Vanderbilt University Political Science - Stats I
Fall 2024 - Prof. Jim Bisbee

Lecture 3

- We've now provided an intellectual foundation for our understanding of probability. We've conceived of probabalistic events in the context of an **experiment** whose results are–in their most basic form–**simple events**. The set of all possible events for a given experiment is its **sample space.**
  - We then specified 3 axioms governing the assignment of probabilities to these events;
  - And then discussed four tools by decompose and compose events of interest into events with conveniently calculated probabilities.
- Before moving on, I want to formalize one additional tool that we use without thinking about it–which is *assigning probabilities in a sample space consisting of equiprobable events*:
  - Last time we had the example of assigning 16 students into 3 teams of 6, 5 and 5 students, with 11 of the students male.
    * As you may recall, we began by considering the probability that the first student picked is male. What is it? $\frac{11}{16}$. Why?
  - Here's how to formalize our intuition.
  - Consider a sample space S consisting of n *equiprobable* simple events E such that $P(E_1) = P(E_2) = ...P(E_n)$.
    * Recall that we define an A as a subset of S – that is, some subset of sample points that make up S.
    * Then $P(A) = \frac{|A|}{n}$,.where the notation $|A|$ stands for the number of elements in the set A – which in this context is the number of sample points in A.
  - Can you see how we get this from the axioms of probability?
    * First let's show that for all events E to be equiprobable, it must be that

$$P(E_i) = \frac{1}{n} \text{ for } i = \{1, 2...n\}$$

* This is because
* $P(S) = P(E_1 \cup E_2 \cup ... \cup E_n) = P(E_1) + P(E_2) + ... + P(E_n) = 1$ [by Axiom 3 and then Axiom 2]
* $P(E_1) + P(E_1) + ... + P(E_1) = nP(E_1) = 1$ [definition of equiprobability]
* $P(E_1) = \frac{1}{n}$; and similarly $P(E_i) = \frac{1}{n}$ for $i = \{1, 2...n\}$
* Finally, since

$$\begin{aligned} P(A) &= \sum_{E_i \in A} P(E_i) \text{ [Axiom 3 again]} \\ P(A) &= \sum_{E_i \in A} \frac{1}{n} = \frac{|A|}{n} \blacksquare. \end{aligned}$$

* [For reference:]
  · Axiom 2: $P(S) = 1$. S occurs every time the experiment is performed.
  · Axiom 3: For any sequence of pairwise mutually exclusive events $A_1, A_2, A_3, ... A_n$, it must be the case that P($A_1 \cup A_2 \cup A_3 ... \cup A_n$) $= \sum_{i=1}^{n} P(A_i)$.

– The point here is that even our intuitive basic notion of assigning probabilities to events is derived from the axioms of probability in logical ways.

0.1 The notion of a random variable

- Now let's discuss a particular kind of experiment: one in which the events of interest are numerical–that is, they are identified in a meaningful way by numbers.
  – Numerical events of interest to political scientists might be the number of seats the Republican Party will hold in the House of Representatives after a midterm election, or the number of signatories to a treaty.
  – In the language of an experiment, we assign a *real number* to each point in the sample space.
  – Call this number the variable Y.
    * Think back: what's a *variable*? [Make them look this up]: Recall that a variable is a logical grouping of attributes. Variables take on values that are exhaustive and

mutually exclusive.

- Each sample point can only take on one value of Y. But the same value may be assigned to multiple sample points.
  - [Draw a circle with S, the sample space. Draw another circle, Y, the variable. Draw lines from several points in S to one point in Y.]
- What do we call such a mapping? A *function*. Thus the variable Y is a *function* of the sample points in S.
- Recall that a function is a mathematical relation assigning each element of one set (the source) to one and only one element of another set (the target).
- In this case, the function's source is S and its target is Y. We can write $f : S \to Y$. This function (and by extension, Y) is defined as a **random variable**. Keep this in mind: whenever we talk about a random variable, *we are really talking about a function* that maps each simple event in a sample space S to a (meaningful) number.
- We write a random variable with a capital letter: Y.
  - Consequently, we write "the probability that $Y = 0$" as $P(Y = 0)$, and so forth.
  - Typically, we refer to any observed or hypothetical value of Y with a lowercase letter denoting the value. So we write $P(Y = y)$, for example.
  - We will still talk about the event of interest A, but we will give it a number, a. The event A thus $\equiv$\{all sample points such that $Y = a$\}.
- More about random variables very shortly. But first:

