STAT 3375Q: Introduction to Mathematical Statistics I

Lecture 7: Special Discrete Distributions: Negative Binomial, Hypergeometric, Poisson

Mary Lai Salvaña, Ph.D.

Department of Statistics
University of Connecticut

February 7, 2024

Outline

1. Quiz 2 Review Exercises Solutions
2. Previously...
   - Bernoulli Distribution
   - Binomial Distribution
   - Geometric Distribution
3. Special Discrete Distributions (cont'd)
   - Negative Binomial Distribution
   - Hypergeometric Distribution
   - Poisson Distribution

Quiz 2 Review Exercises Solutions

Problem 1

Prove that for a random variable X with expected value $E(X) = \lambda$,

$$V(X) = E\{X(X-1)\} + \lambda - \lambda^2.$$

Solution:

Approach 1.

$$\begin{aligned}
V(X) &= E\{(X-\lambda)^2\} \\
&= E(X^2 - 2\lambda X + \lambda^2) \\
&= E(X^2) - E(2\lambda X) + E(\lambda^2) \\
&= E(X^2 - X + X) - 2\lambda E(X) + \lambda^2 \\
&= E(X^2 - X) + E(X) - 2\lambda^2 + \lambda^2 \\
&= E\{X(X-1)\} + \lambda - \lambda^2.
\end{aligned}$$

□

Problem 1

Prove that for a random variable X with expected value $E(X) = \lambda$,

$$V(X) = E\{X(X-1)\} + \lambda - \lambda^2.$$

Solution:

Approach 2.

$$\begin{aligned} V(X) &= E(X^2) - \lambda^2 \\ &= E(X^2) - \lambda + \lambda - \lambda^2 \\ &= E(X^2) - E(X) + \lambda - \lambda^2 \\ &= E(X^2 - X) + \lambda - \lambda^2 \\ &= E\{X(X-1)\} + \lambda - \lambda^2. \end{aligned}$$

□

Problem 2

A manufacturer is sending 10 boxes out for shipment today. Unfortunately, some of the boxes have defective items.

Box #	1	2	3	4	5	6	7	8	9	10
# of defective items	0	0	1	0	2	2	0	0	1	3

a. One of these boxes is to be selected at random for shipment to a particular customer. Let X be the number of defective items in the selected box. What is the probability distribution of X?
b. What is the expected value of defective items?
c. Another manufacturer is known to have X^2 defective items in each of the boxes numbered 1 to 10. If this manufacturer sends out a randomly selected box, what is the expected number of defective items the customer will receive?

Solution:

a.

x	$p(x)$
0	$5/10$
1	$2/10$
2	$2/10$
3	$1/10$

(4 pts)

Problem 2

A manufacturer is sending 10 boxes out for shipment today. Unfortunately, some of the boxes have defective items.

Box #	1	2	3	4	5	6	7	8	9	10
# of defective items	0	0	1	0	2	2	0	0	1	3

a. One of these boxes is to be selected at random for shipment to a particular customer. Let X be the number of defective items in the selected box. What is the probability distribution of X?
b. What is the expected value of defective items?
c. Another manufacturer is known to have X^2 defective items in each of the boxes numbered 1 to 10. If this manufacturer sends out a randomly selected box, what is the expected number of defective items the customer will receive?

Solution:

b. $E(X) = 0\left(\frac{5}{10}\right) + 1\left(\frac{2}{10}\right) + 2\left(\frac{2}{10}\right) + 3\left(\frac{1}{10}\right) = 0.9.$ (3 pts)

Problem 2

A manufacturer is sending 10 boxes out for shipment today. Unfortunately, some of the boxes have defective items.

Box #	1	2	3	4	5	6	7	8	9	10
# of defective items	0	0	1	0	2	2	0	0	1	3

a. One of these boxes is to be selected at random for shipment to a particular customer. Let X be the number of defective items in the selected box. What is the probability distribution of X?
b. What is the expected value of defective items?
c. Another manufacturer is known to have X^2 defective items in each of the boxes numbered 1 to 10. If this manufacturer sends out a randomly selected box, what is the expected number of defective items the customer will receive?

Solution:

c. $E(X^2) = 0^2\left(\frac{5}{10}\right) + 1^2\left(\frac{2}{10}\right) + 2^2\left(\frac{2}{10}\right) + 3^2\left(\frac{1}{10}\right) = 1.9.$ (3 pts)

Problem 3

Suppose the random variable X takes on possible values $x = 0, 1, 2$ and has a probability mass function $f(x) = \frac{2x+3}{k}$, determine the value of k.

