Randomized Algorithms

Prof. Dr. Christian Scheideler

University of Paderborn, SS 2015

1 Introduction

The aim of the lecture is to give an introduction into the fascinating area of randomized algorithms. In this chapter we will start with an overview of some basic definitions in probability theory and present some simple probabilistic proofs. We will then use these techniques to investigate various simple examples: a randomized identity test, randomized text searching, randomized Quicksort, randomized search trees, a randomized algorithm to compute the smallest enclosing circle, and a randomized distributed algorithm for synchronization.

1.1 Basic definitions in probability theory

Consider an arbitrary discrete random experiment (like throwing a coin), and let $\Omega = \{w_1, w_2, w_3, \ldots\}$ be the *sample space*, i.e., the set of all outcomes of this random experiment.

- An *event* is an arbitrary subset of Ω, and
- event A is *true* for some outcome $w \in \Omega$ if and only if $w \in A$.

The function $p : \Omega \to [0, 1]$ is called a *probability distribution* over the sample space if and only if $\sum_{w\in\Omega} p(w) = 1$. In this case, (Ω, p) forms a *probability space*. p naturally extends to events in a sense that for all events $A \subseteq \Omega$ we define $p(A) = \sum_{w\in A} p(w)$. When p is clear from the context, we will use $\Pr[\cdot]$ instead of $p(\cdot)$. The requirements on a probability space imply the following principle.

Theorem 1.1 (Inclusion-Exclusion Principle) *Let $A_1, \ldots, A_n$ be an arbitrary collection of events. Then it holds that*

$$\Pr[\bigcup_{i=1}^{n} A_i] = \sum_{k=1}^{n} (-1)^{k+1} \sum_{i_1<i_2<\ldots<i_k} \Pr[\bigcap_{j=1}^{k} A_{i_j}]$$

Important special cases of this theorem are the so-called Boole's inequalities:

- $\Pr[\bigcup_{i=1}^{n} A_i] \leq \sum_{i=1}^{n} \Pr[A_i]$
- $\Pr[\bigcup_{i=1}^{n} A_i] \geq \sum_{i=1}^{n} \Pr[A_i] - \sum_{1\leq i<j\leq n} \Pr[A_i \cap A_j]$

Example: 2-coloring of hypergraphs

A *hypergraph* $G = (V, E)$ has the property that every hyperedge $e \in E$ can be an arbitrary subset of V. A *2-coloring* of G is an arbitrary mapping $f : V \to \{0, 1\}$ that assigns one out of two colors (here, 0 and 1) to every node. Which class of hypergraphs allows a 2-coloring so that no hyperedge contains only nodes of one color, i.e., it is *monochromatic*?

Theorem 1.2 *For every hypergraph G with m hyperedges of size $\geq \log m + 2$ there is a 2-coloring so that no hyperedge is monochromatic.*

Proof. Let $n = |V|$ and $m = |E|$. Suppose that we use a random experiment in which we have a *uniform* probability distribution on the set F of mappings $f : V \to \{0,1\}$, i.e., $\Omega = F$ with $\Pr[f] = 1/2^n$ for all $f \in \Omega$. For each $i \in \{1,\ldots,m\}$ let A_i be the event (i.e., the set of all mappings $f \in F$ with the property) that the hyperedge e_i is monochromatic. Since there are exactly $2 \cdot 2^{n-|e_i|}$ many mappings f with e_i being monochromatic, $|A_i| = 2 \cdot 2^{n-|e_i|}$ and therefore,

$$\Pr[A_i] = \sum_{f \in A_i} \Pr[f] = 2 \cdot 2^{n-|e_i|} \cdot 2^{-n} = 2^{-|e_i|+1}$$

Moreover, it follows from Boole's inequalities that

$$\Pr[A_1 \cup \ldots \cup A_m] \leq \sum_{i=1}^{m} \Pr[A_i] \leq \sum_{i=1}^{m} 2^{-(\log m+1)} = \frac{1}{2}$$

Thus,

$$\Pr[\bar{A}_1 \cap \ldots \cap \bar{A}_m] = 1 - \Pr[A_1 \cup \ldots \cup A_m] \geq \frac{1}{2}$$

which implies that there exists a 2-coloring for G so that no hyperedge is monochromatic. □

How can we use such a probabilistic proof in order to quickly find such a 2-coloring? This turns out to be relatively easy in this case and will be discussed in the tutorials.

