Probability II

Matthew Gherman and Adam Lott
Inspired by the textbook *Introduction to Probability* by Dimitri Bertsekas and John Tsitsiklis

1 March 2020

4 Random variables

4.1 Basics

In many random experiments, the outcomes themselves are not literally numbers, but we want to associate some numerical information to the outcome because numbers are things that we know how to analyze. Also, depending on what particular aspects of the outcomes we care about, we might want to associate many different pieces of numerical information to them. For example, suppose we flip a coin 100 times. Formally, the outcomes are strings of Hs and Ts of length 100, but there are plenty of numerical data we might care about. For example, maybe we care about the total number of Hs that were flipped, or the length of the longest string of consecutive Ts, or the number of times the particular sequence HTTHT occurred. A **random variable** is a way of assigning numerical information to the outcome of a random experiment. We describe this formally by saying that a random variable X is any function $X : \Omega \to \mathbb{R}$.

Exercise 4.1. Think of a few different random experiments and come up with some examples of random variables for those experiments that you might care about.

Definition 4. If X is a random variable on the probability space $(\Omega, \mathbb{P})$, then we can define what is called the **distribution** of X (also sometimes called the **probability mass function** of X or the **law** of X). The distribution of X, denoted dist_X, can be thought of as a "function" [1] that assigns to each $a \in \mathbb{R}$ the number $\text{dist}_X(a) := \mathbb{P}\{\omega \in \Omega : X(\omega) = a\}$ (this second expression will also be written as $\mathbb{P}(X = a)$ for short). We can also extend the definition to apply to any subset $B \subseteq \mathbb{R}$, meaning that we will define $\text{dist}_X(B) := \mathbb{P}\{\omega \in \Omega : X(\omega) \in B\} = \mathbb{P}(X \in B)$.

Exercise 4.2. (a) Flip three fair coins and let X be the total number of tails. Find $\text{dist}_X(2)$.

(b) Roll two fair six-sided dice and let X be the product of the two rolls. Find $\text{dist}_X(4)$.

[1] You might be wondering why the word "function" appears in quotes, since what we have defined as the distribution is literally a function $\mathbb{R} \to \mathbb{R}$. The answer is that (a) this definition is no longer appropriate when Ω is a continuous probability space, and (b) we want to think of the distribution as something that assigns a numerical value to every *set* in $\mathbb{R}$, not just every point in $\mathbb{R}$. More on this topic coming in week 3.

(c) Roll two fair six-sided dice and let Y be the sum of the two rolls. Find dist_Y. (This means that you should find all possible values for which dist_Y is nonzero, and calculate the dist_Y at those places.)

(d) (B&T page 119) The UCLA soccer team has two games scheduled for one weekend. It has a 0.4 probability of not losing the first game, and a 0.7 probability of not losing the second game, independent of the first. If it does not lose a particular game, the team is equally likely to win or tie, independent of what happens in the other game. The UCLA team will receive two points for a win, one for a tie, and zero for a loss. Let Z be the total number of points the team earns in the weekend. Find dist_Z.

Exercise 4.3 (B & T page 120)**.** You just rented a large house and the realtor gave you five keys, one for each of the five doors of the house. Unfortunately, all keys look identical, so to open the front door, you try them at random.

(a) Find the distribution of the number of trials you need to open the front door if after each unsuccessful trial, you mark the corresponding key so that you never try it again.

(b) Same question, but on each trial you are equally likely to choose any key.

Exercise 4.4. Flip a fair coin 10 times and let X be the total number of heads and Y be the total number of tails. Show that X and Y are not the same random variable but they do have the same distribution.

Exercise 4.5. For each of the following, find an example of a probability model $(\Omega, \mathbb{P})$ and a random variable X which has the given distribution.

(a) $\text{dist}_X(0) = 1/9$, $\text{dist}_X(1) = 4/9$, $\text{dist}_X(2) = 4/9$ (NOTE: there is always an obvious answer – let $\Omega = \{0, 1, 2\}, \mathbb{P}(0) = 1/9, \mathbb{P}(1) = \mathbb{P}(2) = 4/9$, and $X(\omega) = \omega$. Try to find an example that corresponds to a physical situation.)

(b) $\text{dist}_X(k) = (1/2)^k$ for all $k = 1, 2, 3, \ldots$

(c) $\text{dist}_X(k) = \binom{10}{k}(1/4)^k(3/4)^{10-k}$ for all $k = 0, 1, 2, \ldots, 10$.

