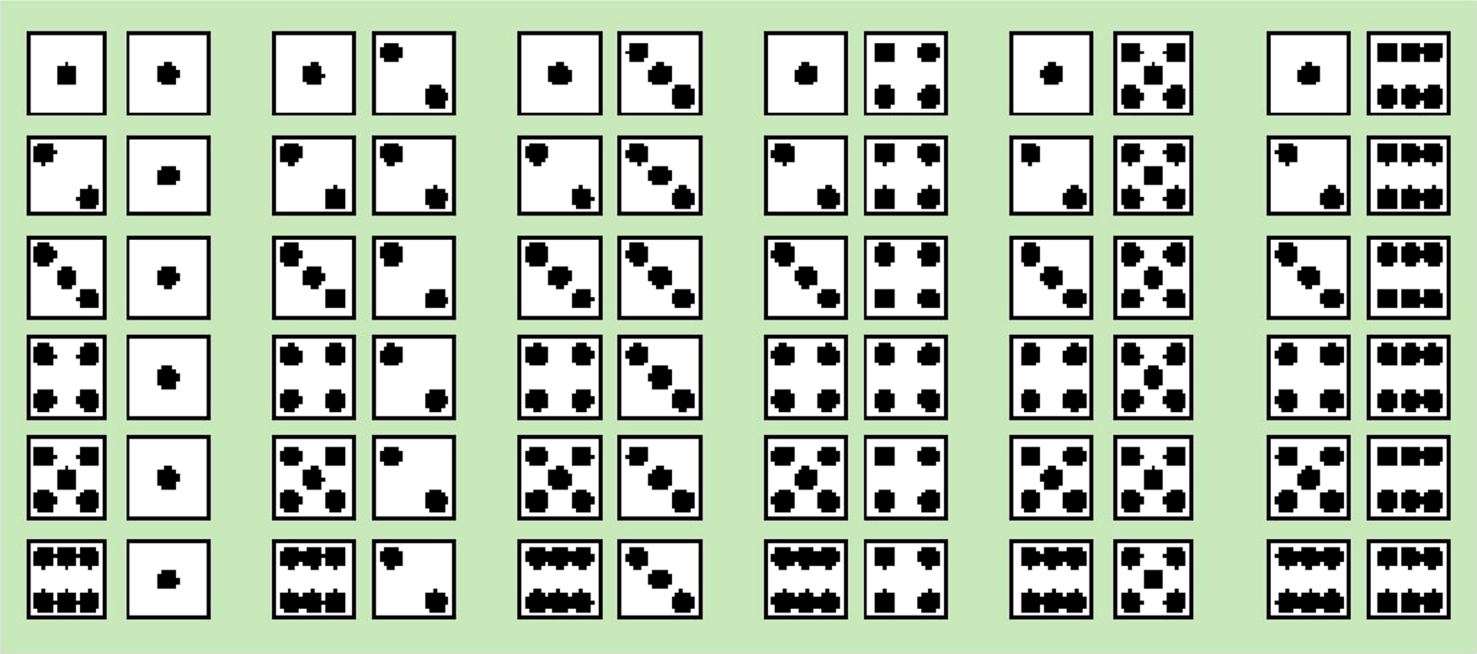
Suppose that a pair of fair dice are thrown.

a) Let E="rolling a seven", compute the probability of rolling a seven, i.e., P(E).

b) Let E="rolling a two " (called 'snake eyes'), compute the probability of rolling "snake eyes", i.e., P(E).

c) Let *E* = "the first dice is a two" and let *F* = "the sum of the dices is less than or equal to 5". Find *P*(*E* or *F*) directly by counting the number of ways *E* or *F* could occur and dividing this result by the number of possible outcomes.