0.2 Quick detour: The notion of a random *sample*

- In our subsequent discussions of probability, we will often invoke the notion of a **random sample**. Let's take a moment to place this idea in the language of our probabilistic model of an experiment.
- In this framework:
  - our experiment is the drawing of a **sample** (the units selected for analysis) from a **population** (the group of units about which we wish to make inferences)

- the **design** of our experiment is the method of sampling.
  - one big decision we make, for example, is whether to sample *with replacement* (units selected for the sample are "placed back into the population" and may therefore again be sampled) or *without replacement* (units selected are set aside and cannot be selected again). Keep these terms in mind; we'll revisit them later.
- the most common way to draw a sample is called *random sampling.* Let N and n represent the numbers of elements in the population and sample, respectively. In this context, a total of [how many?] $\binom{N}{n}$ [the binomial coefficient] different samples without replacement (how many is that? $\frac{N!}{n!(N-n)!}$) may be drawn. If sampling is conducted in such a way that each of these samples has an equal probability of being selected, then we have engaged in *random sampling*, and the result is said to be a *random sample*.

0.3 Back to random variables: probability distributions

- We'll begin our discussion of random variables by focusing on those that are **discrete**. A random variable Y is discrete if it can take on only a finite or countably infinite (a one-to-one correspondence with the integers) number of distinct values.
- In making inferences from samples to populations,
  - we need to know the probability of having observed a particular event.
  - events of interest are frequently numerical events that correspond to values y of discrete random variables Y.
    - An example of such an event might be that 45 out of 100 people we surveyed said they intended to vote for Mitt Romney.
  - In other words, we need to know $P(Y = y)$ for all the values the RV Y can take on.
  - This collection of probabilities is called the **probability distribution** of the RV Y.
  - As an example, consider the experiment: roll a pair of six-sided dice and record the sum of the pips on their faces.
  - The sample space S consists of 36 simple events.

- - * Draw a square, label it S, divide into grid of 6 x 6.
  - Define the random variable Y = the sum of the pips appearing on the faces of a random roll of a pair of six-sided dice.
    * Draw another square, label it Y, and map a few of the sample points in S to Y.
    * Formally, $P(Y = y) = \sum_{E_i : Y(E_i) = y} P(E_i)$. The probability $Y = y$ is defined as the sum of the probabilities of the sample points in S assigned the value y. Sometimes we write this $p(y)$.
    * It its most primitive form, Y's probability distribution is a *table* in which each y is listed alongside $P(Y = y)$ But it may also be a *formula* mapping each y to its $P(Y = y)$, or a *graph* doing the same thing.
    * Table: (entitled "The Probability Distribution of the RV Y")

y	# of sample points	$P(Y = y)$
2	1	$\frac{1}{36}$
3	2	$\frac{2}{36}$
4	3	$\frac{3}{36}$
5	4	$\frac{4}{36}$
6	5	$\frac{5}{36}$
7	6	$\frac{6}{36}$
8	5	$\frac{5}{36}$
9	4	$\frac{4}{36}$
10	3	$\frac{3}{36}$
11	2	$\frac{2}{36}$
12	1	$\frac{1}{36}$
totals	**36**	**1**

- - * Graph: [draw on board]
    * A function (here's one way, there may be others/better). Here, $p(y)$ is known as a **probability function**.

$$P(Y = y) = p(y) = \frac{6 - |7 - y|}{36}, y = \{1, 2, ...6\}.$$

  - The probability function of a discrete random variable is also called its **probability mass function, or PMF.** The "mass of a random variable at y" is the PMF evaluated at y, or $p(y)$.
  - The probabilities associated with distinct values of y sum to 1, and in fact they always do, as the second axiom of probability requires.

0.4 Random variables: expected values

- The probability distribution of an RV is a *theoretical model* for the empirical distribution of data associated with a real population. If we were to roll a pair of dice over and over again, recording the sum of the faces each time, the empirical distribution would look very much like the theoretical probability distribution of Y we just specified.

- As we do with empirical data, we can describe a random variable by talking about its central tendency and dispersion. (At what level of measurement are we implicitly working here? Interval or higher.) So we will be concerned about the mean and variance of a random variable.

