Appendix A

Mathematical Background

A.1 Identities and Inequalities

We list some standard identities and inequalities that are used at various points throughout the text.

THEOREM A.1 (Binomial expansion theorem) *Let x, y be real numbers, and let n be a positive integer. Then*

$$(x+y)^n = \sum_{i=0}^{n} \binom{n}{i} x^i \, y^{n-i}.$$

PROPOSITION A.2 *For all $x \geq 1$ it holds that $(1 - 1/x)^x \leq e^{-1}$.*

PROPOSITION A.3 *For all x it holds that $1 - x \leq e^{-x}$.*

PROPOSITION A.4 *For all x with $0 \leq x \leq 1$ it holds that*

$$e^{-x} \leq 1 - \left(1 - \frac{1}{e}\right) \cdot x \leq 1 - \frac{x}{2}.$$

A.2 Asymptotic Notation

We use standard notation for expressing asymptotic behavior of functions.

DEFINITION A.5 *Let $f(n), g(n)$ be functions from non-negative integers to non-negative reals. Then:*

- *$f(n) = \mathcal{O}(g(n))$ means that there exist positive integers c and n' such that for all $n > n'$ it holds that $f(n) \leq c \cdot g(n)$.*

- *$f(n) = \Omega(g(n))$ means that there exist positive integers c and n' such that for all $n > n'$ it holds that $f(n) \geq c \cdot g(n)$.*
- *$f(n) = \Theta(g(n))$ means that there exist positive integers c_1, c_2, and n' such that for all $n > n'$ it holds that $c_1 \cdot g(n) \leq f(n) \leq c_2 \cdot g(n)$.*
- *$f(n) = o(g(n))$ means that $\lim_{n\to\infty} \frac{f(n)}{g(n)} = 0$.*
- *$f(n) = \omega(g(n))$ means that $\lim_{n\to\infty} \frac{f(n)}{g(n)} = \infty$.*

Example A.6
Let $f(n) = n^4 + 3n + 500$. Then:

- $f(n) = \mathcal{O}(n^4)$.
- $f(n) = \mathcal{O}(n^5)$. In fact, $f(n) = o(n^5)$.
- $f(n) = \Omega(n^3 \log n)$. In fact, $f(n) = \omega(n^3 \log n)$.
- $f(n) = \Theta(n^4)$.

$\diamondsuit$

A.3 Basic Probability

We assume the reader is familiar with basic probability theory, on the level of what is covered in a typical undergraduate course on discrete mathematics. Here we simply remind the reader of some notation and basic facts.

If E is an event, then $\bar{E}$ denotes the complement of that event; i.e., $\bar{E}$ is the event that E does *not* occur. By definition, $\Pr[E] = 1 - \Pr[\bar{E}]$. If E_1 and E_2 are events, then $E_1 \wedge E_2$ denotes their conjunction; i.e., $E_1 \wedge E_2$ is the event that *both* E_1 and E_2 occur. By definition, $\Pr[E_1 \wedge E_2] \leq \Pr[E_1]$. Events E_1 and E_2 are said to be *independent* if $\Pr[E_1 \wedge E_2] = \Pr[E_1] \cdot \Pr[E_2]$.

If E_1 and E_2 are events, then $E_1 \vee E_2$ denotes the disjunction of E_1 and E_2; that is, $E_1 \vee E_2$ is the event that *either* E_1 *or* E_2 occurs. It follows from the definition that $\Pr[E_1 \vee E_2] \geq \Pr[E_1]$. The *union bound* is often a very useful upper bound of this quantity.

PROPOSITION A.7 (Union Bound)

$$\Pr[E_1 \vee E_2] \leq \Pr[E_1] + \Pr[E_2].$$

Repeated application of the union bound for any events $E_1, \ldots, E_k$ gives

$$\Pr\left[\bigvee_{i=1}^{k} E_i\right] \leq \sum_{i=1}^{k} \Pr[E_i].$$

The *conditional probability of* E_1 *given* E_2, denoted $\Pr[E_1 \mid E_2]$, is defined as

$$\Pr[E_1 \mid E_2] \stackrel{\text{def}}{=} \frac{\Pr[E_1 \wedge E_2]}{\Pr[E_2]}$$

as long as $\Pr[E_2] \neq 0$. (If $\Pr[E_2] = 0$ then $\Pr[E_1 \mid E_2]$ is undefined.) This represents the probability that event E_1 occurs, given that event E_2 has occurred. It follows immediately from the definition that

$$\Pr[E_1 \wedge E_2] = \Pr[E_1 \mid E_2] \cdot \Pr[E_2]\,;$$

equality holds even if $\Pr[E_2] = 0$ as long as we interpret multiplication by zero on the right-hand side in the obvious way.

