
Computational Complexity: A Modern Approach

Draft of a book: Dated January 2007
Comments welcome!

Sanjeev Arora and Boaz Barak
Princeton University
complexitybook@gmail.com

Not to be reproduced or distributed without the authors' permission

This is an Internet draft. Some chapters are more finished than others. References and attributions are very preliminary and we apologize in advance for any omissions (but hope you will nevertheless point them out to us).

Please send us bugs, typos, missing references or general comments to
complexitybook@gmail.com — **Thank You!!**

DRAFT

DRAFT

Appendix A

Mathematical Background.

This appendix reviews the mathematical notions used in this book. However, most of these are only used in few places, and so the reader might want to only quickly review Sections [A.1](#), [A.2](#) and [A.3](#), and come back to the other sections as needed. In particular, apart from probability, the first part of the book essentially requires only comfort with mathematical proofs and some very basic notions of discrete math.

The topics described in this appendix are covered in greater depth in many texts and online sources. Almost all of the mathematical background needed is covered in a good undergraduate “discrete math for computer science” course as currently taught at many computer science departments. Some good sources for this material are the lecture notes by Papadimitriou and Vazirani [?], Lehman and Leighton [?] and the book of Rosen [?].

Although knowledge of algorithms is not strictly necessary for this book, it would be quite useful. It would be helpful to review either one of the two excellent recent books by Dasgupta et al [?] and Kleinberg and Tardos [?] or the earlier text by Cormen et al [?]. This book does not require prior knowledge of computability and automata theory, but some basic familiarity with that theory could be useful: see Sipser’s book [?] for an excellent introduction. Mitzenmacher and Upfal [?] and Prabhakar and Raghavan [?] cover both algorithmic reasoning and probability. For more insight on discrete probability, see the book by Alon and Spencer [?].

A.1 Mathematical Proofs

Perhaps *the* mathematical prerequisite needed for this book is a certain level of comfort with mathematical proofs. While in everyday life we might use “proof” to describe a fairly convincing argument, in mathematics a proof is an argument that is convincing *beyond any shadow of a doubt*.¹ For example, consider the following mathematical statement:

Every even number greater than 2 is equal to the sum of two primes.

¹In a famous joke, as a mathematician and an engineer drive in Scotland they see a white sheep on their left side. The engineer says “you see: all the sheep in Scotland are white”. The mathematician replies “All I see is that there exists a sheep in Scotland whose right side is white”.

This statement, known as “Goldbach’s Conjecture”, was conjectured to be true by Christian Goldbach in 1742. In the more than 250 years that have passed since, no one has ever found a counterexample to this statement. In fact, it has been verified to be true for all even numbers from 4 till 100,000,000,000,000,000. Yet still it is not considered proven, since we have not ruled out the possibility that there is some (very large) even number that cannot be expressed as the sum of two primes.

The fact that a mathematical proof has to be absolutely convincing does not mean that it has to be overly formal and tedious. It just has to be clearly written, and contain no logical gaps. When you write proofs try to be clear and concise, rather than using too much formal notation. When you read proofs, try to ask yourself at every statement “am I really convinced that this statement is true?”.

Of course, to be absolutely convinced that some statement is true, we need to be certain of what that statement means. This why there is a special emphasis in mathematics on very precise *definitions*. Whenever you read a definition, try to make sure you completely understand it, perhaps by working through some simple examples. Oftentimes, understanding the meaning of a mathematical statement is more than half the work to prove that it is true.

EXAMPLE A.1

Here is an example for a classical mathematical proof, written by Euclid around 300 B.C. Recall that a *prime number* is an integer $p > 1$ whose only divisors are p and 1, and that every number n is a product of prime numbers. Euclid’s Theorem is the following:

THEOREM A.2

There exist infinitely many primes.

Before proving it, let’s see that we understand what this statement means. It simply means that for every natural number k , there are more than k primes, and hence the number of primes is not finite.

At first, one might think it’s obvious that there are infinitely many primes because there are infinitely many natural numbers, and each natural number is a product of primes. However, this is faulty reasoning: for example, the set of numbers of the form 3^n is infinite, even though their only factor is the single prime 3.

