CS 466/666: Algorithm Design and Analysis **University of Waterloo: Fall 2025**

Lecture 2

September 9, 2024

Instructor: Sepehr Assadi

Topics of this Lecture

This lecture, for the most part, is just a refresher on basic (discrete) probabilistic background that we need for the rest of the course.

1 Probabilistic Background

We review basic probabilistic background that will be used in this course using a simple running example: Consider the probabilistic process of rolling two dice and observing the result.

- **Probability space**: the set of all possible outcomes of the probabilistic process.

  Here, all the 36 combinations of answers, i.e.,

  $$(1,1),(1,2),(1,3),\cdots,(6,5),(6,6).$$

- **Event**: any subset of the probability space.

  Here, one example of an event would be 'sum of the two dice is equal to 7'; another example is 'both dice rolled an even number'.

- **Probability distribution**: an assignment of $\Pr(e) \in [0,1]$ to every element e of the probability space such that $\sum_e \Pr(e) = 1$.

  Here, every one of the 36 elements of the probability space have the same probability $1/36$, i.e.,

  $$\Pr((1,1)) = \Pr((1,2)) = \Pr((1,3)) = \cdots = \Pr((6,6)) = \frac{1}{36}.$$

- **Probability**: for any event E, probability of E is $\Pr(E) = \sum_{e \in E} \Pr(e)$, i.e., the sum of the probabilities assigned by the probability distribution to the elements of this event.

Here, for example,

$$\Pr\left(\text{sum of the two dice is 7}\right) = \sum_{e\in\{(1,6),(6,1),(2,5),(5,2),(3,4),(4,3)\}} \Pr\left(e\right) = 6\cdot\frac{1}{36} = \frac{1}{6};$$

$$\Pr\left(\text{both dice roll even}\right) = \sum_{e\in\{(2,2),(2,4),(2,6),(4,2),(4,4),(4,6),(6,2),(6,4),(6,6)\}} \Pr\left(e\right) = 9\cdot\frac{1}{36} = \frac{1}{4}.$$

- **Random variable**: any function from the elements of the probability space to integers or reals. I.e., a random variable X is a function $X:\Omega\to\mathbb{R}$, where Ω is the probability space.

  Here, an example of a random variable X is the sum of the two dice, e.g., $X((1,3))=4$, $X((2,5))=7$. Another example is a random variable Y that assigns one to the events that both dice roll even and is zero otherwise, e.g. $Y((2,2))=1$ and $Y((2,3))=0$ (this type of random variable that assigns 1 to a particular event and is zero otherwise is called an *indicator* random variable for that event).

- **Independence**: we say two events E_1 and E_2 are independent of each other, denoted by $E_1\perp E_2$, whenever $\Pr(E_1\cap E_2)=\Pr(E_1)\cdot\Pr(E_2)$.

  For instance, the events 'the first die rolls even' and 'the second die rolls odd' are independent of each other, but the events 'the first die rolls even' and 'both dice rolls even' are not independent.

  Similarly, two random variables X and Y are independent, denoted by $X\perp Y$, if for any values a, b, the events $X=a$ and $Y=b$ are independent. Alternatively, conditioning on any value of X (resp. Y), does not change the distribution of Y (resp. X).

- **Expected value**: the expected value of a random variable X, denoted by $\mathbb{E}\left[X\right]$, is the average of its value *weighted* according to the probability distribution, i.e., $\mathbb{E}\left[X\right]=\sum_e \Pr\left(e\right)\cdot X(e)$.

  Here, for the random variables X and Y defined two bullets above, we have,

$$\mathbb{E}\left[X\right] = \sum_e \Pr\left(e\right)\cdot X(e) = \sum_{i\in\{2,\ldots,12\}} \Pr\left(X=i\right)\cdot i = 7;$$

$$\mathbb{E}\left[Y\right] = \sum_e \Pr\left(e\right)\cdot Y(e) = \sum_e \Pr\left(Y(e)=1\right) = \frac{1}{4}.$$

  A very important property of expected value is its *linearity*, often called **linearity of expectation**: for any two random variables X, Y,

$$\mathbb{E}\left[X+Y\right] = \mathbb{E}\left[X\right] + \mathbb{E}\left[Y\right].$$

  This is because

$$\mathbb{E}\left[X+Y\right] = \sum_e \Pr\left(e\right)\cdot\left(X(e)+Y(e)\right) = \sum_e \Pr\left(e\right)X(e) + \sum_e \Pr\left(e\right)Y(e) = \mathbb{E}\left[X\right] + \mathbb{E}\left[Y\right].$$

