

Post(pre)liminaries: What to know before we start

Algorithms

For a general refresher on “classical” algorithms, one cannot recommend enough Tim Roughgarden’s lecture series and textbook, which comes with videos and programming assignments⁵⁵ Another good reference book is Jeff Erickson’s (free) textbook, which covers a lot of material and can be obtained on the author’s website.⁵⁶

Make sure you are familiar with the definition and use of big-Oh (asymptotic) notation: $O()$, $\Omega()$, $\Theta()$. This is, for instance, Chapter 2 of Tim Roughgarden’s book (volume 1).⁵⁷ Below, we point out two particular topics the student may want to reacquaint themselves with.

Median in Linear Time. There is an algorithm, based on the *median-of-median* divide-and-conquer approach, which computes the median of an array of n numbers in time $O(n)$. For more on this, see, Chapter 1.8 of Erickson’s textbook or Chapter 6 of Tim Roughgarden’s book (volume 1). This is also a good occasion to practice solving recurrence relations (with or without the Master Theorem).

Max-Flow and Min-Cut. This course assumes familiarity with algorithms for Max Flow and Min-Cut (in particular in Chapter 5). For a refresher, see, Chapters 10 and 11 of Erickson’s textbook.

(Some) discrete maths

The point of this section is not to be comprehensive, but to give a few facts that you either should know; and, for most, be able to prove. If some seem unfamiliar (or you would be at a loss trying to establish them), it is alright! But please ask questions during the class or on the course online forum, so that we can help you with them (and related concepts). First, the factorial grows as $n! = n^{\Theta(n)}$, or, equivalently,

$$\log n! = \Theta(n \log n).$$

More precisely, *Stirling’s approximation* (harder to show) gives

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

⁵⁵ The textbooks are not free, but the videos and resources can be found at <https://www.algorithmsilluminated.org/>.

⁵⁶ Jeff Erickson. *Algorithms*. 2019

⁵⁷ We might encounter $o()$ at some point, even a couple paragraphs below, but this one will be explained when (or if) needed.

Stirling’s approximation

where $o(1)$ means “something that goes to 0 as $n \rightarrow \infty$.” From this, you can deduce for instance that⁵⁸

$$\binom{2n}{n} = \Theta\left(\frac{2^{2n}}{\sqrt{n}}\right).$$

⁵⁸ Exercise: check it!

Other useful statements include that the n -th Harmonic number, $H_n = \sum_{k=1}^n \frac{1}{k}$, is asymptotically

Harmonic numbers

$$H_n = \ln n + O(1)$$

while $\sum_{k=1}^n \frac{1}{k^2} = O(1)$, since

$$\sum_{k=1}^n \frac{1}{k^2} \leq \sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6}$$

(this one is actually quite difficult to show), and

$$\sum_{k=1}^n \frac{1}{\sqrt{k}} = \Theta(\sqrt{n}).$$

We will also take for granted the formula for geometric series: for $r \neq 1$,

$$\sum_{k=0}^n r^k = \frac{r^{n+1} - 1}{r - 1}$$

and so, for $|r| < 1$, $\sum_{k=0}^{\infty} r^k = \frac{1}{1-r}$. More complicated, it will be helpful to know how to show, for $|r| < 1$, that

$$\sum_{k=1}^{\infty} k r^k = \frac{r}{(1-r)^2}$$

Try it as a (very much not obvious) exercise: if you are stuck, don't worry – but *do ask* how to do it!

To conclude this sample of things worth knowing or remembering, recall the *inequality of arithmetic and geometric means (AM-GM)*: for any $x_1, \dots, x_n \geq 0$,

AM-GM inequality

$$(x_1 x_2 \cdots x_n)^{1/n} \leq \frac{x_1 + \cdots + x_n}{n}$$

and in particular, if $a, b \geq 0$, then $\sqrt{ab} \leq \frac{a+b}{2}$.

Probability

First, some notation. Throughout, we'll use

- $\Pr[\]$ for probabilities
- $\mathbb{E}[\]$ for expectations
- $\text{Var}[\]$ for variances
- $X \sim \pi$ to denote that a random variable (r.v.) has probability distribution π

so, for instance, $\Pr[X \geq 5]$ is the probability that the random variable X is at least 5 (assuming the probability distribution of X is known or clear from context), while $\Pr_{X \sim \pi}[X \geq 5]$ or $\Pr_{\pi}[X \geq 5]$ specifies in addition that X is distributed according to some probability distribution π .

