Repetition
Probability Theory

Algorithm Theory
WS 2018/19

Fabian Kuhn

Randomized Algorithms

Randomized Algorithms

- An algorithm that uses (or can use) random coin flips in order to make decisions
- randomization can be a powerful tool to make algorithms faster or simpler

First: Short Repetition of Basic Probability Theory

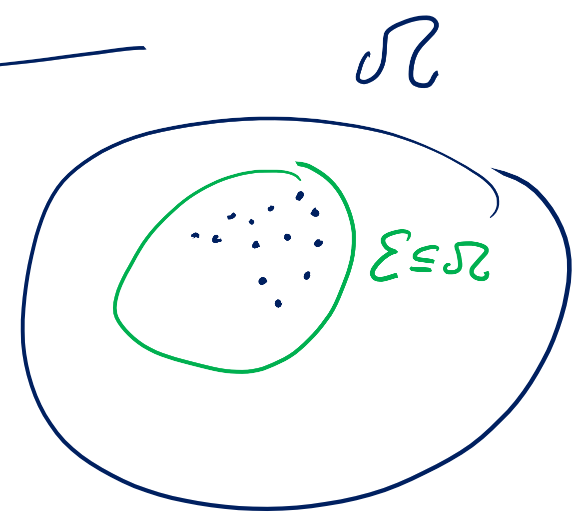
- We need: basic discrete probability theory
  - probability spaces, probability events, independence, random variables, expectation, linearity of expectation, Markov inequality
- Literature, for example
  - your old probability theory book / lecture notes / ...
  - Appendix C of book of Cormen, Rivest, Leiserson, Stein
  - http://www.ti.inf.ethz.ch/ew/courses/APC15/material/ra.pdf

Probability Space and Events

Definition: A (discrete) **probability space** is a pair $(\Omega, \mathbb{P})$, where

- Ω: (countable) set of elementary events
- $\mathbb{P}$: assigns a probability to each $\omega \in \Omega$

$$\mathbb{P} : \Omega \to \mathbb{R}_{\geq 0} \quad \text{s.t.} \quad \sum_{\omega \in \Omega} \mathbb{P}(\omega) = 1$$



Definition: An **event** $\mathcal{E}$ is a subset of Ω

- Event $\mathcal{E} \subseteq \Omega$: set of basic events
- Probability of $\mathcal{E}$

$$\mathbb{P}(\mathcal{E}) := \sum_{\omega \in \mathcal{E}} \mathbb{P}(\omega)$$

Example: Probability Space, Events

flip (biased) coin $\longrightarrow$ $\{H, T\}$

prob. to get $\underline{\underline{H}}$ is equal to $\underline{p}$

__experiment:__ flip coins until we get H

$$\Omega = \{\underset{e_0}{H}, \underset{e_1}{TH}, \underset{e_2}{TTH}, \dots, \underset{e_\infty}{\underbrace{TT\dots T}_{\infty}}\}$$

$$\underline{\mathbb{P}(e_i) = (1-p)^i \cdot p}$$

$$\sum_{i=0}^{\infty} \mathbb{P}(e_i) = p \underbrace{\sum_{i=0}^{\infty} (1-p)^i}_{\frac{1}{1-(1-p)}} = p \cdot \frac{1}{p} = 1$$

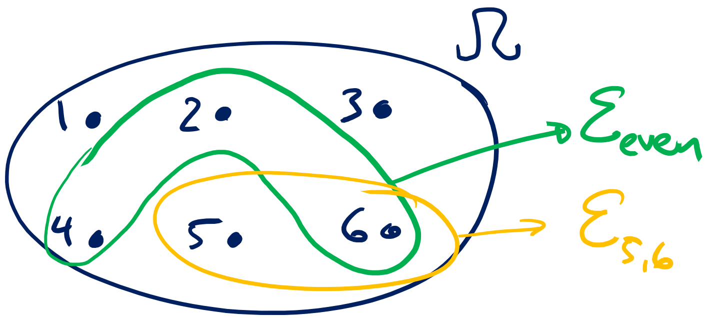
$$\mathcal{E} = \{e_i \mid i \text{ is even}\}$$

$$\mathbb{P}(\mathcal{E}) = \sum_{e_i \in \mathcal{E}} \mathbb{P}(e_i) = \sum_{j=0}^{\infty} \mathbb{P}(e_{2j}) = p \cdot \sum_{j=0}^{\infty} \underbrace{(1-p)^{2j}}_{((1-p)^2)^j} = p \cdot \frac{1}{\underbrace{1-(1-p)^2}_{\underbrace{1-(1-2p+p^2)}_{2p-p^2}}} = \frac{p}{2p-p^2} = \underline{\underline{\frac{1}{2-p}}}$$

