Math-UA.233: Theory of Probability
Lecture 9

Tim Austin

tim@cims.nyu.edu
cims.nyu.edu/∼tim

From last time... 1

Given an experiment with sample space S, a **random variable** ('**RV**') is a real number which depends on the outcome of the experiment. That is, a function $S \longrightarrow \mathbb{R}$.

If a RV X contains all the information we want about our experiment, then we can restrict our attention to events such as

$$\{a < X \leq b\} \quad \text{for real values } a < b$$

and their probabilities. A convenient way to express these is in terms of the CDF of X:

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

From last time... 2

A RV X is **discrete** if all its possible values may be written as a (finite or infinite) list $x_1, x_2, \ldots$.

In this case, the probabilities of all events associated to X may be expressed in terms of its PMF:

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

Using this, for any set $A \subset \mathbb{R}$, we get

$$\boxed{P(X \in A) = \sum_{i \text{ such that } x_i \in A} p(x_i)}.$$

Expectation (Ross Section 4.3)

Now suppose that $x_1, x_2, \ldots, x_n$ are some 'statistics' — i.e., real numbers.

Their **mean** is

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

This is a kind of representative value around which these '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 avoid some technicalities of the general case.

Definition (Ross p119)

Let X be a discrete RV with pos values $x_1, x_2, \ldots$ and PMF p.

*Its **expectation** (or **expected value** or **mean**) is the number*

$$\underbrace{E[X]}_{\text{notation}} = \sum_i p(x_i)x_i = \sum_{a \text{ such that } p(a)>0} p(a)a.$$

Example

If 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.)

MATHEMATICAL REMARK

If the set of possible values taken by X is an *infinite* list x_1, x_2, ..., 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, *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. 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!

MATHEMATICAL REMARK, contd.

Here's the 'correct' definition for the infinite-series case:

> *If X is a discrete RV with an infinite list of possible values $x_1, x_2, \ldots$ and PMF p, then its* ***expectation*** *is the number*
>
> $$\underbrace{E[X]}_{\text{notation}} = \sum_{i=1}^{\infty} p(x_i)x_i,$$
>
> *provided the infinite series on the right converges absolutely, (i.e. $\sum_{i=1}^{\infty} p(x_i)|x_i| < \infty$).*
>
> *Otherwise, we say X* ***does not have*** *an expectation.*

We insist on *absolute* convergence, because this condition lets us sum the terms in any order we like and be sure of getting the same limit. This is essential for many calculations later.

Example (Ross E.g. 4.3a)

Find $E[X]$ when X is the outcome from rolling a fair die.

Example (Ross E.g. 4.3b)

Let $E \subset S$ be an event and let I_E be its indicator variable. Then

$$E[I_E] = P(E) \cdot 1 + P(E^c) \cdot 0 = P(E).$$

Simplest case of all: suppose $S = \{s_1, \ldots, s_n\}$ itself is finite.

Proposition

For each $j = 1, 2, \ldots, n$, let $X(s_j)$ denote the value that X takes when the outcome is s_j. Then

$$\boxed{E[X] = P(\{s_1\}) \cdot X(s_1) + \cdots + P(\{s_n\}) \cdot X(s_n).}$$

OBSERVE: Very similar to the *definition* of $E[X]$, except that *several different outcomes s_j may give the same value $X(s_j)$*.

IDEA FOR PROOF: Group together s_j's which give the same value of X.

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).

Other quantities can also be used this way (such as 'median' and 'mode' in statistics). But the expectation has 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.

In spite of the intuitive connection to the 'mean' of a bunch of statistics, sometimes calculations of expectations have to be treated with care.

Example (Ross E.g. 4.3d)

A school class of 120 students travels in 3 buses. The first bus contains 36 students, the second 40, and the third 44. When the buses arrive, one student is chosen at random. Let X be the number of students on the bus that they traveled on. Find $E[X]$.

$$\text{ANS} = 1208/30 \approx 40.2667 \ (\neq 40)$$

(See also Ross self-test problem 4.4.)

Expectation of a function of a random variable (Ross Sec 4.4)

Suppose that X is a RV, and also that $g : \mathbb{R} \longrightarrow \mathbb{R}$ is some function. Then we may define a new RV $g(X)$: if the outcome of the experiment is s, then this new RV returns the value $g(X(s))$.