Possible outcomes



Answers

- a) P(E) = N(E)/N(S) = 6/36 = 1/6
- b) 1/6
- c) N(E) = 6, N(F)=4+3+2+1 =10,
- N(E and F) =3 , so N(E or F) =13

Probability – some properties

- **Conditional probability** is probabilty that given that event *A* happened event *B* happens

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} \qquad \text{for A, such that } P(A) > 0.$$

- $P(A \cap B) = P(A) \times P(B \mid A) = P(B) \times P(A \mid B)$.
- **Bayes' rule:** $P(B \mid A) = \frac{P(A \mid B) \times P(B)}{P(A)}$
- Chain rules:

$$P(\cap_{i=1}^{N} E_i) = \prod_{i=1}^{N} P(E_i \mid \cap_{j=1}^{i-1} E_j)$$
$$= P(E_1) \times P(E_2 \mid E_1) \times P(E_3 \mid E_1 \cap E_2) \times \ldots$$
$$\times P(E_N \mid E_1 \cap \ldots \cap E_{N-1})$$

Example - conditional probabilities

- A math teacher gave his class two tests. 25% of the class passed both tests and 42% of the class passed the first test.

  What percent of those who passed the first test also passed the second test?

$$P(A \mid B) = P(Second \mid First) = \frac{P(Second, First)}{P(First)} = \frac{0.25}{0.42} = 0.6$$

Random variables- example

- Standard well-shuffled deck of 52 playing cards. Choosing the first card taken from the deck induce random variable X and second card Y.
- What is a probability that we get two aces assuming that the first card is an ace, i.e.,

p(Y=ace | X=ace)= ?

Direct way: Since the first ace is chosen then we have 3 aces and 51 cards left, p (Y=ace | X=ace) = 3/51

Using conditional probability:

$$p(Y = ace \mid X = ace) = \frac{p(Y = ace, X = ace)}{p(X = ace)} = \frac{\frac{2 \cdot \binom{4}{2}}{52 \cdot 51}}{\frac{4}{52}} = \frac{3}{51}$$

Example: Bayes‘ rule

- Quiz with three doors (Monty Hall problem):
  - 3 doors and behind one of these is award. The doors are denoted by 1, 2 in 3.
    Competitor selects one door.
    Quizmaster opens one door, not chosen by competitor. However, no award behind the opened door.

    Example: if competitor selects door 1, then quizmasterpotem opens either door 2 or 3, but not the one hiding the award.

    Quizmaster offers to the competitor to change his initial choice.

    Competitor can keep the initial choice or select other door which has not been opened yet.

  - What is better:
    a. Stay with the initial choice
    b. Change the door
    c. Does not matter

Random variables

- Random variable X is a function mapping Ω to real numbers, i.e. $X : \Omega \to \mathbb{R}$
- Distribution of X is determined (in general) by probability P so that

$$P_X(A) = Pr[X \in A] = Pr(\omega : X(\omega) \in A), \quad A \subseteq \mathbb{R}$$

- BUT commonly we assign $X(\omega)$ numbers, letters colors etc.

Random variables: discrete, continuous

- We can represent distribution of a random variable X using *cumulative probability function* so that $F_X(x) = Pr(X \leq x)$
- **Discrete random variable**:
  - is a variable X which takes values in some finite alphabet $\mathcal{X}$
  - In this case we define the probability mass function (pmf) simply as p_X where $p_X(x) = Pr(X = x)$.
- **Continuous random variable**:
  - is a variable Y where we define *probability density function* (pdf) f_Y as

$$Pr(Y \in A) = \int_A f_Y(y)dy.$$

- There are also random variables which are mixture of those above but we do not treat these.

Functions using random variables

Since random variables are also functions we can define other functions using random variables.

Example: For instance, taking a function f and random variables X and Y then $f(X) : \Omega \to \mathbb{R}$ and $X + Y$ are also random variables.

Example

Let X represent the number obtained when casting a dice.