Example: Probability Space, Events

roll a die

$\Omega = \{1,2,3,4,5,6\}$, $\mathbb{P}(1) = \mathbb{P}(2) = \ldots = \mathbb{P}(6) = 1/6$



$\mathbb{P}(\mathcal{E}_{even}) = 1/2$

$\mathbb{P}(\mathcal{E}_{5,6}) = 1/3$

roll 2 dice

$\Omega = \{(1,1), (1,2), \ldots, (1,6), (2,1), \ldots, (6,6)\}$ $\quad \mathbb{P}((i,j)) = \frac{1}{36}$

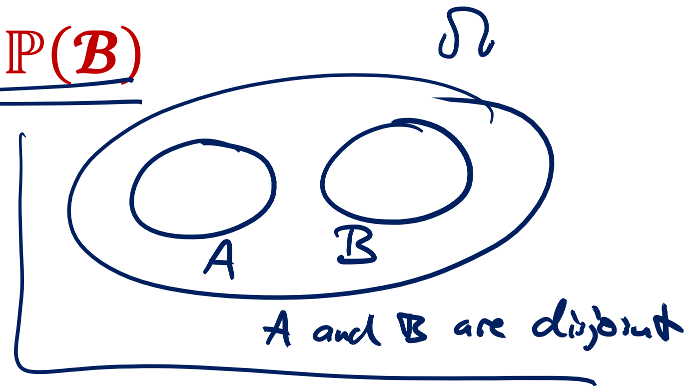
$A_{=} = \{(1,1), (2,2), (3,3), \ldots, (6,6)\}$ $\quad \mathbb{P}(A_{=}) = 6 \cdot \frac{1}{36} = \frac{1}{6}$

$A_{\neq} = \Omega \setminus A_{=} = \overline{A_{=}}$ $\quad \mathbb{P}(\overline{A_{=}}) = 1 - \mathbb{P}(A_{=}) = \frac{5}{6}$

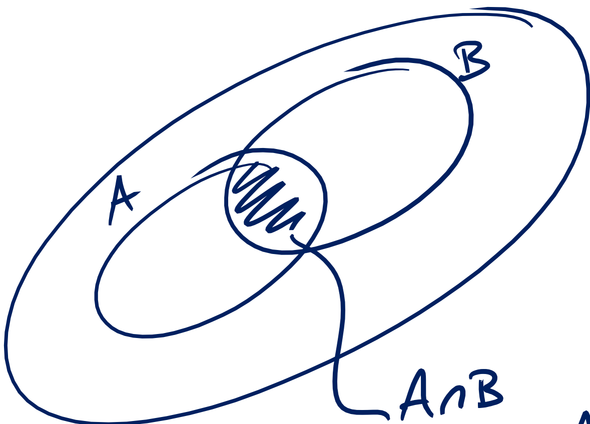
Independent Events

Definition: Events $\mathcal{A} \subseteq \Omega$ and $\mathcal{B} \subseteq \Omega$ are **independent** iff

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



Example: $\Omega = \Omega_1 \times \Omega_2$



roll 2 dice

A: first die even $\quad \mathbb{P}(A) = 1/2$

B: second die is odd $\quad \mathbb{P}(B) = 1/2$

$A \cap B = \{(2,1), (2,3), (2,5), (4,1), \dots, (6,5)\}$

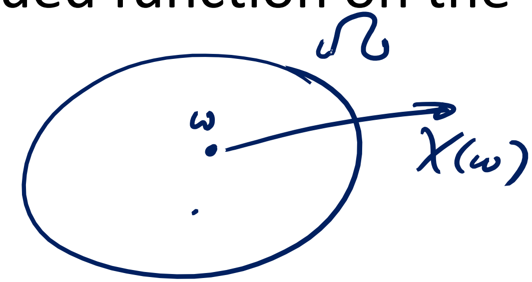
$|A \cap B| = 9 \quad \mathbb{P}(A \cap B) = \frac{9}{36} = \frac{1}{4}$