Conditional probability

The *conditional probability* that the event B is true under the assumption that A is true is given by

$$\Pr[B \mid A] = \frac{\Pr[A \cap B]}{\Pr[A]}$$

From this it follows that

$$\Pr[A \cap B] = \Pr[A] \cdot \Pr[B \mid A]$$

and, in general,

$$\Pr[A_1 \cap \ldots \cap A_n] = \prod_{i=1}^{n} \Pr[A_i \mid A_1 \cap \ldots \cap A_{i-1}]$$

Since

$$\Pr[A \cap B] = \Pr[A] \cdot \Pr[B \mid A] = \Pr[B] \cdot \Pr[A \mid B]$$

we obtain Bayes' formula:

$$\Pr[A \mid B] = \frac{\Pr[A] \cdot \Pr[B \mid A]}{\Pr[B]}$$

Two events A and B are

- *independent* if $\Pr[B \mid A] = \Pr[B]$,
- *negatively correlated* if $\Pr[B \mid A] \leq \Pr[B]$, and
- *positively correlated* if $\Pr[B \mid A] \geq \Pr[B]$.

According to Bayes' formula these properties are symmetric. Hence, for independent events, $\Pr[A \cap B] = \Pr[A] \cdot \Pr[B]$.

Suppose that the sample space Ω can be represented as $\Omega = \Omega_1 \times \ldots \times \Omega_k$ with probability distributions $p_1 : \Omega_1 \to [0,1], \ldots, p_k : \Omega_k \to [0,1]$ so that for each outcome $w = (w_1,\ldots,w_k) \in \Omega$ it holds that $\Pr[w] = \prod_{i=1}^{k} p_i(w_i)$. Then it is easy to show that the outcomes for different subspaces Ω_i are independent and therefore, events over different subspaces are independent. That is, for arbitrary events $A_1 \subseteq \Omega_1$ and $A_2 \subseteq \Omega_2$ it holds for $A_1' = A_1 \times \Omega_2$ and $A_2' = \Omega_1 \times A_2$ that

$$\Pr[A_1' \cap A_2'] = \Pr[A_1'] \cdot \Pr[A_2'] \,.$$

Example: balls into bins

Suppose that we have n balls and n bins. Consider the random experiment that every ball is thrown uniformly and independently at random into one of these bins.

Theorem 1.3 *The probability that bin 1 contains at least one ball is at least* $1/2$.

Proof. In our case, the sample space Ω can be represented as $\Omega = \Omega_1 \times \ldots \times \Omega_n$ with $\Omega_i = \{1, \ldots, n\}$ and probability distributions $p_i : \Omega_i \to [0,1]$ with $p_i(w) = 1/n$ for all $w \in \Omega_i$ (because the balls are thrown *uniformly* at random). Also, for any outcome $w = (w_1, \ldots, w_n) \in \Omega$ it holds that $\Pr[w] = \prod_{i=1}^{n} p_i(w_i)$ (because the balls are thrown *independently* at random). Let A_i be the event that ball i is thrown into bin 1. Then it holds that $\Pr[A_i] = 1/n$ and therefore, $\Pr[A_i \cap A_j] = \Pr[A_i] \cdot \Pr[A_j] = 1/n^2$ for all $i \neq j$. Thus,

$$\begin{aligned}
\Pr[\bigcup_{i=1}^{n} A_i] &\geq \sum_{i=1}^{n} \Pr[A_i] - \sum_{1 \leq i < j \leq n} \Pr[A_i \cap A_j] \\
&= \sum_{i=1}^{n} \frac{1}{n} - \sum_{1 \leq i < j \leq n} \frac{1}{n^2} \\
&= 1 - \binom{n}{2} \frac{1}{n^2} \geq 1 - \frac{1}{2} = \frac{1}{2}
\end{aligned}$$

□

Note that the exact value of the probability is $1 - (1 - 1/n)^n = 1 - 1/e$ for $n \to \infty$.

Random variables

A function $X : \Omega \to \mathbb{R}$ is called a *random variable*. If $X : \Omega \to \{0, 1\}$, we call X a *binary* random variable or simply *indicator*. In order to simplify notation, we define

$$\Pr[X = x] = \Pr[\{w \in \Omega : X(w) = x\}]$$

Analogously,

$$\Pr[X \leq x] = \Pr[\{w \in \Omega : X(w) \leq x\}] \quad \text{und} \quad \Pr[X \geq x] = \Pr[\{w \in \Omega : X(w) \geq x\}]$$

For two random variables X and Y we say that X *stochastically dominates* Y if and only if $\Pr[X \geq z] \geq \Pr[Y \geq z]$ for all z.