- pmf of X is $p_X(x) = 1/6$ for all $x \in \{1, 2, 3, 4, 5, 6\}$
- pmf of X^2 is $p_{X^2} = 1/6$ for all $x \in \{1, 4, 9, 16, 25, 36\}$

Notice

Since pmf p_X is also a function, then $p_X(x)$ is also a random variable and also $p_X^2(x)$ or $lnp_X(x)$

Cummulative function of continous RV

Suppose $U = X^2$, then

$$\begin{aligned} F_U(u) &= P(U \leq u) \\ &= P(X^2 \leq u) \\ &= P(-\sqrt{u} \leq X \leq \sqrt{u}) \\ &= \int_{-\sqrt{u}}^{\sqrt{u}} f(x)dx \\ &= F_X(\sqrt{u}) - F_X(-\sqrt{u}). \end{aligned}$$

To find $f_U(u)$ we need to differentiate $F_U(u)$ over *u*

$$\begin{aligned} f_U(u) &= f_X(\sqrt{u})\left(\frac{1}{2\sqrt{u}}\right) + f_X(-\sqrt{u})\left(\frac{1}{2\sqrt{u}}\right) \\ &= \frac{1}{2\sqrt{u}}\left[f_X(\sqrt{u}) + f_X(-\sqrt{u})\right]. \end{aligned}$$

Joint and marginal distributions

- **Joint distribution** for random variables X and Y is defined as

$$P_{X,Y} = Pr[X \in A \wedge Y \in B] = P(\{\omega : X(\omega) \in A, Y(\omega) \in B\})$$

- For any joint distribution $P_{X,Y}$ there is a **marginal distributions** P_X and P_Y defined as

$$P_X(A) = P(A, \mathbb{R}), \quad P_Y(B) = P(\mathbb{R}, B)$$

$$pmf: \ p_Y(y) = \sum_{x \in \mathcal{X}} p_{X,Y}(x, y) \quad pdf: \ f_Y(y) = \int_{\mathbb{R}} f_{X,Y}(x, y) dx$$

Conditional distributions

Pogojna porazdelitev je definirana podobno kot pogojna verjetnost

$$P_{Y|X}(B|A) = \frac{P_{X,Y}(A,B)}{P_X(A)} \quad \text{za } A, \text{ za katerega velja } P_X(A) > 0.$$

Za diskretne/zvezne naključne spr. to pomeni

- diskretne n.spr.:

$$p_{Y|X}(y|x) = \frac{p_{X,Y}(x,y)}{p_X(x)}, \quad p_X(x) > 0,$$

- zvezne n.spr.:

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}, \quad f_X(x) > 0.$$

Independent random variables

Random variable X is independent from Y (denoted as $X \perp Y$) if we have

$$P_{X,Y}(A, B) = P_X(A) \cdot P_Y(B) \quad \text{for any } A, B \subseteq \mathbb{R}$$

This is equivalent to :

$$P_{X|Y}(A|B) = P_{X|Y}(A) \text{ for any } B, \text{ with } P(B) > 0$$

Similarly:

$$P_{Y|X}(B|A) = P_{Y|X}(B) \text{ for any } A, \text{ with } P(A) > 0$$

Conclusion: The knowledge about one random variable does not tell us anything about the other one. Due to symmetry $X \perp Y \Leftrightarrow Y \perp X$

Mathematical expectation

Matematično upanje diskretne naključne spremenljivke X je definirano kot

$$E[X] = \sum_{x \in \mathcal{X}} p(x)x.$$

Matematično upanje zvezne naključne spremenljivke X je definirano kot

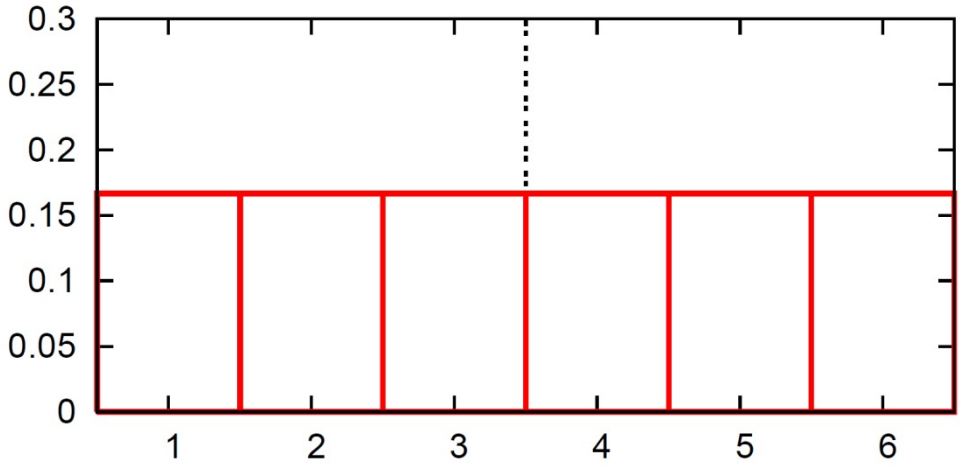
$$E[X] = \int_{\mathbb{R}} f(x)x dx.$$

Lastnosti:

