Repetition Probability Theory

Algorithm Theory
WS 2015/16

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

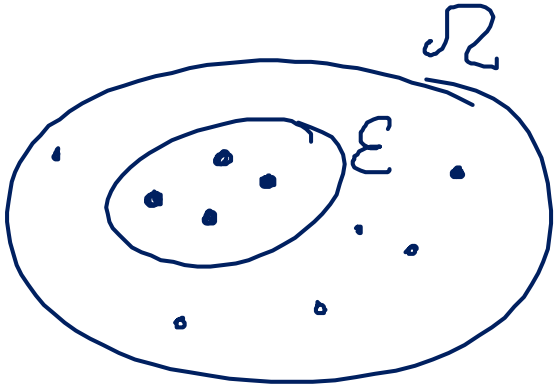
- Ω: set of elementary events
- $\mathbb{P}$: assigns a probability to each $\omega \in \Omega$

$\mathbb{P}(\omega) \geq 0$

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

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

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



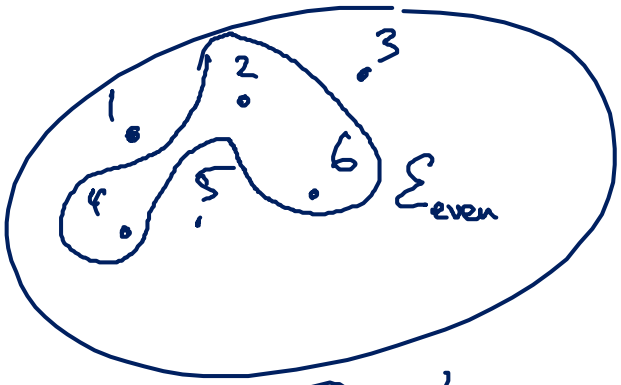
$$\mathbb{P}(\mathcal{E}) := \sum_{\omega \in \mathcal{E}} \mathbb{P}(\omega)$$

Example: Probability Space, Events

roll a die:

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

$\mathbb{P}(\mathcal{E}_{even}) = 3 \cdot \frac{1}{6} = \frac{1}{2}$



roll 2 dice $\Omega = \{(1,1),(1,2),\ldots,(6,6)\}$

$A_= = \{(1,1),(2,2),(3,3),(4,4),(5,5),(6,6)\}$ $\mathbb{P}(A_=) = 6 \cdot \frac{1}{36} = \frac{1}{6}$

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

Example: Probability Space, Events

flip (biased) coin → $\{H, T\}$

prob. to get H is p

experiment: flip coins until we get H

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

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

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

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

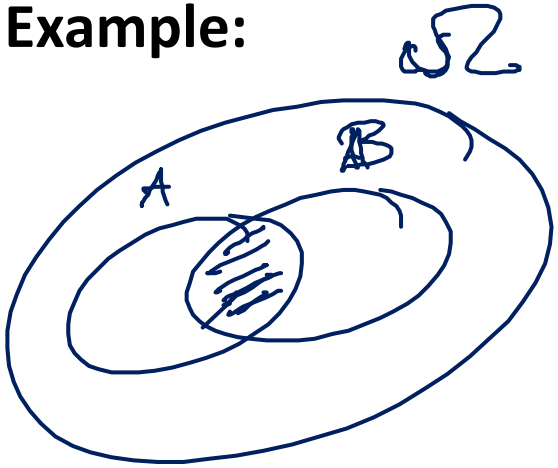
$\mathbb{P}(\mathcal{E}) = \sum_{i=0}^{\infty} \mathbb{P}(e_{2i}) = p \sum_{i=0}^{\infty} \underbrace{(1-p)^{2i}}_{((1-p)^2)^i} = p \frac{1}{1-(1-p)^2} = \frac{1}{2-p}$

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:



roll 2 dice

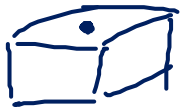
A: first die is even

B: second die is odd

Random Variables

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

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



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

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

Examples: roll a die

- X^{top}: $X^{top}(1) = 1, X^{top}(2) = 2, \dots, X^{top}(6) = 6$
- 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**.

$Y(1) = 0$
$Y(2) = 1$
$Y(3) = 0$
$Y(4) = 1$
$Y(5) = 0$
$Y(6) = 1$

Independent Random Variables

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

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

Independent Random Variables

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

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

not the same as pairwise indep

Example: 2 coin flips $\Omega = \{TT, TH, HT, HH\}$

$X_1 = 1 \iff$ first coin is H

$X_2 = 1 \iff$ second coin is H

$X_3 = 1 \iff$ exactly one coin shows H

$\mathbb{P}(X_1 = 1 \wedge X_3 = 0) = \frac{1}{4}$

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

$\mathbb{P}(\{\omega \mid X_1 = 1\} \cap \{\omega \mid X_2 = 1\} \cap \{\omega \mid X_3 = 1\})$

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

$$\mathbb{E}[X^{top}] = \sum_{i=1}^{6} i \cdot \frac{1}{6} = \frac{21}{6} = 3.5$$

$$\mathbb{E}[X^2] \neq \mathbb{E}[X]^2 \qquad (X^{top}(\omega))^2$$

$$\mathbb{E}[X^{top\,2}] = \sum_{i=1}^{6} i^2 \cdot \frac{1}{6} = \frac{1+4+9+\ldots+36}{6} = \frac{91}{6} = 15.16\ldots$$

$$\mathbb{E}[X^{top} X^{bot}] = \frac{1 \cdot 6 + 2 \cdot 5 + 3 \cdot 4 + \ldots}{6} = \frac{28}{3} = 9.33\ldots$$

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

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

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

- 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: Example

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

- Stop as soon as the first H turns up

H, TH, TTH, ...

Random variable X: number of T before first H

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

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

$$X = X_1 + X_2 + X_3 + \cdots \qquad \mathbb{E}[X_i] = (1-p)^i$$

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

Markov's Inequality

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

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

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$$

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

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

Chebyshev's ineq.

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 $(\Omega \setminus \mathcal{B}, \mathbb{P}')$

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

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 variable on the same probability space $(\Omega, \mathbb{P})$. We then have

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

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

Important Discrete Prob. Distributions

Bernoulli Random Variable $X : \Omega \to \{0, 1\}$

$$\mathbb{P}(X = 1) = p, \mathbb{P}(X = 0) = 1 - p, \qquad \mathbb{E}[X] = p$$

Binomial Random Variable $X \sim \mathbf{Bin}(n, p)$

$$\forall k \in \{0, \dots, n\} : \ \mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad \mathbb{E}[X] = np$$

- measures number of ones in n independent biased coin flip

Geometric Random Variables $X \sim \mathbf{Geom}(p)$

$$\forall k \geq 1 : \ \mathbb{P}(X = k) = p(1-p)^{k-1}, \qquad \mathbb{E}[X] = \frac{1}{p}$$

- measures number independent biased coin flips are necessary to get one “heads”

Happy Holidays!

$X = 0$ w.p. 0.99

$X = 10^{100}$ w.p. 0.01