We can now easily derive Bayes' theorem.

THEOREM A.8 (Bayes' Theorem) *If* $\Pr[E_2] \neq 0$ *then*

$$\Pr[E_1 \mid E_2] = \frac{\Pr[E_2 \mid E_1] \cdot \Pr[E_1]}{\Pr[E_2]}.$$

PROOF This follows because

$$\Pr[E_1 \mid E_2] = \frac{\Pr[E_1 \wedge E_2]}{\Pr[E_2]} = \frac{\Pr[E_2 \wedge E_1]}{\Pr[E_2]} = \frac{\Pr[E_2 \mid E_1] \cdot \Pr[E_1]}{\Pr[E_2]}.$$

■

Let $E_1, \ldots, E_n$ be events such that $\Pr[E_1 \vee \cdots \vee E_n] = 1$ and $\Pr[E_i \wedge E_j] = 0$ for all $i \neq j$. That is, the $\{E_i\}$ *partition* the space of all possible events, so that with probability 1 exactly one of the events E_i occurs. Then for any F

$$\Pr[F] = \sum_{i=1}^{n} \Pr[F \wedge E_i].$$

A special case is when $n = 2$ and $E_2 = \bar{E}_1$, giving

$$\begin{aligned}\Pr[F] &= \Pr[F \wedge E_1] + \Pr[F \wedge \bar{E}_1] \\ &= \Pr[F \mid E_1] \cdot \Pr[E_1] + \Pr[F \mid \bar{E}_1] \cdot \Pr[\bar{E}_1].\end{aligned}$$

Taking $F = E_1 \vee E_2$, we get a tighter version of the union bound:

$$\begin{aligned}\Pr[E_1 \vee E_2] &= \Pr[E_1 \vee E_2 \mid E_1] \cdot \Pr[E_1] + \Pr[E_1 \vee E_2 \mid \bar{E}_1] \cdot \Pr[\bar{E}_1] \\ &\leq \Pr[E_1] + \Pr[E_2 \mid \bar{E}_1].\end{aligned}$$

Extending this to events $E_1, \ldots, E_n$ we obtain

PROPOSITION A.9

$$\Pr\left[\bigvee_{i=1}^{k} E_i\right] \le \Pr[E_1] + \sum_{i=2}^{k} \Pr[E_i \mid \bar{E}_1 \wedge \cdots \wedge \bar{E}_{i-1}].$$

* Useful Probability Bounds

We review some terminology and state probability bounds that are standard, but may not be encountered in a basic discrete mathematics course. The material here is used only in Section 7.3.

A (discrete, real-valued) random variable X is a variable whose value is assigned probabilistically from some finite set S of real numbers. X is non-negative if it does not take negative values; it is a 0/1-random variable if $S = \{0, 1\}$. The 0/1-random variables $X_1, \ldots, X_k$ are *independent* if for all $b_1, \ldots, b_k$ it holds that $\Pr[X_1 = b_1 \wedge \cdots \wedge X_k = b_k] = \prod_{i=1}^{k} \Pr[X_i = b_i]$.

We let $\mathsf{Exp}[X]$ denote the expectation of a random variable X; if X takes values in a set S then $\mathsf{Exp}[X] \stackrel{\text{def}}{=} \sum_{s \in S} s \cdot \Pr[X = s]$. One of the most important facts is that expectation is *linear*; for random variables $X_1, \ldots, X_k$ (with arbitrary dependencies) we have $\mathsf{Exp}[\sum_i X_i] = \sum_i \mathsf{Exp}[X_i]$. If X_1, X_2 are independent, then $\mathsf{Exp}[X_i \cdot X_j] = \mathsf{Exp}[X_i] \cdot \mathsf{Exp}[X_j]$.

