STAT 88: Lecture 7

Contents

Last time

Sec 3.3 The binomial distribution has 2 parameters, Binomial(n, p):

- $n = \#$ independent trials
- $p =$ probability of success
- $X = \#$ successes out of n trials

Binomial formula:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Warm up: 13 cards are dealt from a deck with replacement:

(a) Find the chance that the hand contains two aces.

(b) Find the chance that the hand contains more than two aces.

(c) Find the chance that the hand contains six face cards.

3.4. The Hypergeometric Distribution

When you are sampling at random from a finite population, it is more natural to draw without replacement than with replacement.

Example: Five cards are dealt at the top of a deck. Find the chance of getting exactly 3 diamonds.

Let $X = \#$ diamonds out of 5 cards. We want to choose 3 diamonds out of 13(=52/4). There are $\binom{13}{3}$ ways to do this. For each of these we want to choose 2 nondiamonds out of 39 $\longrightarrow \binom{39}{2}$.

Since all $\binom{52}{5}$ sample are equally likely we get

$$P(X = 3) = \frac{\binom{13}{3}\binom{39}{2}}{\binom{52}{5}}.$$

More generally the ingredients of a hypergeometric distribution are:

- $N =$ population size (52 card deck)
- $G = \#$ good elements in your population ($B = N - G$ is the number of bad elements) (13 aces) (39 non aces)
- $n = \#$ sample size (5 cards)

Let $X = \#$ good elements in your sample. Then the hypergeometric formula is (3 cards)

$$P(X = g) = \frac{\binom{G}{g}\binom{B}{b}}{\binom{N}{n}}.$$

We say $X \sim \text{HG}(N, G, n)$.

Example: (Exercise 3.6.6) In a population of 200 voters, 70 are registered with Party A and the other 130 are registered with Party B. A simple random sample of 40 voters is drawn from this population. Let V be the number of sampled voters who are registered with Party A, and let $W = 40 - V$ be the number of sampled voters who are registered with Party B. Find:

(a) $P(V = 10)$

(b) $P(V > 10)$

(c) $P(W < 3V)$

Hypergeometric Probabilities in Python You can use the stats module of SciPy to calculate hypergeometric probabilities, just as you used it to calculate binomial probabilities.

Hypergeometric formula:

$$P(X = g) = \frac{\binom{G}{g}\binom{B}{b}}{\binom{N}{n}}.$$

$\frac{\binom{70}{10}\binom{130}{30}}{\binom{200}{40}}$ →

```
from scipy import stats
import numpy as np
```

g N G n

```
stats.hypergeom.pmf(10, 200, 70, 40)
```

```
0.05054861360578296
```

```
sum(stats.hypergeom.pmf(np.arange(11,71), 200, 70, 40))
```

```
0.9043345335065547
```

$\sum_{g=11}^{70} \frac{\binom{70}{g}\binom{130}{40-g}}{\binom{200}{40}}$

3.5. Examples

Problem solving techniques:

- Breaking the problem down into smaller pieces
- Examining the assumptions and hence deciding which distributions can be used
- Organizing the information to identify the parameters of the distributions
- Partitioning events into component pieces
- Using the addition and multiplication rules carefully

Randomized Controlled Experiments (RCE) In a RCE, a simple random sample of half the participants will be assigned to the treatment group (T) and the other half to control (C).

Experiment 1 has 100 participants of whom 20 are men.

Question 1. What is the chance that the treatment and control groups in Experiment 1 contain the same number of men?

– Organize the info to identify parameters

Experiment 1 has 100 participants of whom 20 are men.

Experiment 2 has 90 participants of whom 30 are men.

Question 2. What is the chance that the treatment groups in the two experiments have the same number of men?

- Organize the info to identify parameters
- Partition events into component pieces
- Use addition and multiplication rules

Fisher Exact Test In a randomized controlled experiment with 100 participants, 60 participants are in the treatment group and 40 are in the control group. In the treatment group, 50 out of the 60 participants recover after the treatment. In the control group, 30 out of the 40 participants recover.

A total of 80 patients recovered out of 100.

Question. Suppose the treatment is not effective. What is the chance that 50 or more of the recovered patients are randomly assigned to the treatment group?

Start with: What is the chance that 50 of the recovered patients are randomly assigned to the treatment group?