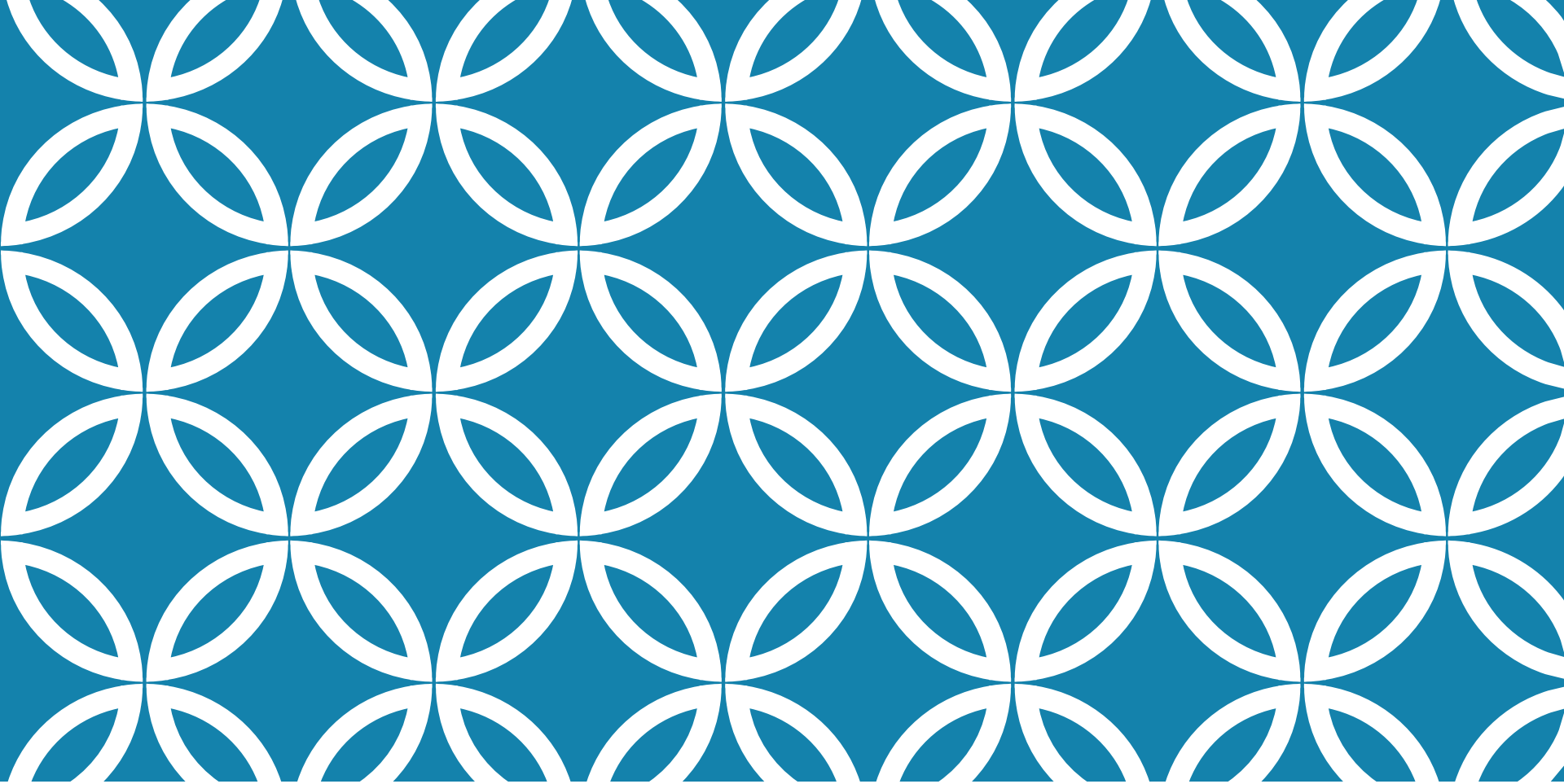


HYPER-GEOMETRIC PROBABILITY DISTRIBUTION

Chapter 5

5-1

HYPER-GEOMETRIC DISTRIBUTION

The hyper-geometric probability distribution is closely related to the binomial distribution. The two-probability distribution differ in two key ways.

