Math-UA.233: Theory of Probability
Lecture 10

Tim Austin

tim@cims.nyu.edu
cims.nyu.edu/∼tim

From last time... 1

Suppose X is a discrete RV with possible values $x_1, x_2 \ldots$ and PMF p.

Its **expectation** is

$$\underbrace{E[X]}_{\text{notation}} = \underbrace{\sum_i p(x_i) \cdot x_i}_{\text{definition}} \qquad \Bigl(\text{provided } \sum_i p(x_i) \cdot |x_i| < \infty\Bigr).$$

It gives a first indication of where the values of X lie.

Its **variance** is

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

It indicates how 'spread out' X is.

From last time... 2

Two useful facts:

- If the sample space $S = \{s_1, \ldots, s_n\}$ is finite, then we have the alternative formula

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

- LOTUS: if $g : \mathbb{R} \longrightarrow \mathbb{R}$, then

$$E[g(X)] = \sum_i p(x_i) g(x_i).$$

Expectation of a sum of several RVs (Ross Sec 4.9)

Sometimes, the information we care about from an experiment is described by *several* random variables, X, Y, $\ldots$.

Working with many RVs at once is tricky, because they can lead to all sorts of 'composite' events such as

$$\{X \leq 5 \text{ and } Y > 7 \text{ and } X^2 + Y^2 \leq 100\},$$

or whatever. The tools for handling such things will be developed later in the course.

But it turns out that the *expectation of a sum* of RVs, such as $E[X + Y]$, behaves quite simply, and this greatly simplifies some calculations.

Proposition (Ross Corollary 4.9.2)

For random variables $X_1, X_2, \ldots, X_n$, we have

$$\boxed{E\left[X_1 + X_2 + \cdots + X_n\right] = E[X_1] + E[X_2] + \cdots + E[X_n]}.$$

This fact is often called **linearity of expectation.**

It holds for non-discrete RVs as well, once one defines expectation appropriately. We'll see a version for some other RVs later in the course.

IN CLASS: will prove it just in the special case when $S = \{s_1, \ldots, s_n\}$ is finite, using the alternative formula

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

(See Pishro-Nik p235 for a more general proof)

Example (Ross E.g. 4.9c)

Find the expected value of the sum of n rolls of a fair die.

Example (Ross E.g. 4.9d)

Suppose that n tests are performed, and that test i is a success with probability p_i. Find the expected number of successes.

OBSERVE: Did we assume that the tests are independent?

The previous example can be turned into the following very general and useful fact.

Proposition (Ross equation (7.3.1))

Let $A_1, A_2, \ldots, A_n$ be a list of events, and let X be the RV which gives the number of these events that occur (that is, for a given outcome s, $X(s)$ equals the number of $i = 1, 2, \ldots, n$ for which $s \in A_i$). Then

$$\boxed{E[X] = P(A_1) + P(A_2) + \cdots + P(A_n)}.$$

IDEA: Realize that $X = I_{A_1} + I_{A_2} + \cdots + I_{A_n}$.

Example (Ross E.g. 7.2h)

Suppose n people throw their hats into the centre of a room. The hats are mixed up, and each person randomly selects one. Find the expected number of people who get their own hat back.

Example (Ross E.g. 7.2j)

Ten hunters are waiting for ducks to fly by. When a flock of ducks fly by, the hunters fire at the same time, but each chooses their target at random, independently of the others. Each hunter hits their target with probability p and these events are independent. Compute the expected number of ducks that escape when a flock of ten flies by.

Bernoulli and binomial random variables (Ross Sec 4.6)

Most of today's class will be given to these important examples.

Definition

A RV X is ***Bernoulli*** *if its possible values are 0 and 1. This means there is some p, $0 \le p \le 1$, such that its PMF is*

$$P(X = i) = \begin{cases} p & \text{if } i = 1 \\ 1 - p & \text{if } i = 0. \end{cases}$$

We call X **Bernoulli**(p) if we want to make the value of p explicit.

