MATH 3339
Statistics for the Sciences
Sec 4.4-4.7

Wendy Wang, Ph.D.
wwang60@central.uh.edu

Lecture 7 - 3339

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Outline

1. Bernoulli Distribution
2. Cumulative Distributions
3. Hypergeometric Distribution

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Special cases for discrete distribution

Special cases for Discrete distribution:

- Bernoulli distribution
- Binomial distribution
- Hypergeometric distribution
- Poisson distribution
- Jointly distributed Variables

The Bernoulli Probability Distribution

Definition: A **Bernoulli** random varilable is a random experiment (a Bernoulli Trial) with the following characteristics:

1. The outcome can be classified as either **success** or **failure** (where these are mutually exclusive and exhaustive).

2. The probability of **success** is **p**, so the probability of **failure** is **q=1-p** . e.g. a coin is flipped (heads or tails), someone is pulled over for speeding (ticket or warning), etc.

Suppose that a coin is flipped. Let X be the random variable that indicates that heads was flipped. Here heads represents “success" and tails represents “failure" so that X is a Bernoulli random variable.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

$E(X) = \sum x P(X=x)$

$= 0 * (1-p)$
$+ 1 * p$
$= p$

(1)

X	0 (tail)	1 (Head)
$P(x)$	$1-p$	p

(2)

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

A compact way of writing this is:

(3)

$$f(x) = P(X = x) = p^x(1-p)^{1-x}$$

$f(0) = P(X=0) = p^0(1-p)^{1-0} = 1-p$

$f(1) = P(X=1) = p^1(1-p)^{(1-1)} = p$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Mean and Variance of a Bernoulli Distribution

If X has the Bernoulli distribution with probability of success p, the **mean** and **variance** of X are

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

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

Then the standard deviation is the square root of the variance.

$$Var(X) = E(X^2) - [E(X)]^2$$
$$= 1^2 * p + \cancel{0^2(1-p)} - p^2$$
$$= p - p^2 = p(1-p)$$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

The Binomial Probability Distribution

Suppose that we flip a coin 10 times. This is a sequence of Bernoulli trials. We are interested in calculating the probability of obtaining a certain number of heads. Let X_i indicate heads on the i-th flip.

$X_i = \begin{cases} 1 & (\text{head}) \\ 0 & (\text{tail}) \end{cases}$

Define $Y = X_1 + X_2 + ... + X_{10}$. What does Y represent?

total # of "head" in this 10 trials.

What is the probability that $Y = 0$? Y can be 0, 1, 2, ..., 10

$0.5 * 0.5 \cdots * 0.5$

What is the probability that $Y = 1$? $P(Y=1) = C_1^{10}\, 0.5^1 * (1-0.5)^{(10-1)}$

HTTTT....T ⟶ $0.5 * (1-0.5) * (1-0.5) * \cdots$

THTTT....T ⟶ $(1-0.5) * 0.5 * (1-0.5) \cdots$

TTHTT....T ⟶ $(1-0.5) * (1-0.5) * 0.5 * (1-0.5) \cdots$

TTTTTTTTTH

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

The Binomial Probability Distribution

Suppose that we flip a coin 10 times. This is a sequence of Bernoulli trials. We are interested in calculating the probability of obtaining a certain number of heads. Let X_i indicate heads on the i-th flip.

Define $Y = X_1 + X_2 + ... + X_{10}$. What does Y represent?

What is the probability that $Y = 2$?

What is the probability that $Y = n$?

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

The Binomial Probability Distribution

X	0	$\cdots$	n
$P(X=x)$			

Here Y is the sum of 10 independent Bernoulli trials. We call this type of random variable a **Binomial random variable.**

A random variable X is a Binomial random variable if the following conditions are satisfied:

1. X represents the number of successes on n Bernoulli trials.
2. The probability of success for each trial is p.
3. The trials are mutually independent.

If X is a binomial random variable with probability p of success on each of n trials, we write $X \sim Binomial(n, p)$

X can be 0, 1, ..., n

If $X \sim Binomial(n, p)$, then $P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}$ where $x = 0, 1, 2, ..., n$

$P(X=x) = C_x^n \, p^x (1-p)^{(n-x)}$

Mean and Variance of a Binomial Distribution

If a count X has the Binomial distribution with number of observations n and probability of success p, the **mean** and **variance** of X are

$\mu_x = \sum x P(X = x)$

$$\begin{aligned} \mu_X &= E[X] = np \\ \sigma_X^2 &= Var[X] = np(1-p) \end{aligned}$$

Then the standard deviation is the square root of the variance.