With the hyper-geometric distribution, the trials are not independent, and the probability of success changes from trial to trial.

In the usual notation for the Hyper-geometric distribution, N_1 denotes the number of elements in the population of size N labeled success, and $(N-N_1)$denotes the number of elements in the population labeled failure.

HYPERGEOMETRIC DISTRIBUTION

When sampling from relatively small populations without replacement, we use the hypergeometric distribution

HYPERGEOMETRIC PROBABILITY EXPERIMENT

1. An outcome on each trial of an experiment is classified into one of two mutually exclusive categories — a success or a failure.
2. The random variable is the number of successes in a fixed number of trials.
3. The trials are not independent.
4. We assume that we sample from a finite population without replacement and $n/N > 0.05$. So, the probability of a success changes for each trial.

HYPER-GEOMETRIC PROBABILITY FUNCTION

The Hyper-geometric probability function is used to compute the probability that in a random selection of n elements, selected without replacement, we obtain x elements labeled success and $(n-x)$ elements labeled failure. For this outcome to occur we must obtain x successes from the $N1$ successes in the population and $n-x$ failures from $(N-N1)$ failures

THE FORMULA

$$P(x)= \frac{C_{N_1}^{x} C_{N-N_1}^{n-x}}{C_{N}^{n}} \text{ .....[5-1]}$$

Where:

xx = the number of successes

nn = the number of trials.

P(x)= the probability of x successes in n trials.

N= the number of elements in the population.

NN_1= the number of elements in the population labeled success.

NOTE:

C_N^n represents the number of ways n can be selected from a population of size N.

$C_{N_1}^x$ represents the number of ways that x successes can be selected from a total of N_1 successes in the population.

$C_{N-N_1}^{n-x}$ represents the number of ways that n-x failures can be selected from a total of $(N-N1)$ failures in the population.

For the hyper-geometric probability distribution, x is a discrete random variable and the probability function p(x) given by equation [5-1] is usually applicable for values x= 0, 1, 2…n

NOTE:(CONT.)

However only values of x where the number of observed successes is less than or equal to the number of successes in the population $(x \leq N1)$ and where the number of observed failures is less than or equal to the number of failures in the population $(n-x) \leq (N-N1)$ are valid. if these conditions do not hold for one or more values of x, the corresponding p(x)= 0 indicates that the probability of this value of x is zero.

EXAMPLE

Let us consider the following quality control application, Electric battery produced by Ontario Electric are packaged in boxes of 12 units each. Suppose an inspector randomly selects three of the 12 batteries in a box for testing. If the box contains exactly 5 defective batteries.

What is the probability that the inspector will find exactly 01 of the 03 batteries defective?

Solution:

In this example we have: n=3 ,N=12, N1 = 5 ,so the probability of finding x=1 defective batteries is:

EXAMPLE (CONT.)

$$P(x) = \frac{C_{N_1}^{x} C_{N-N_1}^{n-x}}{C_N^n} = \frac{C_5^1 C_{12-5}^{3-1}}{C_{12}^3} = \frac{\left(\frac{5!}{1!4!}\right)\left(\frac{7!}{2!5!}\right)}{\frac{12!}{3!9!}} = 0{,}4773$$

Now suppose we wanted to know the probability of finding at least 01 defective battery?