Expectation

The *expectation* of a random variable $X : \Omega \to \mathbb{R}$ is defined as

$$\mathbb{E}[X] = \sum_{w \in \Omega} X(w) \cdot \Pr[w]$$

Therefore, also $\mathbb{E}[X] = \sum_{x \in X(\Omega)} x \cdot \Pr[X = x]$. For the special case that $X : \Omega \to \mathbb{N}$, we obtain

$$\mathbb{E}[X] = \sum_{x \in \mathbb{N}} \Pr[X \geq x]$$

and for an indicator X, $\mathbb{E}[X] = \Pr[X = 1]$. Basic properties of the expectation are:

- X is non-negative: $\mathbb{E}[X] \geq 0$
- $|\mathbb{E}[X]| \leq \mathbb{E}[|X|]$
- $\mathbb{E}[c \cdot X] = c \cdot \mathbb{E}[X]$

- $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$, which is also known as the *linearity* of expectation.

Two random variables X and Y are *(stochastically) independent* if for all $x, y \in \mathbb{R}$ it holds that

$$\Pr[X = x \mid Y = y] = \Pr[X = x]$$

Theorem 1.4 *If X and Y are stochastically independent, then $\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$.*

The proof is an exercise.

Probability bounds

The most basic probability bound is the following:

Theorem 1.5 *For any random variable X,*

$$\Pr[X < \mathbb{E}[X]] < 1 \quad \textit{and} \quad \Pr[X > \mathbb{E}[X]] < 1$$

This theorem suffices to prove the existence of certain combinatorial objects.

Example: MaxCUT

Let $G = (V, E)$ be an undirected graph. For a subset $U \subseteq V$ we call $\bar{U} = V \setminus U$ the *complement* of U and

$$(U, \bar{U}) = \{\{v, w\} \in E \mid v \in U \ \wedge \ w \in \bar{U}\}$$

the *cut* separating U from $\bar{U}$ in G. In the MaxCUT problem we are given a graph $G = (V, E)$, and the task is to find a subset $U \subseteq V$ that maximizes $|(U, \bar{U})|$. This problem is known to be NP-hard in general, but there is always a cut that is relatively close to $|E|$, which is the maximum cut size we can expect.

Theorem 1.6 *For every undirected graph $G = (V, E)$ with m edges there is a cut of size at least $m/2$.*

Proof. Suppose that we toss a coin independently for each node in V with $\Pr[\text{heads}] = \Pr[\text{tails}] = 1/2$. All nodes with outcome "heads" are assigned to U and all other nodes are assigned to $\bar{U}$. For each edge $e = \{v, w\} \in E$ let the binary random variable X_e be 1 if and only if $e \in (U, \bar{U})$. Since the outcomes of the coin tosses for v and w are independent,

$$\Pr[X_e = 1] = \Pr[(\text{heads,tails})] + \Pr[(\text{tails,heads})] = 1/4 + 1/4 = 1/2 \ .$$

Let X be the size of the cut $(U, \bar{U})$. Then it holds that $X = \sum_{e \in E} X_e$ and therefore,

$$\mathbb{E}[X] = \sum_{e \in E} \mathbb{E}[X_e] = m \cdot 1/2 = m/2 \ .$$

From Theorem 1.5 it follows that there is a cut of size at least $m/2$. □

Often concrete probability bounds are needed for the deviation from the expectation. The most well-known inequality for this is Markov's inequality.

Theorem 1.7 (Markov's Inequality) *Let X be an arbitrary non-negative random variable. Then it holds for all $k > 0$ that*

$$\Pr[X \geq k] \leq \frac{\mathbb{E}[X]}{k}$$

Proof.

$$\mathbb{E}[X] = \sum_{x \in X(\Omega)} x \cdot \Pr[X = x] \geq \sum_{x \in X(\Omega), x \geq k} x \cdot \Pr[X = x] \geq k \cdot \Pr[X \geq k]$$

□

This inequality can be generalized in the following way.