Solution:

x	$p(x)$
0	$3/k$
1	$5/k$
2	$7/k$

$\frac{3}{k} + \frac{5}{k} + \frac{7}{k} = 1 \Rightarrow k = 15.$ □

Problem 4

A box contains 5 red and 5 blue marbles. Two marbles are drawn randomly. If they are the same color, then you win \$1.10. If they are different colors, you lose \$1.00. Compute

- (a) the expected value of the amount you win
- (b) the variance of the amount you win

Solution:

Let X be a random variable for the amount you win. Let R be the event that a red marble is drawn and B be the event that a blue marble is drawn.

Event	x	$p(x)$
BB	1.1	$\left(\frac{5}{10}\right)\left(\frac{4}{9}\right) = \frac{2}{9}$ or $C_2^5/C_2^{10} = \frac{2}{9}$
RB $\cup$ BR	-1	$\left(\frac{5}{10}\right)\left(\frac{5}{9}\right) + \left(\frac{5}{10}\right)\left(\frac{5}{9}\right) = \frac{5}{9}$
RR	1.1	$\left(\frac{5}{10}\right)\left(\frac{4}{9}\right) = \frac{2}{9}$ or $C_2^5/C_2^{10} = \frac{2}{9}$

- (a) $E(X) = 1.1(2/9) + 1.1(2/9) - 1(5/9) = -0.6/9 = -0.067.$
- (b) $E(X^2) = 1.1^2(2/9) + 1.1^2(2/9) + (-1)^2(5/9) = 9.84/9 = 1.093$
  $V(X) = E(X^2) - \{E(X)\}^2 = 1.093 - (-0.067)^2 = 1.089.$

Problem 5

If $E(X) = 1$ and $V(X) = 5$, find

a. $E\{(2+X)^2\}$
b. $V(4+3X)$

Solution:

a. $E\{(2+X)^2\} = E(4+4X+X^2) = 4+4E(X)+E(X^2) = 4+4+E(X^2) = 8+E(X^2)$.
To solve for $E(X^2)$, we know
$V(X) = E(X^2) - \{E(X)\}^2 \Rightarrow 5 = E(X^2) - 1^2 \Rightarrow E(X^2) = 6$.
Thus, $E\{(2+X)^2\} = 8+6 = 14$.
b. $V(4+3X) = 3^2 V(X) = 9(5) = 45$.

□

Previously...

Bernoulli Distribution

- Notation: $Y \sim \text{Bern}(p)$ or $Y \sim \text{Be}(p)$
- Usage: single trial, two outcomes, success or fail?
- Parameter: p (probability of success)
- PMF: $p(y) = p^y(1-p)^{1-y}$, $y = 0, 1$ and $0 \leq p \leq 1$.
- Mean or Expected Value: p
- Variance: $p(1-p)$

Binomial Distribution

- Notation: $Y \sim \text{B}(n, p)$ or $Y \sim \text{Bin}(n, p)$
- Usage: multiple trials, two outcomes, counting num of successes in n trials
- Parameters: p (probability of success) and n (num. of trials)
- PMF: $p(y) = \binom{n}{y} p^y (1-p)^{n-y}$, $y = 0, 1, \ldots, n$ and $0 \leq p \leq 1$.
- Mean or Expected Value: np
- Variance: $np(1-p)$

Geometric Distribution

- Notation: $Y \sim G(p)$ or $Y \sim \text{Geo}(p)$ or $Y \sim \text{Geom}(p)$
- Parameter: p (probability of success)
- There are actually 2 types of Geometric distribution:

	Type 1 (textbook) (also called Shifted Geometric Distribution)	**Type 2**
Usage	counting the num. of trials needed to get first success	counting the num. of failures before first success
Values Y can take	$y = 1, 2, 3, \ldots$	$y = 0, 1, 2, \ldots$
PMF	$p(y) = (1-p)^{y-1} p$ y is the num of trials INCLUDING the trial that is a success	$p(y) = (1-p)^{y} p$ y is the num of failures NOT including the trial that is a success
Mean	$\frac{1}{p}$ (how many trials to expect until we get the first success)	$\frac{1-p}{p}$ (how many failures to expect before we get the first success)
Variance	$\frac{1-p}{p^2}$	$\frac{1-p}{p^2}$

Which type to use depends on how you want to solve the problem...