Markov's inequality is useful when little is known about X.

PROPOSITION A.10 (Markov's inequality) *Let X be a non-negative random variable and $v > 0$. Then $\Pr[X \ge v] \le \mathsf{Exp}[X]/v$.*

PROOF Say X takes values in a set S. We have

$$\begin{aligned}
\mathsf{Exp}[X] &= \sum_{s \in S} s \cdot \Pr[X = s] \\
&\ge \sum_{x \in S,\, x < v} \Pr[X = s] \cdot 0 + \sum_{x \in S,\, x \ge v} v \cdot \Pr[X = s] \\
&= v \cdot \Pr[X \ge v].
\end{aligned}$$

■

The variance of X, denoted $\mathsf{Var}[X]$, measures how much X deviates from its expectation. We have $\mathsf{Var}[X] \stackrel{\text{def}}{=} \mathsf{Exp}[(X - \mathsf{Exp}[X])^2] = \mathsf{Exp}[X^2] - \mathsf{Exp}[X]^2$, and one can easily show that $\mathsf{Var}[aX + b] = a^2 \mathsf{Var}[X]$. For a 0/1-random variable X_i, we have $\mathsf{Var}[X_i] \le 1/4$ because in this case $\mathsf{Exp}[X_i] = \mathsf{Exp}[X_i^2]$ and so $\mathsf{Var}[X_i] = \mathsf{Exp}[X_i](1 - \mathsf{Exp}[X_i])$, which is maximized when $\mathsf{Exp}[X_i] = \frac{1}{2}$.

PROPOSITION A.11 (Chebyshev's inequality) *Let X be a random variable and $\delta > 0$. Then:*

$$\Pr[|X - \mathsf{Exp}[X]| \geq \delta] \leq \frac{\mathsf{Var}[X]}{\delta^2}.$$

PROOF Define the non-negative random variable $Y \stackrel{\text{def}}{=} (X - \mathsf{Exp}[X])^2$ and then apply Markov's inequality. So,

$$\begin{aligned}\Pr[|X - \mathsf{Exp}[X]| \geq \delta] &= \Pr[(X - \mathsf{Exp}[X])^2 \geq \delta^2] \\ &\leq \frac{\mathsf{Exp}[(X - \mathsf{Exp}[X])^2]}{\delta^2} = \frac{\mathsf{Var}[X]}{\delta^2}.\end{aligned}$$

■

The 0/1-random variables $X_1, \ldots, X_m$ are *pairwise independent* if for every $i \neq j$ and every $b_i, b_j \in \{0, 1\}$ it holds that

$$\Pr[X_i = b_i \ \wedge \ X_j = b_j] = \Pr[X_i = b_i] \cdot \Pr[X_j = b_j].$$

If $X_1, \ldots, X_m$ are pairwise independent then $\mathsf{Var}[\sum_{i=1}^m X_i] = \sum_{i=1}^m \mathsf{Var}[X_i]$. (This follows since $\mathsf{Exp}[X_i \cdot X_j] = \mathsf{Exp}[X_i] \cdot \mathsf{Exp}[X_j]$ when $i \neq j$, using pairwise independence.) An important corollary of Chebyshev's inequality follows.

COROLLARY A.12 *Let $X_1, \ldots, X_m$ be pairwise-independent random variables with the same expectation μ and variance σ^2. Then for every $\delta > 0$,*

$$\Pr\left[\left|\frac{\sum_{i=1}^m X_i}{m} - \mu\right| \geq \delta\right] \leq \frac{\sigma^2}{\delta^2 m}.$$

PROOF By linearity of expectation, $\mathsf{Exp}[\sum_{i=1}^m X_i/m] = \mu$. Applying Chebyshev's inequality to the random variable $\sum_{i=1}^m X_i/m$, we have

$$\Pr\left[\left|\frac{\sum_{i=1}^m X_i}{m} - \mu\right| \geq \delta\right] \leq \frac{\mathsf{Var}\left[\frac{1}{m} \cdot \sum_{i=1}^m X_i\right]}{\delta^2}.$$

Using pairwise independence, it follows that

$$\mathsf{Var}\left[\frac{1}{m} \cdot \sum_{i=1}^m X_i\right] = \frac{1}{m^2} \sum_{i=1}^m \mathsf{Var}[X_i] = \frac{1}{m^2} \sum_{i=1}^m \sigma^2 = \frac{\sigma^2}{m}.$$

The inequality is obtained by combining the above two equations. ■

Say 0/1-random variables $X_1, \ldots, X_m$ each provides an estimate of some fixed (unknown) bit b. That is, $\Pr[X_i = b] \geq 1/2 + \varepsilon$ for all i, where $\varepsilon > 0$.

