Hypergeometric Distribution

Situation:

Suppose you are drawing n items without replacement from a population which consists of N items, in which G items are good, and $N - G = B$ items are bad. Let X be the number of good items found in the n draws.

Useful Information

Official Course Notation: The book uses the following notation for the hypergeometric distribution:

- N is the total number of items in the population.
- p is the *proportion* of good items among the N total items.
- n is the total number of draws.

This produces the following notation (which you should use if you are asked to provide notation on exams, quizzes, and homeworks):

Alternate Notation: There are actually many ways of describing the parameters of the hypergeometric distribution. While you do not need to know this notation, it might help to see how to correctly set up problems.

- N is the total number of items in the population.
- G is the total number of "good" items in the population. Note that $G = Np$, and $p = G/N$. To clarify things further, we can label $B = N - G$ to be the number of "bad" items in the population. This parameter is not required (i.e. it is defined in terms of other parameters, but it can be useful in setting up problems.
- n is the number of draws made.

PMF:

Expected Value:

Variance:

Calculation when N is large:

Rule of Thumb:

Examples

1. A box has 10 balls, 4 of which are red. You draw three balls
   (a) with replacement;
   (b) without replacement.

   In each case, find the probability that exactly two of the balls are red. How many red balls do you expect to get on average? What is the variance? How do these numbers change depending on whether the drawing is done with or without replacement?

2. A box has 10000 balls, 4000 of which are red. You draw three balls
   (a) with replacement;
   (b) without replacement.

   In each case, find the probability that exactly two of the balls are red. How many red balls do you expect to get on average? What is the variance? How do these numbers change depending on whether the drawing is done with or without replacement?

3. What is the probability of getting a heart flush in a 5-card poker hand?

4. What is the probability of getting a heart flush in a 7-card poker hand?

Poisson Distribution

Situation:

Suppose you are interested in how many events occur in a specified time period. You know how many will occur *on average* in that time period, and you know that the number of events in non-overlapping time periods are independent of each other. Some typical examples of this include:

- Number of calls into a calling center in an hour.
- Number of babies born in a particular hospital in a day.
- Number of cars to drive past a particular landmark in a week.

Length, area, and width can also be described in the same way:

- Number of trees in an acre of forest.
- Number of exits in a 10-mile stretch of highway.
- Number of molecules in a cubic centimeter of air.

All of these examples make use of the Poisson distribution. It is used to study situations where the *rate* at which events occur is important.

Useful Information

Key Features:

Notation:

PMF:

Expected Value:

Variance:

Summing Poisson Random Variables:

Approximating the Binomial

Examples

1. Let X be the number of telephone calls you receive in a day. Suppose that on average, you receive 8 phone calls per day.
   (a) What is the probability that you receive exactly one phone call tomorrow?

   (b) What is the probability that you receive at least two phone calls tomorrow?

2. Let X be the number of people visiting an art show on a given day. Suppose that the average number of people to visit is 10 per day.

   (a) What is the probability that exactly 8 people come?

   (b) What is the probability that 20 people come over the next three days?

   (c) What is the probability that 8 people come on exactly one of the next three days?

   (d) Suppose that the art show lasts a total of 5 days. What is the probability that more than 8 people attend the art show at least four of the five days?

   (e) Suppose each visitor is charged \$10 as an admission fee, and it takes \$80 per day to run the show. Should you expect to make or lose money in the process of this art show? Why or why not?

3. The infant mortality rate in Sweden is 4.5 per 1000 live births. What is the probability of seeing no infant deaths if we randomly examine 500 live births?

4. Using the information from above, what is the probability of seeing at most three infant deaths?

5. The old faithful geyser at Yellowstone National Park has an average of 15 eruptions a day. Only 40% of these eruptions last longer than 2 minutes. For each of the following random variables, write down the distribution and the PMF. (Identify all parameters) What are the expectation and variance?

   (a) Let X be the number of eruptions the geyser has in one day.

   (b) Let Y be the number of eruptions lasting longer than 2 minutes out of the next 15 eruptions.

   (c) Let Z be the number of eruptions until five eruptions last longer than 2 minutes.

   (d) Let $U = 1$ if at least one of the next three eruptions lasts longer than 2 minutes, 0 otherwise.

(e) Let V be the number of days on which the geyser has exactly 10 eruptions out of the next 30 days.

(f) Let W be the number of days until the geyser has exactly 6 eruptions lasting over 2 minutes.

(g) Let R be the number of days until the geyser has at least 10 eruptions, given that there are no eruptions in the next 10 days.

(h) Let S be the number of eruptions lasting longer than 2 minutes in the next 10 eruptions, given that there are at least 5 eruptions lasting longer than 2 minutes.

6. Suppose that a stretch of road has an average of 20 potholes per mile. Suppose that Johnny, a rather uncoordinated biker, has a 30% chance of hitting each pothole while riding his bike on a three-mile stretch of this road.

   (a) What is the probability that the three mile stretch of road has exactly 50 potholes?

   (b) What is the probability that the first half-mile stretch has fewer than 5 potholes?

   (c) What is the expected number of potholes that Johnny hits, given that there are exactly 50 potholes?

   (d) What is the probability that Johnny successfully navigates the first 10 potholes?

   (e) What is the probability that Johnny successfully navigates the first 25 potholes, given that he successfully navigates the first 10 potholes?

   (f) What is the expected number of potholes that Johnny hits?