To prove Theorem A.2, we use the technique of *proof by contradiction*. That is, we assume it is false and try to derive a contradiction from that assumption. Indeed, assume that all the primes can be enumerated as p_1, p_2, \dots, p_k for some number k . Define the number $n = p_1 p_2 \cdots p_k + 1$. Since we assume that the numbers p_1, \dots, p_k are *all* the primes, all of n ’s prime factors must come from this set, and in particular there is some i between 1 and k such that p_i divides n . That is, $n = p_i m$ for some number m . Thus,

$$p_i m = p_1 p_2 \cdots p_k + 1$$

or equivalently,

$$p_i m - p_1 p_2 \cdots p_k = 1.$$

DRAFT

But dividing both sides of this equation by p_i , we will get a whole number on the left hand side (as p_i is a factor of $p_1 p_2 \cdots p_k$) and the fraction $1/p_i$ on the right hand side, deriving a contradiction. This allows us to rightfully place the QED symbol ■ and consider Theorem A.2 as proven.

A.2 Sets, Functions, Pairs, Strings, Graphs, Logic.

A *set* contains a finite or infinite number of elements, without repetition or respect to order, for example $\{2, 17, 5\}$, $\mathbb{N} = \{1, 2, 3, \dots\}$ (the set of natural numbers), $[n] = \{1, 2, \dots, n\}$ (the set of natural numbers from 1 to n), \mathbb{R} (the set of real numbers). For a finite set A , we denote by $|A|$ the number of elements in A . Some operations on sets are: **(1) union**: $A \cup B = \{x : x \in A \text{ or } x \in B\}$, **(2) intersection**: $A \cap B = \{x : x \in A \text{ and } x \in B\}$, and **(3) subtraction**: $A \setminus B = \{x : x \in A \text{ and } x \notin B\}$.

We say that f is a *function* from a set A to B , denoted by $f : A \rightarrow B$, if it maps any element of A into an element of B . If B and A are finite, then the number of possible functions from A to B is $|B|^{|A|}$. We say that f is *one to one* if for every $x, w \in A$ with $x \neq w$, $f(x) \neq f(w)$. If A, B are finite, the existence of such a function implies that $|A| \leq |B|$. We say that f is *onto* if for every $y \in B$ there exists $x \in A$ such that $f(x) = y$. If A, B are finite, the existence of such a function implies that $|A| \geq |B|$. We say that f is a *permutation* if it is both one-to-one and onto. For finite A, B , the existence of a permutation from A to B implies that $|A| = |B|$.

If A, B are sets, then the $A \times B$ denotes the set of all ordered pairs $\langle a, b \rangle$ with $a \in A, b \in B$. Note that if A, B are finite then $|A \times B| = |A| \cdot |B|$. We can define similarly $A \times B \times C$ to be the set of ordered triples $\langle a, b, c \rangle$ with $a \in A, b \in B, c \in C$. For $n \in \mathbb{N}$, we denote by A^n the set $A \times A \times \cdots \times A$ (n times). We will often use the set $\{0, 1\}^n$, consisting of all length- n sequences of bits (i.e., length n strings), and the set $\{0, 1\}^* = \cup_{n \geq 0} \{0, 1\}^n$ ($\{0, 1\}^0$ has a single element: a binary string of length zero, which we call the empty word and denote by ε).

A *graph* G consists of a set V of *vertices* (which we often assume is equal to the set $[n] = \{1, \dots, n\}$ for some $n \in \mathbb{N}$) and a set E of *edges*, which consists of unordered pairs (i.e., size two subsets) of elements in V . We denote the edge $\{u, v\}$ of the graph by \overline{uv} . For $v \in V$, the *neighbors* of v are all the vertices $u \in V$ such that $\overline{uv} \in E$. In a *directed graph*, the edges consist of *ordered pairs* of vertices, to stress this we sometimes denote the edge $\langle u, v \rangle$ in a directed graph by \overrightarrow{uv} . One can represent an n -vertex graph G by its *adjacency matrix* which is an $n \times n$ matrix A such that $A_{i,j}$ is equal to 1 if the edge \overrightarrow{ij} is present in G and is equal to 0 otherwise. One can think of an undirected graph as a directed graph G that satisfies that for every u, v , G contains the edge \overrightarrow{uv} if and only if it contains the edge \overrightarrow{vu} . Hence, one can represent an undirected graph by an adjacency matrix that is *symmetric* ($A_{i,j} = A_{j,i}$ for every $i, j \in [n]$).

A *Boolean variable* is a variable that can be either TRUE or FALSE (we sometimes identify TRUE with 1 and FALSE with 0). We can combine variables via the logical operations AND (\wedge), OR (\vee) and NOT (\neg , sometimes also denoted by an overline), to obtain *Boolean formulae*. For example, the following is a Boolean formulae on the variables u_1, u_2, u_3 : $(u_1 \wedge \overline{u_2}) \vee \neg(u_3 \wedge \overline{u_1})$. The definitions of the operations are the usual: $a \wedge b = \text{TRUE}$ if $a = \text{TRUE}$ and $b = \text{TRUE}$ and is equal to FALSE

otherwise; $\bar{a} = \neg a = \text{TRUE}$ if $a = \text{FALSE}$ and is equal to FALSE otherwise; $a \vee b = \neg(\bar{a} \wedge \bar{b})$. If φ is a formulae in n variables u_1, \dots, u_n , then for any *assignment* of values $u \in \{\text{FALSE}, \text{TRUE}\}^n$ (or equivalently, $\{0, 1\}^n$), we denote by $\varphi(u)$ the value of φ when its variables are assigned the values in u . We say that φ is *satisfiable* if there exists a u such that $\varphi(u) = \text{TRUE}$.