We can estimate b by looking at the value of X_1; this estimate will be correct with probability $\Pr[X_1 = b]$. A better estimate can be obtained by looking at the values of $X_1, \ldots, X_m$ and taking the value that occurs the majority of the time. We analyze how well this does when $X_1, \ldots, X_m$ are pairwise independent.

PROPOSITION A.13 *Fix $\varepsilon > 0$ and $b \in \{0,1\}$, and let $\{X_i\}$ be pairwise-independent, 0/1-random variables for which $\Pr[X_i = b] \geq \frac{1}{2} + \varepsilon$ for all i. Consider the process in which m values $X_1, \ldots, X_m$ are recorded and X is set to the value that occurs a strict majority of the time. Then*

$$\Pr[X \neq b] \leq \frac{1}{4 \cdot \varepsilon^2 \cdot m}.$$

PROOF Assume $b = 1$; by symmetry, this is without loss of generality. Then $\mathsf{Exp}[X_i] = \frac{1}{2} + \varepsilon$. Let X denote the strict majority of the $\{X_i\}$ as in the proposition, and note that $X \neq 1$ if and only if $\sum_{i=1}^m X_i \leq m/2$. So

$$\begin{aligned}
\Pr[X \neq 1] &= \Pr\left[\sum_{i=1}^m X_i \leq m/2\right] \\
&= \Pr\left[\frac{\sum_{i=1}^m X_i}{m} - \frac{1}{2} \leq 0\right] \\
&= \Pr\left[\frac{\sum_{i=1}^m X_i}{m} - \left(\frac{1}{2} + \varepsilon\right) \leq -\varepsilon\right] \\
&\leq \Pr\left[\left|\frac{\sum_{i=1}^m X_i}{m} - \left(\frac{1}{2} + \varepsilon\right)\right| \geq \varepsilon\right].
\end{aligned}$$

Since $\mathsf{Var}[X_i] \leq 1/4$ for all i, applying the previous corollary shows that $\Pr[X \neq 1] \leq \frac{1}{4\varepsilon^2 m}$ as claimed. ■

A better bound is obtained if the $\{X_i\}$ are independent:

PROPOSITION A.14 (Chernoff bound) *Fix $\varepsilon > 0$ and $b \in \{0,1\}$, and let $\{X_i\}$ be independent 0/1-random variables with $\Pr[X_i = b] = \frac{1}{2} + \varepsilon$ for all i. The probability that their majority value is not b is at most $e^{-\varepsilon^2 m/2}$.*

A.4 The "Birthday" Problem

If we choose q elements $y_1, \ldots, y_q$ uniformly from a set of size N, what is the probability that there exist distinct i, j with $y_i = y_j$? We refer to the stated event as a *collision*, and denote the probability of this event by $\mathsf{coll}(q, N)$. This problem is related to the so-called *birthday problem*, which asks what size group of people we need such that with probability 1/2 some pair of people in the group share a birthday. To see the relationship, let y_i denote the birthday of the ith person in the group. If there are q people in the group then we have q values $y_1, \ldots, y_q$ chosen uniformly from $\{1, \ldots, 365\}$, making the simplifying assumption that birthdays are uniformly and independently distributed among the 365 days of a non-leap year. Furthermore, matching birthdays correspond to a collision, i.e., distinct i, j with $y_i = y_j$. So the desired solution to the birthday problem is given by the minimal (integer) value of q for which $\mathsf{coll}(q, 365) \geq 1/2$. (The answer may surprise you—taking $q = 23$ people suffices!)

In this section, we prove lower and upper bounds on $\mathsf{coll}(q, N)$. Taken together and summarized at a high level, they show that if $q < \sqrt{N}$ then the probability of a collision is $\Theta(q^2/N)$; alternately, for $q = \Theta(\sqrt{N})$ the probability of a collision is constant.