Theorem 1.8 (General Markov's Inequality) *Let X be an arbitrary random variable and g be an arbitrary function that is non-negative and monotonically increasing on the values in $X(\Omega)$. Then it holds for all $k \in X(\Omega)$ that*

$$\Pr[X \geq k] \leq \frac{\mathbb{E}[g(X)]}{g(k)}$$

Proof.

$$\mathbb{E}[g(X)] = \sum_{x \in X(\Omega)} g(x) \cdot \Pr[X = x] \geq \sum_{x \in X(\Omega), x \geq k} g(x) \cdot \Pr[X = x] \geq g(k) \cdot \Pr[X \geq k]$$

□

From the Markov inequality we can also derive the well-known Chebychev inequality. The *variance* of a random variable X is defined as $\mathbb{V}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2]$.

Theorem 1.9 (Chebychev's Inequality) *Let X be an arbitrary random variable. For all $k > 0$,*

$$\Pr[|X - \mathbb{E}[X]| \geq k] \leq \frac{\mathbb{V}[X]}{k^2}$$

Proof. From the Markov inequality it follows that

$$\Pr[|X| \geq k] = \Pr[X^2 \geq k^2] \leq \mathbb{E}[X^2]/k^2$$

Substituting X by $X - \mathbb{E}[X]$ results in the theorem. □

More powerful inequalities are the so-called Chernoff bounds.

Theorem 1.10 (Chernoff Bounds) *Let $X_1, \ldots, X_n$ be independent binary random variables. Let $X = \sum_{i=1}^{n} X_i$ and $\mu = \mathbb{E}[X]$. Then it holds for all $\delta > 0$ that*

$$\Pr[X \geq (1+\delta)\mu] \leq \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}} \right)^{\mu} \leq e^{-\delta^2 \mu/(2(1+\delta/3))} \leq e^{-\min\{\delta^2, \delta\}\mu/3}$$

and for all $0 < \delta < 1$ that

$$\Pr[X \leq (1-\delta)\mu] \leq \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}} \right)^{\mu} \leq e^{-\delta^2 \mu/2}$$

Proof. We will only show the first inequality. Let $p_i = \Pr[X_i = 1] = \mathbb{E}[X_i]$ for all i. According to the Markov inequality it holds for every function $g(x) = e^{h \cdot x}$ with $h > 0$ and every $\delta \geq 0$ that

$$\Pr[X \geq (1+\delta)\mu] \quad \leq \quad e^{-h(1+\delta)\mu} \cdot \mathbb{E}[e^{h \cdot X}] \tag{1}$$

Since $X_1, \ldots, X_n$ are independent, it follows from Theorem 1.4 that

$$\begin{aligned}
\mathbb{E}[e^{h \cdot X}] &= \mathbb{E}[e^{h(X_1 + \ldots + X_n)}] = \mathbb{E}[e^{h \cdot X_1} \cdots e^{h \cdot X_n}] = \textstyle\prod_{i=1}^{n} \mathbb{E}[e^{h \cdot X_i}] \\
&= \prod_{i=1}^{n} (p_i e^h + (1 - p_i)) = \textstyle\prod_{i=1}^{n} (1 + p_i(e^h - 1)) \\
&\leq \prod_{i=1}^{n} e^{p_i (e^h - 1)} \quad \text{since } 1 + x \leq e^x \text{ for all } x \\
&= e^{\mu(e^h - 1)} \, .
\end{aligned}$$

Together with inequality (1) this implies that

$$\Pr[X \geq (1+\delta)\mu] \leq e^{-h(1+\delta)\mu} \cdot e^{\mu(e^h - 1)} = e^{-(1 + h(1+\delta) - e^h)\mu} \tag{2}$$

The right hand side of (2) is minimal for $h = h_0$ with $h_0 = \ln(1+\delta)$. Inserted into (2) we obtain

$$\Pr[X \geq (1+\delta)\mu] \leq (1+\delta)^{-(1+\delta)\mu} \cdot e^{\delta \cdot \mu} = \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}} \right)^{\mu}$$

The inequality for $\Pr[X \leq (1-\delta)\mu]$ is an exercise. □

For more details on probability theory see, for example, [1]. Now we have a sufficient foundation to consider some more advanced examples in the next section.

References

[1] C. Scheideler. *Probabilistic Methods for Coordination Problems*. HNI-Verlagsschriftenreihe 78, University of Paderborn, 2000. Siehe `wwwcs.upb.de/cs/scheideler`.