$= \mathbb{P}(A) \cdot \mathbb{P}(B)$

Random Variables

Definition: A **random variable X** is a real-valued function on the elementary events Ω

$$\boldsymbol{X} : \boldsymbol{\Omega} \to \mathbb{R}$$



- We usually write X instead of $X(\omega)$
- We also write

$$\mathbb{P}(X = x) = \mathbb{P}(\{\omega \in \Omega : X(\omega) = x\})$$

Examples:

- $\boldsymbol{X^{top}}: X^{top}(1) = 1, X^{top}(2) = 2, \dots, X^{top}(6) = 6$
- $\boldsymbol{X^{bot}}: X^{bot}(1) = 6, X^{bot}(2) = 5, \dots, X^{bot}(6) = 1$

- Note that for all $\omega \in \Omega$, $X^{top}(\omega) + X^{bot}(\omega) = 7$
- To denote this, we write $X^{top} + X^{bot} = 7$

Indicator Random Variables

A random variable with only takes values 0 and 1 is called a **Bernoulli random variable** or an **indicator random variable**.

roll a die, rand. var. $Y = \begin{cases} 1 & \text{if odd} \\ 0 & \text{if even} \end{cases}$

$Y(1) = 1,\ Y(2) = 0,\ Y(3) = 1,\ \ldots$

$\mathbb{P}(\underbrace{Y=0}_{\mathcal{E}_{even}}) = \frac{1}{2}$

Independent Random Variables

Definition: Two random variables X and Y are called **independent** if

$$\forall x, y \in \mathbb{R}: \ \mathbb{P}(X = x \wedge Y = y) = \mathbb{P}(X = x) \cdot \mathbb{P}(Y = y)$$

two coin flips (fair coin) Bernoulli r.v. X, Y

$X = 1 \iff$ 1st coin flip is H $\quad \mathbb{P}(X=1) = 1/2$

$Y = 1 \iff$ exactly one coin flip is H $\quad \mathbb{P}(Y=1) = 1/2$

prob. space $\Omega = \{(T,T), (T,H), (H,T), (H,H)\}$

$$\mathbb{P}(X=0 \wedge Y=0) = \mathbb{P}((T,T)) = \frac{1}{4}$$

$$\mathbb{P}(X=0 \wedge Y=1) = \mathbb{P}((T,H)) = \frac{1}{4}$$

$$\mathbb{P}(X=1 \wedge Y=0) = \mathbb{P}((H,H)) = \frac{1}{4}$$

$$\mathbb{P}(X=1 \wedge Y=1) = \mathbb{P}((H,T)) = \frac{1}{4}$$

Independent Random Variables

Definition: A collection of andom variables $X_1, X_2, \ldots, X_n$ on a probability space Ω is called **mutually independent** if

$$\forall k \geq 2, 1 \leq i_1 < \cdots < i_k \leq n, \forall x_{i_1}, \ldots, x_{i_k} \in \mathbb{R}:$$
$$\mathbb{P}\left(X_{i_1} = x_{i_1} \wedge \cdots \wedge X_{i_k} = x_{i_k}\right) = \mathbb{P}\left(X_{i_1} = x_{i_1}\right) \cdot \ldots \cdot \mathbb{P}\left(X_{i_k} = x_{i_k}\right)$$

not the same as pairwise independence

example: 2 coin flips X_1, X_2, X_3 Bernoulli r.v.

$X_1 = 1 \iff$ 1st flip is H

$X_2 = 1 \iff$ 2nd flip is H

$X_3 = 1 \iff$ exactly one H

$$\mathbb{P}(X_1 = 1 \wedge X_2 = 1 \wedge X_3 = 1) = 0$$

Expectation

Definition: The **expectation** of a random variable X is defined as

$$\mathbb{E}[X] := \sum_{x \in X(\Omega)} x \cdot \mathbb{P}(X = x) = \sum_{\omega \in \Omega} X(\omega) \cdot \mathbb{P}(\omega)$$

Example:

- recall: X^{top} is outcome of rolling a die

$$\mathbb{E}[X^{top}] = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3.5$$

$$\mathbb{E}[(X^{top})^2] = 1 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 9 \cdot \frac{1}{6} + 16 \cdot \frac{1}{6} + 25 \frac{1}{6} + 36 \cdot \frac{1}{6} = \frac{91}{6} = 15.16\ldots$$

