

4.2: Hypergeometric Distribution

The simplest probability density function is the hypergeometric. This is the most basic one because it is created by combining our knowledge of probabilities from Venn diagrams, the addition and multiplication rules, and the combinatorial counting formula.

To find the number of ways to get 2 aces from the four in the deck we computed:

$$\binom{4}{2} = \frac{4!}{2!(4-2)!} = 6 \quad (4.2.1)$$

And if we did not care what else we had in our hand for the other three cards we would compute:

$$\binom{48}{3} = \frac{48!}{3!45!} = 17,296 \quad (4.2.2)$$

Putting this together, we can compute the probability of getting exactly two aces in a 5 card poker hand as:

$$\frac{\binom{4}{2}\binom{48}{3}}{\binom{52}{5}} = .0399 \quad (4.2.3)$$

This solution is really just the probability distribution known as the Hypergeometric. The generalized formula is:

$$h(x) = \frac{\binom{A}{x}\binom{N-A}{n-x}}{\binom{N}{n}} \quad (4.2.4)$$

where x = the number we are interested in coming from the group with A objects.

$h(x)$ is the probability of x successes, in n attempts, when A successes (aces in this case) are in a population that contains N elements. The hypergeometric distribution is an example of a discrete probability distribution because there is no possibility of partial success, that is, there can be no poker hands with $21/2$ aces. Said another way, a discrete random variable has to be a whole, or counting, number only. This probability distribution works in cases where the probability of a success changes with each draw. Another way of saying this is that the events are NOT independent. In using a deck of cards, we are sampling WITHOUT replacement. If we put each card back after it was drawn then the hypergeometric distribution be an inappropriate Pdf.

For the hypergeometric to work,

1. the population must be dividable into two and only two independent subsets (aces and non-aces in our example). The random variable X = the number of items from the group of interest.
2. the experiment must have changing probabilities of success with each experiment (the fact that cards are not replaced after the draw in our example makes this true in this case). Another way to say this is that you sample without replacement and therefore each pick is not independent.
3. the random variable must be discrete, rather than continuous.

? Exercise 4.2.1

A candy dish contains 30 jelly beans and 20 gumdrops. Ten candies are picked at random. What is the probability that 5 of the 10 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 A in the formula) is gumdrops. The size of the group of interest (first group) is 30. The size of the second group is 20. The size of the sample is 10 (jelly beans or gumdrops). Let X = the number of gumdrops in the sample of 10. X takes on the values $x = 0, 1, 2, \dots, 10$. a. What is the probability statement written mathematically? b. What is the hypergeometric probability density function written out to solve this problem? c. What is the answer to the question "What is the probability of drawing 5 gumdrops in 10 picks from the dish?"

Answer

a. $P(x = 5)$

b. $P(x = 5) = \frac{\binom{30}{5}\binom{20}{5}}{\binom{50}{10}}$

c. $P(x = 5) = 0.215$

Solution

$$P(x = 35)$$

$$\mu = \frac{nr}{r+b} = \frac{(4)(6)}{6+5} = 2.18$$

Try It 4.2.1

A bag contains letter tiles. Forty-four of the tiles are vowels, and 56 are consonants. Seven tiles are picked at random. You want to know the probability that four of the seven tiles are vowels. What is the group of interest, the size of the group of interest, and the size of the sample?

? Exercise 4.2.2

TSA officers claim that they randomly select boarding passengers for additional search of carry-on luggage. During the past hour, of the 100 passengers passing through the security point, 5 were silver-haired grandmothers, of which 3 were chosen for additional screening among the 20 total chosen for “enhanced vetting.” Calculate the relevant probability and comment.

Answer

Step 1. Determine what the random variable is. Or asked another way, what are we interested in? In this case it is the number of silver-haired grandmothers who were stopped for additional search of carry-on luggage. We also determine that this is a discrete random variable because it is a counting number. Further, we determine that X can take on only the values $X = 0, 1, 2, 3, 4, 5$.

Step 2. What is the experiment? Here passengers are selected we are told randomly for additional baggage search. This would be selection without replacement.

Step 3. Can the objects, the passengers here, be divide into at most two categories? Yes, here because we are interested in only the category silver-haired grandmothers and everyone else. These data can be analyzed with the hypergeometric probability density function because it meets all the requirements of that distribution.

Step 4. What is the question asked? Unlike Example 4.1 about drawing candies from a bowl, there is no obvious question being asked. The candy bowl problem is mechanical with an obvious question directly stated concerning the probability of drawing 5 gumdrops. In short, an obvious question with no implications. Do we really care about the probability of drawing 5 gumdrops? Here there is an interesting question because the security office made a claim that we may indirectly test using simple probability. The security office claims they randomly select passengers for enhanced vetting. From the data we see immediately that of the 5 silver-haired grandmother passengers, 3 were selected for additional search. This is 60 percent of all silver-haired grandmothers. We can form our own question to determine the probability of selecting 3 of 5 in this category of passengers if we are selecting randomly. Remember that random selection means that each object has an equal probability of being selecting. Our probability question in words is thus “What is the probability of selecting 3 silver-haired grandmothers from a group of 100 passengers if the group included only 5 silver-haired grandmothers?” Formally, what is $P(X = 3)$ using the hypergeometric probability distribution?

From the formula: $X = 3$, $A = 5$ $N = 100$ (the total number of passengers from which those will be drawn for enhanced vetting) $n = 20$ (the number of passengers that will be selected for enhanced vetting)

$$P(x) = \frac{\binom{5}{3} \binom{95}{17}}{\binom{100}{20}} = 0.048 \quad (4.2.5)$$

Our conclusion is that the probability of the outcome of 3 of 5 silver-haired grandmothers being randomly selected for additional luggage search is only 0.048. This is only one one-thousandth better than drawing 5 cards from a 52-card deck and getting a pair of aces. In short, an unrealistically small probability. A comment here would be that we can with only just under 5% confidence believe that the passengers are selected on a purely random basis.

This example demonstrates the ability to answer questions of some interest and policy implication with the use of the tools of probability analysis.

Try It 4.2.2

There are a total of 200 balls in a bag. This includes 25 red balls. 40 balls are selected randomly out of the bag. 8 of them are found to be red. Find the relevant probability.

This page titled [4.2: Hypergeometric Distribution](#) is shared under a [CC BY 4.0](#) license and was authored, remixed, and/or curated by [OpenStax](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.