We will often use the *quantifiers* \forall (for all) and \exists (exists). That is, if φ is a condition that can be TRUE or FALSE depending on the value of a variable x , then we write $\forall x \varphi(x)$ to denote the statement that φ is TRUE for *every* possible value that can be assigned to x . If A is a set then we write $\forall_{x \in A} \varphi(x)$ to denote the statement that φ is TRUE for every assignment for x from the set A . The quantifier \exists is defined similarly. Formally, we say that $\exists x \varphi(x)$ holds if and only if $\neg(\forall x \neg \varphi(x))$ holds.

A.3 Probability theory

A *finite probability space* is a finite set $\Omega = \{\omega_1, \dots, \omega_N\}$ along with a set of numbers $p_1, \dots, p_N \in [0, 1]$ such that $\sum_{i=1}^N p_i = 1$. A random element is selected from this space by choosing ω_i with probability p_i . If x is chosen from the sample space Ω then we denote this by $x \in_R \Omega$. If no distribution is specified then we use the uniform distribution over the elements of Ω (i.e., $p_i = \frac{1}{N}$ for every i).

An *event* over the space Ω is a subset $A \subseteq \Omega$ and the *probability* that A occurs, denoted by $\Pr[A]$, is equal to $\sum_{i: \omega_i \in A} p_i$. To give an example, the probability space could be that of all 2^n possible outcomes of n tosses of a fair coin (i.e., $\Omega = \{0, 1\}^n$ and $p_i = 2^{-n}$ for every $i \in [2^n]$) and the event A can be that the number of coins that come up “heads” (or, equivalently, 1) is even. In this case, $\Pr[A] = 1/2$ (exercise). The following simple bound—called the *union bound*—is often used in the book. For every set of events A_1, A_2, \dots, A_n ,

$$\Pr[\cup_{i=1}^n A_i] \leq \sum_{i=1}^n \Pr[A_i]. \quad (1)$$

Inclusion exclusion principle. The union bound is a special case of a more general principle. Indeed, note that if the sets A_1, \dots, A_n are not *disjoint* then the probability of $\cup_i A_i$ could be smaller than $\sum_i \Pr[A_i]$ since we are overcounting elements that appear in more than one set. We can correct this by subtracting $\sum_{i < j} \Pr[A_i \cap A_j]$ but then we might be undercounting, since we subtracted elements that appear in at least 3 sets too many times. Continuing this process we get

CLAIM A.3 (INCLUSION-EXCLUSION PRINCIPLE)

For every A_1, \dots, A_n ,

$$\Pr[\cup_{i=1}^n A_i] = \sum_{i=1}^n \Pr[A_i] - \sum_{1 \leq i < j \leq n} \Pr[A_i \cap A_j] + \dots + (-1)^{n-1} \Pr[A_1 \cap \dots \cap A_n].$$

Moreover, this is an alternating sum which means that if we take only the first k summands of the right hand side, then this upperbounds the left-hand side if k is odd, and lowerbounds it if k is even.

DRAFT

We sometimes use the following corollary of this claim:

CLAIM A.4

For every events A_1, \dots, A_n ,

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

Random subsum principle. The following fact is used often in the book:

CLAIM A.5 (THE RANDOM SUBSUM PRINCIPLE)

For $x, y \in \{0, 1\}^n$, denote $x \odot y = \sum_{i=1}^n x_i y_i \pmod{2}$ (that is, $x \odot y$ is equal to 1 if the number of i 's such that $x_i = y_i = 1$ is odd and equal to 0 otherwise). Then for every $y \neq 0^n$,

$$\Pr_{x \in_R \{0,1\}^n} [x \odot y = 1] = \frac{1}{2}$$

PROOF: Suppose that y_j is nonzero. We can think of choosing x as follows: first choose all the coordinates of x other than the j^{th} and only choose the j^{th} coordinate last. After we choose all the coordinates of x other than the j^{th} , the value $\sum_{i:i \neq j} x_i y_i \pmod{2}$ is fixed to be some $c \in \{0, 1\}$. Regardless of what c is, with probability $1/2$ we choose $x_j = 0$, in which case $x \odot y = c$ and with probability $1/2$ we choose $x_j = 1$, in which case $x \odot y = 1 - c$. We see that in any case $x \odot y$ will be equal to 1 with probability $1/2$. ■

A.3.1 Random variables and expectations.