Geometric Distribution

Example 1:

A representative from the National Football League's Marketing Division randomly selects people on a random street in Kansas City, Kansas until he finds a person who attended the last home football game. Let p, the probability that he succeeds in finding such a person, equal 0.20. What is the probability that the marketing representative must select 4 people before he finds one who attended the last home football game? How many people should we expect the marketing representative need to select until he finds one who attended the last home football game?

Solution:

	Type 1	**Type 2**
Random Variable	Let Y denote the number of people he selects until he finds his first success.	Let Y denote the number of people he selects before he finds his first success.
PMF	$p(y) = (1-p)^{y-1}p$	$p(y) = (1-p)^{y}p$
Probability of interest	$P(Y=4) = p(4) = (1-0.2)^{4-1}(0.2) = 0.1024$	$P(Y=3) = p(3) = (1-0.2)^{3}(0.2) = 0.1024$
Mean	$E(Y) = \frac{1}{p} = \frac{1}{0.2} = 5$ He should expect to have to select 5 people until he finds one who attended the last football game.	$E(Y) = \frac{1-p}{p} = \frac{1-0.2}{0.2} = 4$ He should expect to fail 4 times before he finds one who attended the last football game.
Variance	$V(Y) = \frac{1-p}{p^2} = \frac{1-0.2}{0.2^2} = 20$	$V(Y) = \frac{1-p}{p^2} = \frac{1-0.2}{0.2^2} = 20$

Geometric Distribution

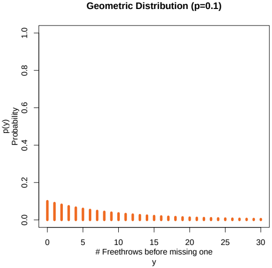
Example 2:

Professional basketball player Steve Nash was a 90% free throw shooter over his career. If Steve Nash starts shooting free throws, how many would he expect to make before missing one? What is the probability that he could make 20 in a row?

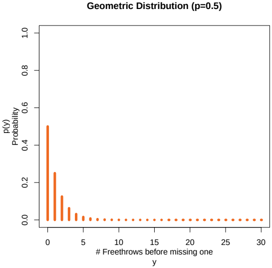
Solution:

- Type 2 Geometric Distribution
- Let Y be the random variable for the number of free throws Steve Nash makes before missing one.
- $p = 0.1$ since the success here is the missed free throw.
- $E(Y) = \frac{1-p}{p} = \frac{1-0.1}{0.1} = \frac{0.9}{0.1} = 9$.
  We expect Steve Nash to make 9 free throws (failures) before missing one.
- To make 20 in a row corresponds to $Y \geq 20$ (num of failures is 20 or more).
  $P(Y \geq 20) = 1 - P(Y \leq 19) = 1 - \sum_{y=0}^{19}(1-p)^y p = 1 - 0.878 = 0.122$.
  This means that Steve Nash could run off 20 (or more) free throws in a row about 12% of the times he wants to try.

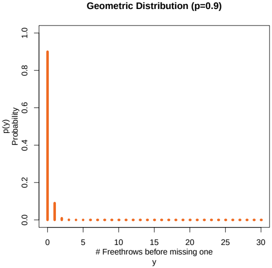
Geometric Distribution



90% free throw shooter



50% free throw shooter



10% free throw shooter

Special Discrete Distributions (cont'd)

Negative Binomial Distribution

Definition 3.9: Negative Binomial Distribution

A random variable Y is said to have a *negative binomial probability distribution* if and only if

$$p(y) = \binom{y-1}{r-1} p^r (1-p)^{y-r},$$

where $y = r, r+1, r+2, \ldots$ and $0 \le p \le 1$.

- the probability distribution of a random variable counting the number of trials needed to get the rth success
- We call $p(y) = \binom{y-1}{r-1} p^r (1-p)^{y-r}$ as the pmf of a neg. binom. random variable.
- Notation: $Y \sim \text{NB}(r, p)$, read as: "Y is a neg. binom. random variable with r successes and probability of success p."
- The Geometric distribution is a special case of the neg. binom. distribution with $r = 1$.

Negative Binomial Distribution

Theorem 3.9

If Y is a random variable with a negative binomial distribution, then

$$\mu = E(Y) = \frac{r}{p} \quad \text{and} \quad \sigma^2 = V(Y) = \frac{r(1-p)}{p^2}.$$

Proof:

Left as an exercise...