Another way of saying this: X is a Bernoulli(p) RV if it is simply the indicator variable of an event whose probability is p. So

$$\text{Bernoulli RVs} = \text{indicator RVs.}$$

Why have another name for them?

The name 'Bernoulli RVs' indicates a *context*. We use it when we have an experiment consisting of a sequence of n independent trials. When we let $E_1, \ldots, E_n$ be the events of successes on those trials, and then we let $X_1, \ldots, X_n$ be the indicator variables of those events. These are a 'sequence of Bernoulli RVs', or sometimes of 'Bernoulli trials'.

Usually we assume that each trial has the same probability of success. So these would all be Bernoulli(p) RVs for some common value of p.

Now suppose n independent trials are performed, and that each results in success with probability p. Let X be the number of successes that actually occur (often the most important feature of an outcome).

For each i, let E_i be the event of success on the i^{th} trial, and let X_i be the indicator of this event. Then $X_1, \ldots, X_n$ are Bernoulli(p) RVs, and

$$\boxed{X = X_1 + X_2 + \cdots + X_n.}$$

The possible values of X are 0, 1, ..., n.

Using the independence of the trials, we obtain a formula for the PMF p of X (see Ross equation (4.6.2)):

$$\boxed{p(i) = P(X = i) = \binom{n}{i} p^i (1-p)^{n-i} \quad \text{for } i = 0, 1, 2, \ldots, n.}$$

(REMARK: By the binomial theorem, we know that

$$\sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} = (p + (1-p))^n = 1^n = 1,$$

— as it should be!)

Example (Ross E.g. 4.6a)

Five fair coins are flipped. If the outcomes are assumed independent, find the PMF of the number of heads obtained.

Example (Ross E.g. 4.6b)

A company produces screws, each of which is defective with probability 1% independently of the others. They sell the screws in packets of 10, and will replace a packet if it contains more than one defective screw. What proportion of packets must the company replace?

Example (Ross E.g. 4.6d)

According to a simple model, right- or left-handedness is determined by a pair of genes in a person's DNA. Each of these genes may say 'right' ('R') or 'left' ('L'). If either of the genes says R, then the person is right-handed; they are left-handed only if both genes say L. A person with one gene of each type is called 'hybrid'. Children (usually) have a copy of one gene from each parent, and each is equally likely to be a copy of either of that parent's genes.

If two hybrid parents have four children, what is the probability that three of them will be right-handed? (The genes passed to different children are independent.)

The calculations above are so important that we make another definition around them.

Definition

A RV X is ***binomial with parameters*** (n, p) *(or just '****binom****(n, p)') if its possible values are* 0*,* 1*, . . . , n and if its PMF is given by*

$$P(X = i) = \binom{n}{i} p^i (1-p)^{n-i} \quad \textit{for } i = 0, 1, 2, \ldots, n$$

(the formula for numbers of successes from before).

So this definition is just giving a name to something we already know, in order to help us talk about it in the future.

BUT there's an important conceptual shift behind that definition.

We introduced the binomial RVs by counting successes in a sequence of trials.

But formally, X is binomial if and only if it has the right PMF: it doesn't matter at all *what kind of story or model X comes from*.

In principle, a binom(n, p) RV could arise in a *different* setting — not a sequence of trials. But its associated probabilities, expectation, etc. can all be computed from its PMF, so the 'origin' of X doesn't matter for these calculations.

This is why the definition of binomial RVs doesn't mention what the underlying experiment was at all.

IMPORTANT CONSEQUENCE: some properties of binomial RVs are easier to understand if you think about a sequence of trials, and some are easier to extract directly from the PMF.

Either way is correct!

The first place where we can see the effect of these two ways of thinking about binomial RVs is the following.

Proposition (Ross p132 and E.g. 7.2e)

If X is binom(n, p), then

$$\boxed{E[X] = np.}$$

IDEA: Either compute using the PMF, or use linearity of expectation.