- We specify and manipulate formulas describing random variables using the *expectations operator*, which is defined as follows:

  - Let Y be a discrete RV with the probability function $p(y)$. The **expected value** of Y is written $E(Y)$, and defined to be:

$$E(Y) \equiv \sum_y y p(y).$$

  - Note that this is each possible value of Y times the probability that Y takes on the value y, i.e., $p(y)$, summed up over all y. [Calculate this value for the two dice example.]
  - So the expectations operator tells us to consider all the possible values of a RV, multiply each of these values by their probability of occurrence, and to sum up these products.
  - The expected value is the way we talk about the central tendency of a random variable with a theoretical probability distribution. It is equivalent to the idea of the mean of an empirical frequency distribution.

- Now we take what is a very powerful step as we move into inferential statistics. Recall that the probability distribution of an RV is a *theoretical model* for the empirical distribution of data associated with a real population. If this theoretical model is accurate, then $E(Y) = \mu$, the *population mean*.

$$E(Y) \equiv \sum_y y p(y) = \mu.$$

- Here, μ is a *parameter*: a characteristic of the distribution of Y in the population that we never actually observe but about which we are often keenly interested. It is the first of many such parameters we will encounter in this class.

- We are often interested in the expected value of *functions of random variables.* (You'll see why in a minute.) Consider as usual the discrete RV, Y with probability function $p(y)$. Now consider any real-valued function of Y, $g(Y)$. Then the expected value of $g(Y)$is given by:

$$E[g(Y)] = \sum_y g(y)p(y).$$

- That is, the expected value of a function of a RV is given by:
  - evaluating the function for each value of Y
  - multiplying it by the probability that $Y = y$
  - and summing up over all possible values of Y.

- Now this is not a definition, but rather can be proven based upon the definition of $E\left(Y\right)$ .Proof:
  - Consider a discrete RV Y taking on a finite number, n, of values $y_1, y_2, ... y_n$.
  - Suppose $g(y)$ takes on m different values $g_1, g_2 ... g_m$, $m \leq n$.
    * We're being this picky because it's of course possible that $g\left(\right)$ is not one-to-one; that is, it may take on the same value for multiple y's.
  - Note that $g(Y)$ is itself a random variable (by definition of RV). It is a function mapping the sample space Y to the reals.

- So we can define a new probability function, p-star, for g.
- Write the probability that g takes on a value g_i as

$$\begin{aligned} p^*(g_i) &= P[g(Y) = g_i] \\ &= \sum_{y_j : g(y_j) = g_i} p(y_j), \text{ for } i = 1, 2, ...m \end{aligned}$$

- where $y_j : g(y_j) = g_i$ means "all y_j such that g equals g_i when evaluated at y_j."
- Now the definition of expected value tells us that:

$$E[g(Y)] = \sum_{i=1}^{m} g_i \, p^*(g_i)$$

- (Note that we're doing the same thing as before:
  * Taking each possible value of g, that is all the g_i's
  * Multiplying it by its associated probability (which we've defined in this case as $p^*(g_i)$)
  * And summing up these products.)
- Moving on with proof:

$$\begin{aligned} &= \sum_{i=1}^{m} g_i \left\{ \sum_{y_j : g(y_j) = g_i} p(y_j) \right\} \text{ [substituting]} \\ &= \sum_{i=1}^{m} \left\{ \sum_{y_j : g(y_j) = g_i} g_i \, p(y_j) \right\} \text{ [can move } g_i \text{ inside because of nested summations]} \\ &= \sum_{j=1}^{n} g(y_j) p(y_j). \text{[since } g_i = g(y_j) \text{ for any } y_j; \text{ we change index to signify we're} \\ &\quad \text{now interested in all } n \text{ values that } Y \text{ can take on]} \\ &= \sum_{y} g(y) p(y). \text{ [writing more simply]} \end{aligned}$$

- Again, the intuition behind $E[g(Y)] = \sum_y g(y)p(y)$ is straightforward. We proceed by:
  * evaluating the function for each value of Y
  * multiplying it by the probability that $Y = y$

        * and summing up over all possible values of Y.

- We can now define the variance of a random variable Y. Recall before that we defined the variance of an empirical variable as $s^2 = \frac{1}{N}\sum_{i=1}^{N}(y_i - \overline{y})^2$.

  – Well, the variance of a random variable is how it varies about its mean:

$$VAR(Y) \equiv E[(Y - E(Y))^2]$$

  – And if the RV Y accurately describes the population distribution then

$$VAR(Y) = E[(Y - \mu)^2].$$

  – The standard deviation of Y is $+\sqrt{VAR(Y)}$.