Negative Binomial Distribution

There are actually 2 types of Negative Binomial distribution:

	Type 1 (textbook)	**Type 2**
Usage	counting the num. of trials needed to get rth success	counting the num. of failures before rth success
Values Y can take	$y = r, r+1, \ldots$	$y = 0, 1, 2, \ldots$
PMF	$\binom{y-1}{r-1} p^r (1-p)^{y-r}$ y is the num of trials INCLUDING the trial that is the rth success	$\binom{r+y-1}{y} p^r (1-p)^y$ y is the num of failures NOT including the trial that is the rth success
Mean	$\frac{r}{p}$ (how many trials to expect until we get the rth success)	$\frac{r(1-p)}{p}$ (how many failures to expect before we get the rth success)
Variance	$\frac{r(1-p)}{p^2}$	$\frac{r(1-p)}{p^2}$

Negative Binomial Distribution

Example:

Suppose we are counting the number of goals we score (success) for the penalty kicks we make (trials). Given that the probability of making a goal is 0.2, what is the probability that we score our 3rd goal on our 10th penalty kick.

Solution:

	Type 1	**Type 2**
Random Variable	Let Y denote the number of penalty kicks until we got our 3rd goal.	Let Y denote the number of penalty kicks missed before we got our 3rd goal.
PMF	$\binom{y-1}{r-1} p^r (1-p)^{y-r}$	$\binom{r+y-1}{y} p^r (1-p)^y$
Probability of interest	$P(Y = 10) = p(10) = \binom{10-1}{3-1} (0.2)^3 (1-0.2)^{10-3} = 0.0604$	$P(Y = 7) = p(7) = \binom{3+7-1}{7} (0.2)^3 (1-0.2)^7 = 0.0604$
Mean	$E(Y) = \frac{r}{p} = \frac{3}{0.2} = 15$ We should expect to kick 15 penalty kicks until we got the 3rd goal.	$E(Y) = \frac{r(1-p)}{p} = \frac{3(1-0.2)}{0.2} = 12$ We should expect to fail at 12 penalty kicks before we got the 3rd goal.
Variance	$V(Y) = \frac{r(1-p)}{p^2} = \frac{3(1-0.2)}{0.2^2} = 60$	$V(Y) = \frac{r(1-p)}{p^2} = \frac{3(1-0.2)}{0.2^2} = 60$

Hypergeometric Distribution

Definition 3.10: Hypergeometric Distribution

A random variable Y is said to have a *hypergeometric probability distribution* if and only if

$$p(y) = \frac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}},$$

where y is an integer $0, 1, 2, \dots, n$, $y \leq r$ and $n - y \leq N - r$.

- The hypergeometric story:
  - You have a (finite) population of N items.
  - You have a subset of interest with r items. (num. of successes in population)
  - You choose a sample (w/o replacement) of n items from the population with N items.
  - Let Y be the random variable for the number of items in the sample that belongs to the subset of interest. (num. of successes in the sample)

Hypergeometric Distribution

- the probability distribution of a random variable counting the number of successes from a small population without replacement
- We call $p(y) = \frac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}}$ as the pmf of a hypergeometric random variable.
- You can consider
  - y as the num. of successes in the sample
  - r as the num. of successes in the population
  - $N - r$ as the the num. of failures in the population
  - $n - y$ as the the num. of failures in the sample
- Notation: $Y \sim \text{Hyper}(N, r, n)$ of $Y \sim \text{H}(N, r, n)$, read as: "Y is a hypergeometric random variable with population size N, sample size n, and r items from a group of interest."
- We can treat $\frac{r}{N}$ as the probability of success when n is small enough relative to N.
- The Binomial distribution is a special case of the hypergeometric distribution wherein $Y \sim \text{Hyper}(N, r, n)$ can be approximated by $Y \sim \text{B}(n, \frac{r}{N})$.
- As the population size N increases, the hypergeometric distribution more closely approximates the binomial distribution.

Hypergeometric Distribution

Example 1:

There is a class of 20 students with 14 boys and 6 girls. 5 students will be chosen to take part in a math competition. What is the probability that 2 girls will be selected?