- $E[X]$ je lahko tudi $\pm\infty$
- Linearnost: $E[kX] = kE[X]$, $E[X + Y] = E[X] + E[Y]$.
- Neodvisnost: $E[XY] = E[X]E[Y]$, če $X \perp Y$.

Law of large numbers

Let $X_1, X_2, \ldots$ be a sequence of independent random variables corresponding to casting a dice. We have $p_{X_i}(x) = 1/6$ for all X_i, $i \in \mathbb{N}$, with $x \in \{1, 2, 3, 4, 5, 6\}$.



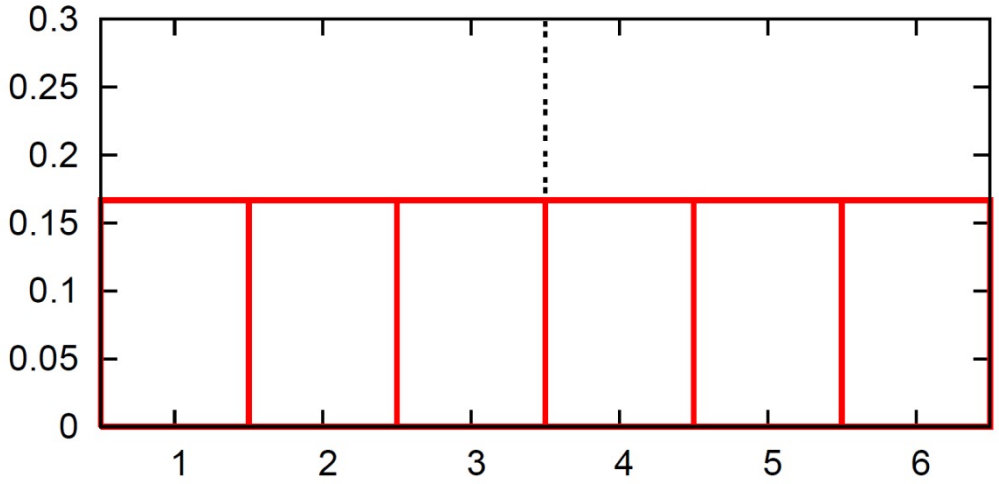
$$E[X_i] = \sum_{x=1}^{6} \frac{1}{6} x = \frac{21}{6} = 3.5 \quad \text{za vsak } i \in \mathbb{N}$$

