

## 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.

### ✓ Example 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?"

#### Solution

a.  $P(x = 5)$

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

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

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](#).

- [4.1: Hypergeometric Distribution](#) \* has no license indicated.