An upper bound for the collision probability is easy to obtain.

LEMMA A.15 *Fix a positive integer N, and say q elements $y_1, \ldots, y_q$ are chosen uniformly and independently at random from a set of size N. Then the probability that there exist distinct i, j with $y_i = y_j$ is at most $\frac{q^2}{2N}$. That is,*

$$\mathsf{coll}(q, N) \leq \frac{q^2}{2N}.$$

PROOF The proof is a simple application of the union bound (Proposition A.7). Recall that a *collision* means that there exist distinct i, j with $y_i = y_j$. Let Coll denote the event of a collision, and let $\mathsf{Coll}_{i,j}$ denote the event that $y_i = y_j$. It is immediate that $\Pr[\mathsf{Coll}_{i,j}] = 1/N$ for any distinct i, j. Furthermore, $\mathsf{Coll} = \bigvee_{i \neq j} \mathsf{Coll}_{i,j}$ and so repeated application of the union bound implies that

$$\begin{aligned} \Pr[\mathsf{Coll}] &= \Pr\left[\bigvee_{i \neq j} \mathsf{Coll}_{i,j}\right] \\ &\leq \sum_{i \neq j} \Pr[\mathsf{Coll}_{i,j}] = \binom{q}{2} \cdot \frac{1}{N} \leq \frac{q^2}{2N}. \end{aligned}$$

■

LEMMA A.16 *Fix a positive integer N, and say $q \leq \sqrt{2N}$ elements $y_1, \ldots, y_q$ are chosen uniformly and independently at random from a set of size N. Then the probability that there exist distinct i, j with $y_i = y_j$ is at least $\frac{q(q-1)}{4N}$. In fact,*

$$\mathsf{coll}(q, N) \geq 1 - e^{-q(q-1)/2N} \geq \frac{q(q-1)}{4N}.$$

PROOF Recall that a *collision* means that there exist distinct i, j with $y_i = y_j$. Let Coll denote this event. Let NoColl_i be the event that there is *no* collision among $y_1, \ldots, y_i$; that is, $y_j \neq y_k$ for all $j < k \leq i$. Then $\mathsf{NoColl}_q = \overline{\mathsf{Coll}}$ is the event that there is no collision at all.

If NoColl_q occurs then NoColl_i must also have occurred for all $i \leq q$. Thus,

$$\Pr[\mathsf{NoColl}_q] = \Pr[\mathsf{NoColl}_1] \cdot \Pr[\mathsf{NoColl}_2 \mid \mathsf{NoColl}_1] \cdots \Pr[\mathsf{NoColl}_q \mid \mathsf{NoColl}_{q-1}].$$

Now, $\Pr[\mathsf{NoColl}_1] = 1$ since y_1 cannot collide with itself. Furthermore, if event NoColl_i occurs then $\{y_1, \ldots, y_i\}$ contains i distinct values; so, the probability that y_{i+1} collides with one of these values is $\frac{i}{N}$ and hence the probability that y_{i+1} does *not* collide with any of these values is $1 - \frac{i}{N}$. This means

$$\Pr[\mathsf{NoColl}_{i+1} \mid \mathsf{NoColl}_i] = 1 - \frac{i}{N},$$

and so

$$\Pr[\mathsf{NoColl}_q] = \prod_{i=1}^{q-1} \left(1 - \frac{i}{N}\right).$$

Since $i/N < 1$ for all i, we have $1 - \frac{i}{N} \leq e^{-i/N}$ (by Inequality A.3) and so

$$\Pr[\mathsf{NoColl}_q] \leq \prod_{i=1}^{q-1} e^{-i/N} = e^{-\sum_{i=1}^{q-1}(i/N)} = e^{-q(q-1)/2N}.$$

We conclude that

$$\Pr[\mathsf{Coll}] = 1 - \Pr[\mathsf{NoColl}_q] \geq 1 - e^{-q(q-1)/2N} \geq \frac{q(q-1)}{4N},$$

using Inequality A.4 in the last step (note that $q(q-1)/2N < 1$). ■

A.5 *Finite Fields

We use finite fields only sparingly in the book, but we include a definition and some basic facts for completeness. Further details can be found in any textbook on abstract algebra.