Lecture 12

Hypergeometric and Poisson Random Variables

Text: Chapter 4

STAT 8010 Statistical Methods I
September 16, 2019

Whitney Huang
Clemson University

Notes

Agenda

1. Bernoulli Trials and Binomial Random Variables
2. Hypergeometric Random Variables
3. Poisson Random Variables

Notes

Bernoulli Trials and Binomial Random Variables

- **Bernoulli Trials**:
  - The result of each trial may be either a success or failure
  - The probability of success, p, is the same in every trial
  - The trials are **independent:** the outcome of one trial has no influence on later outcomes
- **Binomial Random Variables**
  - The number of successes in n Bernoulli trials, where the probability of success in one trial is $p \Rightarrow X \sim \text{Bin}(n, p)$
  - **Probability mass function** $p_X(x) = \binom{n}{x} p^x (1-p)^{n-x}, \; x = 0, 1, \cdots, n$
  - **Mean:** $\mathbb{E}[X] = np$; **Variance:** $Var(X) = np(1-p)$

Notes

Example

Suppose that 95% of consumers can recognize Coke in a blind taste test. Assume consumers are independent of one another. The company randomly selects 4 consumers for a taste test. Let X be the number of consumers who recognize Coke.

1. Write out the pmf table for X
2. What is the probability that X is at least 1?
3. What is the probability that X is at most 3?

Notes

Binomial and Hypergeometric Distributions

The binomial distribution describes the probability of k successes in n trials with replacement.

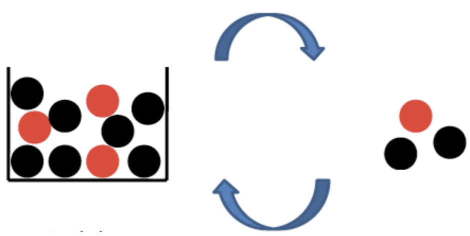
We want a distribution to describe the probability of k successes in n trials without replacement from a finite population of size N containing exactly K successes.

$\Rightarrow$ Hypergeometric Distribution

> Important applications are **quality control** and statistical **estimation of population proportions**. The hypergeometric r.v. is the equivalent of a Binomial r.v. except that sampling is done without replacement.

Notes

Drawing balls from an urn with 2 different colors



We have an urn with 3 red ("success") and 7 black ("failure") balls. Suppose we select 3 balls randomly.

- With replacement: $X \sim \text{Bin}(n = 3, p = 0.3)$
- Without replacement: $X \sim \text{Hypergeo}(n = 3, N = 10, K = 3)$

Notes

Hypergeometric Distributions

Let X be a hypergeometric r.v.

- The definition of X: The number of successes in n trials of a random experiment, where sampling is done without replacement
- The support: $x \in \{\max(0, n+K-N), \cdots, \min(n, K)\}$
- Its parameter(s) and definition(s): N: the population size, n: the sample size, and K: number of success in the population
- The probability mass function (pmf): $p_X(x) = \frac{\binom{K}{k}\times\binom{N-K}{n-k}}{\binom{N}{n}}$
- The expected value: $\mathbb{E}[X] = n\frac{K}{N}$
- The variance: $Var(X) = n\frac{K}{N}\frac{N-K}{N}\frac{N-n}{N-1}$

Notes

Example

There are 100 identical looking 52" TVs at Best Buy in Anderson, SC. Let 10 of them be defective. Suppose we want to buy 8 of the aforementioned TVs (at random). What is the probability that we don't get any defective TVs?

Solution.

Let D be the number of defective TVs in the sample.

$D \sim \text{Hypergeo}(N = 100, n = 8, K = 10)$
$\mathbb{P}(D = 0) = \frac{\binom{10}{0}\binom{90}{8}}{\binom{100}{8}} = 0.4166$

Notes

Poisson Random Variables

So far, we have seen discrete probability distributions of the number of successes in a sequence of random experiments with specified sample size.

- Bernoulli distribution: independent trial (sampling with replacement), sample size $= 1$
- Binomial distribution: independent trials (sampling with replacement), sample size $= n$
- Hypergeometric distribution: dependent trials (sampling without replacement), sample size $= n$

The Poisson distribution is a discrete probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time and/or space $\Rightarrow$ **does not have a (fixed) sample size**

Notes

Poisson Distributions

Let X be a Poisson r.v.

- The definition of X: The number of successes
- The support: $x = 0, 1, 2, \cdots$
- Its parameter(s) and definition(s): λ: the average number of successes
- The probability mass function (pmf): $p_X(x) = \frac{e^{-\lambda}\lambda^x}{x!}$
- The expected value: $\mathbb{E}[X] = \lambda$
- The variance: $Var(X) = \lambda$

Notes

Poisson approximation to Binomial Distribution

If $X \sim \text{Bin}(n, p)$ and $n > 100$ and $p < .01$ then we can approximate the distribution X by using $X^* \sim \text{Poi}(\lambda = n \times p)$

Example

Let $X \sim \text{Bin}(200, 0.005)$. Then,

$$\begin{aligned} \mathbb{E}[X] = 1; &\quad \mathbb{E}[X^*] = 1 \\ Var(X) = 0.995; &\quad Var(X^*) = 1 \\ \mathbb{P}(X = 1) = 0.3688; &\quad \mathbb{P}(X^* = 1) = 0.3679 \end{aligned}$$

Notes

Example

Suppose a certain disease has a 0.14 % of occurring. Let's sample 1,000 people. Find the exact and approximate probabilities that 0 people have the disease and at most 5 people have the disease.

Solution.

Set-up:
Let X be the number of people have the disease in the sample ($n = 1000$).
Which distribution to use?

- The sample size is fixed ($n = 1000$) $\Rightarrow$ Binomial or Hypergeometric

What are the parameters? $n = 1000, p = .0014$

Notes

Example Cont'd

Any approximation?

Since $n = 1000 > 100$ and $p = .0014 < .01$. We can use Poisson X^* to approximate the Binomial distribution X

Exact probabilities

$X \sim \text{Bin}(n = 1000, p = .0014)$

$\mathbb{P}(X = 0) = \binom{1000}{0}(.0014)^0(1 - .0014)^{1000-0} = .2464$

$\mathbb{P}(X \leq 5) = \sum_{x=0}^{5} \binom{1000}{x}(.0014)^x(1 - .0014)^{1000-x} = .9986$

Approximate probabilities

$X^* \sim \text{Poi}(\lambda = n \times p = 1.4)$

$\mathbb{P}(X = 0) \approx \mathbb{P}(X^* = 0) = \frac{e^{-1.4}1.4^0}{0!} = .2466$

$\mathbb{P}(X \leq 5) \approx \mathbb{P}(X^* \leq 5) = \sum_{x=0}^{5} \frac{e^{-1.4}\lambda^x}{x!} = .9986$

Notes

Notes

Notes