Remark

$$\mathbb{E}[X^2] \neq \mathbb{E}[X]^2$$

$$\mathbb{E}[X^{top} \cdot X^{bot}] = \frac{1}{6} \cdot (1 \cdot 6 + 2 \cdot 5 + 3 \cdot 4 + 4 \cdot 3 + 5 \cdot 2 + 6 \cdot 1) = \frac{56}{6} = 9.33\ldots$$

$$\mathbb{E}[X \cdot Y] \neq \mathbb{E}[X] \cdot \mathbb{E}[Y]$$

Expectation: Examples

Sums and Products of Random Variables

Linearity of Expectation:

For random variables X and Y and any $c \in \mathbb{R}$, we have

$$\begin{aligned} \mathbb{E}[cX] &= c \cdot \mathbb{E}[X] \\ \mathbb{E}[X+Y] &= \mathbb{E}[X] + \mathbb{E}[Y] \end{aligned}$$

- holds also if the random variables are not independent

Product of Random Variables:

For two **independent** random variables X and Y, we have

$$\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$$

Linearity of Expectation:

For random variables X and Y and any $c \in \mathbb{R}$, we have

$$\mathbb{E}[cX] = c \cdot \mathbb{E}[X], \quad \boxed{\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]}$$

$$\underline{\mathbb{E}[X+Y]} = \sum_{x,y} (x+y) \cdot \mathbb{P}(X=x \wedge Y=y) \quad \text{(def. of exp.)}$$

$$= \sum_x \sum_y x \cdot \mathbb{P}(X=x, Y=y) + \sum_y \sum_x y \, \mathbb{P}(X=x, Y=y)$$

$$= \underbrace{\sum_x x \cdot \underbrace{\sum_y \mathbb{P}(X=x \wedge Y=y)}_{=\mathbb{P}(X=x)}}_{=\mathbb{E}[X]} + \underbrace{\sum_y y \cdot \underbrace{\sum_x \mathbb{P}(X=x \wedge Y=y)}_{=\mathbb{P}(Y=y)}}_{=\mathbb{E}[Y]}$$

Sums and Products of Random Variables

Product of Random Variables:

For two **independent** random variables X and Y, we have

$$\mathbb{E}[\boldsymbol{X} \cdot \boldsymbol{Y}] = \mathbb{E}[\boldsymbol{X}] \cdot \mathbb{E}[\boldsymbol{Y}]$$

$$\mathbb{E}[X \cdot Y] = \sum_x \sum_y x \cdot y \cdot \underline{\mathbb{P}(X=x) \cdot \mathbb{P}(Y=y)}$$

$$= \sum_x x \cdot \mathbb{P}(X=x) \cdot \underbrace{\sum_y y \cdot \mathbb{P}(Y=y)}_{\mathbb{E}[Y]}$$

$$= \mathbb{E}[Y] \cdot \underbrace{\sum_x x \mathbb{P}(X=x)}_{\mathbb{E}[X]}$$

Linearity of Expectation: Example

$\mathbb{P}(H) = p$

Sequence of coin flips: $C_1, C_2, \ldots \in \{H, T\}$

$\{e_0, e_1, e_2, \ldots\}$

- Stop as soon as the first H turns up

Random variable X: number of T before first H

$\mathbb{P}(X = i) = (1-p)^i \cdot p$

$$\mathbb{E}[X] = \sum_{i=0}^{\infty} i \cdot p \cdot (1-p)^i = \frac{1-p}{p}$$

Indicator random variable X_i ($i \geq 1$):

- $X_i = 1$: i^{th} coin flip happens and its outcome is T
  $X_i = 0$: otherwise

$$X = X_1 + X_2 + X_3 + \ldots + X_\infty$$

$\mathbb{P}(X_i = 1) = (1-p)^i \qquad \mathbb{E}[X_i] = (1-p)^i$

$$\mathbb{E}[X] = \mathbb{E}[X_1 + \ldots + X_\infty]$$

$$\overset{\text{lin. of exp.}}{=} \sum_{i=1}^{\infty} \mathbb{E}[X_i] = \sum_{i=1}^{\infty} (1-p)^i = (1-p)\frac{1}{1-(1-p)} = \frac{1-p}{p}$$