The Binomial Probability Distribution

R commands:

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

$$P(X = x) = \text{dbinom}(x, n, p)$$

$$P(X \le x) = \text{pbinom}(x, n, p)$$

the largest possible value for $X \le x$

$$P(X > x) = 1 - P(X \le x)$$
$$= 1 - \text{pbinom}(x, n, p)$$

The Binomial Probability Distribution

$n = 8$

Example: Suppose that at a 4-way stop in a certain subdivision, only 12% of drivers come to a complete stop. What is the probability that among 8 drivers, at least 6 of them will run the stop sign?

"Success" = "run the stop sign"

$P(\text{success}) = 1 - 0.12 = 0.88$ / X = # of drivers run of stop sign

$P(\text{at least } 6) = P(X \geq 6) = 1 - P(X < 6)$

$= 1 - P(X = 0, 1, 2, 3, 4, 5)$

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

$= 1 - \text{pbinom}(5, 8, 0.88)$

What is the expected number of drivers who will run the stop sign?

$\mu = n \cdot p$

$= 8 * 0.88$

$= \boxed{7.04}$

On ave, we expect 7.04 drivers run the stop sign among 8.

```
> 1-pbinom(5,8,0.88)
[1] 0.9392108
```

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

The Binomial Probability Distribution

$n=12$

Suppose $X \sim Binomial(12, 0.3)$, find the following:

$p=0.3$

$P(2 \leq X < 5) =$ P(X can be 2, 3, 4) $= P(X=2) + P(X=3) + P(X=4)$

$P(X=3) = \binom{12}{3} 0.3^3 (1-0.3)^{(12-3)}$

$=$ dbinom(3, 12, 0.3)

~~$P(X$ [illegible] $5) =$~~

$P(2 \le X < 5)$ $\quad X \sim \text{Binom}(12, 0.3)$

$= P(X \text{ can be } 2, 3, 4)$

$= P(X = 2) + P(X = 3) + P(X = 4)$

$= \text{dbinom}(2, 12, 0.3) + \text{dbinom}(3, 12, 0.3) + \text{dbinom}(4, 12, 0.3)$

$P(2 \le X < 5) = P(X < 5) - P(X < 2)$

$= \text{pbinom}(4, 12, 0.3) - \text{pbinom}(1, 12, 0.3)$

$$\underline{P(X>5)} = 1 - P(X \le 5)$$

$$\doteq 1 - \text{pbinom}(5, 12, 0.3)$$

Example

$n = 5$ $p = 0.80$

Suppose it is known that 80% of the people exposed to the flu virus will contract the flu. Out of a family of five exposed to the virus, what is the probability that:

X = # of ppl in this family contract the flu

$X: 0, 1, 2, 3, 4, 5$

1. No one will contract the flu?

$P(X = 0) = \text{dbinom}(0, 5, 0.8)$

2. All will contract the flu?

$P(X=5) = \text{dbinom}(5, 5, 0.8)$.

3. Exactly two will get the flu?

$P(X=2) = \text{dbinom}(2, 5, 0.8)$

4. At least two will get the flu?

$P(X \geq 2) = P(X \text{ can be } 2, 3, 4, 5)$
$= 1 - P(X \text{ can be } 0, 1) = 1 - \text{pbinom}(1, 5, 0.8)$

5. at most 3 will get flu?

$P(X \text{ at most } 3) = P(X \le 3)$

$= P(X \text{ can be } 0, 1, 2, \underline{\underline{3}})$

$= \text{pbinom}(3, 5, 0.8)$

6. fewer than 4 will get flu?

$P(X < 4) = P(X \text{ can be } 0, 1, 2, \underline{\underline{3}})$

$= \text{pbinom}(3, 5, 0.8)$

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.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

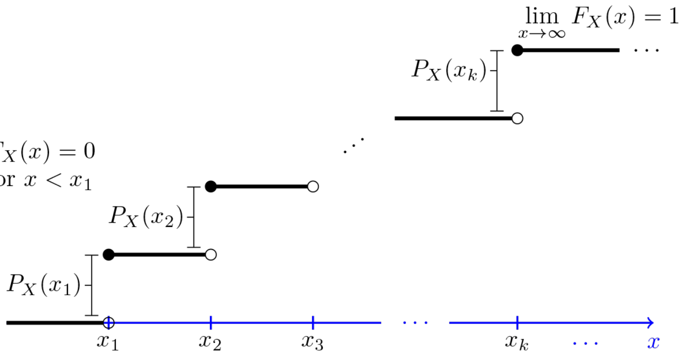
Cumulative Distribution Function Properties

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. $P(X = b) = F(b) - F(b-)$.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Graph of CDF



CDF of a Binomial R.V.

