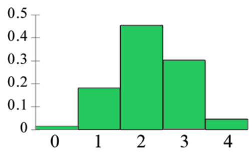
MDM4U: Mathematics of Data Management

Successes Among Dependent Events

Hypergeometric Probability Distributions

J. Garvin



Hypergeometric Probability Distributions

Recap

What is the probability of receiving two hearts in a three-card hand dealt from a standard deck?

The sample space is the number of three-card hands, or ${}_{52}C_3$.

There are 13 hearts in a deck, and 39 cards in the remaining 3 suits. Therefore, a two-heart hand can be dealt in ${}_{13}C_2 \times {}_{39}C_1$ ways.

The probability is $P(2) = \frac{{}_{13}C_2 \times {}_{39}C_1}{{}_{52}C_3} = \frac{3042}{22100} \approx 0.138$.

Hypergeometric Probability Distributions

What would the probability distribution for the number of hearts dealt in a three-card hand look like?

Use a table to calculate the probabilities of receiving 0-3 hearts:

Hearts	Probability
0	$\frac{{}_{13}C_0 \times {}_{39}C_3}{{}_{52}C_3} = \frac{9139}{22100} \approx 0.414$
1	$\frac{{}_{13}C_1 \times {}_{39}C_2}{{}_{52}C_3} = \frac{9633}{22100} \approx 0.436$
2	$\frac{{}_{13}C_2 \times {}_{39}C_1}{{}_{52}C_3} = \frac{3042}{22100} \approx 0.138$
3	$\frac{{}_{13}C_3 \times {}_{39}C_0}{{}_{52}C_3} = \frac{286}{22100} \approx 0.013$

Hypergeometric Probability Distributions

In the previous example, each time a heart was dealt, it was removed from the deck. It could not be selected again.

This is an example of *dependent* events.

We use *combinations* to select items *without replacement*.

We cannot use a binomial probability distribution because Bernoulli trials are independent events.

Hypergeometric Probability Distributions

Probability in a Hypergeometric Probability Distribution

If a is the number of successful outcomes from a total of n possible outcomes, then the probability of x successes in r dependent trials is

$P(x) = \frac{{}_{a}C_x \times {}_{n-a}C_{r-x}}{{}_{n}C_r}$

The proof of this formula comes directly from the definition of probability, $P(E) = \frac{n(E)}{n(S)}$, and the ways in which x items can be selected from a possible choices.

Hypergeometric Probability Distributions

Example

A manager randomly selects 4 employees from 6 men and 4 women to work as a team. Determine the probability that exactly two men are selected.

There are ${}_{6}C_2$ ways to choose the two men and ${}_{4}C_2$ ways to choose the two women to complete the team, from a total of ${}_{10}C_4$ possibilities.

Therefore, the probability that exactly 2 men are selected is $P(2) = \frac{{}_{6}C_2 \times {}_{4}C_2}{{}_{10}C_4} = \frac{15 \times 6}{210} = \frac{90}{210} = \frac{3}{7}$.

Hypergeometric Probability Distributions

Your Turn

You shuffle a standard deck of cards, and deal a hand of five cards from atop the deck. What is the probability that the hand contains exactly three clubs?

There are ${}_{13}C_3$ ways to choose the clubs and ${}_{39}C_2$ ways to choose the remaining two cards in the hand.

There are a total of ${}_{52}C_5$ possible five-card hands.

Therefore, the probability of having exactly 3 clubs is $P(3) = \frac{{}_{13}C_3 \times {}_{39}C_2}{{}_{52}C_5} = \frac{2\,717}{33\,320} \approx 0.08$.

Hypergeometric Probability Distributions

Example

A quality control inspector at a popsicle factory knows that approximately 6% of all popsicles shipped by the factory are broken. He randomly pulls 10 popsicles from a batch of 500 for testing. What is the probability that *at least* 2 are broken?

If 6% of all popsicles are broken, there should be $500 \times 0.06 = 30$ broken popsicles in the batch.

This means there are 30 broken (a) and 470 ($n - a$) non-broken popsicles, from which 10 (x) must be selected from a total of 500 (n) popsicles.

Hypergeometric Probability Distributions

To find the probability that at least 2 are broken, use an indirect method, eliminating the cases where no popsicles or one popsicle is broken.

$$\begin{aligned} P(x \geq 2) &= 1 - P(0) - P(1) \\ &= 1 - \frac{{}_{30}C_0 \times {}_{470}C_{10}}{{}_{500}C_{10}} - \frac{{}_{30}C_1 \times {}_{470}C_9}{{}_{500}C_{10}} \\ &\approx 0.116 \end{aligned}$$

Expected Value

Like the previous probability distributions, there is a formula for expected value.

Example

In a hypergeometric probability distribution consisting of r trials, with a successful outcomes from a total of n possible outcomes, the expected value is $E(X) = \frac{ra}{n}$

Since the proof of this formula requires a great deal of algebraic manipulation, it is left as an exercise for the determined student.

Expected Value

Example

A bag contains 15 red and 20 blue marbles. If 7 marbles are randomly drawn from the bag, what is the expected number of red marbles drawn?

There are $r = 7$ dependent trials, and $a = 15$ successful outcomes from a total of $n = 15 + 20 = 35$ possible outcomes.

So the expected number of red marbles drawn is $E(X) = \frac{7 \times 15}{35} = \frac{105}{35} = 3$.

Expected Value

Your Turn

1000 fish from a small lake were caught and tagged. After one year, a new catch of 500 fish contained 60 that were tagged. What is the estimated fish population in the lake? *Hint: What do the three values given above represent in the equation for expected value?*

There are $r = 500$ trials (the new catch).

There are $a = 1000$ successful outcomes (catching a tagged fish).

The expected value is $E(X) = 60$ (the number of tagged fish caught).

Substituting into the expected value equation, we obtain $60 = \frac{500 \times 1000}{n}$, or $n = \frac{500 \times 1000}{60} \approx 8333$ fish.

Expected Value

A alternative method of solving is to assume that the proportion of tagged fish in the sample corresponds to that of the entire population.

60 tagged fish were caught in a sample of 500, for a proportion $\frac{60}{500}$.

1000 fish from the entire population were originally tagged, for a proportion $\frac{1000}{n}$.

Therefore, $\frac{60}{500} = \frac{1000}{n}$. This can be rearranged as $n = \frac{500 \times 1000}{60} \approx 8333$ fish, as obtained earlier.

Questions?