Remark 3. The previous two exercises point out an important concept. Sometimes random variables which correspond to very different sources of randomness (the underlying experiment) can produce the same distribution. In many cases, the *distribution itself is much more important than the actual source of randomness and choice of variable that produced it.* If this confuses you, don't worry, we'll come back to it later.

Exercise 4.6. Let X be any random variable. Prove that

$$\sum_{a\in\mathbb{R}} \text{dist}_X(a) \;=\; 1.$$

(Recall that since Ω is a countable set, X can take only countably many values. So you should interpret the notation $\sum_{a\in\mathbb{R}} f(a)$ as "sum over those values a for which $f(a)$ is not zero".)

4.2 Expected value

Given a random variable X, its distribution dist_X gives us complete information about what values X can take and how likely each value is. But sometimes it is easier to deal with "coarser" information about X that can be described with a single number. The notion of the **expected value** of a random variable is a way of describing the "average value" that you expect a random variable to take.

As an example, consider spinning a spinner that lands on 1 with probability 1/2, 2 with probability 1/4, and 3 with probability 1/4, and let X be the number the spinner lands on. What does the "average value" of X mean? One way to interpret this is to repeat the same experiment lots and lots of times, and take the average of all the results you get. Suppose we spin the same spinner N times, where N is very large. Then we would expect the spinner to land on 1 about $N/2$ of those times and on 2 and 3 each about $N/4$ times. So the average value of all the results would be about

$$\frac{1 \cdot (N/2) + 2 \cdot (N/4) + 3 \cdot (N/4)}{N} = 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} + 3 \cdot \frac{1}{4}.$$

This says that the average value we expect to get from X over many repeated trials can just be written as a weighted average of the possible values of X, where the weights are given by the probabilities of each value appearing. Let's now turn this idea into a definition.

Definition 5. Let X be a random variable. The **expected value** (or **expectation**) of X is defined to be the number

$$\mathbb{E}[X] := \sum_{a \in \mathbb{R}} a \cdot \mathrm{dist}_X(a).$$

Exercise 4.7. (a) Calculate the expectation of a single roll of a fair six-sided die.

(b) Flip three fair coins independently and let Y be the total numer of heads. Calculate $\mathbb{E}[Y]$.

(c) (B & T page 122) Fix positive integers $a \le b$. Let Z be a random variable that takes as values, with equal probability, the powers of 2 in the interval $[2^a, 2^b]$. Calculate $\mathbb{E}[Z]$.

(d) (CHALLENGE) Suppose you have a biased coin that lands on heads with probability p and tails with probability $1 - p$. Flip the coin repeatedly (each flip is independent) until getting heads for the first time, and let X be the total number of flips required. Calculate $\mathbb{E}[X]$ and interpret the result.

Exercise 4.8 (B & T page 123)**.** A prize is placed uniformly at random in one of N boxes, numbered 1 through N. You search for the prize by asking yes-no questions. Find the expected number of questions until you are sure of the location of the prize under each of the following strategies.

(a) Enumeration – you ask questions of the form "is it in box k?"

(b) Bisection – you ask questions of the form "is it in a box numbered less than or equal to k?"

Exercise 4.9 (B & T page 91)**.** You are on a quiz show and you are given two questions to answer, and you must decide which one to answer first. Question 1 will be answered correctly with probability 0.8 and a correct answer yields a prize of $100. Question 2 will be answered correctly with probability 0.5 and a correct answer yields a prize of $200. If you answer the first question wrong, you lose and don't get to attempt the second question. Assume your answers to the questions are independent of each other. Determine the best strategy if:

(a) Your goal is to maximize your expected earnings.

(b) Your goal is to maximize your chances of winning something.

Exercise 4.10 (The St. Petersburg paradox)**.** Flip a fair coin repeatedly until getting heads for the first time. If the first heads appears on the nth flip, you win 2^n dollars.

(a) Calculate your expected winnings.

(b) Suppose there is an entry fee for this game and you are allowed to play as many times as you want. Based on your answer to part (a), would you be willing to pay $10/game to play? What about $10,000,000,000/game? Explain why the answer dictated by part (a) is not a very good practical strategy.

