MATH 3339
Statistics for the Sciences
Live Lecture Help

James West
jdwest@uh.edu

University of Houston

Session 6

Office Hours: see schedule in the "Office Hours" channel on Teams
Course webpage: www.casa.uh.edu

Email policy

When you email me you **MUST** include the following

- MATH 3339 Section 20024 and a description of your issue in the **Subject Line**
- Your name and ID# in the **Body**
- Complete sentences, punctuation, and paragraph breaks
- Email messages to the class will be sent to your Exchange account (user@cougarnet.uh.edu)

Using R and R-Studio

1. Download R from https://cran.r-project.org/
2. Download R-Studio from https://www.rstudio.com/

Outline

1. Updates and Announcements
2. Recap
3. Examples
4. Student submitted questions

Updates and Announcements

- Check your reservation for Test !
- Use Practice Test 1 and Test 1 Review Sheet for practice!

PMF, Mean, and Variance for Bernoulli Random Variable

If X has the Bernoulli distribution with probability of success p, the pmf for X is:

$$f_X(x) = P(X = x) = \begin{cases} p^x(1-p)^{1-x} & \text{if } x = 0, 1 \\ 0 & \text{if } x \neq 0, 1 \end{cases}$$

The **mean** and **variance** of X are:

$$\mu_X \quad = \quad E[X] = p$$

$$\sigma_X^2 \quad = \quad Var[X] = p(1-p)$$

PMF, Mean, and Variance for Binomial Random Variable

If $X \sim Binom(n, p)$, the pmf is:

$$f_X(x) = P(X = x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{if } x = 0, 1, 2, ..., n \\ 0 & \text{if } x \neq 0, 1, 2, ..., n \end{cases}$$

The **mean** and **variance** of X are:

$$\mu_X = E[X] = np$$

$$\sigma_X^2 = Var[X] = np(1-p)$$

CDF of a Binomial R.V.

If $X \sim Binom(n,p)$

$$F_X(x) = \begin{cases} 0, & \text{if } x < 0 \\ \sum_{k=0}^{x^*} \binom{n}{k} p^k (1-p)^{n-k}, & \text{if } 0 \leq x \leq n \\ 1, & \text{if } n \leq x \end{cases}$$

where $x^* = \lfloor x \rfloor$, the first integer less than or equal to x.

Cumulative Distribution Function

Recall that a quantitative random variable X has a **cumulative distribution function** given by

$$F_X(x) = P(X \leq x)$$

for all $x \in \mathbb{R}$.

When we have a discrete random variable X, the cdf is related to the pmf in the following way:

$$F_X(x) = \sum_{x_i \leq x} f_X(x_i)$$

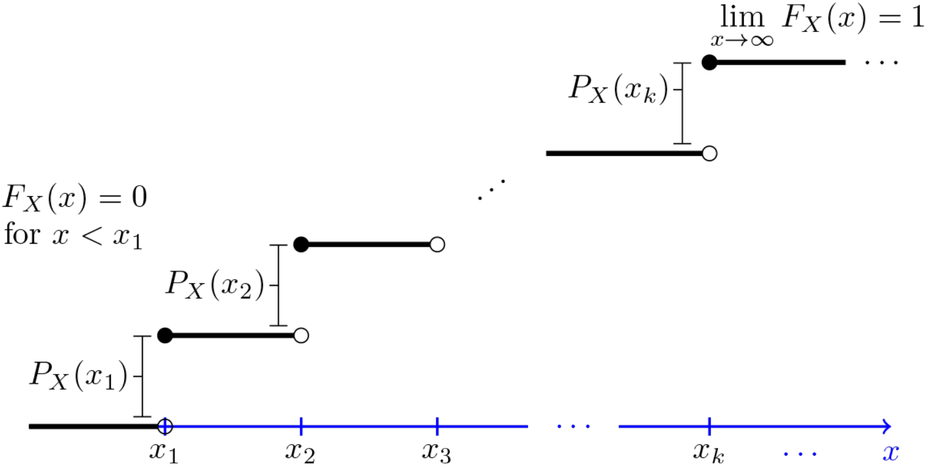
where $x_1, x_2, \ldots$ are the values of X.

Cumulative Distribution Function Properties

Any cdf F has the following properties:

1. F is a non-decreasing function defined on $\mathbb{R}$
2. F is right-continuous, meaning for each a, $F(a) = F(a+) = \lim_{x \to a^+} F(x)$
3. $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$
4. $P(a < X \leq b) = F(b) - F(a)$ for all real a and b, where $a < b$.
5. $P(X > a) = 1 - F(a)$.
6. $P(X < b) = F(b-) = \lim_{x \to b^-} F(x)$.
7. $P(a < X < b) = F(b-) - F(a)$.
8. $P(X = b) = F(b) - F(b-)$.

Graph of a CDF



Hypergeometric Distribution

The **parameters** of a hypergeometric distribution are m, n, k. We write $X \sim Hyper(m, n, k)$. The probability mass function for a hypergeometric is:

$$f_X(x) = P(X = x) = \frac{\binom{m}{x}\binom{n}{k-x}}{\binom{m+n}{k}}$$

Mean and Variance of a Hypergeometric Distribution

Let Y have a hypergeometric distribution with parameter, m, n, and k.

$p = \frac{m}{m+n} = \frac{m}{N}$

$1-p = \frac{n}{m+n} = \frac{n}{N}$

- The mean of Y is:

$$\mu_Y = E(Y) = k\left(\frac{m}{m+n}\right) = kp.$$

- The variance of Y is:

$$\sigma_Y^2 = var(Y) = kp(1-p)\left(1 - \frac{k-1}{m+n-1}\right).$$

- $1 - \frac{k-1}{m+n-1}$ is called the **finite population correction factor**. As, the population increases, this factor will get closer to 1.