  – And (again assuming that our RV variable is an accurate theoretical model of the world) then $VAR(Y) = \sigma^2$, the population variance, with $\sqrt[+]{\sigma^2} = \sigma$ the population standard deviation.

  – Do Example 3.2 (p. 94) on the board.

0.5 Some helpful results

- Go over "Some helpful results about the math of expectations for discrete RVs" handout. Proofs:

- To show $E(c) = c$:

  – Consider the function $g(Y) \equiv c$. By the expectation of a function of a random variable theorem, $E(c) = \sum_y cp(y) = c\sum_y p(y)$. But by Axiom 2, $\sum_y p(y) = 1$. Hence $E(c) = c(1) = c$. $\blacksquare$.

- To show $E[cg(Y)] = cE[g(Y)]$:

- By the expectation of a function of a random variable theorem,

$$
\begin{aligned}
E[cg(Y)] &= \sum_y cg(y)p(y) \\
&= c\sum_y g(y)p(y) \\
&= cE[g(Y)] \quad \blacksquare.
\end{aligned}
$$

- To show that we can distribute expectations, consider the case where $k = 2$:

$$
\begin{aligned}
E[g_1(Y) + g_2(Y)] &= \sum_y [g_1(y) + g_2(y)]p(y) \quad \text{[since } g_1(Y) + g_2(Y) \text{ is a function of } Y\text{]} \\
&= \sum_y [g_1(y)p(y)] + \sum_y [g_2(y)p(y)] \quad \text{[distributing summations]} \\
&= E[g_1(Y)] + E[g_2(Y)] \quad \text{[by definition of the expectations operator]} \\
& \blacksquare.
\end{aligned}
$$

- All of these results help us re-write the formula for the population variance in a very helpful way (proof is on handout).
- [Emphasize: we always treat parameters as constants when applying the expectations operator. They are invariant.]

0.6 Three theoretical probability models

- As illustrations, we will discuss three of the standard discrete probability distributions:
  - the **Bernoulli**
  - the **binomial**
  - the **Poisson**

0.7 The Bernoulli

- A Bernoulli experiment is the observation of an experiment consisting of one trial with two outcomes: zero or one.
- Thus the Bernoulli random variable Y takes on the values $\{0, 1\}$.

- E.g.'s:
  - a coin toss
  - whether an individual approves or disapproves of Barack Obama's performance as president
  - whether a country signs a treaty
- A Bernoulli random variable is characterized by one parameter: π, the probability of success.
- Thus its probability distribution is

$$\begin{aligned} p(y=1) &= \pi \\ p(y=0) &= 1-\pi \end{aligned}$$

- Sometimes it's convenient to write the probability (mass) function

$$p(y) = \pi^y (1-\pi)^{(1-y)}$$

  - Think of the exponents as "switches" which turn on and off (i.e. equal to one) depending on the value of y at which the function is being evaluated.
- For a little practice, let's show that

$$\begin{aligned} E(Y) &= \pi. \\ E(Y) &= \sum_y p(y)\, y \\ &= p(y=0)(0) + p(y=1)(1) \\ &= \pi. \end{aligned}$$

- How about

$$\begin{aligned} VAR\,(Y) &= \pi\,(1-\pi) \\ VAR\,(Y) &\equiv E[(Y-)^2] \\ &= E(Y^2) - [E\,(Y)]^2 \text{ [from handout; see how it's already helpful?]} \end{aligned}$$

- Note that

$$\begin{aligned} E(Y^2) &= \sum_y y^2 p(y = Y) \\ &= 0^2\,(1-\pi) + 1^2\pi \\ &= \pi \end{aligned}$$

- Now substitute

$$\begin{aligned} VAR\,(Y) &= E(Y^2) - [E\,(Y)]^2 \\ &= \pi - \pi^2 = \pi\,(1-\pi)\,. \end{aligned}$$

- Let's talk a bit more about **parameters**. Together with the probability function, parameters determine the shape, location and spread of the distribution. A **location parameter** specifies the location in the center of the distribution. As we change a location parameter we shift the PMF to the left or right. Its empirical referent is, of course, the mean. A **scale parameter** specifies the spread (or scale) of the distribution around its central location. Its empirical referent is the standard deviation. What's notable about the Bernoulli (and, as you will see, the Binomial and the Poisson) is that they only have location parameters; their spread and their location are tied together.

