Chapter 1

Discrete Event Spaces and Probabilities

In Part I, we are interested in algorithmic problems with *discrete* solution spaces that we solve by *randomized* algorithms. We thus need the formal basis of analyzing random processes for discrete event spaces.

We define $\mathbb{N} = \{1, 2, 3, \ldots\}$ to be the set of natural numbers and $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$ to be the set of natural numbers with zero. For $n \in \mathbb{N}$, we denote the set $\{1, \ldots, n\}$ by $[n]$. We use $\mathbb{R}_{\geq 0}$ to denote the set $\{x \in \mathbb{R} \mid x \geq 0\}$. Given a vector $x \in \mathbb{R}^n$, we use $x_1, \ldots, x_n$ to denote its entries, i.e., $x = (x_1, \ldots, x_n)^\mathsf{T}$. We denote the power set of a set M by 2^M.

1.1 Discrete Probability Spaces

This section corresponds to Chapter 1 in [MU05]. We define what we mean by probabilities and see basic notations and rules. Intuitively, we can think of probabilities as frequencies when observing a process for a long time. For example, the probability that we roll a six with a die is $1/6$ – we expect that one sixth of all rolls will be six if we observe the die infinitely long. Notice that we express probabilities by numbers between zero and one.

To define probabilities, we first need a formal notion of *events*: Everything that can happen in the process that we observe. The most basic type of event is the *elementary event*. When we roll a die, one of six elementary events happens. Based on these, we can define more complicated events like „rolling an even number". These can be described by sets of elementary events.

Definition 1.1. *Let Ω be a finite or countable set that we call* sample space. *The elements of Ω are* elementary events. *An* event *is a subset $A \subseteq \Omega$. The* complementary event *of A is $\overline{A} = \Omega \backslash A$.*

Recall that the set of all subsets of Ω is 2^Ω, the power set. Thus, every event A is an element of 2^Ω. If $A, B \in 2^\Omega$ are two events, then $A \cup B$ is the event that at least one

of A and B occurs, and $A \cap B$ is the event that both events occur. Recall that A and B are disjoint if $A \cap B = \emptyset$.

A *probability measure* assigns a value to every event – its probability. When events are disjoint, then their probability has to add up to the probability of the joint event.

Definition 1.2. *A* probability measure *or* probability *on Ω is a mapping $\mathbf{Pr}\colon 2^\Omega \to [0,1]$ that satisfies $\mathbf{Pr}(\Omega) = 1$ and is σ-additive. The latter means that*

$$\mathbf{Pr}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbf{Pr}(A_i)$$

holds for every countably infinite sequence $A_1, A_2, \ldots$ of pairwise disjoint events. We say that $(\Omega, \mathbf{Pr})$ is a discrete probability space*.*

We observe that $\mathbf{Pr}(\emptyset)$ is always zero because $\mathbf{Pr}(\Omega) = \mathbf{Pr}(\Omega \cup \emptyset) = 1 + \mathbf{Pr}(\emptyset)$. In the same manner, we can prove that

$$\mathbf{Pr}(A) = \sum_{a \in A} \mathbf{Pr}(\{a\}) \tag{1.1}$$

is true. Thus, the probability of an event A is the sum of the probabilities of the elementary events that A consists of. In fact, probability measures can equivalently be defined by setting $\mathbf{Pr}(\{a\})$ for all $A \in \Omega$, extending it to events by 4.1 and demanding that $\mathbf{Pr}(\Omega) = 1$.

Example 1.3. *We look at some examples for sample spaces and probability measures.*

- *When we model one roll of a die, we set $\Omega = \{1,2,3,4,5,6\}$ and $\mathbf{Pr}(i) = 1/6$ for all $i \in \Omega$. The event to roll an even number is $A = \{2,4,6\}$, and its probability is $\mathbf{Pr}(A) = \mathbf{Pr}(2) + \mathbf{Pr}(4) + \mathbf{Pr}(6) = 1/2$.*
- *When we model a fair coin toss, we set $\Omega = \{H, T\}$ for the elementary events that we get heads or tail, and let $\mathbf{Pr}(H) = \mathbf{Pr}(T) = 1/2$. We can also model multiple fair coin tosses. For two turns, we set $\Omega = \{(H,H),(H,T),(T,H),(T,T)\}$ for the four possible outcomes. Intuitively, they all have the same probability, i.e. we set $\mathbf{Pr}((H,H)) = \mathbf{Pr}((H,T)) = \mathbf{Pr}((T,H)) = \mathbf{Pr}((T,T)) = 1/4$. We observe that the event A to get heads in the first turn is still $1/2$: It is $A = \{(H,H),(H,T)\}$ and thus $\mathbf{Pr}(A) = \mathbf{Pr}((H,H)) + \mathbf{Pr}((H,T)) = 1/4 + 1/4 = 1/2$.*
- *Now we want to model two fair coin tosses, but we cannot distinguish the coins and throw them at the same time. Thus, we model $\Omega = \{\{H,H\},\{H,T\},\{T,T\}\}$ for the three elementary events that both coins come up heads, one shows heads and one tail, or both show tails. We still want to model fair coin tosses, so we now need different probabilities for the elementary events: We set $\mathbf{Pr}(\{H,H\}) = \mathbf{Pr}(\{T,T\}) = 1/4$ and $\mathbf{Pr}(\{H,T\}) = 1/2$.*

