Probability and Mathematical Statistics in Data Science

Lecture 08: Section 3.5: Examples

Binomial Probability Mass Function

- The pmf of a binomial random variable is

$$b(x; n, p) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x = 0, 1, 2, 3, \ldots, n \\ 0 & \text{otherwise} \end{cases}$$

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

n = number of trials

p = probability of success

$q = 1 - p$ = probability of failure

x = # of successes in n trials

Note: $n! = n \times (n-1) \times \ldots \times 2 \times 1$, and $n!$ is read as "*n*factorial."

Example

Binomial probabilities (Four patients). We have established a scenario in which four patients are treated with an intervention that is successful 75% of the time. The number of patients who respond to treatment is a binomial random variable with parameters $n = 4$ and $p = 0.75$. It follows that probability of a failure $q = 1 - p = 1 - 0.75 = 0.25$.

What is the probability of observing no successes among the four treatments?

Solution: $\Pr(X = 0) = {}_nC_x \cdot p^x \cdot q^{n-x} = {}_4C_0 \cdot 0.75^0 \cdot 0.25^{4-0} = 1 \cdot 1 \cdot 0.0039$
$= 0.0039$

What is the probability of one success?

Solution: $\Pr(X = 1) = {}_nC_x \cdot p^x \cdot q^{n-x} = {}_4C_1 \cdot 0.75^1 \cdot 0.25^{4-1}$
$= 4 \cdot 0.75 \cdot 0.015625 = 0.0469$

The probability of two successes is
$\Pr(X = 2) = {}_4C_2 \cdot 0.75^2 \cdot 0.25^{4-2} = 6 \cdot 0.5625 \cdot 0.0625 = 0.2109$

The probability of three successes is
$\Pr(X = 3) = {}_4C_3 \cdot 0.75^3 \cdot 0.25^{4-3} = 4 \cdot 0.4219 \cdot 0.25 = 0.4219$

The probability of four successes is
$\Pr(X = 4) = {}_4C_4 \cdot 0.75^4 \cdot 0.25^{4-4} = 1 \cdot 0.3164 \cdot 1 = 0.3164$

Table 6.1 lists this *pmf* in tabular form.

TABLE 6.1 Probability mass function for $X \sim b(4, 0.75)$ in tabular form.

No. of successes (x)	0	1	2	3	4
$\Pr(X = x)$	0.0039	0.0469	0.2109	0.4219	0.3164

Question

$$P(X = x) = {}_nC_x\, p^x q^{n-x} \qquad P(X = x) = q^{x-1}p$$

Teen smoking. The Centers for Disease Control and Prevention say that about 30% of high-school students smoke tobacco (down from a high of 38% in 1997). Suppose you randomly select high-school students to survey them on their attitudes toward scenes of smoking in the movies. What's the probability that

a) none of the first 4 students you interview is a smoker?
b) the first smoker is the sixth person you choose?
c) there are no more than 2 smokers among 10 people you choose?

Question

- **Lefties** Assume that 13% of people are left-handed. If we select 5 people at random, find the probability of each outcome.

  a) The first lefty is the fifth person chosen.
  b) There are some lefties among the 5 people.
  c) The first lefty is the second or third person.
  d) There are exactly 3 lefties in the group.
  e) There are at least 3 lefties in the group.

Hypergeometric pmf

- If X is the number of successes in a completely random sample of size *n* drawn from a population consisting of *M* successes and (*N* – *G* failures, then the distribution of X is given by

$$P(X = g) = \frac{\binom{G}{g}\binom{N-G}{n-g}}{\binom{N}{n}}$$

Example

An instructor who taught two sections of engineering statistics last term, the first with 20 students and the second with 30, decided to assign a term project. After all projects had been turned in, the instructor randomly ordered them before grading. Consider the first 15 graded projects.

a. What is the probability that exactly 10 of these are from the second section?

b. What is the probability that at least 10 of these are from the second section?

Randomized Controlled Experiments

In each experiment, **half** the participants will be randomly assigned to the treatment group and the other half to control. The experiment has 100 participants of whom 20 are men.

What is the chance that the treatment and control groups in the experiment contain the same number of men?

Did the treatment have an effect?

- RCE with 100 participants, 60 in Treatment, 40 in Control
- T: 50 recover, out of 60 (83%), C: 30 recover out of 40 (75%)
- Suppose treatment had no effect, and these 80 just happened to recover. What is the chance that 50 out of the 80 would have been assigned to the treatment group by chance?

Example

- A large supermarket chain in Florida occasionally selects employees to receive management training. A group of women there claimed that female employees were passed over for this training in favor of their male colleagues.

- The company denied this claim. (A similar complaint of gender bias was made about promotions and pay for the 1.6 million women who work or who have worked for Wal-Mart. The Supreme Court heard the case in 2011 and ruled in favor of Wal-Mart.)

Example

- Suppose that the large employee pool of the Florida chain (with a 1000 people) that can be tapped for management training is **half male and half female**. Since this program began, none of the 10 employees chosen have been female. What would be the probability of 0 out of 10 selections being female, if there truly was no gender bias?
- Method 1: pretend we are sampling with replacement, use Binomial.
- Method 2: Use Hypergeometric

Hypergeometric but don't know N

- A state has several million households, half of which have annual incomes over 50,000 dollars. In a simple random sample of 400 households taken from the state, what is the chance that more than 215 have incomes over 50,000 dollars?

How should we do this? $n = 400$, $k = 215$, $G=N/2$, $N=???$