  For *independent* random variables, we further have that expectation is *multiplicative*, i.e., for any two random variables X, Y where $X\perp Y$,

$$\mathbb{E}\left[X\cdot Y\right] = \mathbb{E}\left[X\right]\cdot\mathbb{E}\left[Y\right].$$

  This is because

$$\begin{aligned}\mathbb{E}\left[X\cdot Y\right] &= \sum_x\sum_y \Pr\left(X=x\wedge Y=y\right)\cdot\Big(x\cdot y\Big) = \sum_x\sum_y \Pr\left(X=x\right)\Pr\left(Y=y\right)\cdot\Big(x\cdot y\Big)\\ &= \left(\sum_x \Pr\left(X=x\right)\cdot x\right)\cdot\left(\sum_y \Pr\left(Y=y\right)\cdot y\right) = \mathbb{E}\left[X\right]\cdot\mathbb{E}\left[Y\right],\end{aligned}$$

  where we used the independence of X and Y in the second equality.

- **Variance**: the variance of a random variable X, denoted by $\mathrm{Var}[X]$, is a measure of the 'distance' of an average value of X, from the expected value of X. More accurately,

$$\mathrm{Var}[X] := \mathbb{E}\left[(X - \mathbb{E}[X])^2\right].$$

Variance of a random variable is a measure of its 'spread': the larger the variance, the more likely that the random variable takes a value 'far' from its expectation. A simple calculation implies

$$\mathrm{Var}[X] = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathbb{E}\left[X^2\right] - 2 \cdot \mathbb{E}[X \cdot \mathbb{E}[X]] + (\mathbb{E}[X])^2 = \mathbb{E}\left[X^2\right] - (\mathbb{E}[X])^2.$$

Note that unlike expectation, variance in general is *not* linear. However, for two *independent* random variables X, Y, it will be linear, i.e.,

$$\begin{aligned}
\mathrm{Var}[X+Y] &= \mathbb{E}\left[(X+Y)^2\right] - (\mathbb{E}[X+Y])^2 \\
&= \mathbb{E}\left[X^2\right] + \mathbb{E}[Y]^2 + 2 \cdot \mathbb{E}[X \cdot Y] - (\mathbb{E}[X])^2 - (\mathbb{E}[Y])^2 - 2 \cdot \mathbb{E}[X] \cdot \mathbb{E}[Y] \\
&\qquad\qquad \text{(by expanding the terms)} \\
&= \mathbb{E}\left[X^2\right] - (\mathbb{E}[X])^2 + \mathbb{E}\left[Y^2\right] - (\mathbb{E}[Y])^2 \qquad (\text{as } \mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y] \text{ when } X \perp Y) \\
&= \mathrm{Var}[X] + \mathrm{Var}[Y].
\end{aligned}$$

- **Conditioning**: Given two events E and F in the same probability space, the *conditional* probability of E given that F has happened is:

$$\Pr(E \mid F) = \frac{\Pr(E \wedge F)}{\Pr(F)}.$$

For instance, in our example, E can be 'sum of the two dice is 8' and F can be 'both dice roll even', in which case

$$\Pr(E \mid F) = \frac{\sum_{e \in (2,6),(4,4),(6,2)} \Pr(e)}{\sum_{e \in (2,2),(2,4),\ldots,(6,6)} \Pr(e)} = \frac{3/36}{1/4} = \frac{1}{3}.$$

Another way of looking at this is that conditioned on F, we have a new probability distribution q with probability space on 9 elements $\{(2,2),(2,4),(2,6),(4,2),(4,4),(4,6),(6,2),(6,4),(6,6)\}$ and each have the same probability, and thus we are measuring probability of E under this new distribution which is

$$\Pr(E \mid F) = \sum_{e \in (2,6),(4,4),(6,2)} \Pr(e \mid F) = \sum_{e \in (2,6),(4,4),(6,2)} q(e) = \frac{3}{9} = \frac{1}{3}.$$

We can define a conditional expectation of a random variable X given an *event* F also as

$$\mathbb{E}[X \mid F] = \sum_x \Pr(X = x \mid F) \cdot x;$$

that is the expectation of the random variable X under the new probability distribution defined by conditioning on F.