In two of the examples in 1.3, we assigned the same probability to every elementary event. In this case, the elementary events occur *uniformly at random*. When we say that we choose an element uniformly at random from t choices/events, we mean that each of the t events has probability $1/t$.

Application: Polynomial Tester (Part I: The Power of Random Decisions) Our first randomized algorithm tests whether two polynomials $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ are equal. The degree of the polynomials will be important, so let d be the maximum degree of f and g.

Our algorithm cannot see the formulas that describe f and g. Instead, it can send an $x \in \mathbb{R}$ to a *black box* and receives $f(x)$ and $g(x)$. The simplest randomized algorithm we can think of is to send an x chosen uniformly at random from $\mathbb{R}$ and to check if $f(x) = g(x)$ is true. If yes, we output that f and g are equal, if no, then we output that f and g are not equal. In the latter case, our algorithm has no error because we found proof that f and g are not equal. But what is the probability that f and g are not equal, but we still output yes?

We need a bit more knowledge about polynomials. Since f and g are polynomials, we know that $f - g$ is a polynomial, too. Furthermore, the degree of $f - g$ is bounded by d as well. The fact that we now crucially need is that a polynomial of degree at most d (that is not the zero polynomial) has at most d roots. Thus, $f(x) - g(x) = 0$ can only be true for d different values of x. That means that $f(x) = g(x)$ can only be true for d different values of x as well!

What is the probability that we pick one of these d values when choosing an x uniformly at random? In the way that we formulated our algorithm, we need a continuous probability space to model and answer this question. We would rather analyze a discrete probability space. So we change our algorithm: We choose x uniformly at random from a set of t possible values. Formally, we set $\Omega := \{1, \dots, t\}$ and $\mathbf{Pr}(x) = 1/t$. We do *worst case analysis*, so we assume that Ω contains as many roots as possible. To have a chance to give the right answer, we thus need $t \geq d + 1$.

Now the probability that we choose a root and falsely output that f and g are equal is d/t. If we set $t = 100d$, then the failure probability of our algorithm is $1/100$. This is true even though our algorithm only checked a single x! A deterministic algorithm could decide the question without any error by for checking $d + 1$ values whether $f(x) = g(x)$ is true. Our randomized version saves d of these checks at the cost of a small error.

Recall that a negative answer of our algorithm has no error, only a positive answer might be incorrect. We call algorithms of this time *randomized algorithms with one-sided error*.

Union bound and product spaces We often want to analyze the probability that one of some events occurs. This is also helpful when we want to analyze that none of a set of events occurs. For two events, we get the following lemma.

Lemma 1.4. *Let $(\Omega, \mathbf{Pr})$ be a discrete probability space and let $A, B \in 2^\Omega$ be events. It holds that*

$$\mathbf{Pr}(A \cup B) = \mathbf{Pr}(A) + \mathbf{Pr}(B) - \mathbf{Pr}(A \cap B).$$

Proof. We can apply σ-additivity twice to get that

$$\mathbf{Pr}(A \cup B) = \mathbf{Pr}(A) + \mathbf{Pr}(B \setminus A) = \mathbf{Pr}(A) + \mathbf{Pr}(B) - \mathbf{Pr}(A \cap B). \qquad \square$$

We should memorize two things. First, the probability for $A \cup B$ *can* be smaller than $\mathbf{Pr}(A) + \mathbf{Pr}(B)$. The two terms are only equal if A and B are disjoint. Second, $\mathbf{Pr}(A \cup B)$ can *not* be greater than $\mathbf{Pr}(A) + \mathbf{Pr}(B)$. It is always $\mathbf{Pr}(A \cup B) \leq \mathbf{Pr}(A) + \mathbf{Pr}(B)$. This simple fact can be extremely useful. It can be generalized to the union of a family $(A_i)_{i \in \mathbb{N}}$ of events.