Example (Ross E.g. 4.6c)

A player bets on one of the numbers 1, ..., 6. Then three dice are rolled. If i, $i = 1, 2, 3$, of the dice show the player's number, then s/he wins \$i. If none of them show that number, then s/he loses \$1. Is this game fair to the player?

INTERPRETATION: Let \$$X$ be the amount the player wins or loses. We say the game is 'fair' if $E[X] = 0$.

IDEA: This X is *not* binomial, but it is related. How?

We can also choose between two methods in computing the variance.

Proposition (Ross p132 and E.g. 7.3a)

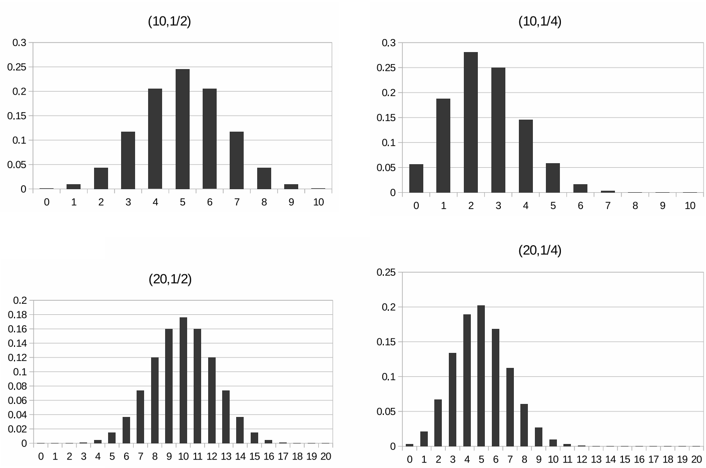
If X is binom(n, p), then

$$\boxed{\mathrm{Var}(X) = np(1-p).}$$

IDEA: After writing X as a sum of Bernoulli RVs, turn that into a sum of RVs that equals X^2.

This result is quite intuitive: the variance goes up as n increases, and for a fixed n it is maximized for $p = 1/2$: the 'most random' or 'least biased' situation.

Sometimes, it is more important to have a rough idea of how the PMF of a binomial RV behaves than to do exact calculations. Some pictures for different values of (n, p):



Some other discrete RVs (Ross Sec 4.8)

Here's another not-at-all-new example.

Proposition (Ross equation (4.8.1))

Suppose that independent Bernoulli(p) trials are performed with $p > 0$. Let X be the number of trials until the first success; let us say that '$X = \infty$' if all trials are failures forever. Then the possible values of X are $1, 2, \ldots$ and '∞', we can compute

$$\boxed{P(X = k) = (1 - p)^{k-1} p \quad \text{for } k = 1, 2, \ldots,}$$

and therefore

$$\boxed{P(X = \infty) = 1 - P(X = 1) - P(X = 2) - \cdots = 0.}$$

CONSEQUENCE: Although '$X = \infty$' is theoretically possible, its probability of occurring is zero.

Example (Ross E.g. 4.8a)

An urn contains n white and m black balls. We select balls at random and then replace them until the first time we get a black ball. What is the probability that

(a) *exactly k draws are needed?*
(b) *at least k draws are needed?*

Definition

*Let $0 < p \leq 1$. A RV X is **geometric with parameter** (p) (or just '**geometric**(p)') if its possible values are 1, 2, ... (and maybe '∞') and if its PMF is given by*

$$P(X = k) = (1-p)^{k-1}p \quad \text{for } k = 1, 2, \ldots, \qquad P(X = \infty) = 0.$$

Once again, we have seen how geometric RVs arise from a story ('waiting for a success'), but the definition only requires that the PMF be a certain function; it doesn't care where the RV 'comes from'.

Proposition (Ross equation (4.8.4))

An urn contains m white balls and $N - m$ black balls. We choose n of the balls at random without replacement. Let X be the number of white balls that we pick. Then the possible values are k for $n - (N - m) \leq k \leq \min(n, m)$, and

$$\boxed{P(X = k) = \frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}}}$$

for those possible ks.

Definition

Any RV with the PMF given above is called ***hypergeometric***(n, N, m).