Finally, we have the following important **law of total probability**: for any set of (disjoint) events $F_1, \ldots, F_k$ that partition the probability space, and any other event E, we have,

$$\Pr(E) = \sum_{i=1}^{k} \Pr(E \mid F_i) \cdot \Pr(F_i).$$

This is simply because

$$\sum_{i=1}^{k} \Pr(E \mid F_i) \cdot \Pr(F_i) = \sum_{i=1}^{k} \Pr(E \wedge F_i) = \sum_{i=1}^{k} \sum_{e \in F_i \cap E} \Pr(e) = \sum_{e \in E} \Pr(e) = \Pr(E),$$

where the first equality is by the definition of conditional probability and the third one is because the sets $E \cap F_i$ are disjoint and partition E entirely.

In our example, we can take F_1 as the event that 'both dice roll even' and F_2 as the event that 'at least one die rolls odd', to partition the probability space and write the probability of any event E as

$$\Pr(E) = \Pr(E \mid F_1) \cdot \Pr(F_1) + \Pr(E \mid F_2) \cdot \Pr(F_2) = \Pr(E \mid F_1) \cdot \frac{1}{4} + \Pr(E \mid F_2) \cdot \frac{3}{4}.$$

And to conclude, we have the **law of total expectation** as a corollary of the above: for any two random variables X, Y:

$$\mathbb{E}[X] = \sum_y \Pr(Y = y) \cdot \mathbb{E}[X \mid Y = y].$$

The above is an extremely short refresher of the probabilistic background. This cannot by any means replace a proper introduction to this amazing concept. You are strongly encouraged to take a look back at the materials from your previous courses on probability, or the further reading materials on the course page.

2 Concentration Inequalities

When working with random variables, perhaps the easiest way to "summarize" the variable is to focus on its expected value. However, expected value on its own can often be misleading: for instance, consider a random variable which is 0 with probability $1/2$ and is 1 with the remaining value. Expected value of X is $1/2$ but of course we do not 'expect' X to ever take the value of $1/2$! This, and many other examples, suggest that summarizing a random variable just down to its expectation may lose too much information.

On the other extreme, a random variable can be uniquely identified by its probability distribution:

$$\mathbb{P}_X : k \to [0, 1] \qquad \text{such that} \qquad \mathbb{P}_X(a) = \Pr(X = a).$$

Yet, the distribution of even very simple random variables can be quite cumbersome to work with. Consider, for instance, the simple example of throwing a fair coin 100 times and defining X to be the number of heads. Here, for every $a \in \{0, \ldots, 100\}$,

$$\Pr(X = a) = \binom{100}{a} \cdot 2^{-100},$$

which, even in this simple form, is rather tedious to work with.

Concentration inequalities are a saving grace between these two extremes: morally speaking (but not strictly speaking true), they allow us to extract (perhaps, the most) "important" information about the distribution of our random variables, without getting to compute the very precise distribution itself. More accurately, they allow us to bound the probability of *deviation* of a random variable from its expectation (as a function of its distance from the expectation).

We will study various concentration inequalities in the course of this term, as they arise quite frequently in the analysis of randomized algorithms (and way beyond). This lecture, includes two of the most basic and highly applicable ones.

2.1 Markov's Inequality

The simplest and most basic variant of concentration results is **Markov's inequality** or **Markov bound**:

Proposition 1 (**Markov Bound**). *For a* non-negative *random variable X and $t > 0$,*

$$\Pr(X \geqslant t \cdot \mathbb{E}[X]) \leqslant \frac{1}{t}.$$

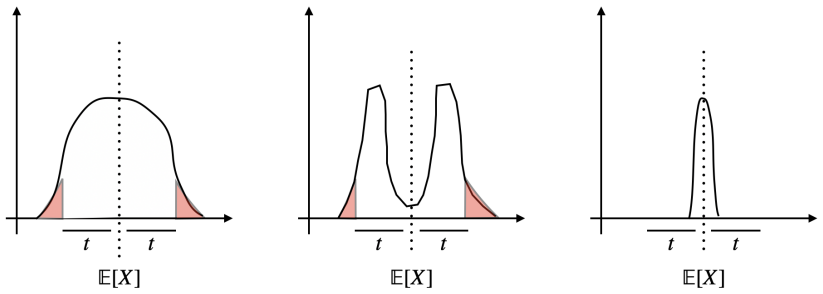

Figure 1: An illustration of three different random variables and their probability distributions. Here, all these variables have the same expected value. Moreover, the first two variables show roughly the same concentration for the particular choice of t; i.e., for the first two variables, the probability that their values are more than t away from the expectation is almost the same (the probability is the part shaded in red) even though the distributions are quite different; however, the third variable is much more concentrated.

