MATH 3339 - 03 15951
Statistics for the Sciences
Discrete Random Variables

Wendy Wang
wwang60@central.uh.edu

Lecture 15 - 3339
Outline

1. The Binomial Probability Distribution
2. Cumulative Distributions
3. Hypergeometric Distribution
Popper Set Up

- Fill in all of the proper bubbles.
- Make sure your ID number is correct.
- Make sure the filled in circles are very dark.
- This is popper number 10.
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 $Y = X_1 + X_2 + \ldots + X_{10}$. What does $Y$ represent?

What is the probability that $Y = 2$?

What is the probability that $Y = n$?
The Binomial Probability Distribution

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

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

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

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

Then the standard deviation is the square root of the variance.
The Binomial Probability Distribution

R commands:
The Binomial Probability Distribution

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?

What is the expected number of drivers who will run the stop sign?
The Binomial Probability Distribution

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

$P(2 \leq X < 5) =$

$P(X > 5) =$
Example

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:

1. No one will contract the flu?

2. All will contract the flu?

3. Exactly two will get the flu?

4. At least two will get the flu?
Example

Suppose it is known that 80% of the people exposed to the flu virus will contract the flu. Suppose a family of 5 has been exposed to the flu.

1. Find the mean number of sick members of this family.

2. Find the variance.
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 $X$. 
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 CDF

\[ F_X(x) = 0 \quad \text{for } x < x_1 \]

\[ F_X(x) \quad \text{for } x_1 \leq x \leq x_2 \]

\[ F_X(x_k) \quad \text{for } x_k \leq x \]

\[ \lim_{x \to \infty} F_X(x) = 1 \]
CDF of a Binomial R.V.

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

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

where $x^* = x$, the first integer less than or equal to $x$. 
Hypergeometric Distribution
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 technician looks at 6 refrigerators, what is the probability that exactly 5 have a defective compressor?
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: $m + n = N$.

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.

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

$$f_X(x) = P(X = x) = \binom{m}{x} \binom{n}{k-x} / \binom{m+n}{k}$$
Conditions for a Hypergeometric Distribution

1. The population or set to be sampled consists of \( N \) individuals, objects or elements (a \textit{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.

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

\[
f_X(x) = P(X = x) = \frac{\binom{m}{x} \binom{n}{k-x}}{\binom{m+n}{k}}
\]
Using R

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

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

  - \( P(X = 5) \)
    
    ```r
    > dhyper(5, 7, 5, 6)
    [1] 0.1136364
    ```

  - \( P(X \leq 4) \)
    
    ```r
    > phyper(4, 7, 5, 6)
    [1] 0.8787879
    ```
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:
  
  $$ \sigma^2_Y = var(Y) = kp(1 - p) \left( 1 - \frac{k - 1}{m + n - 1} \right). $$

- $1 - \frac{k - 1}{m + n - 1}$ is called the **finite population correction factor**. As the population increases, this factor will get closer to 1.
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 $X$ be the number of 3-megapixel cameras among the 5 selected for behind the counter storage.

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

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

3. Calculate the mean and standard deviation of $X$. 