Exercise 4.11. Let X be a random variable that only takes positive integer values. Prove that

$$\mathbb{E}[X] = \sum_{k=1}^{\infty} \mathbb{P}(X \geq k).$$

4.3 Functions of random variables

Given a random variable X, we can apply some transformation $f : \mathbb{R} \to \mathbb{R}$ to it to create a new random variable $f(X)$. For example, if X is the number of heads seen in a sequence of 10 coin flips, then X^2, $2X + 5$, and $\log(|\sin X|)$ are all just other random variables on the same probability space. In this section we will see that if you know information about X, then it's not too hard to determine information about any $f(X)$ also.

Exercise 4.12 (B & T page 122)**.** Let X be a random variable which takes the values 0 through 9 each with probability 1/10. Find the distributions of the random variables

(a) $X \bmod 3$

(b) $5 \bmod (X + 1)$

Exercise 4.13. Let X be any random variable and $f : \mathbb{R} \to \mathbb{R}$ be any transformation. Prove that

$$\mathbb{E}[f(X)] = \sum_{a \in \mathbb{R}} f(a)\, \mathrm{dist}_X(a)$$

and interpret this formula.
NOTE: this formula is *extremely* useful because it says that as long as you know the distribution of X, you can calculate the expectation of any function of X without having to recalculate the new distribution.

We will now introduce another very important statistic associated to a random variable, called the **variance**. While the purpose of the expected value is to determine the average "size" of X, the purpose of the variance is to quantify how "spread out" the distribution of X is by looking at how far away X is from its expected value.

Definition 6. The **variance** of a random variable X is defined as

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

Exercise 4.14. Let X take the values ± 1 each with probability $1/2$, and let Y take the values ± 100 each with probability $1/2$. Verify that $\mathbb{E}[X] = \mathbb{E}[Y] = 0$ but that $\text{var}(Y)$ is much larger than $\text{var}(X)$. This example shows how variance captures the notion of how spread out the distribution of a random variable is.

Exercise 4.15. Prove that $\text{var}(X) \geq 0$ and $\text{var}(X) = 0$ if and only if X is deterministic, that is there is some $a \in \mathbb{R}$ so that $\mathbb{P}(X = a) = 1$.

Exercise 4.16. Prove the alternate formula

$$\text{var}(X) \ = \ \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2.$$

Use this to deduce the interesting inequality $\mathbb{E}\left[X^2\right] \geq \mathbb{E}[X]^2$. NOTE: in practice, this formula for the variance is usually much easier to calculate than the original defintion.

Exercise 4.17. Let X be any random variable and let $c > 0$ be a number. Calculate the expectation and variance of the random variables cX and $X + c$ in terms of $\mathbb{E}[X]$ and $\text{var}(X)$.

Exercise 4.18 (OPTIONAL)**.** Let Y take the values $k = 1, 2, 3, \ldots$ with probability Z/k^3 (where Z is just the normalization constant). Show that $\mathbb{E}[Y]$ is finite but $\text{var}(Y)$ is infinite.

5 Collections of many random variables

5.1 Joint distributions

Often in an experiment, there are many different random variables that we care about, and we also want to keep track of how they relate to each other. In this section we will learn about how to extend the theory from above to study many random variables simultaneously.

Definition 7. Let X and Y be two random variables on the same probability space. We can consider the pair (X, Y) to be a kind of "random variable" that takes values in $\mathbb{R}^2$ instead of in $\mathbb{R}$. We then can define the **joint distribution** of X and Y to be the function $\text{dist}_{X,Y}$ which assigns to each pair $(a, b) \in \mathbb{R}^2$ the probability $\text{dist}_{X,Y}(a, b) = \mathbb{P}(X = a \text{ and } Y = b)$.

Exercise 5.1. Flip three fair coins. Let X be the total number of heads and Y be the length of the longest string of consecutive tails. Calculate $\text{dist}_{X,Y}$. It may be easiest to organize this information in a table.

Exercise 5.2. Given the joint distribution of two random variables $\text{dist}_{X,Y}$, show that we can recover the individual distributions of X and Y by the formulas

$$\text{dist}_X(a) \ = \ \sum_{b \in \mathbb{R}} \text{dist}_{X,Y}(a, b), \qquad \text{dist}_Y(b) \ = \ \sum_{a \in \mathbb{R}} \text{dist}_{X,Y}(a, b).$$

The distributions of the individual random variables are sometimes called the **marginal** distributions of the joint distribution.

We can also perform transformations on pairs of random variables in the same way that we did on single random variables. If X, Y are two random variables and $g : \mathbb{R}^2 \to \mathbb{R}$ is any function, then $g(X, Y)$ is a new random variable. For example, $X^2 + Y^3$ or $\frac{X-Y}{2}$.