A *random variable* is a mapping from a probability space to \mathbb{R} . For example, if Ω is as above, the set of all possible outcomes of n tosses of a fair coin, then we can denote by X the number of coins that came up heads.

The *expectation* of a random variable X , denoted by $E[X]$, is its weighted average. That is, $E[X] = \sum_{i=1}^N p_i X(\omega_i)$. The following simple claim follows from the definition:

CLAIM A.6 (LINEARITY OF EXPECTATION)

For X, Y random variables over a space Ω , denote by $X + Y$ the random variable that maps ω to $X(\omega) + Y(\omega)$. Then,

$$E[X + Y] = E[X] + E[Y]$$

This claim implies that the random variable X from the example above has expectation $n/2$. Indeed $X = \sum_{i=1}^n X_i$ where X_i is equal to 1 if the i^{th} coin came up heads and is equal to 0 otherwise. But clearly, $E[X_i] = 1/2$ for every i .

For a real number α and a random variable X , we define αX to be the random variable mapping ω to $\alpha \cdot X(\omega)$. Note that $E[\alpha X] = \alpha E[X]$.

A.3.2 The averaging argument

We list various versions of the “averaging argument.” Sometimes we give two versions of the same result, one as a fact about numbers and one as a fact about probability spaces.

LEMMA A.7

If a_1, a_2, \dots, a_n are some numbers whose average is c then some $a_i \geq c$.

LEMMA A.8 (“THE PROBABILISTIC METHOD”)

If X is a random variable which takes values from a finite set and $E[X] = \mu$ then the event “ $X \geq \mu$ ” has nonzero probability.

LEMMA A.9

If $a_1, a_2, \dots, a_n \geq 0$ are numbers whose average is c then the fraction of a_i ’s that are greater than (resp., at least) kc is less than (resp., at most) $1/k$.

LEMMA A.10 (“MARKOV’S INEQUALITY”)

Any non-negative random variable X satisfies

$$\Pr(X \geq kE[X]) \leq \frac{1}{k}.$$

COROLLARY A.11

If $a_1, a_2, \dots, a_n \in [0, 1]$ are numbers whose average is $1 - \gamma$ then at least $1 - \sqrt{\gamma}$ fraction of them are at least $1 - \sqrt{\gamma}$.

Can we give any meaningful upperbound on $\Pr[X < c \cdot E[X]]$ where $c < 1$? Yes, if X is bounded.

LEMMA A.12

If a_1, a_2, \dots, a_n are numbers in the interval $[0, 1]$ whose average is ρ then at least $\rho/2$ of the a_i ’s are at least as large as $\rho/2$.

PROOF: Let γ be the fraction of i ’s such that $a_i \geq \rho/2$. Then $\gamma + (1 - \gamma)\rho/2$ must be at least $\rho/2$, so $\gamma \geq \rho/2$. ■ More generally, we have

LEMMA A.13

If $X \in [0, 1]$ and $E[X] = \mu$ then for any $c < 1$ we have

$$\Pr[X \leq c\mu] \leq \frac{1 - \mu}{1 - c\mu}.$$

EXAMPLE A.14

Suppose you took a lot of exams, each scored from 1 to 100. If your average score was 90 then in at least half the exams you scored at least 80.

A.3.3 Conditional probability and independence

If we already know that an event B happened, this reduces the space from Ω to $\Omega \cap B$, where we need to scale the probabilities by $1/\Pr[B]$ so they will sum up to one. Thus, the probability of an event A *conditioned* on an event B , denoted $\Pr[A|B]$, is equal to $\Pr[A \cap B]/\Pr[B]$ (where we always assume that B has positive probability).

We say that two events A, B are *independent* if $\Pr[A \cap B] = \Pr[A] \Pr[B]$. Note that this implies that $\Pr[A|B] = \Pr[A]$ and $\Pr[B|A] = \Pr[B]$. We say that a set of events A_1, \dots, A_n are *mutually independent* if for every subset $S \subset [n]$,

$$\Pr[\cap_{i \in S} A_i] = \prod_{i \in S} \Pr[A_i]. \quad (2)$$

We say that A_1, \dots, A_n are *k-wise independent* if (2) holds for every $S \subseteq [n]$ with $|S| \leq k$.

