S520

Chad Burrus

2010/02/01

1 4.4 Binomial Distribution

1. Binomial Experiments

   $X_1, X_2, \ldots, X_n$ ~Bernoulli(p)

   Implicit assumptions:

   (a) Trials are independent-when we distribute a random variable like this, we assume the trials are independent.

   (b) p is the same for each trial
   "Independent and identically distributed" or "iid"-so commonly used it has it's own acronym

$$X_i = \begin{cases} 1, & \text{trial is a success} \\ 0, & \text{trial is a failure} \end{cases}$$

$$\begin{aligned} Y &= \text{number of successes in } n \text{ trials} \\ &= \sum_{i=1}^{n} X_i \end{aligned}$$

This last formula makes it easier for us to find properties.

$$Y(S) = \{0, 1, 2, \ldots, n\}$$

Recall: for $k = \{0, 1, 2, \ldots, n\}$, $f(k) = P(Y = k) = P(k \text{ successes and } (n-k) \text{ failures})$.

Example: What's the probability of running n trials and getting k successes? So:

$$n = 5,\ k = 2$$

	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Total Probability
Experiment 1	0	1	1	0	0	
Probability	$(1-p)$	p	p	$(1-p)$	$(1-p)$	$= p^2(1-p)^3$
Experiment 2	0	0	1	0	1	
Probability	$(1-p)$	$(1-p)$	p	$(1-p)$	p	$= p^2(1-p)^3$

Figure 1: Two possible outcomes from 5 trials with two successes.

So, what can we say?

$$P(k \text{ successes}, (n-k) \text{ failures}) = \boxed{\binom{n}{k} p^k (1-p)^{n-k}}$$

2. Expected Value of Binomial Distributions

$$\begin{aligned} EX_i &= 1 \times P(X_i = 1) + 0 \times P(X_i = 0) \\ &= P(X_i = 1) = p \Leftarrow \text{Bernoulli trial expected value} \end{aligned} \tag{1}$$

$$EY = E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} EX_i = \sum_{i=1}^{n} p = np \Leftarrow \text{Binomial expected value}$$

3. Variance

$$\begin{aligned} VarX_i &= EX_i^2 - (EX_i)^2 \\ &= EX_i - (EX_i)^2 \Leftarrow \text{since Bernoulli trials are only } \{0,1\},\ 1^2 = 1, \text{ and } 0^2 = 0 \\ &= p - p^2 \\ &= p(1-p) \Leftarrow \text{variance of Bernoulli} \end{aligned}$$

Y ~Binomial($n; p$), so:

$$\begin{aligned} VarY &= Var\left(\sum_{i=1}^{n} X_i\right) \\ &= \sum_{i=1}^{n} X_i \Leftarrow \text{since the trials are independent} \\ &= \sum_{i=1}^{n} p(1-p) \\ &= np(1-p) \end{aligned}$$

The techniques are easy—building the model is usually much harder.

Exercise 4.5.11: We have a hotel with 100 rooms preparing for a statistics conference. Normally, there's a 10% no-show rate, so it accepts 110 reservations for the conference. However, there's only a 4% no-show rate for statisticians. What is the probability that at least 1 person with a reservation won't get a room?

Solution:

(a) Let each person with a reservation be a Bernoulli trial, so $n = 110$.

(b) Define rv X as:

$$X_i = \begin{cases} 1, & \text{person } i \text{ shows} \\ 0, & \text{person } i \text{ no-shows} \end{cases}$$

(c) Assumption: Are the trials independent? Pretty much—there may be some overlap, but in general they should be independent.

(d) Assumption: Is the probability p really the same for everyone? Again, close enough for our purposes.

(e) Then we have Y such that:

$$Y = \sum_{i=1}^{110} X_i = \text{number of people who claim their rooms}$$

What we're looking for $\Rightarrow P(Y > 100) = 1 - P(Y \leq 100) \Leftarrow$ easy to find if I have the cdf—that's where R comes in

There's an R function called *pbinom* that's typical of the way R functions are defined—the first letter is what you want, the remainder are the name of the probability family we want to use. Example:

$$
\begin{aligned}
&pbinom(\\
&\qquad q, \Leftarrow \text{Argument of the cdf (100, in this case)} \\
&\qquad size, \Leftarrow \text{Number of trials (110)} \\
&\qquad prob, \Leftarrow \text{Probability of success (0.96)} \\
&\qquad)
\end{aligned}
$$

$$
\begin{aligned}
&> 1 - pbinom(100, 110, 0.96) \\
&[1] 0.9870095
\end{aligned}
$$