Exercise 5.3. As you might expect (refer back to exercise 4.13), we can calculate the expected value of any function of X and Y by just knowing the joint distribution. Prove the formula

$$\mathbb{E}\left[g(X, Y)\right] \ = \ \sum_{a,b \in \mathbb{R}} g(a, b) \, \text{dist}_{X,Y}(a, b).$$

Exercise 5.4. Let X and Y be any two random variables on the same probability space. Prove that $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$ and give an example to show that $\mathbb{E}[XY] \neq \mathbb{E}[X]\mathbb{E}[Y]$ in general. The first property is known as the **linearity of expectation**.

5.2 Independence

Just as we had notions of conditioning and independence for pairs of events, we also have similar ideas for pairs of random variables.

Definition 8. Two random variables X and Y are **independent** if for every $a \in \mathbb{R}$ and $b \in \mathbb{R}$, the events $\{X = a\}$ and $\{Y = b\}$ are independent events.

Exercise 5.5. Prove that X and Y are independent if and only if $\text{dist}_{X,Y}(a,b) = \text{dist}_X(a) \cdot \text{dist}_Y(b)$ for all a, b.

Exercise 5.6. Prove that if X and Y are independent random variables, then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$. Give an example to show that even if X and Y are not independent, $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ is still possible. Random variables X, Y for which that formula holds are called **uncorrelated**. Also prove that if X and Y are independent, then $\text{var}(X+Y) = \text{var}(X) + \text{var}(Y)$, and give an example to show that this is not necessarily true if X and Y are not independent.

Exercise 5.7. So far we have defined everything only for two random variables X and Y. But everything we have done can be defined in a very similar way for any number of random variables $X_1, X_2, \ldots, X_n$.

(a) Prove that if $X_1, \ldots, X_n$ are independent random variables each with distribution $\text{Ber}(p)$, then $Y = X_1 + \ldots + X_n$ has distribution $\text{Binom}(n,p)$.

(b) Use the above fact to recalculate the expectation and variance of $\text{Binom}(n,p)$ in a much easier way.

Exercise 5.8 (OPTIONAL)**.** Prove that if $X \sim \text{Poi}(\lambda_1)$ and $Y \sim \text{Poi}(\lambda_2)$ and X and Y are independent, then $X + Y \sim \text{Poi}(\lambda_1 + \lambda_2)$.

6 Common distributions

6.1 Bernoulli distribution

Definition 9. Let p be a fixed number between 0 and 1. We say that a random variable X has the **Bernoulli distribution** with parameter p if

$$\mathbb{P}(X = 1) \; = \; p, \qquad \mathbb{P}(X = 0) \; = \; 1 - p.$$

This is often notated $X \sim \text{Ber}(p)$. This definition and notation are reflective of the philosophy mentioned above, that the distribution is more important than the random variable itself. Formally, X is defined on some probability space Ω, but we don't care if Ω has two, two billion, or infinity elements in it, and we don't care what the elements of Ω look like. All that matters is that when we look at the values of the random variable X, it comes out to 1 with probability p and 0 with probability $1 - p$.

Exercise 6.1. What physical situation does a Bernoulli distribution correspond to?

Exercise 6.2. Calculate the expectation and variance of the distribution $\text{Ber}(p)$ and interpret the expectation.

6.2 Geometric distribution

Definition 10. Let p be a number between 0 and 1. The **geometric distribution** with parameter p, denoted $\text{Geom}(p)$, is the distribution of a random variable X such that

$$\mathbb{P}(X = k) \; = \; p(1 - p)^{k-1} \qquad k = 1, 2, 3, \ldots$$

Exercise 6.3. Verify that this is a valid probability distribution.

Exercise 6.4. What physical situation does a geometric distribution correspond to?

Exercise 6.5 (CHALLENGE)**.** Calculate the expectation and variance of the distribution Geom(p) and interpret the expectation.

6.3 Binomial distribution

Definition 11. Let n be a positive integer and p be a number between 0 and 1. The **binomial distribution** with parameters n and p, denoted Binom(n, p), is the distribution of a random variable X such that

$$\mathbb{P}(X = k) \;=\; \binom{n}{k} p^k (1-p)^{n-k} \qquad k = 0, 1, 2, \ldots, n.$$

Exercise 6.6. Verify that this is a valid probability distribution.

Exercise 6.7. What physical situation does a binomial distribution correspond to?