We say that two random variables X, Y are *independent* if for every $x, y \in \mathbb{R}$, the events $\{X = x\}$ and $\{Y = y\}$ are independent. We generalize similarly the definition of mutual independence and *k-wise independence* to sets of random variables X_1, \dots, X_n . We have the following claim:

CLAIM A.15

If X_1, \dots, X_n are mutually independent then

$$\mathbb{E}[X_1 \cdots X_n] = \prod_{i=1}^n \mathbb{E}[X_i]$$

PROOF:

$$\begin{aligned} \mathbb{E}[X_1 \cdots X_n] &= \sum_x x \Pr[X_1 \cdots X_n = x] = \\ &= \sum_{x_1, \dots, x_n} x_1 \cdots x_n \Pr[X_1 = x_1 \text{ and } X_2 = x_2 \cdots \text{ and } X_n = x_n] = \text{(by independence)} \\ &= \sum_{x_1, \dots, x_n} x_1 \cdots x_n \Pr[X_1 = x_1] \cdots \Pr[X_n = x_n] = \\ &= \left(\sum_{x_1} x_1 \Pr[X_1 = x_1] \right) \left(\sum_{x_2} x_2 \Pr[X_2 = x_2] \right) \cdots \left(\sum_{x_n} x_n \Pr[X_n = x_n] \right) = \prod_{i=1}^n \mathbb{E}[X_i] \end{aligned}$$

where the sums above are over all the possible real numbers that can be obtained by applying the random variables or their products to the finite set Ω . ■

A.3.4 Deviation upperbounds

Under various conditions, one can give upperbounds on the probability of a random variable “straying too far” from its expectation. These upperbounds are usually derived by clever use of Markov’s inequality.

The *variance* of a random variable X is defined to be $\text{Var}[X] = \mathbb{E}[(X - \mathbb{E}(X))^2]$. Note that since it is the expectation of a non-negative random variable, $\text{Var}[X]$ is always non-negative. Also, using

linearity of expectation, we can derive that $\text{Var}[X] = \text{E}[X^2] - (\text{E}[X])^2$. The *standard deviation* of a variable X is defined to be $\sqrt{\text{Var}[X]}$.

The first bound is Chebyshev's inequality, useful when only the variance is known.

LEMMA A.16 (CHEBYSHEV INEQUALITY)

If X is a random variable with standard deviation σ , then for every $k > 0$,

$$\Pr[|X - \text{E}[X]| > k\sigma] \leq 1/k^2$$

PROOF: Apply Markov's inequality to the random variable $(X - \text{E}[X])^2$, noting that by definition of variance, $\text{E}[(X - \text{E}[X])^2] = \sigma^2$. ■

Chebyshev's inequality is often useful in the case that X is equal to $\sum_{i=1}^n X_i$ for pairwise independent random variables X_1, \dots, X_n . This is because of the following claim, that is left as an exercise:

CLAIM A.17

If X_1, \dots, X_n are pairwise independent then

$$\text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i)$$

The next inequality has many names, and is widely known in theoretical computer science as the *Chernoff bound*. It considers scenarios of the following type. Suppose we toss a fair coin n times. The expected number of heads is $n/2$. How tightly is this number concentrated? Should we be very surprised if after 1000 tosses we have 625 heads? The bound we present is slightly more general, since it concerns n different coin tosses of possibly different expectations (the expectation of a coin is the probability of obtaining "heads"; for a fair coin this is $1/2$). These are sometimes known as Poisson trials.

THEOREM A.18 ("CHERNOFF" BOUNDS)

Let X_1, X_2, \dots, X_n be mutually independent random variables over $\{0, 1\}$ (i.e., X_i can be either 0 or 1) and let $\mu = \sum_{i=1}^n \text{E}[X_i]$. Then for every $\delta > 0$,

$$\Pr\left[\sum_{i=1}^n X_i \geq (1 + \delta)\mu\right] \leq \left[\frac{e^\delta}{(1 + \delta)^{(1 + \delta)}}\right]^\mu. \quad (3)$$

$$\Pr\left[\sum_{i=1}^n X_i \leq (1 - \delta)\mu\right] \leq \left[\frac{e^{-\delta}}{(1 - \delta)^{(1 - \delta)}}\right]^\mu. \quad (4)$$

Often, what we use need is only the corollary that under the above conditions, for every $c > 0$

$$\Pr\left[\left|\sum_{i=1}^n X_i - \mu\right| \geq c\mu\right] \leq 2^{-c^2 n/2}$$

