Math-UA.233: Theory of Probability
Lecture 12

Tim Austin

tim@cims.nyu.edu
cims.nyu.edu/∼tim
We finished the previous lecture with linearity of expectation. Here’s another example as a warmup for today.

**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 kinds of random variable.

**Definition**

A RV $X$ is **Bernoulli** if its possible values are 0 and 1. This means there is some $p$, $0 \leq p \leq 1$, 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 RV if it is simply the indicator variable of the event $\{X = 1\}$. So Bernoulli RVs are just 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 $n$ (usually large) independent trials, 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 $\text{Bernoulli}(p)$ RVs for some common value of $p$. 
So 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 important feature of the outcome of this experiment.

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

\[
X = X_1 + X_2 + \cdots + X_n.
\]
The possible values of \( X \) are 0, 1, \ldots, \( n \). Using the independence of the trials, we obtain a formula for the PMF \( p \) of \( X \) (see Ross equation (4.6.2)):

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

(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, left- or right-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.)
This calculation is so important that we make another definition around it.

**Definition**

A RV $X$ is **binomial with parameters** $(n, p)$ (or just ‘$\text{binom}(n, p)$’) if its possible values are $0, 1, \ldots, n$ and if its PMF is given by

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

(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, the definition of binomial RVs just specifies their PMF: it doesn’t refer at all to *where they come from*.

This is because, once you know your RV is binom\((n, p)\), you can get any information you want about it from the PMF alone. In principle, a binom\((n, p)\) RV could arise in a *different* setting — not a sequence of trials — and its associated probabilities, expectation, etc. would all be just the same.

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 see 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*

\[ E[X] = np. \]

**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*

\[
\text{Var}(X) = np(1 - p).
\]

**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 \((n, p)\):
The previous examples illustrate the following precise statement.

**Proposition (Ross Prop 4.6.1)**

If $0 < p < 1$ and $X$ is binom$(n, p)$, then as $k$ runs from 0 to $n$, $P\{X = k\}$ increases monotonically and then decreases monotonically, reaching its maximum when $k$ is the largest integer less than or equal to $(n + 1)p$. 
In the state of Tennefornia, \( n = 2k + 1 \) citizens will vote between Alice and Bob to be the next president. No-one likes either candidate, so all voters will choose their vote by independently flipping a fair coin.

\( X \) is one of the voters. We say her vote is ‘decisive’ if exactly half of the other \( 2k \) citizens vote for each candidate. What is the approximate probability that \( X \)’s vote is decisive?

IDEA: Remember Stirling’s approximation:

\[
k! \sim \sqrt{2\pi k} \frac{k^k}{e^k}.
\]

(And see Ross for a further discussion of the consequences in the US electoral college system.)
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. 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 ‘$\infty$’, and

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

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

A RV $X$ is **geometric with parameter** $(p)$ (or just ‘**geometric** $(p)$’) if its possible values are $1, 2, \ldots$ (and maybe ‘$\infty$’) and if its PMF is given by

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

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

\[
P\{X = k\} = \frac{\binom{m}{k} \binom{N-m}{n-k}}{\binom{N}{n}}
\]

for those possible \( k \)s.

Definition

Any RV with the PMF given above is called \textbf{hypergeometric}(n, N, m).
Example (Ross E.g. 4.8h)

An unknown number $N$ of fish live in a pond. An ecologist catches $m$ of them at random, tags them and returns them to the pond. After giving the fish time to disperse, s/he catch another $n$ fish at random. Let $X$ be the number of tagged fish contained in the second catch. How can we use $X$ to make a guess about the size of $N$?