Math-UA.233: Theory of Probability
Lecture 10

Tim Austin

tim@cims.nyu.edu
cims.nyu.edu/~tim
Random Variables (Ross Ch 4)

For many experiments, we’re not really interested in knowing every possible detail about the outcome. Often the feature we really care about can be described by some numerical value.

**Definition**

*Let an experiment be described by sample space S. Then a random variable (or ‘RV’) is a real number whose value is determined by the outcome of the experiment. Mathematically, it is a function from S to the real numbers.*

Put another way, a random variable is ‘some quantity depending on the outcome that we can measure’.
Let $S$ be a sample space, let $P$ be a probability distribution on $S$, and let $X$ be a RV for this experiment.

For any real number $x$, we now have an associated event: the event that the value given by $X$ is equal to $x$. This is usually written $\{X = x\}$. Many basic questions involve calculating the probabilities of events of this kind. These may need any of the techniques we have learned so far in the course.
Example (Ross E.g. 4.1a)

Toss 3 fair coins. If \( Y \) is the number of heads that appear, then \( Y \) is a RV whose possible values are 0, 1, 2, and 3. We can compute the probabilities of the corresponding events:

\[
P(Y = 0) = P(\{TTT\}) = 1/8
\]

\[
P(Y = 1) = P(\{TTH, THT, HTT\}) = 3/8
\]

\[
P(Y = 2) = P(\{THH, HTH, HHT\}) = 3/8
\]

\[
P(Y = 3) = P(\{HHH\}) = 1/8.
\]

Observe:

\[
P(Y = 0 \text{ or } 1 \text{ or } 2 \text{ or } 3) \overset{\text{axiom}}{=} 3 \left( \frac{1}{8} + \frac{3}{8} + \frac{3}{8} + \frac{1}{8} \right) = 1.
\]
Example (Ross E.g. 4.1b)

A life insurance agent has two elderly clients. Each has a policy which pays $100,000 upon death. Let $Y$ (respectively, $O$) be the event that the younger (respectively, older) one dies in the following year. Assume these events are independent, with probabilities 0.05 and 0.1, respectively. Let $X$ denote the total amount of money (in units of $100,000) that will be paid out this year. Find the possible values that $X$ can take, and their associated probabilities.
Example (Ross E.g. 4.1d)

An urn contains 20 balls numbered 1 through 20. Four are selected at random without replacement. Let $X$ be the largest of their four numbers. Find the possible values that $X$ can take, and their associated probabilities.
Discussion: events defined in terms of RVs

Suppose an experiment is described by $S$ and $P$, but all we really care about is the value taken by some RV $X$.

For each real number $x$, there may be many individual outcomes in the event $\{X = x\}$: that is, many different ways that the value $x$ can come out of the experiment.

But if we don’t care about that extra information, then $X$ gives a natural choice for a ‘reduced’ sample space and probability distribution. Simply let $S'$ be the set of real numbers $\{-\infty < x < \infty\}$, and for $A \subset \{-\infty < x < \infty\}$ define

$$P'(A) = P(\{X \text{ takes a value in the set } A\}).$$
Some RVs have a huge range of possible values — indeed, some can take any real number value at all! In that case, this ‘reduced’ description of the experiment is still quite complicated.

But the key idea is that, for any subset $A$ of the real numbers you can think of, you get an associated event

$$\{X \text{ takes a value in the set } A\},$$

and these are all the events that can be defined by ‘a story you can tell only in terms of $X$’.
In practice, we use this point of view mostly for certain simple and natural sets $A$. We’ve already seen the events $\{X = x\}$. An alternative is to choose a real number $a$ as a ‘threshold’, and consider the event that $X$ takes a value no larger than $a$: that is, $\{X \leq a\}$. These appear often enough that we make the following definition.

**Definition (Ross p116)**

*If $X$ is a RV, then is cumulative distribution function (or ‘CDF’) is the function*

$$F(a) = P\{X \leq a\},$$

*where $a$ is any real number.*
Many other subsets of the real line can be expressed in terms of the sets \( \{ -\infty < x \leq a \} \) for different values of the threshold \( a \). As a result, we can find many other probabilities that concern \( X \) in terms of its CDF. For example:

**Lemma (Ross equation (10.1))**

If \( X \) is a RV, \( F \) is its CDF, and \( a < b \) are real values then

\[
P\{ a < X \leq b \} = F(b) - F(a).
\]
Random variables are intuitively simple, but some examples are technically very complicated. Special complications arise for variables that can take a continuous range of possible values.

We will overcome some of these difficulties later in the course, but for now we focus on a simpler class.

**Definition (Discrete random variables)**

A random variable $X$ is **discrete** if there is a (finite or infinite) list of real numbers $x_1, x_2, \ldots$ such that $X$ must always take a value from this list.
IMPORTANT MATHEMATICAL REMARK

Let $A$ be any set of real numbers: $A \subset \{-\infty < x < \infty\}$. Then we call $A$ **countable** if its elements can be ordered into a finite or infinite list.

Some sets have this property and others don’t. For instance, finite sets obviously have this property, and so does the subset of positive whole numbers: $A = \{1, 2, 3, \ldots \}$. How about the set of all whole numbers? Or the set of rational numbers?

It’s a remarkable fact that the set of all real numbers cannot be written as a list in any possible way. This is a theorem of Georg Cantor from 1891. The Wikipedia entry on ‘countable set’ is a good place to start reading about this.

A random variable is discrete if the set of values that it can possibly take is countable.
Definition

If $X$ is a discrete RV, then its probability mass function (or ‘PMF’) is the function

$$p(a) = P\{X = a\},$$

where $a$ is any real number.