Proof. Let $\mu := \mathbb{E}[X]$. We can use the law of total expectation to have:

$$\begin{aligned}\mathbb{E}[X] &= \mathbb{E}[X \mid X \geqslant t \cdot \mu] \cdot \Pr(X \geqslant t \cdot \mu) + \mathbb{E}[X \mid X < t \cdot \mu] \cdot \Pr(X < t \cdot \mu) \\ &\geqslant t \cdot \mu \cdot \Pr(X \geqslant t \cdot \mu) + 0.\end{aligned}$$

(the first term since we conditioned on $X \geqslant t \cdot \mu$ and the second term since X is non-negative)

Thus, $\Pr(X \geqslant t \cdot \mathbb{E}[X]) = \Pr(X \geqslant t \cdot \mu) \leqslant 1/t$, otherwise the RHS above will be larger than the LHS. □

Markov bound only bounds the *upper tail* of the distribution[1]: the probability that a random variable takes value t times larger than its expectation is at most $1/t$. This is a basic but extremely useful property. One can alternatively state the Markov bound as follows.

Corollary 2. *For a* non-negative *random variable X and $b > 0$,*

$$\Pr(X \geqslant b) \leqslant \frac{\mathbb{E}[X]}{b}.$$

Proof. The proof is by simply picking $t = b/\mathbb{E}[X]$ in Proposition 1. □

Note that a random variable X, which, with probability $\mathbb{E}[X]/b$ takes the value b and otherwise is 0 will be a tight example for Markov bound; i.e., one cannot expect to improve Markov bound in general.

> **Remark.** Even though Markov bound may sound almost trivial (and it is indeed straightforward), it is the basis for proving all other concentration inequalities that we use in this course; moreover, Markov bound is used one way or another in analysis of almost every randomized algorithm.

2.2 Chebyshev's Inequality

We now consider our second concentration inequality: **Chebyshev's inequality**. Unlike Markov bound that only required a knowledge of the expected value of the random variable to bound its deviation probability, Chebyshev's inequality applies to the settings in which we could additionally bound the *variance* of the random variable as well.[2]

[1] Although one can use it to bound the lower tail in special cases as well, but the bounds there are generally very weak.

[2] As we showed earlier, Markov bound *can* be tight for certain random variables; thus, naturally whenever we need a stronger bound we should show that our variable satisfies additional guarantees that what is only required by Markov bound.

Recall that for a random variable X, **variance** of X is:

$$\mathrm{Var}[X] := \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathbb{E}\left[X^2\right] - (\mathbb{E}[X])^2.$$

Chebyshev inequality allows us to bound deviation of a random variable based on its variance.

Proposition 3 (**Chebyshev's Inequality**). *For any random variable X and $t > 0$,*

$$\Pr\left(|X - \mathbb{E}[X]| \geqslant t \cdot \mathbb{E}[X]\right) \leqslant \frac{\mathrm{Var}[X]}{\mathbb{E}[X]^2 \cdot t^2}.$$

Proof. Define a new random variable $Y := (X - \mathbb{E}[X])^2$. Clearly, Y is non-negative. Moreover, $|X - \mathbb{E}[X]| \geqslant t \cdot \mathbb{E}[X]$ if and only if $Y = (X - \mathbb{E}[X])^2 \geqslant t^2 \cdot \mathbb{E}[X]^2$. Hence,

$$\begin{aligned}
\Pr\left(|X - \mathbb{E}[X]| \geqslant t \cdot \mathbb{E}[X]\right) = \Pr\left(Y \geqslant t^2 \cdot \mathbb{E}[X]^2\right) &\leqslant \frac{\mathbb{E}[Y]}{\mathbb{E}[X]^2 \cdot t^2} && \text{(by Markov bound of Corollary 2)} \\
&= \frac{\mathrm{Var}[X]}{\mathbb{E}[X]^2 \cdot t^2}. && \text{(as } \mathbb{E}[Y] = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathrm{Var}[X] \text{ by definition)}
\end{aligned}$$

□

A useful variant of Chebyshev's inequality is the following.

Corollary 4. *For a random variable X and $b > 0$,*

$$\Pr\left(|X - \mathbb{E}[X]| \geqslant b\right) \leqslant \frac{\mathrm{Var}[X]}{b^2}.$$

Proof. The proof is by simply picking $t = b/\mathbb{E}[X]$ in Proposition 3. □