The Probability Function of the Poisson Distribution

A random variable X with nonnegative integer values has a Poisson distribution if its frequency function is:

$$f_X(x) = P(X = x) = e^{-\mu}\frac{\mu^x}{x!}$$

for $x = 0, 1, 2, \ldots$, where $\mu > 0$ is a constant. If X has a Poisson distribution with parameter μ, we can write $X \sim Pois(\mu)$.

The Mean and Variance of the Poisson Distribution

Let $X \sim Pois(\mu)$

- The mean of X is μ per unit of measure. By the conditions of the Poisson distribution.
- The variance of X is also μ per unit of measure.
- The standard deviation of X is $\sqrt{\mu}$.

Hypergeometric Distribution

Example: A fish tank in a pet store has 25 fish in it. 9 are orange and 16 are white. Determine the probability that if we select 6 fish from the tank, at least 3 will be white.

$$f_X(x) = P(X = x) = \frac{\binom{m}{x}\binom{n}{k-x}}{\binom{m+n}{k}}$$

here $m = 16,\ n = 9,\ k = 6,\ N = 16+9 = 25$

$X \sim \text{Hyper}(16, 9, 6)$

$$P(X \geq 3) = \sum_{x=3}^{6} P(X = x) = \sum_{x=3}^{6} f(x) = \sum_{x=3}^{6} \frac{\binom{16}{x}\binom{9}{6-x}}{\binom{25}{6}}$$

or $P(X \geq 3) = 1 - P(X < 3) = 1 - P(X \leq 2)$

$= 1 - \text{phyper}(2, 16, 9, 6)$

```
> 1-phyper(2,16,9,6)
[1] 0.9027668
```

Hypergeometric Distribution

Example: A fish tank in a pet store has 25 fish in it. 9 are orange and 16 are white. Determine the probability that if we select 6 fish from the tank, we will find 4 orange fish.

Hypergeometric Distribution

Example: A fish tank in a pet store has 25 fish in it. 9 are orange and 16 are white. Determine the probability that if we select 6 fish from the tank, we will find fewer than 4 orange fish.

Example: Let X be the number of flaws on the surface of a randomly selected boiler of a certain type that has a Poisson distribution with parameter $\mu = 4$. Find $P(2 \leq X < 6)$

$$X \sim Pois(4)$$

$$P(2 \leq X < 6) = P(2 \leq X \leq 5) = \sum_{x=2}^{5} f(x)$$

$$= \sum_{x=2}^{5} e^{-4} \cdot \frac{4^x}{x!} = e^{-4}\left[\frac{4^2}{2!} + \frac{4^3}{3!} + \frac{4^4}{4!} + \frac{4^5}{5!}\right]$$

or $P(2 \leq X \leq 5) = P(X \leq 5) - P(X \leq 1)$

```
> ppois(5,4)-ppois(1,4)
[1] 0.6935522
```

Poisson Distribution

Example: Calls to a toll-free telephone hotline service are made randomly and independently at an expected rate of two per minute. The hotline service has five customer service representatives, none of whom is currently busy. Using a Poisson distribution, determine the probability that the hotline receives fewer than five calls in the next minute.

Poisson Distribution

Example: Calls to a toll-free telephone hotline service are made randomly and independently at an expected rate of two per minute. The hotline service has five customer service representatives, none of whom is currently busy. Using a Poisson distribution, determine the probability that the hotline receives fewer than five calls in the next minute.

Example

Binomial w/ $n=9$, $p=0.75$

Suppose 75% of the customers at carrier A opt for an unlimited plan. There are 9 customers currently in the store. Find the probabilities below:

1. Exactly 5 will get an unlimited plan?

$X \sim \text{Binom}(9, 0.75)$

$$P(X=5) = \binom{9}{5}(0.75)^5(0.25)^4$$

```
> dbinom(5,9,.75)
[1] 0.1167984
> choose(9,5)*.75^5*.25^4
[1] 0.1167984
```

2. At least two will get an unlimited plan?

$$P(X \geq 2) = 1 - P(X<2) = 1 - P(X \leq 1)$$
$$= 1 - F(1) = 1 - \text{pbinom}(1, 9, 0.75)$$

```
> 1-pbinom(1,9,0.75)
[1] 0.9998932
```

10

A 3

7 3

① A w/ $p = \frac{7}{10}$ and get A

② A w/ $p_1 = \frac{6}{9}$

The decline of salmon fisheries along the Columbia River in Oregon has caused great concern among commercial and recreational fishermen. The paper 'Feeding of Predaceous Fishes on Out-Migrating Juvenile Salmonids in John Day Reservoir, Columbia River' (Trans. Amer. Fisheries Soc. (1991: 405-420) gave the accompanying data on y = maximum size of salmonids consumed by a northern squaw fish (the most abundant salmonid predator) and x = squawfish length, both in mm. Here is the computer software printout of the summary:

Coefficients:				
	Estimate	Std. Error	t value	Pr(> |t|)
(Intercept)	−90.020	16.702	−5.390	0.000
Length	0.705	0.049	14.358	0.000

Using this information, give the equation of the least squares regression line.

A company makes sports bikes. 90% pass final inspection (and 10% fail and need to be fixed). What is the probability that the 10^{th} bike will not pass final inspection?

Using R and R-Studio

1. Download R from https://cran.r-project.org/
2. Download R-Studio from https://www.rstudio.com/