DRAFT

PROOF: Surprisingly, the Chernoff bound is also proved using the Markov inequality. We only prove the first inequality; a similar proof exists for the second. We introduce a positive dummy variable t , and observe that

$$\mathbb{E}[\exp(tX)] = \mathbb{E}[\exp(t \sum_i X_i)] = \mathbb{E}[\prod_i \exp(tX_i)] = \prod_i \mathbb{E}[\exp(tX_i)], \quad (5)$$

where $\exp(z)$ denotes e^z and the last equality holds because the X_i r.v.s are independent. Now,

$$\mathbb{E}[\exp(tX_i)] = (1 - p_i) + p_i e^t,$$

therefore,

$$\begin{aligned} \prod_i \mathbb{E}[\exp(tX_i)] &= \prod_i [1 + p_i(e^t - 1)] \leq \prod_i \exp(p_i(e^t - 1)) \\ &= \exp\left(\sum_i p_i(e^t - 1)\right) = \exp(\mu(e^t - 1)), \end{aligned} \quad (6)$$

as $1 + x \leq e^x$. Finally, apply Markov's inequality to the random variable $\exp(tX)$, viz.

$$\Pr[X \geq (1 + \delta)\mu] = \Pr[\exp(tX) \geq \exp(t(1 + \delta)\mu)] \leq \frac{\mathbb{E}[\exp(tX)]}{\exp(t(1 + \delta)\mu)} = \frac{\exp((e^t - 1)\mu)}{\exp(t(1 + \delta)\mu)},$$

using lines (5) and (6) and the fact that t is positive. Since t is a dummy variable, we can choose any positive value we like for it. Simple calculus shows that the right hand side is minimized for $t = \ln(1 + \delta)$ and this leads to the theorem statement. ■

By the way, if all n coin tosses are fair (Heads has probability $1/2$) then the the probability of seeing N heads where $|N - n/2| > a\sqrt{n}$ is at most $e^{-a^2/2}$. The chance of seeing at least 625 heads in 1000 tosses of an unbiased coin is less than 5.3×10^{-7} .

A.3.5 Some other inequalities.

Jensen's inequality.

The following inequality, generalizing the inequality $\mathbb{E}[X^2] \geq \mathbb{E}[X]^2$, is also often useful:

CLAIM A.19

We say that $f : \mathbb{R} \rightarrow \mathbb{R}$ is convex if for every $p \in [0, 1]$ and $x, y \in \mathbb{R}$, $f(px + (1 - p)y) \leq p \cdot f(x) + (1 - p) \cdot f(y)$. Then, for every random variable X and convex function f , $f(\mathbb{E}[X]) \leq \mathbb{E}[f(X)]$.

Approximating the binomial coefficient

Of special interest is the *Binomial* random variable B_n denoting the number of coins that come up “heads” when tossing n fair coins. For every k , $\Pr[B_n = k] = 2^{-n} \binom{n}{k}$ where $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ denotes the number of size- k subsets of $[n]$. Clearly, $\binom{n}{k} \leq n^k$, but sometimes we will need a better estimate for $\binom{n}{k}$ and use the following approximation:

CLAIM A.20

For every $n, k < n$,

$$\left(\frac{n}{k}\right)^k \leq \binom{n}{k} \leq \left(\frac{ne}{k}\right)^k$$

The best approximation can be obtained via Stirling's formula:

LEMMA A.21 (STIRLING'S FORMULA)

For every n ,

$$\sqrt{2\pi n} \left(\frac{n}{e}\right)^n e^{\frac{1}{12n+1}} < n! < \sqrt{2\pi n} \left(\frac{n}{e}\right)^n e^{\frac{1}{12n}}$$

It can be proven by taking natural logarithms and approximating $\ln n! = \ln(1 \cdot 2 \cdots n) = \sum_{i=1}^n \ln i$ by the integral $\int_1^n \ln x \, dx = n \ln n - n + 1$. It implies the following corollary:

COROLLARY A.22

For every $n \in \mathbb{N}$ and $\alpha \in [0, 1]$,

$$\binom{n}{\alpha n} = (1 \pm O(n^{-1})) \frac{1}{\sqrt{2\pi n \alpha(1-\alpha)}} 2^{H(\alpha)n}$$

where $H(\alpha) = \alpha \log(1/\alpha) + (1-\alpha) \log(1/(1-\alpha))$ and the constants hidden in the O notation are independent of both n and α .

More useful estimates.

