

Negative Hypergeometric Distribution

A random variable Y has a negative hypergeometric distribution if

- Objects are selected without replacement from a population of size N .
- Each member of the population belongs to one of two groups; success or failure. The number of successes in the population is denoted as r and therefore the number of failures in the population is $N - r$.
- Objects are selected from the population until the k th success occurs.
- The random variable Y is defined as the number of the trial on which the k th success occurs.
- The parameters for the hypergeometric random variable Y are value of k , the population size N and the number of successes in the population r .
- The probability distribution function of Y is

$$P(Y = y) = p(y) = \frac{\binom{y-1}{k-1} \binom{N-y}{r-k}}{\binom{N}{r}} \quad y = k, k + 1, \dots, N - r + k$$

- The theoretical mean of the negative hypergeometric random variable Y is

$$\mu = E(Y) = k \left(\frac{N + 1}{r + 1} \right)$$

- The theoretical variance of the negative hypergeometric random variable Y is

$$\sigma^2 = V(Y) = \frac{k(N + 1)(N - r)(r + 1 - k)}{(r + 1)^2(r + 2)}$$

Working with negative hypergeometric random variables in R.

R does not include the negative hypergeometric random variable as a built-in function. In order to work with this random variable, you will need to copy and paste the following function into R.

```
dneghyper<- function(y,r,N,k){choose(y-1,k-1)*choose(N-y,r-k)/choose(N,r)}
```

To find a probability $P(Y = y) = p(y)$ for a single value y , the command in R is

```
dneghyper(y,r,N,k)
```

To find the probability $P(Y \leq y)$, use the sum command to add up all $p(y)$ values for y between and including k and y .

```
sum(dneghyper(k:y,r,N,k))
```

To find the probability $P(y_1 \leq Y \leq y_2)$, use the sum command to add up all $p(y)$ values for y between and including y_1 and y_2 .

```
sum(dneghyper(y1:y2,r,N,k))
```

To find the probability $P(Y \geq y) = 1 - P(Y < y) = 1 - P(Y \leq y - 1)$, use the sum command to find $P(Y \leq y - 1)$ and subtract this value from 1.

```
1 - sum(dneghyper(k:y-1,r,N,k))
```

Problems.

1. How is the probability distribution function $p(y)$ derived?
2. A box contains 40 balls: 10 red, 15 yellow, and 15 green. Balls are selected from this box without replacement.
 - (a) What is the probability that the third yellow ball will be the seventh ball chosen?
 - (b) What is the probability that the third yellow ball will be the 10th ball chosen?
 - (c) Find the expected number of draws needed to obtain the third yellow ball.
 - (d) Find the variance of the number of draws needed to obtain the third yellow ball.
3. A company receives a shipment of 50 condensers, of which 5 are defective.
 - (a) When sampling without replacement, what is the probability the second defective condenser will be 10th condensers chosen?
 - (b) When sampling without replacement, what is the probability the third defective condenser will be chosen within the first 25 draws.
 - (c) Find the expected number of draws needed to obtain the second defective condenser.
 - (d) Find the variance of the number of draws needed to obtain the second defective condenser.