Of course, if $a$ is not one of the possible values that $X$ can take, then $\{X = a\} = \emptyset$, and so $p(a) = 0$. 


Proposition (Ross p117)

If the possible values for $X$ are $x_1, x_2, \ldots, \text{then}$

(a) $p(x_i) \geq 0$ for every $i = 1, 2, \ldots$, and

(b) $\sum_i p(x_i) = 1$, where the sum is over that whole list of possible values.

IDEA: (a) comes from axiom 1. For part (b), the sets \{ $X = x_1$ \}, \{ $X = x_2$ \}, $\ldots$ form a partition of $S$. 
The thing that makes discrete RVs simple is the following:

**Proposition**

If $X$ is a discrete RV with possible values $x_1, x_2, \ldots$, and $A$ is any subset of the real numbers, then

$$P\left(\{X \text{ takes a value in the set } A\}\right) = \sum_{\text{all } i \text{ such that } x_i \text{ is in } A} p(x_i) = \sum_{\text{all values } a \text{ in } A \text{ such that } p(a)>0} p(a).$$

In particular, in this case the CDF $F$ of $X$ is given by

$$F(a) = \sum_{\text{all } i \text{ such that } x_i \leq a} p(x_i).$$
So for a *discrete* RV $X$, if we know its PMF (i.e. the probability of getting each individual possible value), then we can work out the probability of any other event that can be described in terms of $X$.

This is *not true* for arbitrary RVs. When we study ‘continuous’ RVs later in the course, we will meet many examples where the PMF is useless, and a different idea is needed.

Ultimately, this property is special to discrete RVs because axiom 3 of probability applies only to families of events that can be written in a *sequence*. Otherwise, we cannot make sense of the sum of probabilities as a convergent series.
Example (The sum-of-two-dice again)

Let $X$ be the sum when two fair dice are rolled. Its possible values are 2, 3, ..., 12. Plot its PMF as a graph.
Example (The sum-of-two-dice again, cont.)

Let $X$ be the sum when two fair dice are rolled. Plot its CDF as a function on the real line.
Example (Ross p 118)

Let $X$ be a discrete RV with PMF given by

$$p(1) = \frac{1}{4}, \quad p(2) = \frac{1}{2}, \quad p(3) = \frac{1}{8}, \quad p(4) = \frac{1}{8}.$$ 

Plot its CDF as a function on the real line.
Example (Ross E.g. 4.2a)

Let $X$ be a discrete RV with PMF given by $p(i) = c\lambda^i / i!$, $i = 0, 1, 2, \ldots$, where $c$ and $\lambda$ are two positive real parameters. Find (a) $P\{X = 0\}$ and (b) $P\{X > 2\}$.

(This is called a Poisson random variable, and we will study it in much more detail later.)
Suppose that $x_1, x_2, \ldots, x_n$ are some statistics (that is, just some real numbers). Their mean is

$$\bar{x} = \frac{x_1 + \cdots + x_n}{n}.$$

This is a kind of representative value around which the statistics are distributed.

This idea has an important abstraction in the setting of general experiments and RVs.

For now, we will develop it only for discrete RVs, which are avoid some technicalities of the general case.
Definition (Ross p119)

Let $X$ be a discrete RV with possible values $x_1, x_2, \ldots$ and PMF $p$. Then its **expectation** (or **expected value** or **mean**) is the number

$$E[X] = \sum p(x_i)x_i = \sum_{a \text{ such that } p(a)>0} p(a)a.$$
The relation to the ‘mean’ in statistics is clearest when $X$ has only finitely many possible values $x_1, x_2, \ldots, x_n$. Then

$$E[X] = p(x_1) \cdot x_1 + p(x_2) \cdot x_2 + \cdots + p(x_n) \cdot x_n.$$  

Since $p(x_1) + p(x_2) + \cdots + p(x_n) = 1$, this is just the average of the possible values that $X$ can take, *weighted* by the probabilities of getting those values.

**NOTE:** $X$ may have only finitely many possible values, even if $S$ is infinite. For instance, consider an infinite sequence of rolls of a die, and let $X$ be the value shown by the second roll. There are infinitely many possible outcomes for the experiment (i.e., sequences of numbers shown by the rolls), but this $X$ has six possible values.
IRRITATING MATHEMATICAL REMARK

If the set of possible values taken by $X$ is an \textit{infinite} list $x_1, x_2, \ldots$, then the expectation

$$E[X] = \sum_{i=1}^{\infty} p(x_i)x_i$$

must be interpreted as a convergent series. But how do we know that it converges?

Well, \textit{actually it might not}. For instance, $X$ could take the values 2, $-4$, 8, $-16$, \ldots with probabilities $1/2$, $1/4$, $1/8$, $1/16$, \ldots respectively, and then the above series is

$$\frac{2}{2} - \frac{4}{4} + \frac{8}{8} - \frac{16}{16} + \cdots = 1 - 1 + 1 - 1 + \ldots.$$ 

There's no clever way out of this: we just have to realize that some discrete RVs don't have well-defined expectations.

In this course, we will simply make sure not to work with those!
Example (Ross E.g. 4.3a)

Find $E[X]$ when $X$ is the outcome from rolling a fair die.
Digression: Why is expectation important?

(a) As with the mean in statistics, we can think of $E[X]$ as indicating where the values taken by $X$ ‘typically’ lie (even though $E[X]$ may not actually equal any of the possible values of $X$ — recall the die example). There are plenty of other quantities that can be used this way (such as ‘median’ and ‘mode’ in statistics). But the expectation has a better theory and more computational tools available, making it more useful to solve problems.

(b) Expectations turn out to be directly connected with long-run averages when we perform an experiment many independent times — this will follow from the Law of Large Numbers later in the course.