The following inequalities can be obtained via elementary calculus:

- For every $x \geq 1$, $(1 - \frac{1}{x})^x \leq \frac{1}{e} \leq (1 - \frac{1}{x+1})^x$
- For every k , $\sum_{i=1}^n i^k = \Theta\left(\frac{n^{k+1}}{k+1}\right)$
- For every $k > 1$, $\sum_{i=1}^{\infty} n^{-k} < O(1)$.
- For every $c, \epsilon > 0$, $\sum_{i=1}^{\infty} \frac{n^c}{(1+\epsilon)^n} < O(1)$.
- For every n , $\sum_{i=1}^n \ln i = \ln n! \pm O(1)$

A.4 Finite fields and groups

A *field* is a set \mathbb{F} that has an addition (+) and multiplication (\cdot) operations that behave in the expected way: satisfy associative, commutative and distributive laws, have both additive and multiplicative inverses, and neutral elements 0 and 1 for addition and multiplication respectively. Familiar fields are the real numbers (\mathbb{R}), the rational numbers (\mathbb{Q}) and the complex numbers (\mathbb{C}), but there are also *finite* fields.

If q is a prime, then we denote by $\text{GF}(q)$ the field consisting of the elements $\{0, \dots, q-1\}$ with addition and multiplication performed modulo q . For example, the numbers $\{0, \dots, 6\}$ yield a field

if addition and multiplication are performed modulo 7. We leave it to the reader to verify $\text{GF}(q)$ is indeed a field for every prime q . The simplest example for such a field is the field $\text{GF}(2)$ consisting of $\{0, 1\}$ where multiplication is the AND (\wedge) operation and addition is the XOR operation.

Every finite field \mathbb{F} has a number ℓ such that for every $x \in \mathbb{F}$, $x + x + \dots + x$ (ℓ times) is equal to the zero element of \mathbb{F} (exercise). This number ℓ is called the *characteristic* of \mathbb{F} . For every prime q , the characteristic of $\text{GF}(q)$ is equal to q .

A.4.1 Non-prime fields.

One can see that if n is not prime, then the set $\{0, \dots, n-1\}$ with addition and multiplication modulo n is not a field, as there exist two non-zero elements x, y in this set such that $x \cdot y = n = 0 \pmod{n}$. Nevertheless, there are finite fields of size n for non-prime n . Specifically, for every prime q , and $k \geq 1$, there exists a field of q^k elements, which we denote by $\text{GF}(q^k)$. We will very rarely need to use such fields in this book, but still provide an outline of their construction below.

For every prime q and k there exists an *irreducible* degree k polynomial P over the field $\text{GF}(q)$ (P is irreducible if it cannot be expressed as the product of two polynomials P', P'' of lower degree). We then let $\text{GF}(q^k)$ be the set of all $k-1$ -degree polynomials over $\text{GF}(q)$. Each such polynomial can be represented as a vector of its k coefficients. We perform both addition and multiplication modulo the polynomial P . Note that addition corresponds to standard vector addition of k -dimensional vectors over $\text{GF}(q)$, and both addition and multiplication can be easily done in $\text{poly}(n, \log q)$ time (we can reduce a polynomial S modulo a polynomial P using a similar algorithm to long division of numbers). It turns out that no matter how we choose the irreducible polynomial P , we will get the same field, up to renaming of the elements. There is a deterministic $\text{poly}(q, k)$ -time algorithm to obtain an irreducible polynomial of degree k over $\text{GF}(q)$. There are also probabilistic algorithms (and deterministic algorithms whose analysis relies on unproven assumptions) that obtain such a polynomial in $\text{poly}(\log q, k)$ time.

For us, the most important example of a finite field is $\text{GF}(2^k)$, which consists of the set $\{0, 1\}^k$, with addition being component-wise XOR, and multiplication being polynomial multiplication via some irreducible polynomial which we can find in $\text{poly}(k)$ time. In fact, we will mostly not even be interested in the multiplicative structure of $\text{GF}(2^k)$ and only use the addition operation (i.e., use it as the vector space $\text{GF}(2)^k$, see below).

A.4.2 Groups.

A *group* is a set that only has a single operation, say \star , that is associative and has an inverse. That is, (G, \star) is a group if

1. For every $a, b, c \in G$, $(a \star b) \star c = a \star (b \star c)$
2. There exists a special element $id \in G$ such that $a \star id = a$ for every $a \in G$, and for every $a \in G$ there exists $b \in G$ such that $a \star b = b \star a = id$.

If G is a finite group, it is known that for every $a \in G$, $a \star a \star \dots \star a$ ($|G|$ times) is equal to the element id . A group is called *commutative* or Abelian if its operation satisfies $a \star b = b \star a$ for every $a, b \in G$. For every number $n \geq 2$, the set $\{0, \dots, n-1\}$ with the operation being addition

modulo n is an Abelian group. Also, the set $\{k : k \in [n-1], \gcd(k, n) = 1\}$ with the operation being multiplication modulo n is an Abelian group.