Now, a *probability distribution* π over some set \mathcal{X} is a beast that satisfies a collection of mathematical axioms, which we will not recall in detail here.⁵⁹ When \mathcal{X} is discrete (finite or countable), as it will mostly be the case here, we can for our purposes think of a probability distribution π as a non-negative function defined on subsets of \mathcal{X} , such that

$$\pi(\emptyset) = 0, \pi(\mathcal{X}) = 1$$

(the probability of “nothing” is 0, the probability of “everything” is 1); and, for any countable collection of pairwise disjoint subsets $S_1, \dots, S_k, \dots \subseteq \mathcal{X}$,

$$\pi\left(\bigcup_{k=1}^{\infty} S_k\right) = \sum_{k=1}^{\infty} \pi(S_k)$$

where a set $S \subseteq \mathcal{X}$ is also called an *event* (so “the probabilities of disjoint events add up”). In particular, if S, T are disjoint events ($S \cap T = \emptyset$), then

$$\pi(S \cup T) = \pi(S) + \pi(T).$$

More generally, if S, T are arbitrary (not necessarily disjoint), then⁶⁰

$$\pi(S \cup T) = \pi(S) + \pi(T) - \pi(S \cap T).$$

When \mathcal{X} is continuous, for instance \mathbb{R} , and π is “well-behaved” (almost always) we have similar properties, replacing sums (\sum) with integrals (\int).

Random variables. A random variable (r.v.) X over domain \mathcal{X} will be specified by its probability distribution π , where, for any (measurable) event $S \subseteq \mathcal{X}$,

$$\pi(S) = \Pr[X \in S]$$

In the discrete case, say, when $\mathcal{X} = \{1, 2, \dots, n\}$ or $\mathcal{X} = \mathbb{N}$, we can fully specify the probability distribution of X by its *probability mass function* (pmf), writing

$$\pi(i) = \Pr[X = i],$$

and then (from the axioms above), we have $\pi(S) = \sum_{i \in S} \pi(i)$.

Most of the time, we will deal with random variables taking either integer or real values, that is, $\mathcal{X} \subseteq \mathbb{R}$. For such an r.v. X , we define the *cumulative distribution function* (cdf)

$$F_X(t) = \Pr[X \leq t], \quad t \in \mathbb{R}$$

which fully characterizes the distribution of the random variable X . (If two r.v.’s have the same cdf, they have the same probability distribution.)

⁵⁹ For a start, if we want to be precise, it’s not “just” a set \mathcal{X} , but a measurable space (\mathcal{X}, Σ) , where Σ is a σ -algebra (a collection of subsets of \mathcal{X} closed under complement, countable unions, and countable intersections). We won’t go there, as in our case the whole thing will be clear from context.

⁶⁰ Exercise: prove it using the previous rule, applied to (1) $S \cup T = (S \setminus T) \cup T$ and (2) $S = (S \setminus T) \cup (S \cap T)$.

For instance, “the probability that X is an even number” is $\pi(\{0, 2, 4, 6, \dots\}) = \pi(0) + \pi(2) + \pi(4) + \pi(6) + \dots$

Indicator random variables An important specific case of random variables are those “indicating an event” by taking a value in $\{0, 1\}$: that is, for an event $A \subseteq \mathcal{X}$, we write $\mathbb{1}_A$ for the random variable which is 1 if A happens, and 0 otherwise. As an example: say X is a random variable uniformly distributed on $\{1, 2, 3, 4\}$. Then

$$\mathbb{1}_{X \text{ even}}$$

is equal to 1 if $X = 2$ or $X = 4$, and 0 if $X = 1$ or $X = 3$.

Independence. Two random variables X, Y are *independent* if

$$\Pr[X \in S, Y \in T] = \Pr[X \in S] \cdot \Pr[Y \in T] \quad \forall S, T.$$

This is equivalent to having

$$\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)]\mathbb{E}[g(Y)]$$

for *all* functions f, g (whenever the expectations are well defined).

In particular, that implies that $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.

This extends to more than two r.v.’s in the “straightforward way”: X_1, \dots, X_n are *mutually independent* (or just *independent*) if

$$\Pr[X_1 \in S, X_2 \in S_2, \dots, X_n \in S_n] = \prod_{k=1}^n \Pr[X_k \in S_k] \quad \forall S_1, \dots, S_n.$$

This is equivalent to having

$$\mathbb{E}\left[\prod_{k=1}^n f_k(X_k)\right] = \prod_{k=1}^n \mathbb{E}[f_k(X_k)]$$

for *all* functions f_1, \dots, f_n (again, whenever the expectations are well defined).