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

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

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

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Hypergeometric Distribution

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Each of 12 refrigerators of a certain type has been returned to a distributor because of an audible, high-pitched, oscillating noise when the refrigerator is running. Suppose that 7 of these refrigerators have a defective compressor and the other 5 have less serious problems. The technition looks at 6 refrigerators, what is the probability that exactly 5 have a defective compressor?

$m = 7$ def. $\leftarrow$ 5

$n = 5$ good $\leftarrow$ 1

$N = 12$ 6

X = # of def. compressor. among the 6.

$X \sim$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Conditions for a Hypergeometric Distribution

1. The population or set to be sampled consists of N individuals, objects or elements (a *finite* population).
2. Each individual can be characterized as a "success" or "failure." There are m successes in the population, and n failures in the population. Notice: $m + n = N$. $m=7$, $n=5$
3. A sample size of k individuals is selected without replacement in such a way that each subset of size k is equally likely to be chosen. $k=6$ $P(X=5)=?$

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

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

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Conditions for a Hypergeometric Distribution

binomial : $Y \sim binom(n, p)$

1. The population or set to be sampled consists of N individuals, objects or elements (a *finite* population).
2. Each individual can be characterized as a "success" or "failure." There are m successes in the population, and n failures in the population. Notice: $m + n = N$.
3. A sample size of k individuals is selected without replacement in such a way that each subset of size k is equally likely to be chosen.

$m+n=N$

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

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

dbinom

- R commands: $P(X = x) =$ dhyper(x,m,n,k) and $P(X \leq x) =$ phyper(x,m,n,k)

pbinom

- Going back to the refrigerator example, $m = 7$, $n = 5$, $k = 6$. $P(X = 5)$

```
> dhyper(5,7,5,6)
[1] 0.1136364
```

- $P(X \leq 4)$

P(X can be 0, 1, 2, 3, 4)

```
> phyper(4,7,5,6)
[1] 0.8787879
```

$P(X<4) = P(X \text{ can be } 0,1,2,3)$

$= \text{phyper}(3,7,5,6)$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Mean and Variance of a Hypergeometric Distribution

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

- 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.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Digital Cameras

3-mega 6
4-mega 9
15

A certain type of digital camera comes in either a 3-megapixel version or a 4-megapixel version. A camera store has received a shipment of 15 of these cameras, of which 6 have 3-megapixel resolution. Suppose that 5 of these cameras are randomly selected to be stored behind the counter; the other 10 are placed in a storeroom. Let X = the number of 3-megapixel cameras among the 5 selected for behind the counter storage.

$X \sim hyper(6, 9, 5)$

1. What is the probability that exactly 2 of the 3-megapixel cameras are stored behind the counter?

$P(X=2) = dhyper(2, 6, 9, 5)$

```
> dhyper(2,6,9,5)
[1] 0.4195804
```

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Digital Cameras

A certain type of digital camera comes in either a 3-megapixel version or a 4-megapixel version. A camera store has received a shipment of 15 of these cameras, of which 6 have 3-megapixel resolution. Suppose that 5 of these cameras are randomly selected to be stored behind the counter; the other 10 are placed in a storeroom. Let X = the number of 3-megapixel cameras among the 5 selected for behind the counter storage.

2. What is the probability that at least one of the 3-megapixel cameras are stored behind the counter?

$$P(X \text{ at least } 1) = P(X \geq 1)$$
$$= 1 - P(X = 0)$$
$$= 1 - \text{dhyper}(0, 6, 9, 5)$$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Digital Cameras

A certain type of digital camera comes in either a 3-megapixel version or a 4-megapixel version. A camera store has received a shipment of 15 of these cameras, of which 6 have 3-megapixel resolution. Suppose that 5 of these cameras are randomly selected to be stored behind the counter; the other 10 are placed in a storeroom. Let X = the number of 3-megapixel cameras among the 5 selected for behind the counter storage.

3. Calculate the mean and standard deviation of X.

$m = 6 \quad n = 9 \quad k = 5$