A
... Primer to Randomness

Version: 150

In this course we design several randomized algorithms, which requires basic knowledge of randomness and the analysis of random variables. In this section we give a short overview of the necessary concepts, omitting some proofs.

Recall that our machine model is the random access machine. For randomized algorithms, we assume access to a method RAND() that returns a random bit. Sometimes we also assume access to a method RAND(n), which returns a random number in $\{1, \dots, n\}$.

To formally argue about randomized algorithms, we need to define an underlying *probability space*.

Definition A.1. *A (finite)* ***probability space*** *consists of a finite set of outcomes Ω as well as a probability measure $\Pr\colon \Omega \to [0,1]$ satisfying*

$$\sum_{x \in \Omega} \Pr[x] = 1.$$

The function $\Pr$ is also called a ***distribution****. The* ***uniform*** *distribution over Ω is the measure $\Pr$ with $\Pr[x] = 1/|\Omega|$ for all $x \in \Omega$.*

Note that after k calls to RAND(), the probability space is the uniform distribution over $\Omega = \{0,1\}^k$, since every k-bit string is generated with the same probability. More generally, after calls RAND(n_1), ..., RAND(n_k) the probability space is the uniform distribution over $\{1, \dots, n_1\} \times \dots \times \{1, \dots, n_k\}$.

Definition A.2. *An* ***event*** *is a set $\mathcal{E} \subseteq \Omega$. The probability of event $\mathcal{E}$ is $\Pr[\mathcal{E}] = \sum_{x \in \mathcal{E}} \Pr[x]$.*

For example, for $\Omega = \{0,1\}^k$, an event would be that an even number of generated bits are 1, i.e., $\mathcal{E} = \{x \in \{0,1\}^k \mid \sum_{j=1}^{k} x_i \text{ is even}\}$.

For the uniform distribution over $\{0,1\}^k$, this event has probability $\frac{1}{2}$.

A probability space $D = (\Omega, \Pr)$ is also sometimes called a **distribution**. For an event $\mathcal{E}$, instead of $\Pr[\mathcal{E}]$ we often write

$$\Pr_{x \sim D}[x \in \mathcal{E}],$$

in order to stress that D is the underlying distribution and that x is the only source of randomness. Here, $x \sim D$ means that x is drawn at random with distribution D (i.e., x attains any $y \in \Omega$ with probability $\Pr[y]$).

All randomized algorithm that we design in this course are of the following form. The only source of randomness are calls to RAND() (in particular, the input to the algorithm is assumed to be worst-case, not random). For any instance I, let $\mathcal{E}_I$ be the event that the algorithm correctly solves I. Let $\mathcal{I}_n$ be the set of instances of size n. The **success probability** p_n of the algorithm is the minimum over all inputs I of size n of the probability that the algorithm correctly solves I, i.e., $p_n = \min_{I \in \mathcal{I}_n} \Pr[\mathcal{E}_I]$. We always ensure that **the success probability is at least** $\frac{2}{3}$. Typically, we even want to succeed **with high probability**, meaning that $p_n \geq 1 - \frac{1}{n}$. See Lemma A.11 below for how to boost a success probability from $\frac{2}{3}$ to $1 - \frac{1}{n}$.

Lemma A.3 (Union Bound)**.** *For any events $\mathcal{E}, \mathcal{E}'$ we have* $\Pr[\mathcal{E} \cup \mathcal{E}'] \leq \Pr[\mathcal{E}] + \Pr[\mathcal{E}']$.

Proof. We have

$$\sum_{x \in \mathcal{E} \cup \mathcal{E}} \Pr[x] \leq \sum_{x \in \mathcal{E}} \Pr[x] + \sum_{x \in \mathcal{E}'} \Pr[x],$$

since each element $x \in (\mathcal{E} \setminus \mathcal{E}') \cup (\mathcal{E}' \setminus \mathcal{E})$ is counted once on the left and once on the right and each element $x \in \mathcal{E} \cap \mathcal{E}'$ is counted once on the left and twice on the right. □

In the analysis of randomized algorithms, we typically consider **error events**, in which certain desired properties do not hold. We want to show that the combined probability of all error events is small, in order to show that the algorithm has a large success probability. To this end, it often suffices to bound the probability of each single error event, and then to use the union bound to get an upper bound on the probability that any error event occurs.

Definition A.4. *A **random variable** is a function $X\colon \Omega \to \mathbb{N}$.*
*The **expectation** of random variable X is*

$$\mathrm{Ex}[X] := \sum_{x\in\Omega} X(x)\cdot \Pr[x].$$

We denote $\mathbb{N} = \{0,1,2,\ldots\}$.

For instance, the number of times we have to flip a coin until we see heads is a random variable. Its expectation is 2.

Lemma A.5. *For $n \in \mathbb{N}$ we use '$X = n$' as shorthand notation for the event $\{x \in \Omega \mid X(x) = n\}$, similarly for '$X \geq n$'. With this notation, the expectation satisfies the following identities:*

$$\mathrm{Ex}[X] = \sum_{n=0}^{\infty} n\cdot \Pr[X=n] = \sum_{n=1}^{\infty} \Pr[X\geq n].$$

The following is a very useful property of expectations.

