Hypergeometric Distributions

What is the expected number of aces you can expect get in a typical poker hand?

What is probability distribution for starting line-up for a basketball team?

These are not binomial distributions. How do we figure them out?

Consider the following…

In Ontario, a citizen may be called for jury duty every three years. Although most juries have 12 members, those for civil trials usually require only 6 members. Suppose a civil-court jury is being selected from a pool of 18 citizens, 8 of whom are men.

a) What is the probability distribution for the number of women on the jury?

b) How many women could you expect to have on the jury?

Let’s think about this…

- Are the probabilities the same for each jury member selected?
- Once a jury member is chosen, do you return him/her to the pool?
- That is, can a jury member be chosen twice?
- No replacement!
- The probabilities are *dependent*

Jury Selection

- This is an example of a hypergeometric distribution
- Success or failure only possible outcomes
- The probabilities are dependent
  - Probability of success changes as each trial made
- Let's calculate the theoretical probability distribution:

x	0	1	2	3	4	5	6
P(x)							

x	0	1	2	3	4	5	6
$P(x)$	$\frac{\binom{10}{0}\binom{8}{6}}{\binom{18}{6}}$ ≈ 0.00151	$\frac{\binom{10}{1}\binom{8}{5}}{\binom{18}{6}}$ ≈ 0.03017	$\frac{\binom{10}{2}\binom{8}{4}}{\binom{18}{6}}$ ≈ 0.16968	$\frac{\binom{10}{3}\binom{8}{3}}{\binom{18}{6}}$ ≈ 0.36199	$\frac{\binom{10}{4}\binom{8}{2}}{\binom{18}{6}}$ ≈ 0.31674	$\frac{\binom{10}{5}\binom{8}{1}}{\binom{18}{6}}$ ≈ 0.10860	$\frac{\binom{10}{6}\binom{8}{0}}{\binom{18}{6}}$ ≈ 0.01131

What is the expected number of women on the jury?

$$E(X) = \sum_{i=0}^{6} x_i P(x_i)$$

$$= 0 \cdot (0.00151) + 1 \cdot (0.03017) + 2 \cdot (0.16968) + 3 \cdot (0.36199)$$

$$+ 4 \cdot (0.31674) + 5 \cdot (0.10860) + 6 \cdot (0.01131)$$

$$\approx 3.33$$

The expected number of women on the jury is approximately 3.33.

Hypergeometric Distributions

- Success or failure only possible outcomes
- The probabilities are dependent
  - Probability of success changes as each trial made
- Random variable is number of successes
- The probability of obtaining *x* successes based on a random sample of size *n* from a population of size *N* is given by

$$\boxed{P(x) = \frac{C_k^x \cdot C_{N-k}^{n-x}}{C_N^n} = \frac{\binom{k}{x} \cdot \binom{N-k}{n-x}}{\binom{N}{n}}}$$

where *k* is the number of successes in the population.

Mean and Standard Deviation of a Hypergeometric Random Variable

A hypergeometric random variable X has mean and standard deviation given by the formulas

$$\boxed{\mu = E(X) = \frac{nk}{N}}$$

$$\boxed{\sigma = \sqrt{\left(\frac{N-n}{N-1}\right) \cdot n \cdot \frac{k}{N} \cdot \frac{N-k}{N}}}$$

where n is the sample size

k is the number of successes in the population

N is the size of the population

Check:

Calculate the expected number of women on the jury:

$N = 18$

$n = 6$

k = number of successes = number of women = 10

$$E(X) = \frac{6(10)}{18} \approx 3.33$$

Classify the following distributions as binomial, hypergeometric, or neither

Are there only two outcomes (success/failure)?

Are the trials indepndent/dependent?

(a) Drawing names out of a hat without replacement and recording the number of names that begin with a constant

(b) Generating random numbers on a calculator and counting the number of 5s

(c) Counting the number of hearts in a hand of five cards dealt from a well-shuffled deck.

(d) Asking all students in a class whether they prefer cola or ginger ale

Classify the following distributions as binomial, hypergeometric, or neither

(a) Drawing names out of a hat without replacement and recording the number of names that begin with a constant hypergeometric

(b) Generating random numbers on a calculator and counting the number of 5s binomial

(c) Counting the number of hearts in a hand of five cards dealt from a well-shuffled deck. hypergeometric

(d) Asking all students in a class whether they prefer cola or ginger ale binomial

Classify the following distributions as binomial, hypergeometric, or neither

(e) Predicting the number of green frogs from a randomly selected group of amphibians.