0.8 The Binomial

- A binomial experiment is the observation of a sequence of identical and independent Bernoulli trials. Specifically:

- A fixed number, n, of trials with one of two outcomes: success S or failure F.
- As with the Bernoulli, the probability of success on any single trial is π. The probability of failure is thus $1 - \pi$.
- The trials are independent.
- The random variable of interest is Y, the # of successes observed during the n trials.

- E.g.'s:
  - # of heads observed in a certain # of coin tosses
  - # of people approving of Barack Obama out of a certain # of people
  - # of countries signing a treaty out of n eligible countries

- Let's find the binomial probability distribution. Recall that a probability distribution (table, graph, formula) tells us with what probability our RV of interest, Y, takes on all possible values y, i.e. $P(Y = y)$, or simply $p(y)$.

- In the context of a probability model, we consider the event "the number of successes equals y," or $Y = y$, to be our "event of interest."

- Let's assign a probability $p(y)$ to each of these events.

- One such event might be $SSFSFFFSFS...FS$, where there are n such positions.

- If we were to count the S's and F's, we might see

$$S_1SSSS...SSS_y \qquad F_1FF....FF_{n-y}$$

- In this context, the number of S's equals what? y. And so the number of F's equals what? $n - y$.

- Think about our laws of probability. The event of interest we have witnessed is the **intersection** of n simple events: $S_1 \cap S_2 \cap ... \cap S_y \cap F_1 \cap F_2 ... \cap F_{n-y}$. These are **independent** events (by definition of the binomial experiment). So $P(S_1 \cap S_2 \cap ... \cap S_y \cap F_1 \cap F_2 ... \cap F_{n-y})$ simply equals $P(S_1)P(S_2)...P(S_y)P(F_1)P(F_2)P(F_{n-y})$.

- So what's that? It's $\pi^y (1-\pi)^{n-y}$.

- But this is *not* the probability of seeing the event $Y = y$. Why? Because the event $Y = y$ could happen in many more ways than the order in which we saw it. How many different ways are there to order y S's and $n-y$ F's? It's the number of ways of choosing y elements from a total of n elements. Or "n, choose y." From math camp, you'll remember that this is equal to (ha!) the binomial coefficient $\binom{n}{y}$,or $\frac{n!}{y!(n-y)!}$.

- Thus the probability of observing $Y = y$ is

$$p(y) = \frac{n!}{y!(n-y)!}\pi^y(1-\pi)^{n-y}.$$

This can be written as $Binomial(\pi; n)$

- In the parlance of probability distributions, the binomial has two **parameters**: π and n.

- So (very quickly). 9 students in the class, 5 male. I pick six students at random with replacement. What's the chance that I pick the same number of males and females?

  - Call 'success' a female. $\pi = \frac{4}{9}$.
  - We have # of trials $n = 6$. We have # of successes as $y = 3$.
  - We thus evaluate the probability distribution for $Bi(\frac{4}{9}; 6)$ at $y = 3$:

$$\begin{aligned} p(Y &= 3) = \frac{6!}{3!(6-3)!}\left(\frac{4}{9}\right)^3\left(1-\frac{4}{9}\right)^{6-3} \\ &\approx .30 \end{aligned}$$

- Let's say I turned this question around a bit. I draw six students at random (with replacement). On average, how many females will I pick? And how much will this number vary over repeated draws of six?

- Answering questions like this requires a formula for the expected value and the variance of the Bernoulli random variable Y.

- And in fact, we can prove that if Y is a binomial random variable,

$$E(Y) = n\pi = \mu \text{ and } VAR(Y) = n\pi(1-\pi) = \mu(1-\pi) = \sigma^2.$$

- We'll omit the proofs in class because we need to move on. But the proofs (pp. 107-108) are informative.

- Thus $E(Y|p = \frac{4}{9}; n = 6) = 6 \times \frac{4}{9} = 2.\overline{6}$. $VAR(Y) \approx 2.7\left(1 - \frac{4}{9}\right) = 1.5$. The s.d. is $\sqrt{1.5} \approx 1.2$. The typical # of females I will draw in a sample of 6 is 2.7, but the typical value will fall 1.2 females away from 2.7.

- In practice, Binomials are not conveniently calculated. We typically use statistical software to calculate the probability that the Binomial random variable Y takes on some value y. In Stata, we type `.di binomialp`(n, y, π) to get $P(Y = y)$.

