Discrete Random Variables: Hypergeometric (optional)*

Susan Dean
Barbara Illowsky, Ph.D.

Abstract

This module describes the properties of a hypergeometric experiment and hypergeometric probability distribution. This module is included in the Collaborative Statistics textbook/collection as an optional lesson.

The characteristics of a hypergeometric experiment are:

1. You take samples from **2** groups.
2. You are concerned with a group of interest, called the first group.
3. You sample **without replacement** from the combined groups. For example, you want to choose a softball team from a combined group of 11 men and 13 women. The team consists of 10 players.
4. Each pick is **not** independent, since sampling is without replacement. In the softball example, the probability of picking a women first is $\frac{13}{24}$. The probability of picking a man second is $\frac{11}{23}$ if a woman was picked first. It is $\frac{10}{23}$ if a man was picked first. The probability of the second pick depends on what happened in the first pick.
5. You are **not** dealing with Bernoulli Trials.

The outcomes of a hypergeometric experiment fit a **hypergeometric probability** distribution. The random variable X = the number of items from the group of interest. The mean and variance are given in the summary.

Example 1

A candy dish contains 100 jelly beans and 80 gumdrops. Fifty candies are picked at random. What is the probability that 35 of the 50 are gumdrops? The two groups are jelly beans and gumdrops. Since the probability question asks for the probability of picking gumdrops, the group of interest (first group) is gumdrops. The size of the group of interest (first group) is 80. The size of the second group is 100. The size of the sample is 50 (jelly beans or gumdrops). Let X = the number of gumdrops in the sample of 50. X takes on the values $x = 0$, 1, 2, ..., 50. The probability question is $P(x = 35)$.

Example 2

Suppose a shipment of 100 VCRs is known to have 10 defective VCRs. An inspector randomly chooses 12 for inspection. He is interested in determining the probability that, among the 12, at

*Version 1.15: May 27, 2012 4:17 am -0500
†http://creativecommons.org/licenses/by/3.0/

most 2 are defective. The two groups are the 90 non-defective VCRs and the 10 defective VCRs. The group of interest (first group) is the defective group because the probability question asks for the probability of at most 2 defective VCRs. The size of the sample is 12 VCRs. (They may be non-defective or defective.) Let X = the number of defective VCRs in the sample of 12. X takes on the values 0, 1, 2, ..., 10. X may not take on the values 11 or 12. The sample size is 12, but there are only 10 defective VCRs. The inspector wants to know $P(x \leq 2)$ ("At most" means "less than or equal to").

Example 3

You are president of an on-campus special events organization. You need a committee of 7 to plan a special birthday party for the president of the college. Your organization consists of 18 women and 15 men. You are interested in the number of men on your committee. If the members of the committee are randomly selected, what is the probability that your committee has more than 4 men?

This is a hypergeometric problem because you are choosing your committee from two groups (men and women).

Problem 1 *(Solution on p. 4.)*

Are you choosing with or without replacement?

Problem 2 *(Solution on p. 4.)*

What is the group of interest?

Problem 3 *(Solution on p. 4.)*

How many are in the group of interest?

Problem 4 *(Solution on p. 4.)*

How many are in the other group?

Problem 5 *(Solution on p. 4.)*

Let X = __________ on the committee. What values does X take on?

Problem 6 *(Solution on p. 4.)*

The probability question is P(_______).

1 Notation for the Hypergeometric: H = Hypergeometric Probability Distribution Function

$X \sim \mathrm{H}(\mathrm{r},\ \mathrm{b},\ \mathrm{n})$

Read this as "X is a random variable with a hypergeometric distribution." The parameters are r, b, and n. r = the size of the group of interest (first group), b = the size of the second group, n = the size of the chosen sample

Example 4

A school site committee is to be chosen randomly from 6 men and 5 women. If the committee consists of 4 members chosen randomly, what is the probability that 2 of them are men? How many men do you expect to be on the committee?

Let X = the number of men on the committee of 4. The men are the group of interest (first group).

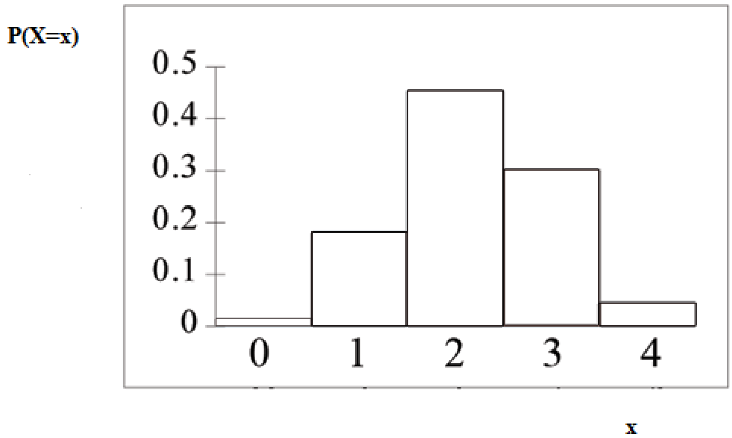
X takes on the values 0, 1, 2, 3, 4, where $r = 6$, $b = 5$, and $n = 4$. $X \sim H(6, 5, 4)$

Find $P(x = 2)$. $P(x = 2) = 0.4545$ (calculator or computer)

NOTE: Currently, the TI-83+ and TI-84 do not have hypergeometric probability functions. There are a number of computer packages, including Microsoft Excel, that do.

The probability that there are 2 men on the committee is about 0.45.

The graph of $X \sim H(6,5,4)$ is:



The y-axis contains the probability of X, where X = the number of men on the committee.

You would expect $m = 2.18$(about 2) men on the committee.

The formula for the mean is $\mu = \frac{n \cdot r}{r+b} = \frac{4 \cdot 6}{6+5} = 2.18$

The formula for the variance is fairly complex. You will find it in the Summary of the Discrete Probability Functions Chapter[1].

[1]"Discrete Random Variables: Summary of the Discrete Probability Functions" <http://cnx.org/content/m16833/latest/>

Solutions to Exercises in this Module

Solution to Example 3, Problem 1 (p. 2)
Without
Solution to Example 3, Problem 2 (p. 2)
The men
Solution to Example 3, Problem 3 (p. 2)
15 men
Solution to Example 3, Problem 4 (p. 2)
18 women
Solution to Example 3, Problem 5 (p. 2)
Let X = **the number of men** on the committee. x = 0, 1, 2, …, 7.
Solution to Example 3, Problem 6 (p. 2)
P(x>4)

Glossary

Definition 1: Hypergeometric Distribution
A discrete random variable (RV) that is characterized by

- A fixed number of trials.
- The probability of success is not the same from trial to trial.

We sample from two groups of items when we are interested in only one group. X is defined as the number of successes out of the total number of items chosen. Notation: $X\tilde{}H\left(r,b,n\right)$., where r = the number of items in the group of interest, b = the number of items in the group not of interest, and n = the number of items chosen.