Lemma 1.5 (Union Bound)**.** *For a finite or countably infinite sequence $A_1, A_2, \ldots$ of events, it holds that*

$$\mathbf{Pr}\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} \mathbf{Pr}(A_i).$$

Proof. Recall that $A \backslash B = \{a \in \Omega \mid a \in A, a \notin B\}$. Observe that $\mathbf{Pr}(A \backslash B) \leq \mathbf{Pr}(A)$ holds for any two events $A, B \in 2^{\Omega}$ by definition. Thus, we can use σ-additivity to obtain that

$$\mathbf{Pr}\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \mathbf{Pr}\left(\bigcup_{i=1}^{\infty} \Big(A_i \backslash \Big(\bigcup_{j=1}^{i-1} A_j\Big)\Big)\right) = \sum_{i=1}^{\infty} \mathbf{Pr}\Big(A_i \backslash \Big(\bigcup_{j=1}^{i-1} A_j\Big)\Big) \leq \sum_{i=1}^{\infty} \mathbf{Pr}(A_i).$$

□

Let $B_1, B_2, \ldots, B_\ell$ be bad events that we all want to *not* occur. The probability that at least one of them occurs is bounded by

$$\mathbf{Pr}\left(\bigcup_{i=1}^{\ell} B_i\right) \leq \sum_{i=1}^{\ell} \mathbf{Pr}(B_i).$$

by the Union Bound. Thus, the probability that no bad event occurs, which is the complementary event, has a probability of at most

$$1 - \sum_{i=1}^{\ell} \mathbf{Pr}(B_i).$$

Similarly, assume that we have a set of good events $G_1, G_2, \ldots, G_\ell$ that we all want to occur. For any B_i, the (bad) complementary event $\overline{G_i}$ has probability $1 - \mathbf{Pr}(G_i)$. The probability that none of them occurs is at most

$$\mathbf{Pr}\left(\bigcup_{i=1}^{\ell} \overline{G_i}\right) \leq \sum_{i=1}^{\ell} (1 - \mathbf{Pr}(G_i)).$$

Thus, the probability that the event 'at least one of $\overline{G_i}$ occurs' does *not* occur is at most $1 - \sum_{i=1}^{\ell}(1 - \mathbf{Pr}(G_i))$.

Application: Polynomial Tester (Part II: More than one bad event) Again, we want to test if two polynomials f and g are equal. This time, we only have access to a black box *with error*. Our algorithm can send a value $x \in \mathbb{R}$ to the black box and ask whether $f(x) = g(x)$ is true. With probability 0.9, the black box gives a correct

answer, and with probability 0.1, it replies incorrectly. Our algorithm does not change: It still chooses one of t possible values for x, sends x to the black box and then repeats the answer of the black box.

We model this situation with two probability spaces. The probability space $(\Omega_1, \mathbf{Pr}_1)$ models the random behavior of our algorithm. It is defined by $\Omega_1 := \{1, \ldots, t\}$ and $\mathbf{Pr}_1(x) = 1/t$ for all $x \in \Omega_1$. The random behavior of the black box is described by $(\Omega_2, \mathbf{Pr}_2)$, $\Omega_2 = \{R, W\}$ and $\mathbf{Pr}_2(R) = 0.9$, $\mathbf{Pr}_2(W) = 0.1$.

To analyze the whole process, we need to combine $(\Omega_1, \mathbf{Pr}_1)$ and $(\Omega_2, \mathbf{Pr}_2)$. We assume that the error of the black box is *independent* of the random behavior of our algorithm. We model this by using the *product space* $(\Omega, \mathbf{Pr})$.

Fact 1.6. *Let $(\Omega_1, \mathbf{Pr}_1)$ and $(\Omega_2, \mathbf{Pr}_2)$ be discrete probability spaces. Define the* product space *$(\Omega, \mathbf{Pr})$ by $\Omega = \Omega_1 \times \Omega_2$ and $\mathbf{Pr}(x, y) = \mathbf{Pr}(x) \cdot \mathbf{Pr}(y)$ for all $(x, y) \in \Omega$. Then $(\Omega, \mathbf{Pr})$ is a discrete probability space. Furthermore,*

$$\mathbf{Pr}_1(x) = \mathbf{Pr}(\{x\} \times \Omega_2) \text{ and } \mathbf{Pr}_2(y) = \mathbf{Pr}(\Omega_1 \times \{y\})$$

is true for all $x, y \in \Omega$.

The random behavior of our algorithm and the black box is jointly described by the product space of $(\Omega_1, \mathbf{Pr}_1)$ and $(\Omega_2, \mathbf{Pr}_2)$. Notice that the outcome of our algorithm now has *two-sided error*. When f and g are equal, $f(x) = g(x)$ is always true, independent of the x that we choose. However, the event $W \in \Omega_2$ that black box replies incorrectly might occur. Thus, the error probability of the algorithm in this case is $\mathbf{Pr}(\Omega_1 \times \{W\}) = \mathbf{Pr}(\{W\}) = 0.1$.