If \mathbb{F} is a field and $k \geq 1$, then the set of k -dimensional vectors of \mathbb{F} (i.e., \mathbb{F}^k) together with the operation of componentwise addition, yields an Abelian group. As mentioned above, the most interesting special case for us is the group $\text{GF}(2)^k$ for some k . Note that in this group the identity element is the vector 0^k and for every $x \in \text{GF}(2)^k$, $x + x = 0^k$. This group is often referred to as the *Boolean cube*.

A.5 Vector spaces and Hilbert spaces

A.6 Polynomials

We list some basic facts about univariate polynomials.

THEOREM A.23

A nonzero polynomial of degree d has at most d distinct roots.

PROOF: Suppose $p(x) = \sum_{i=0}^d c_i x^i$ has $d+1$ distinct roots $\alpha_1, \dots, \alpha_{d+1}$ in some field \mathbb{F} . Then

$$\sum_{i=0}^d \alpha_j^i \cdot c_i = p(\alpha_j) = 0,$$

for $j = 1, \dots, d+1$. This means that the system $\mathbf{A}\mathbf{y} = \mathbf{0}$ with

$$\mathbf{A} = \begin{pmatrix} 1 & \alpha_1 & \alpha_1^2 & \cdots & \alpha_1^d \\ 1 & \alpha_2 & \alpha_2^2 & \cdots & \alpha_2^d \\ \dots & \dots & \dots & \dots & \dots \\ 1 & \alpha_{d+1} & \alpha_{d+1}^2 & \cdots & \alpha_{d+1}^d \end{pmatrix}$$

has a solution $\mathbf{y} = \mathbf{c}$. The matrix \mathbf{A} is a *Vandermonde* matrix, and it can be shown that

$$\det \mathbf{A} = \prod_{i>j} (\alpha_i - \alpha_j),$$

which is nonzero for distinct α_i . Hence $\text{rank} \mathbf{A} = d+1$. The system $\mathbf{A}\mathbf{y} = \mathbf{0}$ has therefore only a trivial solution — a contradiction to $\mathbf{c} \neq \mathbf{0}$. ■

THEOREM A.24

For any set of pairs $(a_1, b_1), \dots, (a_{d+1}, b_{d+1})$ there exists a unique polynomial $g(x)$ of degree at most d such that $g(a_i) = b_i$ for all $i = 1, 2, \dots, d+1$.

PROOF: The requirements are satisfied by *Lagrange Interpolating Polynomial*:

$$\sum_{i=1}^{d+1} b_i \cdot \frac{\prod_{j \neq i} (x - a_j)}{\prod_{j \neq i} (a_i - a_j)}.$$

DRAFT

If two polynomials $g_1(x), g_2(x)$ satisfy the requirements then their difference $p(x) = g_1(x) - g_2(x)$ is of degree at most d , and is zero for $x = a_1, \dots, a_{d+1}$. Thus, from the previous theorem, polynomial $p(x)$ must be zero and polynomials $g_1(x), g_2(x)$ identical. ■

The following elementary result is usually attributed to Schwartz and Zippel in the computer science community, though it was certainly known earlier (see e.g. DeMillo and Lipton [?]).

LEMMA A.25

If a polynomial $p(x_1, x_2, \dots, x_m)$ over $F = GF(q)$ is nonzero and has total degree at most d , then

$$\Pr[p(a_1..a_m) \neq 0] \geq 1 - \frac{d}{q},$$

where the probability is over all choices of $a_1..a_m \in F$.

PROOF: We use induction on m . If $m = 1$ the statement follows from Theorem A.23. Suppose the statement is true when the number of variables is at most $m - 1$. Then p can be written as

$$p(x_1, x_2, \dots, x_m) = \sum_{i=0}^d x_1^i p_i(x_2, \dots, x_m),$$

where p_i has total degree at most $d - i$. Since p is nonzero, at least one of p_i is nonzero. Let k be the largest i such that p_i is nonzero. Then by the inductive hypothesis,

$$\Pr_{a_2, a_3, \dots, a_m} [p_i(a_2, a_3, \dots, a_m) \neq 0] \geq 1 - \frac{d - k}{q}.$$

Whenever $p_i(a_2, a_3, \dots, a_m) \neq 0$, $p(x_1, a_2, a_3, \dots, a_m)$ is a nonzero univariate polynomial of degree k , and hence becomes 0 only for at most k values of x_1 . Hence

$$\Pr[p(a_1..a_m) \neq 0] \geq \left(1 - \frac{k}{q}\right) \left(1 - \frac{d - k}{q}\right) \geq 1 - \frac{d}{q},$$

and the induction is completed. ■

DRAFT