Probability and Statistics I

11. The Hypergeometric and Negative Binomial Distributions

3.6 The Hypergeometric and Negative Binomial Distributions

3.6a The Hypergeometric Distribution

Application

The hypergeometric distribution models the number of successes when sampling *without replacement* from a fixed population.

Assumptions

1. The population consists of a finite number of individuals, N.
2. M individuals can be classified as successes and $N - M$ as failures.
3. A sample of n individuals is chosen so *without replacement* so that each sample is equally likely.

Mathematically, we write

$$X \sim \text{Hypergeometric}(n, M, N).$$

PMF and CDF

- PMF: $h(x; n, M, N) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$
  $$x = \max(0, n - N + M), \ldots, \min(n, M)$$
- CDF: Requires special functions.

Properties

- Mean: $E(X) = nM/N$
- Variance: $V(X) = \left(\frac{N-n}{N-1}\right)\left(\frac{nM}{N}\right)\left(1 - \frac{M}{N}\right)$

Calculator

`https://stattrek.com/online-calculator/hypergeometric`

Example 11.1

An accountant is conducting a financial audit on a big, multinational company. The company has 200 accounts, of which 20 have errors.

Suppose that the accountant checks 10 randomly selected accounts. Let Y be the number of accounts with errors in her sample.

a) What is the distribution of Y?
b) What is the probability that $Y = 2$?
c) What are the mean and variance of Y?

Binomial Approximation

Let $p = M/N$ be the proportion of successes in the population. Then the mean and variance can be rewritten as

- Mean: $E(X) = np$
- Variance: $V(X) = \left(\frac{N-n}{N-1}\right) np\,(1-p)$

If N is much larger than n ($n/N \leq .05$), then $V(X) \approx np(1-p)$, which is the variance of a binomial distribution. In this case we can approximate the distribution of X as

$$X \dot{\sim} \text{Binomial}(n, p).$$

The symbol $\dot{\sim}$ is read as "approximately distributed as".

Your book suggests that this is appropriate when $n/N \leq .05$.

Example 11.2

An accountant is conducting a financial audit on a big, multinational company. The company has 200 accounts, of which 20 have errors.

Suppose that the accountant checks 10 randomly selected accounts. Let Y be the number of accounts with errors in her sample.

a) Is it appropriate to approximate the distribution of Y by a binomial?
b) What is the approximate distribution of Y?
c) Approximate the probability that $Y = 2$.
d) Approximate the mean and variance of Y.

3.6b The Negative Binomial Distribution

Application

The negative binomial distribution models the number of failures in a binomial experiment if trials are repeated until the r^{th} success.

Assumptions

1. The experiment consists of independent and identical trials.
2. Each trial has two outcomes (called success and failure).
3. The probability of success, p, is the same for all trials.
4. The experiment continues until the r^{th} success occurs.

Mathematically, we write

$$X \sim \text{Negative Binomial}(r, p).$$

PMF and CDF

- PMF: $nb(z; r, p) = \binom{z+r-1}{r-1} p^r (1-p)^z, \quad z = 0, 1, 2, \ldots$
- CDF: requires special functions

Properties

- Mean: $E(Z) = \frac{r(1-p)}{p}$
- Variance: $V(Z) = \frac{r(1-p)}{p^2}$

Calculator

`https://stattrek.com/online-calculator/negative-binomial`

Note that the calculator uses a different parametrization of the negative binomial distribution. It counts the number of trials $(x = z + r)$ until the r^{th} success.

Example 11.3

An accountant is conducting a financial audit on a big, multinational company. The company has 200 accounts, of which 20 have errors.

Suppose that the accountant samples accounts *with replacement* until she finds 2 with errors. Let Z denote the number of accounts without errors in her sample.

a) What is the distribution of Z?
b) What is the probability that $Z = 10$?
c) What are the mean and variance of Z?

Application

The geometric distribution is the special case of the negative binomial distribution that models the number of failures in a binomial experiment if trials are repeated until the 1^{st} success.

Assumptions

1. The experiment consists of independent and identical trials.
2. Each trial has two outcomes (called success and failure).
3. The probability of success, p, is the same for all trials.
4. The experiment continues until the 1^{st} success occurs.

Mathematically, we write

$$X \sim \text{Geometric}(p).$$

PMF and CDF

- PMF: $nb(z; 1, p) = p^1(1-p)^z, \quad z = 0, 1, 2, \ldots$
- CDF: requires special functions

Properties

- Mean: $E(Z) = \frac{(1-p)}{p}$
- Variance: $V(Z) = \frac{(1-p)}{p^2}$

Binomial vs Hypergeometric vs Negative Binomial

The Binomial, Hypergeometric, and Negative Binomial distributions all apply to experiments with repeated trials in which repeated trials result in one of two outcomes (labelled success and failure).

	Binomial	Hypergeometric	Neg. Binomial
# of Trials	Fixed	Fixed	Random
Trials are independent	Yes	No	Yes
Prob. of Success	Fixed	Varies	Fixed

Questions?

Exercise 11.1

Suppose that packages of Smarties each contain 30 smarties and that there is a .25 probability that each Smartie is red. For each of the following problems: i) identify the distribution of the random variable, ii) compute the mean and variance, iii) compute the probability provided.

a) The number of red Smarties in a package and the probability that the package contains more than 10 red Smarties.

b) The number of red Smarties you pick if you draw 5 Smarties without replacement from a pack containing exactly 8 red Smarties. The probability this number is less than 3.

c) The number of package you must open until you find a package with no red smarties. The probability you open exactly 5000 boxes until finding one with no red candies.