0.9 The Poisson

- A final discrete probability distribution we will examine is the Poisson.
- A Poisson experiment is the observation of a *count of events* that occur in an *interval, broadly defined* – a given space, time period, or any other dimension. It is particularly helpful when modifying relatively *infrequent* events (as more frequent events can be modeled with more generic distributions).
  - Environmental laws per Congressional session.
  - Errors per page.
  - Government shutdowns per decade.
  - Homeless shelters per census tract.
- We discuss the Poisson after the Binomial because it can actually be conceived of the Binomial experiment as the number of trials approaches infinity. (What? Let's consider:)
  - Split the interval into n subintervals, each so small that at most one event could occur in it. That is, in each subinterval a Bernoulli trial takes place:

$$\begin{aligned} P(y=1) &= \pi \\ P(y=0) &= 1-\pi \\ \text{AND } P(y>1) &= 0. \end{aligned}$$

  - In this subinterval,

$$p(y) = \pi^y (1-\pi)^{(1-y)} \text{ [Bernoulli]}$$

  - And if we have n subintervals, then

$$p(y) = \frac{n!}{y!(n-y)!}\pi^y(1-\pi)^{n-y} \text{ [Binomial]}$$

  - How many such subintervals are needed? Who knows. But we can handle this problem by making the subintervals infinitely small by taking the limit of the Binomial probability function as n goes to infinity. Our parameter of interest–the number of successes over the interval–is $n\pi$. We call this parameter $\lambda = n\pi$. So:

$$\begin{aligned} & \lim_{n\longrightarrow\infty} \frac{n!}{y!(n-y)!}\pi^y(1-\pi)^{n-y} \\ = & \lim_{n\longrightarrow\infty} \frac{n!}{y!(n-y)!}\left(\frac{\lambda}{n}\right)^y\left(1-\frac{\lambda}{n}\right)^{n-y} \\ = & \lim_{n\longrightarrow\infty} \frac{n!}{y!(n-y)!}\frac{\lambda^y}{n^y}\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-y} \end{aligned}$$

– Noting that by definition of e,

$$\lim_{n\longrightarrow\infty}\left(1-\frac{\lambda}{n}\right)^n = e^{-\lambda}$$

– And so

$$\begin{aligned} = & \lim_{n\longrightarrow\infty} \frac{n!}{y!(n-y)!}\frac{\lambda^y}{n^y}e^{-\lambda}\,(1) \\ = & \frac{e^{-\lambda}\lambda^y}{y!}\lim_{n\longrightarrow\infty}\frac{n!}{(n-y)!n^y} \quad \text{[pulling out everything not related to } n\text{]} \\ = & \frac{\lambda^y}{y!}e^{-\lambda}\lim_{n\longrightarrow\infty}\frac{n\,(n-1)\,(n-2)\,...(n-y+1)}{n^y} \quad \text{[shaving off } n-y \text{ terms from numerator]} \\ = & \frac{\lambda^y}{y!}e^{-\lambda}\lim_{n\longrightarrow\infty}\frac{n}{n}\times\frac{(n-1)}{n}\times\frac{(n-2)}{n}\times ...\times\frac{(n-y+1)}{n} \\ = & \frac{\lambda^y}{y!}e^{-\lambda}\lim_{n\longrightarrow\infty} 1\times\left(1-\frac{1}{n}\right)\times\left(1-\frac{2}{n}\right)\times ...\times\left(1-\frac{y+1}{n}\right) \\ = & \frac{\lambda^y}{y!}e^{-\lambda}\,(1) \end{aligned}$$

– And so

$$p\,(y) = \frac{\lambda^y}{y!}e^{-\lambda}.$$

- In practice, Poissons are not conveniently calculated. We typically use statistical software to calculate the probability that the Binomial random variable Y takes on some value y. In Stata, we type `.di poissonp`(λ, y) to get $P\,(Y=y)$.

- It is the case [proofs are in your book] that for a Poisson RV,

$$\begin{aligned}\mu &= E(Y) = \lambda \\ \sigma^2 &= VAR(Y) = \lambda.\end{aligned}$$

The take-home point is that there are many theoretical models of specific experiments. Each of them has a probability distribution $p(y)$, and each has a mean ($E(Y)$) and a variance ($VAR(Y)$). Get to know these a bit by browsing the remainder of Chapter 3. See also the inside back cover of your text for a quick summary.