Markov's Inequality

Lemma: Let X be a nonnegative random variable.
Then for all $c > 0$

$$\mathbb{P}(X \geq c \cdot \mathbb{E}[X]) \leq \frac{1}{c}$$

$$\mathrm{Var}(X) := \mathbb{E}\big[\underbrace{(X - \mathbb{E}[X])^2}_{Z \geq 0}\big] \qquad \mathbb{P}\big(Z \geq c^2 \cdot \underbrace{\mathbb{E}[Z]}_{\mathrm{Var}(X)}\big) \leq \frac{1}{c^2}$$

$$\mathbb{P}\big((X - \mathbb{E}[X])^2 \geq c^2 \cdot \mathrm{Var}(X)\big) \leq \frac{1}{c^2}$$

$$\boxed{\mathbb{P}\big(|X - \mathbb{E}[X]| \geq c \cdot \sigma(X)\big) \leq \frac{1}{c^2}}$$

Chebyshev's inequality

$\sigma(X) := \sqrt{\mathrm{Var}(X)}$

std. deviation

Conditional Probabilities

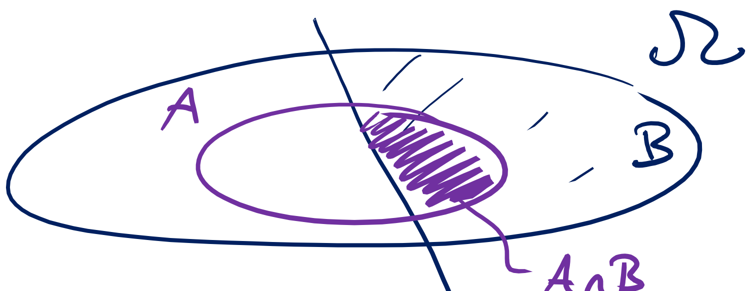
For events $\mathcal{A} \subseteq \Omega$ and $\mathcal{B} \subseteq \Omega$, the **conditional probability** of $\mathcal{A}$ given $\mathcal{B}$ is defined as

$$\mathbb{P}(\boldsymbol{\mathcal{A}}|\boldsymbol{\mathcal{B}}) := \frac{\mathbb{P}(\boldsymbol{\mathcal{A}} \cap \boldsymbol{\mathcal{B}})}{\mathbb{P}(\boldsymbol{\mathcal{B}})}$$

Conditioning on event $\mathcal{B}$ defines a new probability space $(\mathcal{B}, \mathbb{P}')$

$$\forall \omega \in B : \ \mathbb{P}'(\omega) = \frac{\mathbb{P}(\omega)}{\mathbb{P}(\mathcal{B})}.$$

Two events are **independent iff** $\mathbb{P}(\boldsymbol{\mathcal{A}}|\boldsymbol{\mathcal{B}}) = \mathbb{P}(\boldsymbol{\mathcal{A}})$

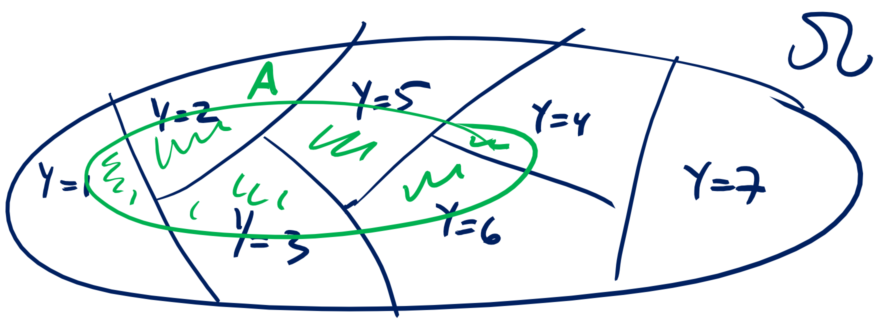


Law of Total Probability / Expectation

Lemma: Let X and Y be two random variables on the same probability space $(\Omega, \mathbb{P})$. We then have

$$\forall x \in \mathbb{R} : \mathbb{P}(\underbrace{X = x}_{A}) = \sum_{y \in Y(\Omega)} \mathbb{P}(\underbrace{X = x}_{A} \mid Y = y) \cdot \mathbb{P}(Y = y).$$

$$\mathbb{E}[X] = \sum_{y \in Y(\Omega)} \mathbb{E}[X \mid Y = y] \cdot \mathbb{P}(Y = y)$$



$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}(A \cap Y=1) \\ &\quad + \mathbb{P}(A \cap Y=2) \\ &\quad \vdots \\ &\quad + \mathbb{P}(A \cap Y=7) \\ &= \mathbb{P}(A \mid Y=1) \cdot \mathbb{P}(Y=1) \\ &\quad + \ldots\end{aligned}$$