Probability Distributions:

Hypergeometric Distribution

Consider this scenario: a hat contains 20 employee names, 12 of which are female. If five names are drawn from the hat, what is the probability that there is one female name drawn?

The difference between hypergeometric and binomial is:

In hypergeometric, the value of p changes with each trial.

So how do we solve this then?

? Consider this scenario: a hat contains 20 employee names, 12 of which are female. If five names are drawn from the hat, what is the probability that there is one female name drawn?

By Problem Solving

$$\frac{\text{Favourable Outcomes}}{\text{Total Outcomes}}$$

By Following Rules

a = the number of female names in the hat (12)
n = the population size (20)
r = the number of trials (5)
x = the number of favourable outcomes (p)

$$P(1) = \frac{{}_aC_x \times {}_{n-a}C_{r-x}}{{}_nC_r}$$
$$= \frac{{}_{12}C_1 \times {}_8C_4}{{}_{20}C_5}$$
$$= \frac{12 \times 70}{15504}$$
$$\cong 0.0542$$

Expected Value

$$E(X) = r\,\frac{a}{n}$$

Achtung! - r

nP

Example: A box of thumbtacks contains 15 red and 20 blue tacks. If 10 tacks are withdrawn at random from the box, what is the expected number of red tacks drawn?

Let a success be that a red thumbtack is drawn. We are drawing without replacement and we have success/failure outcomes, so the probability distribution function is hypergeometric where
a = the number of red thumbtacks (15)
n = the population size (35)
r = the number of trials (10)

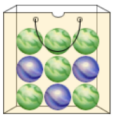
The expected value is:

$$E(X) = \frac{ra}{n}$$
$$= \frac{10 \times 15}{35}$$
$$= \frac{150}{35}$$
$$\cong 4$$

One would expect to draw 4 tacks on average.

Achtung!

What is the probability of drawing out 2 blue and 1 green marble in any order (drawing one at a time without replacement)?



Drawing them all at once (Hypergeometric Probability)

$$\frac{{}_3C_2 \cdot {}_6C_1}{{}_9C_3} = 21.4\%$$

Dependent Probability (Using techniques from unit two)

$$\frac{3}{9} \cdot \frac{2}{8} \cdot \frac{6}{7} = \frac{36}{504} \cong 7.1\%$$

1. Which of these random variables have a hypergeometric distribution? Explain why.
   a) the number of clubs dealt from a deck
   b) the number of attempts before rolling a six with a die
   c) the number of 3s produced by a random-number generator
   d) the number of defective screws in a random sample of 20 taken from a production line that has a 2% defect rate
   e) the number of male names on a page selected at random from a telephone book
   f) the number of left-handed people in a group selected from the general population
   g) the number of left-handed people selected from a group of 10 comprised equally of left-handed and right-handed people

2. Consider this scenario: a hat contains 20 employee names, 12 of which are female. If five names are drawn from the hat, what is the probability that there are **at least two** female names drawn?

3. There are five cats and seven dogs in a pet shop. Four pets are chosen at random for a visit to a children's hospital.
   a) What is the probability that exactly two of the pets will be dogs?
   b) What is the expected number of dogs chosen?

4. In a swim meet, there are 16 competitors, 5 of whom are from the Eastern Swim Club.
   a) What is the probability that 2 of the 5 swimmers in the first heat are from the Eastern Swim Club?
   b) What is the expected number of Eastern Swim Club members in the first heat?

5. A 12-member jury for a criminal case will be selected from a pool of 14 men and 11 women.
   a) What is the probability that the jury will have 6 men and 6 women?
   b) What is the probability that at least 3 jurors will be women?
   c) What is the expected number of women?

6. In a study of Canada geese, 200 of a known population of 1200 geese were caught and tagged. Later, another 50 geese were caught.
   a) Develop a simulation to determine the expected number of tagged geese in the second sample.
   b) Use the formula from this lesson to verify your simulation results.

7. Earlier this year, 520 seals were caught and tagged. On a recent survey, 30 out of 125 seals had been tagged.
   a) Estimate the size of the seal population.
   b) Explain why you cannot calculate the exact size of the seal population.

Think It Through, Talk It Over:

8. Under what conditions would a binomial distribution be a good approximation for a hypergeometric distribution.

Answer Clues

1a) yes
b) no
c) no
d) no
e) technically yes, but effectively no
f) technically yes, but effectively no
g) yes

2) 94.2%

3a) 42.5%
b) between 2 and 3 dogs

4a) 37.8%
b) Between 1 and 2

5a) 26.7%
b) 98.1%
c) Between five and six

6a) Should be between 8 and 9

7a) About 2,166
b) Population changes and you can't account for statistical discrepancy

8) When the population of objects is very large as to be hardly affected by drawing objects without replacement.