Solution:

- Let Y be the number of girls selected.
- Y is a hypergeometric random variable since each pick is not independent.
  - The probability of picking a girl first is $\frac{6}{20}$.
  - The probability of picking a boy second is $\frac{14}{19}$ if a girl was picked first. It is $\frac{13}{19}$ if a boy was picked first.
  - The probability of "success" changes every pick.
  - The probability of "success" depends on the first pick.

Hypergeometric Distribution

Example 1:

There is a class of 20 students with 14 boys and 6 girls. 5 students will be chosen to take part in a math competition. What is the probability that 2 girls will be selected?

Solution:

- Let Y be the number of girls selected.
- Y is a hypergeometric random variable since each pick is not independent.
- Compute $P(Y = 2)$.
- Given: $N = 20$, $r = 6$, and $n = 5$.
- $p(y) = \dfrac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}} \Rightarrow p(2) = \dfrac{\binom{6}{2}\binom{20-6}{5-2}}{\binom{20}{5}} = \dfrac{\frac{6!}{2!(6-2)!}\frac{14!}{3!(14-3)!}}{\frac{20!}{5!(20-5)!}} = \dfrac{(15)(364)}{15504} =$ 0.3522.
- The probability of selecting 2 girls is 0.3522.

Hypergeometric Distribution

Example 2:

From a group of 20 Ph.D. engineers, 10 are randomly selected for employment. What is the probability that the 10 selected include all the 5 best engineers in the group of 20?

Solution:

- Let Y be the number of best engineers among the 10 selected.
- Y is a hypergeometric random variable since each pick is not independent.
- Compute $P(Y = 5)$.
- Given: $N = 20$, $r = 5$, and $n = 10$.
- $p(y) = \dfrac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}} \Rightarrow p(5) = \dfrac{\binom{5}{5}\binom{20-5}{10-5}}{\binom{20}{10}} = \dfrac{\frac{5!}{5!(5-5)!}\,\frac{15!}{5!(15-5)!}}{\frac{20!}{10!(20-10)!}} = \dfrac{21}{1292} = 0.0162.$
- The probability of selecting all 5 best engineers is 0.0162.

Theorem 3.10

If Y is a random variable with a hypergeometric distribution, then

$$\mu = E(Y) = \frac{nr}{N} \quad \text{and} \quad \sigma^2 = V(Y) = n\left(\frac{r}{N}\right)\left(\frac{N-r}{N}\right)\left(\frac{N-n}{N-1}\right).$$

Proof:

Left as an exercise...

Hypergeometric Distribution

Proof:

$$E(Y) = \sum_y y p(y) \quad \text{defn of expected value}$$

$$= \sum_{y=0}^{n} y \frac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}} \quad \text{pmf of hypergeometric distribution}$$

$$= \sum_{y=1}^{n} y \frac{\frac{r!}{y!(r-y)!}\frac{(N-r)!}{(n-y)!(N-r-n+y)!}}{\frac{N!}{n!(N-n)!}}$$

start index at $y = 1$ since $y = 0$ does not contribute to the summation

$$= \sum_{y=1}^{n} y \frac{r!}{y!(r-y)!}\frac{(N-r)!}{(n-y)!(N-r-n+y)!}\frac{n!(N-n)!}{N!}$$

$$= n\sum_{y=1}^{n} y \frac{r!}{y!(r-y)!}\frac{(N-r)!}{(n-y)!(N-r-n+y)!}\frac{(n-1)!(N-n)!}{N!}$$

(continued next slide...)

Hypergeometric Distribution

Proof:

$$\begin{aligned}
E(Y) &= n \sum_{y=1}^{n} y \frac{r!}{y!(r-y)!} \frac{(N-r)!}{(n-y)!(N-r-n+y)!} \frac{(n-1)!(N-n)!}{N!} \\
&= nr \sum_{y=1}^{n} y \frac{(r-1)!}{y!(r-y)!} \frac{(N-r)!}{(n-y)!(N-r-n+y)!} \frac{(n-1)!(N-n)!}{N!} \\
&= \frac{nr}{N} \sum_{y=1}^{n} y \frac{(r-1)!}{y!(r-y)!} \frac{(N-r)!}{(n-y)!(N-r-n+y)!} \frac{(n-1)!(N-n)!}{(N-1)!} \\
&= \frac{nr}{N} \sum_{y=1}^{n} y \frac{(r-1)!}{y!(r-y)!} \frac{(N-r)!}{(n-y)!(N-r-n+y)!} \frac{(n-1)!(N-n)!}{(N-1)!} \\
&= \frac{nr}{N} \sum_{y=1}^{n} \frac{(r-1)!}{(y-1)!(r-y)!} \frac{(N-r)!}{(n-y)!(N-r-n+y)!} \frac{(n-1)!(N-n)!}{(N-1)!}
\end{aligned}$$

(continued next slide...)