Law of large numbers

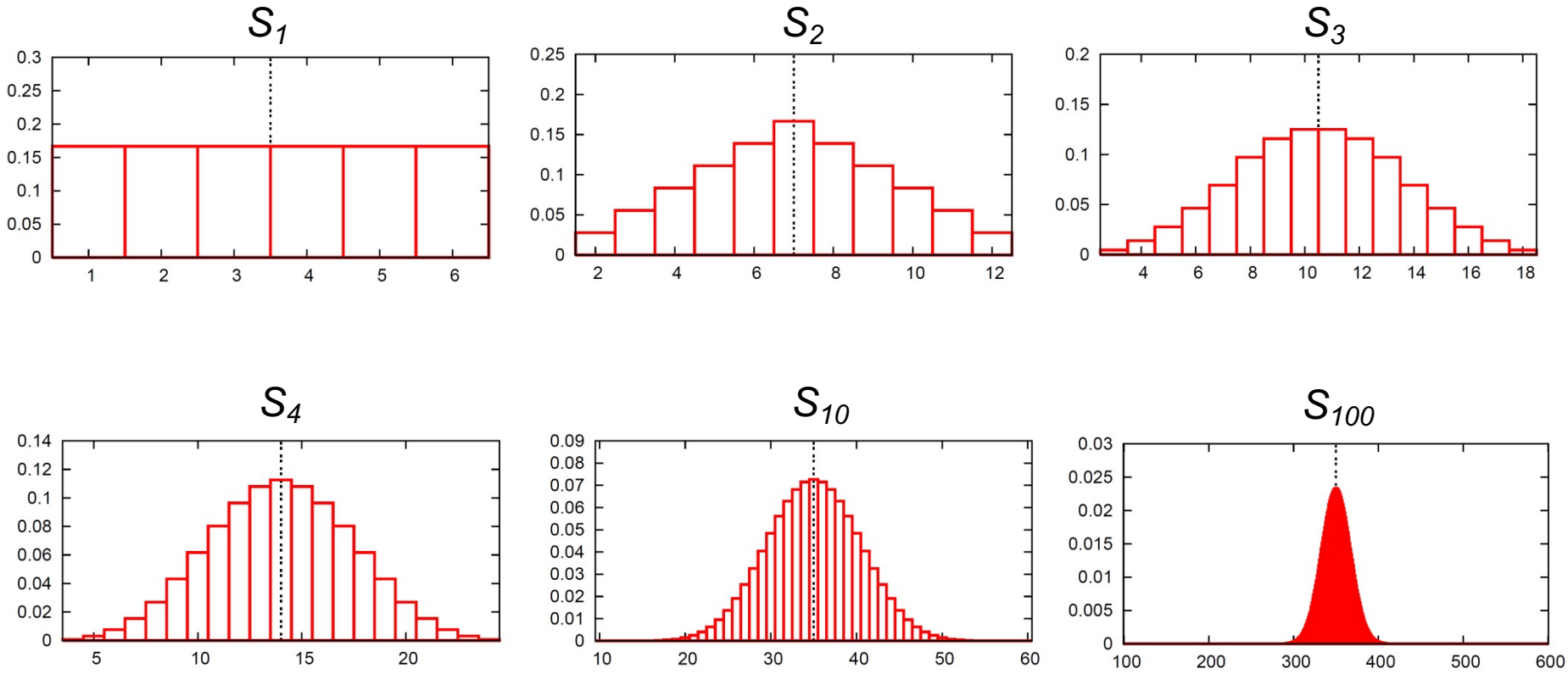
Let $S_n = \sum_{i=1}^{n} X_i$ be the sum of dots after casting a dice n times.
Distribution of a random variable S_n can be computed using

$$P_{S_n}(x) = \frac{\text{number of ways of getting sum } \mathbf{x} \text{ after } \mathbf{n} \text{ throws}}{6^n}$$

distribution S_1



Law of large numbers



Law of large numbers

Theorem

If we have a sequence of Independent and Identically Distributed (IID) Random Variables (RV) with mathematical expectation μ, then the quantity $\frac{1}{n}S_n$ converges to μ:

$$\lim_{n\to\infty} Pr\left[\left|\frac{S_n}{n} - \mu\right| < \epsilon\right] = 1 \text{ for any } \epsilon > 0.$$

This will be useful in proving AEP later, one of the most important results in information theory