Lemma A.6 (Linearity of Expectation)**.** *For any random variables X, Y and constants $\alpha, \beta \in \mathbb{N}$ we have* $\mathrm{Ex}[\alpha\cdot X + \beta\cdot Y] = \alpha\cdot\mathrm{Ex}[X] + \beta\cdot\mathrm{Ex}[Y]$.

Proof. We have

$$\begin{aligned}\mathrm{Ex}[\alpha X+\beta Y] &= \sum_{x\in\Omega}(\alpha X(x)+\beta Y(x))\Pr[x]\\ &= \alpha\sum_{x\in\Omega} X(x)\Pr[x] + \beta\sum_{x\in\Omega} Y(x)\Pr[x]\\ &= \alpha\mathrm{Ex}[X]+\beta\mathrm{Ex}[Y].\end{aligned}$$

□

We often want to bound the probability that a random variable attains a very large value. Markov's inequality yields a basic bound of this kind.

Lemma A.7 (Markov Inequality)**.** *For any random variable $X\colon \Omega \to \mathbb{N}$ and any $t > 0$ we have*

$$\Pr[X\geq t] \leq \frac{\mathrm{Ex}[X]}{t}.$$

Proof. This follows from

$$\mathrm{Ex}[X] = \sum_{n=1}^{\infty}\Pr[X\geq n] \geq \sum_{n=1}^{t}\Pr[X\geq n] \geq \sum_{n=1}^{t}\Pr[X\geq t] = t\cdot\Pr[X\geq t].$$

□

Definition A.8. *Two events $\mathcal{E}, \mathcal{E}'$ are called **independent** if* $\Pr[\mathcal{E}\cap\mathcal{E}'] = \Pr[\mathcal{E}]\cdot\Pr[\mathcal{E}']$.
*Two random variables X, Y are called **independent** if for any $x, y \in \mathbb{N}$ the events $X \leq x$ and $Y \leq y$ are independent.*

The outcomes of different calls to RAND() are assumed to be independent. In particular, running the same algorithm twice (with fresh randomness) yields independent outcomes.

Definition A.9. *An **indicator random variable** (or short indicator variable) is a random variable $I\colon \Omega \to \{0,1\}$. We have $\mathrm{Ex}[I] = \Pr[I = 1]$.*

In other words, an indicator variable is a random variable that never takes values greater than 1. Note that an indicator variable I is essentially the same as the event $I = 1$ (one could say that I indicates the event $I = 1$). In fact, indicator variables and events are two notations for the same objects. However, it is sometimes useful to work with indicator variables, since this enables statements such as the following.

Lemma A.10 (Chernoff Bound)**.** *Let $I_1, \ldots, I_n$ be independent indicator variables and let $X := \sum_{j=1}^{n} I_j$. Then for all $t > 0$ we have*

$$\Pr[X \geq \mathrm{Ex}[X] + t] \leq \exp\left(-\frac{2t^2}{n}\right).$$

By symmetry, the same holds for $\Pr[X \leq \mathrm{Ex}[X] - t]$.

The Chernoff bound is a very strong concentration inequality; we omit its proof. Many variations with different error bounds as well as generalization that try to relax the independence assumption are known. We will often use the following corollary.

Lemma A.11 (Boosting)**.** *Let $I_1, \ldots, I_n$ be independent indicator variables with $\mathrm{Ex}[I_j] \geq \frac{2}{3}$ for all j. Then the majority of $I_1, \ldots, I_n$ is 1 with probability at least $1 - \exp(-\frac{n}{20})$.*

Proof. By linearity of expectation, for $X := \sum_{j=1}^{n} I_j$ we have

$$\mathrm{Ex}[X] = \mathrm{Ex}[I_1] + \ldots + \mathrm{Ex}[I_n] \geq \frac{2n}{3}.$$

Observe that the majority is different from 1 only if $X \leq \frac{n}{2}$. This happens only if $X - \mathrm{Ex}[X] \leq \frac{n}{2} - \frac{2n}{3} = -\frac{n}{6}$. The Chernoff bound with $t := \frac{n}{6}$ now yields a bound of $\exp(-\frac{2n}{36}) \leq \exp(-\frac{n}{20})$. □

We typically use boosting as follows. Suppose that we have a randomized algorithm with success probability at least $\frac{2}{3}$. Run this algorithm $20\ln(1/\delta)$ times and use majority vote among all outcomes. Then the resulting success probability is at least $1 - \exp(-\frac{20\ln(1/\delta)}{20}) = 1 - \delta$. This success probability can be made arbitrarily close to 1 by picking an appropriate δ. In particular, we obtain a with-high-probability guarantee by picking $\delta = \frac{1}{n}$, and

this only requires $\mathcal{O}(\log n)$ repetitions of the original algorithm.

Bibliography

Version: 150