Hypergeometric Distribution

Proof:

Recall: pmf of hypergeometric r.v. is $p(y) = \dfrac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}} = \dfrac{\frac{r!}{y!(r-y)!}\frac{(N-r)!}{(n-y)!(N-r-n+y)!}}{\frac{N!}{n!(N-n)!}}$.

$$E(Y) = \frac{nr}{N}\sum_{y=1}^{n} \frac{(r-1)!}{(y-1)!(r-y)!}\frac{(N-r)!}{(n-y)!(N-r-n+y)!}\frac{(n-1)!(N-n)!}{(N-1)!}$$

$$= \frac{nr}{N}\sum_{y=1}^{n} \frac{\frac{(r-1)!}{(y-1)!(r-y)!}\frac{(N-r)!}{(n-y)!(N-r-n+y)!}}{\frac{(N-1)!}{(n-1)!(N-n)!}}$$

$$= \frac{nr}{N}\sum_{z=0}^{n-1} \frac{\frac{(r-1)!}{z!(r-z-1)!}\frac{(N-r)!}{(n-z-1)!(N-r-n+z+1)!}}{\frac{(N-1)!}{(n-1)!(N-n)!}}$$

Let $z = y - 1$.

$$= \frac{nr}{N}\sum_{z=0}^{n-1} \frac{\binom{r-1}{z}\binom{N-r}{n-1-z}}{\binom{N-1}{n-1}}$$

hypergeometric pmf with parameters: $N-1$, $n-1$, and $r-1$

$$= \frac{nr}{N}\sum_{z=0}^{n-1} p(z) = \frac{nr}{N}.$$

probabilities sum to 1

Hypergeometric Distribution

Proof: To derive the variance, we derive first $E\{Y(Y-1)\}$.

$$E\{Y(Y-1)\} = \sum_y y(y-1)p(y) \quad \text{defn of expected value}$$

$$= \sum_{y=0}^{n} y(y-1)\frac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}} \quad \text{pmf of hypergeometric distribution}$$

$$= \sum_{y=2}^{n} y(y-1)\frac{\frac{r!}{y!(r-y)!}\frac{(N-r)!}{(n-y)!(N-r-n+y)!}}{\frac{N!}{n!(N-n)!}}$$

start index at $y = 2$ since $y = 0, 1$ don't contribute to the summation

$$= \sum_{y=2}^{n} y(y-1)\frac{r!}{y!(r-y)!}\frac{(N-r)!}{(n-y)!(N-r-n+y)!}\frac{n!(N-n)!}{N!}$$

$$= \sum_{y=2}^{n} \frac{r!}{(y-2)!(r-y)!}\frac{(N-r)!}{(n-y)!(N-r-n+y)!}\frac{n!(N-n)!}{N!}$$

(continued next slide...)

Hypergeometric Distribution

Proof:

$$
\begin{aligned}
E\{Y(Y-1)\} &= \sum_{y=2}^{n} \frac{r!}{(y-2)!(r-y)!} \frac{(N-r)!}{(n-y)!(N-r-n+y)!} \frac{n!(N-n)!}{N!} \\
&= n(n-1) \sum_{y=2}^{n} \frac{r!}{(y-2)!(r-y)!} \frac{(N-r)!}{(n-y)!(N-r-n+y)!} \frac{(n-2)!(N-n)!}{N!} \\
&= n(n-1)r(r-1) \sum_{y=2}^{n} \frac{(r-2)!}{(y-2)!(r-y)!} \frac{(N-r)!}{(n-y)!(N-r-n+y)!} \frac{(n-2)!(N-n)!}{N!} \\
&= \frac{n(n-1)r(r-1)}{N(N-1)} \sum_{y=2}^{n} \frac{(r-2)!}{(y-2)!(r-y)!} \frac{(N-r)!}{(n-y)!(N-r-n+y)!} \frac{(n-2)!(N-n)!}{(N-2)!} \\
&= \frac{n(n-1)r(r-1)}{N(N-1)} \sum_{y=2}^{n} \frac{\frac{(r-2)!}{(y-2)!(r-y)!} \frac{(N-r)!}{(n-y)!(N-r-n+y)!}}{\frac{(N-2)!}{(n-2)!(N-n)!}} \\
&= \frac{n(n-1)r(r-1)}{N(N-1)} \sum_{z=0}^{n-2} \frac{\frac{(r-2)!}{z!(r-2-z)!} \frac{(N-r)!}{(n-2-z)!(N-r-n+z+2)!}}{\frac{(N-2)!}{(n-2)!(N-n)!}} \qquad \text{Let } z = y-2.
\end{aligned}
$$

(continued next slide...)