Exercise 6.8 (CHALLENGE)**.** Calculate the expectation and variance of the distribution Binom(n, p) and interpret the expectation.

6.4 Optional – Poisson distribution

Definition 12. Let λ be a positive number. The **Poisson distribution** with parameter λ, denoted Poi(λ), is the distribution of a random variable X such that

$$\mathbb{P}(X = k) \;=\; e^{-\lambda} \cdot \frac{\lambda^k}{k!} \qquad k = 0, 1, 2, \ldots.$$

Exercise 6.9. Verify that this is a valid probability distribution.

Exercise 6.10. Calculate the expectation and variance of the distribution $\text{Poi}(\lambda)$.

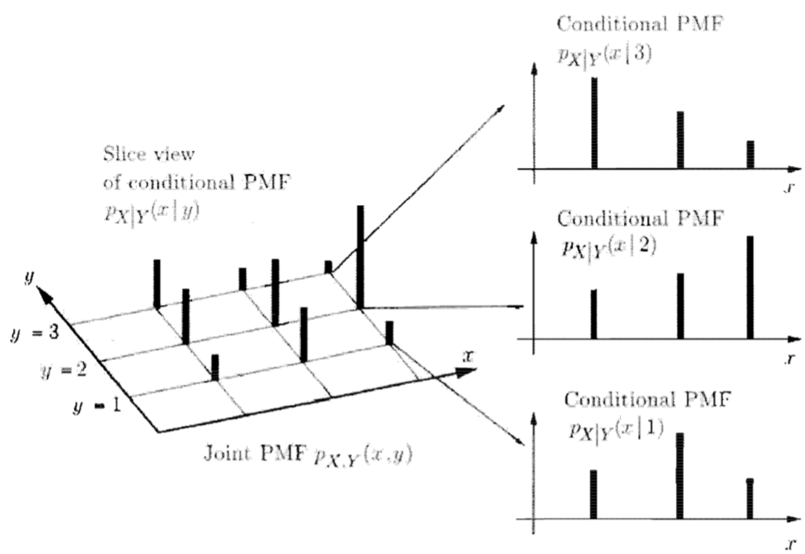
7 Conditioning

Definition 13. Let X and Y be two random variables defined on the same probability space. Let $b \in \mathbb{R}$ be a fixed value for which $\mathbb{P}(Y = b)$ is positive. We define the **conditional distribution of X given the event $\{Y = b\}$** to be the function $\text{dist}_{X|Y}(\cdot|b)$ [2] which assigns to each $a \in \mathbb{R}$ the number

$$\text{dist}_{X|Y}(a|b) \;=\; \frac{\mathbb{P}(X = a, Y = b)}{\mathbb{P}(Y = b)} \;=\; \frac{\text{dist}_{X,Y}(a, b)}{\text{dist}_Y(b)}.$$

You should think of the conditional distribution as saying: "Suppose we already know that Y took the value b. How does this change the likelihoods of X taking all of its different values?"

The picture below may help you visualize the situation.



Now that we have a notion of conditional distribution, we can define conditional expectation. The **conditional expectation of X given** $\{Y = b\}$ is defined by

$$\mathbb{E}[X|Y = b] \;=\; \sum_a a \cdot \text{dist}_{X|Y}(a|b).$$

Exercise 7.1. Roll two fair six-sided dice. Let X be the sum of the rolls and let Y be the result of the first roll. Calculate $\mathbb{E}[X|Y = b]$ for each $b = 1, 2, \ldots, 6$.

Exercise 7.2. Let X and Y be two random variables. Prove the **tower property of conditional expectation**:

$$\mathbb{E}[X] \;=\; \sum_b \mathbb{E}[X|Y = b] \cdot \text{dist}_Y(b)$$

[2]The "$\cdot$" is just a symbol that tells you where the argument of the function goes.

and interpret the formula.

Exercise 7.3 (B & T page 128)**.** A coin that has probability of heads equal to p is tossed successively and independently until a head comes twice in a row or a tail comes twice in a row. Find the expected value of the number of tosses.

Exercise 7.4 (B & T page 128)**.** A spider and a fly move along a straight line. At each second, the fly moves a unit step to the right or to the left with equal probability p, and stays where it is with probability $1 - 2p$. The spider always takes a unit step in the direction of the fly. The spider and the fly start D units apart, where D is a random number with distribution Geom$(1/2)$. If the spider lands on top of the fly, it's the end. What is the expected value of the time it takes for this to happen?