Math-UA.233: Theory of Probability
Lecture 8

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*** *('**RV**') is a real number X whose value is determined by the outcome of the experiment. Mathematically, it is a function*

$$X : S \longrightarrow \mathbb{R}.$$

Put another way, a RV is 'some quantity depending on the outcome that we can measure'.

Let

- S be a sample space,
- P be a probability distribution on S, and
- X be a RV for this experiment.

For any $x \in \mathbb{R}$, we now have an associated event: the event that the value given by X is equal to x. This may be written

$$\{s \in S \mid X(s) = x\} \quad \text{or just} \quad \{X = x\}.$$

We call x a **possible value** of X if $\{X = x\} \neq \emptyset$.

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:

$$P(Y=0) = P(\{\mathrm{TTT}\}) = 1/8$$

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

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

$$P(Y=3) = P(\{\mathrm{HHH}\}) = 1/8.$$

Observe:

$$P(Y = 0 \textit{ or } 1 \textit{ or } 2 \textit{ or } 3) \overset{\text{axiom } 3}{=} \frac{1}{8} + \frac{3}{8} + \frac{3}{8} + \frac{1}{8} = 1.$$

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.

Very simple, but slightly more abstract:

Example (Ross E.g. 4.3b)

*For any event $E \subset S$, its **indicator variable** I_E is defined by*

$$I_E = \begin{cases} 1 & \text{if } E \text{ occurs} \\ 0 & \text{if } E^c \text{ occurs.} \end{cases}$$

Thus, I_E returns a 1 or a 0 to indicate whether or not the event E occurs.

Conversely, if X is any RV whose only possible values are 0 and 1, then it follows that

$$X = I_E \quad \text{when we let } E = \{X = 1\}.$$

Indicator variables are perhaps the simplest possible RVs, but they have many uses. They will reappear often in the course.

Events defined in terms of RVs and their probabilities

Consider S, P, X as before.

For each possible value 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 suppose we don't care about that extra information. Then X gives a natural choice for a 'reduced' description of the experiment:

- Let $S' = \mathbb{R}$, and
- for $A \subset \mathbb{R}$ define

$$P'(A) = P(X \in A) = P(\{X \text{ takes a value in the set } A\}).$$

Then S' and P' still satisfy the axioms of probability.

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 is still quite complicated.

But the key idea is this: for any subset $A \subset \mathbb{R}$, you get an associated event

$$\underbrace{\{X \in A\}}_{\text{shorthand notation}} = \{X \text{ takes a value in the set } A\}.$$

These are the events that can be defined by 'a story you can tell only in terms of X'.

In practice, we mostly focus on quite simple choices for A.

We've already seen the events $\{X = x\}$ (for which $A = \{x\}$).

Very useful alternative: choose a real number a as a 'threshold', and consider the event that X takes a value no larger than a: that is, $\{X \leq a\}$. For this we choose $A = (-\infty, a]$.

Definition (Ross p116)

*If X is a RV, then its **cumulative distribution function** ('**CDF**') is the function*

$$F(a) = P(X \leq a), \quad \textit{defined for } a \in \mathbb{R}.$$

(Sometimes write F_X to make it clear that it's a feature of X.)

First basic properties of CDFs:

Proposition (See Ross Sec 4.10)

If X is a RV, then its CDF F is a function $\mathbb{R} \longrightarrow [0,1]$ such that:

1. *(monotonicity) $F(a) \leq F(b)$ whenever $a \leq b$;*
2. *$F(a) \longrightarrow 1$ as $a \longrightarrow +\infty$;*
3. *$F(a) \longrightarrow 0$ as $a \longrightarrow -\infty$;*
4. *(right-hand limits) $F(x) \longrightarrow F(a)$ as $x \longrightarrow a+$.*

Part 1 follows from the monotonicity of probability values. The others are more subtle, using the full strength of *countable* additivity of P — see Ross Sec 4.10 for proofs.