Hypergeometric Distribution

Proof:

Recall: pmf of hypergeometric r.v. is $p(y) = \frac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}} = \frac{\frac{r!}{y!(r-y)!}\frac{(N-r)!}{(n-y)!(N-r-n+y)!}}{\frac{N!}{n!(N-n)!}}$.

$$E\{Y(Y-1)\} = \frac{n(n-1)r(r-1)}{N(N-1)} \sum_{z=0}^{n-2} \frac{\frac{(r-2)!}{z!(r-2-z)!}\frac{(N-r)!}{(n-2-z)!(N-r-n+z+2)!}}{\frac{(N-2)!}{(n-2)!(N-n)!}}$$

$$= \frac{n(n-1)r(r-1)}{N(N-1)} \sum_{z=0}^{n-2} \frac{\binom{r-2}{z}\binom{N-r}{n-2-z}}{\binom{N-2}{n-2}}$$

hypergeometric pmf with parameters: $N-2$, $n-2$, and $r-2$

$$= \frac{n(n-1)r(r-1)}{N(N-1)} \sum_{z=0}^{n-2} p(z)$$

$$= \frac{n(n-1)r(r-1)}{N(N-1)}.$$ probabilities sum to 1

Proof:

We have $E\{Y(Y-1)\} = \frac{n(n-1)r(r-1)}{N(N-1)}$.

But we need $E(Y^2)$ which is:

$$\begin{aligned} E(Y^2) &= E\{Y(Y-1)\} + E(Y) \\ &= \frac{n(n-1)r(r-1)}{N(N-1)} + \frac{nr}{N}. \end{aligned}$$

We can now solve for $V(Y)$ as follows:

$$\begin{aligned} V(Y) &= E(Y^2) - \{E(Y)\}^2 \\ &= \frac{n(n-1)r(r-1)}{N(N-1)} + \frac{nr}{N} - \left(\frac{nr}{N}\right)^2 \\ &= \frac{n(n-1)r(r-1)}{N(N-1)} + \frac{nr}{N} - \frac{n^2r^2}{N^2} \\ &= \frac{Nn(n-1)r(r-1) + nrN(N-1) - n^2r^2(N-1)}{N^2(N-1)} \\ &= \frac{(Nn^2 - Nn)(r^2 - r) + nrN^2 - nrN - n^2r^2N + n^2r^2}{N^2(N-1)} \\ &= \frac{Nn^2r^2 - Nnr^2 - Nn^2r + Nnr + nrN^2 - nrN - n^2r^2N + n^2r^2}{N^2(N-1)} \end{aligned}$$

(continued next slide...)

Hypergeometric Distribution

Proof:

$$
\begin{aligned}
V(Y) &= \frac{Nn^2r^2 - Nnr^2 - Nn^2r + Nnr + nrN^2 - nrN - n^2r^2N + n^2r^2}{N^2(N-1)} \\
&= \frac{nrN^2 - nr^2N - n^2rN + n^2r^2}{N^2(N-1)} \\
&= \frac{nr(N^2 - rN - nN + nr)}{N^2(N-1)} \\
&= \frac{nr(N-r)(N-n)}{N^2(N-1)} \\
&= n\left(\frac{r}{N}\right)\left(\frac{N-r}{N}\right)\left(\frac{N-n}{N-1}\right).
\end{aligned}
$$

□

Poisson Distribution

Definition 3.11: Poisson Distribution

A random variable Y is said to have a *Poisson probability distribution* if and only if

$$p(y) = \frac{\lambda^y}{y!}e^{-\lambda},$$

where $y = 0, 1, 2, \ldots$ and $\lambda > 0$.