The easiest way to find this probability is to first compute the probability that the inspector does not find any defective battery. The probability of (x=0).

$$P(x=0) = \frac{C_{N_1}^{x} C_{N-N_1}^{n-x}}{C_N^n} = \frac{C_5^0 C_{12-5}^{3-0}}{C_{12}^3} = \frac{\frac{5!}{0!5!}\,\frac{7!}{3!4!}}{\frac{12!}{3!9!}} = 0{,}1591$$

$P(x \geq 1) = 1 - P(x < 1) = 1 - P(x=0) = 1 - 0{,}1591 = 0{,}8409$

The probability of finding at least 1 battery defective is 0.8409

THE EXPECTED VALUE AND VARIANCE FOR HYPER GEOMETRIC PROBABILITY DISTRIBUTION:

The formula of the expected value E(x):

$$E(x) = \mu = n*\left(\frac{N_1}{N}\right) = n*p \quad \text{.....[5-2]}$$

Where: $p = \frac{N_1}{N}$ = The probability of succes on the first trial.

For the variance V(x):

$$V(x) = \partial^2 = \mathbf{n}\left(\frac{N_1}{N}\right)\left(1 - \frac{N_1}{N}\right)\left(\frac{N-n}{N-1}\right) = \mathbf{n.p.q.}\left(\frac{N-n}{N-1}\right) \text{ .. [5-3]}$$

Where: $q = \left(1 - \frac{N_1}{N}\right) = (\mathbf{1 - p})$

In the preceding example: n=3, $N1=5$ and N=12 thus,

$$p = \frac{N_1}{N} = \frac{5}{12} = 0.4167$$

$$E(x) = n*p = 3*0{,}4167 = 1{,}25$$

$$V(x) = n*p*q*\left(\frac{N-n}{N-1}\right) = 3*0{,}4167*(1-0{,}4167)*\left(\frac{12-3}{12-1}\right) = 0{,}6$$

The standard deviation $\sigma = \sqrt{V(x)} = \sqrt{0{,}6} = 0{,}77$

NOTE

If the population size is large, the term $\left(\frac{\mathbf{N-n}}{\mathbf{N-1}}\right)$ in equation (5-2) approaches .

As result $E(x) = n.p$ and $Var(x) = n.p.q$ note that these expressions are the same as the expressions used to compute the expected value and variance of a binomial distribution. When the population size is large, a hypergeometric distribution can be approximated a binomial distribution with n trials and a probability of success $p = \frac{N_1}{N}$.

EXAMPLE

Business Faculty students' Grades:

According to the Higher ministry of Algeria, 7 of the top 10 graduate faculties of business have students with an average undergraduate grade of 15 or higher. Suppose that we randomly select 2 of the top 10 graduate faculties of business.

Answer to these questions :

1. What is the probability that exactly one faculty has students with an average undergraduate grade of 15 or higher?
2. What is the probability that both faculties have students with an average undergraduate grade of 15 or higher.
3. What is the probability that neither faculty has students with an average undergraduate grade of 15 or higher?
4. find E(x) and Var(x).

EXAMPLE (CONT.)

Solution:

1. $P(x) = \frac{C_{N_1}^{x} C_{N-N_1}^{n-x}}{C_N^n} = \frac{C_7^1 C_{10-7}^{2-1}}{C_{10}^2} = \frac{C_7^1 C_3^1}{C_{10}^2} = 0{,}467$
2. $P(x=2) = \frac{C_{N_1}^{x} C_{N-N_1}^{n-x}}{C_N^n} = \frac{C_7^2 C_{10-7}^{2-2}}{C_{10}^2} = 0{,}467$
3. $P(x=2) = \frac{C_{N_1}^{x} C_{N-N_1}^{n-x}}{C_N^n} = \frac{C_7^0 C_{10-7}^{2-0}}{C_{10}^2} = 0{,}0067$
4. $E(x) = n^*p = 2 * \frac{7}{10} = 1{,}4$
5. $Var(x) = \left(\frac{N-n}{N-1}\right) = 2 * \frac{7}{10} * \left(\frac{10-2}{10-1}\right) = 0{,}373$