Inequality

- Jensen' inequality
- Gibbs' inequality

Jensen's inequality

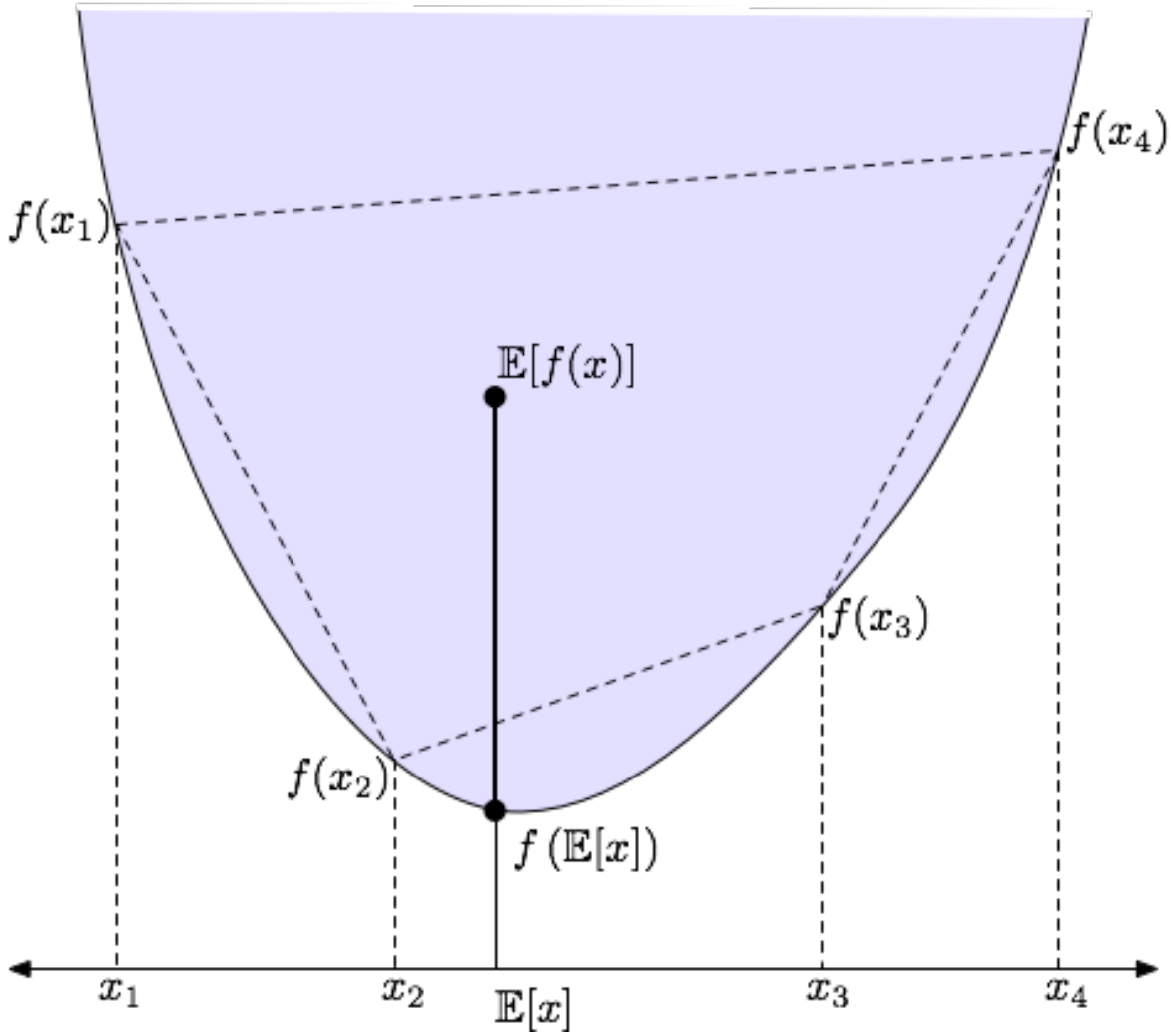
Jensen's inequality

if f is a convex function and X a random variable then we have

$$E[f(X)] \geq f(E[X]).$$

If f is strictly convex then the equality is obtained iff $X = E(X)$ with probability 1 (thus X is a constant).

Jensen's inequality: meaning?



Jensen's inequality: example

- Whiteboard (?):
  - Average surface of 3 squares equals to $A = 100m^2$.
    Average length of the squares' side is $l = 10$ m.
  - What can we tell about the surface of the largest square?
  - Answer (MacKay book, page 35). Hint: What is interesting function f(x) here ? Is it convex, strictly ...?

Gibbs‘ inequality

Gibbs' inequality

For any discrete distributions p and q the following inequality is valid

$$\sum_{x \in \mathcal{X}} p(x) \ln p(x) \geq \sum_{x \in \mathcal{X}} p(x) \ln q(x).$$

The equality happens IFF $p(x) = q(x)$ for any $x \in \mathcal{X}$.