Bayes’ Rule, Conditional Expectation, and Law of Total Expectation.

Given two events A, B (and a probability distribution), the *conditional probability of A knowing B* (or *conditioned on B*) is given by

$$\Pr[A | B] = \frac{\Pr[A \cap B]}{\Pr[B]} \quad (\text{Bayes’ Rule})$$

For instance, if X is a random variable uniform on $\{1, 2, \dots, 2n\}$, then the “probability that X is prime conditioned on X being even” is

$$\Pr[X \text{ prime} | X \text{ even}] = \frac{\Pr[X \text{ is an even prime}]}{\Pr[X \text{ even}]} = \frac{\Pr[X = 2]}{\Pr[X \in \{2, 4, 6, \dots, 2n\}]} = \frac{\frac{1}{2n}}{n \cdot \frac{1}{2n}} = \frac{1}{n}$$

(note that 2 is the only even prime.)

This can be used to define *conditional expectations*:⁶¹ for an event A ,

$$\mathbb{E}[X | A] = \sum_{x \in \mathcal{X}} x \cdot \Pr[X = x | A] = \frac{\mathbb{E}[X\mathbb{1}_A]}{\Pr[A]}$$

(this is for the discrete case: again, the continuous case is similar, with \int instead of \sum and a bit of extra care.)

⁶¹ We are *not* giving here a full formal treatment of conditional expectations: that would require measure theory and a lot more paper. See, e.g., MATH4069 at the University (or ask after the class) if you’re curious.

We can even go one step further and define conditional expectations with respect to *random variables*, not events. Given two random variables X, Y , we can write

$$\mathbb{E}[X | Y]$$

for the new random variable “ X conditioned on Y .” We will not define it formally here, but this random variable corresponds to “averaging over whatever randomness remains in X once you know Y ”, and satisfies a few important properties: the first is the *Law of Total Expectation*,

Law of Total Expectation

$$\mathbb{E}[\mathbb{E}[X | Y]] = \mathbb{E}[X]$$

which basically states that “taking expectations step by step” gives the same result as doing it in one go.

The other is that, whenever X and Y are independent then knowing Y does not change anything about X , and so conditioning on Y still requires you to “average over all the randomness of X ”:

$$\mathbb{E}[X | Y] = \mathbb{E}[X] \quad (\text{if } X, Y \text{ independent})$$

The other extreme is when X is a deterministic function of Y . Then knowing Y means there is no other randomness left at play, and $f(Y)$ acts like a constant (fixed number) as far as the conditional expectation is concerned:

$$\mathbb{E}[f(Y) | Y] = f(Y) \quad (\text{for any function } f)$$

For instance, $\mathbb{E}[Y^2 | Y] = Y^2$; or, if X, Y are independent, combining both facts above, $\mathbb{E}[Y^2 X | Y] = Y^2 \mathbb{E}[X | Y] = Y^2 \mathbb{E}[X]$. Besides the above, conditional expectations basically behave like “normal” expectations, but adding “ $| \cdot]$ ” at the end.

Some probability distributions. For $p \in (0, 1]$, the *Geometric distribution with parameter p* , $\text{Geom}(p)$, is the discrete distribution over $\mathbb{N} = \{1, 2, \dots, n, \dots\}$ with probability mass function

$$\pi(k) = p(1 - p)^{k-1}$$

for $k \in \mathbb{N}$. It has expectation $1/p$ and variance $(1 - p)/p^2$.

For $\lambda \geq 0$, the *Poisson distribution with parameter λ* , $\text{Poi}(\lambda)$, is the discrete distribution over $\mathbb{N} \cup \{0\} = \{0, 1, 2, \dots, n, \dots\}$ with probability mass function

$$\pi(k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

for $k \in \mathbb{N} \cup \{0\}$. It has expectation λ and variance λ .

For $p \in [0, 1]$, the *Bernoulli distribution with parameter p* , $\text{Bern}(p)$, is the discrete distribution over $\{0, 1\}$ with probability mass function

$$\pi(0) = 1 - p, \quad \pi(1) = p$$

It has expectation p and variance $p(1 - p)$.

For integer $n \geq 1$ and $p \in [0, 1]$, the *Binomial distribution* with parameters n and p , $\text{Bin}(n, p)$, is the discrete distribution over $\{0, 1, 2, \dots, n\}$ with probability mass function

$$\pi(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

for $0 \leq k \leq n$. It has expectation np and variance $np(1-p)$.

The Bernoulli distribution is a “baby” Binomial distribution with $n = 1$.