(f) Selecting the winning ticket in a lottery

(g) Predicting the expected number of heads when flipping a coin 100 times

(h) Predicting the number of boys among five children randomly selected from a group of eight boys and six girls

Classify the following distributions as binomial, hypergeometric, or neither

(e) Predicting the number of green frogs from a randomly selected group of amphibians.

hypergeometric

(f) Selecting the winning ticket in a lottery

Neither (uniform)

(g) Predicting the expected number of heads when flipping a coin 100 times

binomial

(h) Predicting the number of boys among five children randomly selected from a group of eight boys and six girls

hypergeometric

Example 2

A box contains seven yellow, three green, five purple, and six red candies jumbled together.

(a) What is the expected number of red candies among five candies poured from the box?

(b) Verify that the expectation formula for a hypergeometric distribution gives same result as the general equation for the expectation of any probability distribution.

Example 2a

$N = 7 + 3 + 5 + 6 = 21$

$n = 5$

k = number of successes = number of red candies = 6

$$E(X) = \frac{5(6)}{21} \approx 1.429$$

One would expect to have approximately 1.4 red candies among the 5 candies.

Example 2b

Using the general formula for expected value:

$$E(X) = \sum x_i P(x_i)$$

$$= 0 \cdot \frac{\binom{6}{0}\binom{15}{5}}{\binom{21}{5}} + 1 \cdot \frac{\binom{6}{1}\binom{15}{4}}{\binom{21}{5}} + 2 \cdot \frac{\binom{6}{2}\binom{15}{3}}{\binom{21}{5}} + 3 \cdot \frac{\binom{6}{3}\binom{15}{2}}{\binom{21}{5}} + 4 \cdot \frac{\binom{6}{4}\binom{15}{1}}{\binom{21}{5}} + 5 \cdot \frac{\binom{6}{5}\binom{15}{0}}{\binom{21}{5}}$$

$$\approx 1.429$$

Again, the expected number of red candies is approximately 1.4.

Example 3

In the spring, the Ministry of the Environment caught and tagged 500 raccoons in a wilderness area. The raccoons were released after being vaccinated against rabies. To estimate the raccoon population in the area, the ministry caught 40 raccoons during the summer. Of these, 15 had tags.

(a) Determine whether this situation can be modelled with a hypergeometric distribution.

(b) Estimate the raccoon population in the wilderness area.

Example 3a

- Hypergeometric: 2 outcomes, dependent trials
- Raccoons either tagged (success) or not tagged (failure)
- The 40 raccoons caught were all different from each other
  - i.e., no repetitions
  - Trials dependent
- This situation has all the characteristics of hypergeometric distribution

Example 3b

Assume that number of tagged raccoons caught during summer is expected value

- That is, if we catch 40 raccoons, how many can we expected to be tagged?

$E(X) = 15$

Solve for population size, N

n = number of trials = number raccoons caught

= 40

k = number of successes in population = number of tagged raccoons in population = 500

Example 3b

$$E(X) = \frac{nk}{N}$$

$$15 = \frac{40(500)}{N}$$

$$N = \frac{40(500)}{15}$$

$$N \approx 1333.3$$

There are approximately 1333 raccoons in the area.

Example 4

Suppose that a researcher goes to a small college of 200 faculty, 12 of which have blood type O-negative. She obtains a simple random sample of $n = 20$ of the faculty. Determine the mean and standard deviation of the number of randomly selected faculty that will have blood type O-negative.

Example 4

$$N = 200$$
$$n = 20$$
$$k = 12$$

$$\mu = E(X) = \frac{nk}{N} = \frac{20 \cdot 12}{200} = 1.2$$

$$\sigma = \sqrt{\left(\frac{N-n}{N-1}\right) \cdot n \cdot \frac{k}{N} \cdot \frac{N-k}{N}} = \sqrt{\left(\frac{200-20}{199}\right) \cdot 20 \cdot \frac{12}{200} \cdot \frac{200-12}{200}} = 1.01$$

Interpretation: We expect that, in a random sample of 20 faculty members, 1.2 will have blood type O-negative. If we take many different samples of size 20 from this population, the mean number of faculty that have blood type O-negative will approach 1.2.