Many other subsets of $\mathbb{R}$ can be expressed in terms of the sets $(-\infty, a]$ for different values of a. As a result, we can find many other probabilities that concern X in terms of its CDF. E.g.:

Lemma (Ross equation (4.10.1))

If X is a RV, F is its CDF, and $a < b$ then

$$P\{a < X \leq b\} = F(b) - F(a).$$

IDEA: we have

$$(-\infty, b] = \underbrace{(-\infty, a] \cup (a, b]}_{\text{disjoint}},$$

and therefore

$$\{X \leq b\} = \underbrace{\{X \leq a\} \cup \{a < X \leq b\}}_{\text{disjoint}}$$

Discrete random variables; Ross Section 4.2

Random variables are intuitively simple, but some examples are technically very complicated. Special complications arise for RVs 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 its possible values can be written as a (finite or infinite) list $x_1, x_2, \ldots$.*

MATHEMATICAL REMARK

Let $A \subset \mathbb{R}$. Then A is **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 can*not* 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 (Ross p116)

*If X is a discrete RV, then its **probability mass function** ('**PMF**') is the function*

$$p(a) = P\{X = a\}, \quad \textit{defined for } a \in \mathbb{R}.$$

(Sometimes write p_X to make it clear that it's a feature of X.)

Of course, if a is not one of the possible values that X can take, then $\{X = a\} = \emptyset$, and so $p(a) = 0$. We often ignore these choices of a when we write down p.

Proposition (Ross p117)

If the possible values for X are $x_1, x_2, \ldots$, 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 — now apply axioms 2 and 3.

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 $A \subset \mathbb{R}$, then

$$P(X \in A) = \sum_{\text{all } i \text{ such that } x_i \in A} p(x_i) = \sum_{\text{all } a \in A \text{ such that } p(a) > 0} p(a).$$

In particular, in this case the CDF 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 $c > 0$ and $\lambda > 0$ be real parameters, and let X be a discrete RV with possible values 0,1,2,... and with PMF given by

$$p(i) = c\frac{\lambda^i}{i!}, \quad i = 0, 1, 2, \ldots.$$

Find c in terms of λ, and then 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.)

Pause to reflect: what kinds of object have we met so far?

Name	Usual notation	Kind of object
Sample space	S	abstract set
Outcome	s	element of S
Event	$E, F, \ldots,$ etc	subset of S
Prob vals/dist	P	$\mathrm{f}^{\underline{\mathrm{n}}}$: $\{$subsets of $S\} \longrightarrow [0,1]$
Random variables	$X, Y, \ldots,$ etc	$\mathrm{f}^{\underline{\mathrm{n}}}$: $S \longrightarrow \mathbb{R}$
CDF of a RV X	F_X	$\mathrm{f}^{\underline{\mathrm{n}}}$: $\mathbb{R} \longrightarrow [0,1]$
Possible vals of a disc RV	$x_1, x_2, \ldots$ etc	real number
PMF of a disc RV X	p_X	$\mathrm{f}^{\underline{\mathrm{n}}}$: $\{$possible vals$\} \longrightarrow [0,1]$

(Here '$\mathrm{f}^{\underline{\mathrm{n}}}$' is short for 'function')

(Except sometimes we regard p_X as a function $\mathbb{R} \longrightarrow [0,1]$ where $p_X(a) = 0$ if a isn't a possible value of X.)

So if a question asks you to find the PMF of some discrete RV X, your answer *must* be a function

$$\{\text{possible values of } X\} \longrightarrow [0, 1].$$

If it's not written down yet, you'll have to start by stating/explaining what are the possible values of X.

Anything else — e.g., a final answer which is just a single number — *cannot* be correct.

Getting this wrong is called a **category error**. It's worthwhile to check your work for these regularly.

Another practice example

Example (Ross 4.1d)

Independent trials consisting of flipping a p-biased coin are performed until either a head occurs or a total of n flips is reached. Let X be the number of times the coin is flipped. Find its possible values, PMF and CDF.