If f and g are not equal, then the probability that we choose one of the $k \leq d$ roots $r_1, \ldots, r_k$ as x from $\{1, \ldots, t\}$ is bounded by d/t. Let $A = \{r_1, \ldots, r_k\} \times \Omega_2$ be the event that this happens. Furthermore, let $B = \Omega_1 \times \{W\}$ be the event that the black box answers incorrectly. The events A and B are bad events, we want both of them to not occur. As we argued above, we can use the Union Bound to obtain

$$\mathbf{Pr}(A \cup B) \leq \mathbf{Pr}(A) + \mathbf{Pr}(B) = \mathbf{Pr}_1(\{r_1, \ldots, r_k\}) + \mathbf{Pr}_2(\{W\}) = \frac{k}{100d} + 0.1 \leq 0.11.$$

where we again set $t = 100d$.

In this example, we could have computed the failure probability exactly. However, we have learned a simple yet powerful tool that is often helpful when bounding the error probability of more complex randomized algorithms.

1.2 Independent Events & Conditional Probability

In the last section, we intuitively used the term *independent events*. We now formally define what we mean by independence.

Definition 1.7. *Let $(\Omega, \mathbf{Pr})$ be a disrete probability space. We say that two events $A \in 2^\Omega$ and $B \in 2^\Omega$ are* independent *if $\mathbf{Pr}(A \cap B) = \mathbf{Pr}(A) \cdot \mathbf{Pr}(B)$ holds. A sequence $A_1, \ldots, A_k$ of events is independent if*

$$\mathbf{Pr}\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} \mathbf{Pr}(A_i)$$

holds for every $I \subset \{1, \ldots, k\}$. The sequence $A_1, \ldots, A_k$ is pairwise independent *if A_i and A_j are independent for every $i, j \in \{1, \ldots, k\}$ with $i \neq j$.*

Intuitively, the independence of two events A and B means that we gain no information about B when we get to know whether A occurred, and we gain no information about A when we find out that B occurred.

Example 1.8. *Consider the probability space $(\Omega, \mathbf{Pr})$ which models two independent rolls of a die where we set $\Omega = \{(i, j) \mid i, j \in \{1, \ldots, 6\}\}$ and $\mathbf{Pr}(i, j) = 1/36$.*

- *Intuitively, the outcome of the first roll of a fair die has no consequence for the outcome of the second roll. Assume that we know that event $A = \{(2, j) \mid j \in \{1, \ldots, 6\}\} \cup \{(4, j) \mid j \in \{1, \ldots, 6\}\} \cup \{(6, j) \mid j \in \{1, \ldots, 6\}\}$ occurred, which is that the first roll gives an even number. The probability of this event is $\mathbf{Pr}(A) = 1/2$ because $|A| = 18$ and all elementary events have the same probability. Let B be the event that the second roll gives a 3, which has probability $1/6$. We observe that $A \cap B = \{(2, 3), (4, 3), (6, 3)\}$ has probability $1/12$. This confirms our intuitive idea that A and B are independent because*

$$\mathbf{Pr}(A \cap B) = 1/12 = (1/2) \cdot (1/6) = \mathbf{Pr}(A) \cdot \mathbf{Pr}(B).$$

- *Consider the event C that the sum of the two rolls is 8. This event is not independent of the event D that the first roll is a 1, since the occurrence of D means that C has probability zero. Consider the event E that the first roll is a 5. Again, E and C are not independent, because rolling a 5 and thus not rolling a 1 increases the probability that the sum is 8. We verify this intuition. We have $E \cap C = \{(5, 3)\}$ with probability $1/36$. Event E has probability $1/6$. Finally, the tuples $(i, 8 - i)$ for $i \in \{2, \ldots, 6\}$ are the possible outcomes with sum 8, so $\mathbf{Pr}(C) = 5/36$. We see that*

$$\mathbf{Pr}(E \cap C) = \frac{1}{36} > \mathbf{Pr}(E) \cdot \mathbf{Pr}(C) = \frac{5}{6 \cdot 36}$$

- *Now assume that F is the event that the sum is 7, and G is the event that we rolled a 6. From the last example, we might get the intuition that these events are not independent. However, at a second glance we see that $\mathbf{Pr}(F) = 1/6$, $\mathbf{Pr}(G) = 1/6$ and $\mathbf{Pr}(F \cup G) = 1/36$. Events F and G are indeed independent. We should always verify our intuition about independence by computing the probabilities, in particular in more complex scenarios.*