Mathematically, we have formed the *composition* of the functions

$$X : S \longrightarrow \mathbb{R} \quad \text{and} \quad g : \mathbb{R} \longrightarrow \mathbb{R}.$$

Often, we know something about X, and want to turn that into information about $g(X)$. This is often quite easy for the expectation $E[g(X)]$.

Example (Ross E.g. 4.4a)

Let X be a RV that takes the possible values -1, 0 and 1 with probabilities 0.2, 0.5 and 0.3, respectively. Compute $E[X^2]$.

Same ideas give the following general fact:

Proposition (Ross Prop 4.4.1)

If X is discrete with pos vals $x_1, x_2, \ldots$ and with PMF p_X, then

- *$g(X)$ is discrete with possible values $g(x_1), g(x_2), \ldots$ (except that this list may contain REPEATS), and*
- *we have*

$$\boxed{E[g(X)] = \sum_i p_X(x_i) g(x_i)}.$$

(Sometimes: 'Law of the Unconscious Statistician', 'LOTUS'.)

IDEA: The only tricky thing is that several *different* values x_i might give the *same* value of $g(x_i)$. For these, we must use that

$$P(g(X) = y) = \sum_{i \text{ such that } g(x_i)=y} p_X(x_i).$$

Here is a basic use of an expectation as a 'representative' value for a RV. As such, it is a natural choice for 'the thing to maximize' in this problem.

Example (Ross E.g. 4.4b)

A store orders umbrellas in September and then sells them until April. They sell each umbrella for b dollars, and they lose ℓ dollars for each umbrella they don't end up selling. In a given year, the number of customers who want umbrellas is a random variable X with PMF $p(i)$, $i = 0, 1, 2, \ldots$.

How many umbrellas should the store buy in September to maximize their expected profit?

Another simple consequence of LOTUS:

Corollary (Ross Corollary 4.4.1)

If a and b are constants then

$$E[aX + b] = aE[X] + b.$$

(Will use this often later.)

Variance and standard deviation (Ross Sec 4.5)

Another important quantity in statistics is the following. If x_1, x_2, $\ldots$, $x_n \in \mathbb{R}$ and $\overline{x}$ is their mean, then their **variance** is

$$v = \frac{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + \cdots + (x_n - \overline{x})^2}{n}.$$

Their **standard deviation** is its square root, $\sqrt{v}$.

The mean serves as a kind of 'representative' value for the whole list of statistics, and then the variance measures how 'spread out' the statistics are around that representative value.

Like the mean, variance has an important generalization to RVs. Once again we focus on discrete RVs for now.

To motivate the generalization, consider the following RVs:

- $W = 0$ always;,
- $Y = 1$ or -1, each with probability $1/2$;
- $Z = 100$ or -100, each with probability $1/2$.

Then $E[W] = E[Y] = E[Z] = 0$, but Z is much more 'spread out' than Y, which is more 'spread out' than W. Variance gives a way to measure this fact.

Definition (Variance)

For any discrete RV X, if we let $\mu = E[X]$, then the ***variance*** *of X is*

$$\underbrace{\mathrm{Var}(X)}_{\text{notation}} = E\left[(X-\mu)^2\right].$$

Its ***standard deviation*** *is the square root:*

$$\underbrace{\mathrm{SD}(X)}_{\text{notation}} = \sqrt{\mathrm{Var}(X)}$$

An alternative formula for variance:

$$\boxed{\mathrm{Var}(X) = E[X^2] - (E[X])^2}.$$

Very often, computing $\mathrm{Var}(X)$ is easy if we use LOTUS with the function

$$g(x) = (x - \mu)^2.$$

Example (Ross E.g. 4.5a)

Calculate $\mathrm{Var}(X)$ *if* X *represents the outcome when a fair die is rolled.*

Unlike expectation, finding $\mathrm{Var}(g(X))$ in terms of X can be very complicated — there's no good analog of LOTUS. But here's on important special case where things are simple.

Proposition (See Ross p126)

For any constants $a, b \in \mathbb{R}$ we have

$$\boxed{\mathrm{Var}(aX + b) = a^2 \mathrm{Var}(X)}.$$

This is quite intuitive:

- Adding the constant b just shifts the values taken by our RV, it doesn't effect how 'spread out' they are — and the variance stays the same.
- Multiplying by a 'dilates' the values taken, so it makes the variance bigger if $|a| > 1$ or smaller if $|a| < 1$.

The following property gives more insight into what variance means:

Proposition

Let X be a discrete RV and let $\mu = E[X]$. If $\mathrm{Var}(X) = 0$, then

$$P(X \neq \mu) = 0.$$