- the probability distribution of a random variable counting the number of times an event occurs within a specified time interval
- Examples: traffic flow, fault prediction on electric cables, randomly occurring accidents, calls coming into a telephone switch board
- We call $p(y) = \frac{\lambda^y}{y!}e^{-\lambda}$ as the pmf of a Poisson random variable.
- The parameter λ represents the average or expected rate of occurrence.
- Notation: $Y \sim \text{Pois}(\lambda)$ or $Y \sim \text{P}(\lambda)$, read as: "Y is a Poisson random variable with average or expected rate of occurrence λ."
- typically used for problems which cannot be solved using the binomial distribution (large n and small p)

Poisson Distribution: Approximation to the Binomial Dist.

Suppose we have a binomial random variable Y with n trials and probability of success p...

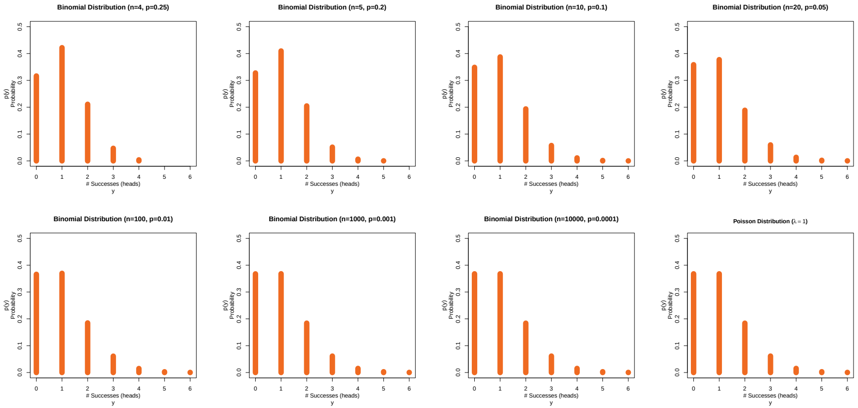
y	$np=1$, $n=4$, $p=0.25$, $p(y)$	$np=1$, $n=5$, $p=0.20$, $p(y)$	$np=1$, $n=10$, $p=0.10$, $p(y)$	$np=1$, $n=20$, $p=0.05$, $p(y)$	$np=1$, $n=100$, $p=0.01$, $p(y)$	$np=1$, $n=1000$, $p=0.001$, $p(y)$	$np=1$, $n=10000$, $p=0.0001$, $p(y)$
0	0.316	0.328	0.349	0.359	0.366	0.368	0.368
1	0.422	0.410	0.387	0.377	0.370	0.368	0.368
2	0.211	0.205	0.194	0.189	0.185	0.184	0.184
3	0.047	0.051	0.057	0.060	0.061	0.061	0.061
4	0.004	0.006	0.011	0.013	0.015	0.015	0.015
5		0.000	0.001	0.002	0.003	0.003	0.003
6			0.000	0.000	0.000	0.001	0.001

$\lambda=np=1$, $p(y)$
0.368
0.368
0.184
0.061
0.015
0.003
0.001

Notice the probability values of Y computed using the binomal pmf approaching the probability values computed using the Poisson pmf as n increases...

Poisson Distribution: Approximation to the Binomial Dist.

Visualizing the table in the previous slide...



- It can be seen that the Poisson distribution can approximate the binomial distribution when n is large and p is small.
- Helpful for situations such as examining the number of defective items in a large batch and the defective rate is small...
- Poisson is pmf is easier to compute than the binomial pmf with large n and small p.

Poisson Distribution

Example 1:

The manufacturer of the disk drives in one of the well-known brands of microcomputers expects 2% of the disk drives to malfunction during the microccomputer's warranty period. Calculate the probability that in a sample of 100 disk drives, not more than three will malfunction.

Solution: Let Y be the number of disk drives malfunctioning.

	Binomial	Poisson
	$p(y) = \binom{n}{y} p^y (1-p)^{n-y}$	$p(y) = \frac{\lambda^y}{y!} e^{-\lambda}$
	n=100	λ=np
	p=0.02	λ=100(0.02) = 2
y	$p(y) = \binom{100}{y} (0.02)^y (0.98)^{100-y}$	$p(y) = \frac{2^y}{y!} e^{-2}$
0	0.13262	0.13534
1	0.27065	0.27067
2	0.27341	0.27067
3	0.18228	0.18045
Total	0.85890	0.85713

Poisson Distribution

Theorem 3.11

If Y is a random variable with a Poisson distribution with parameter λ, then

$$\mu = E(Y) = \lambda \quad \text{and} \quad \sigma^2 = V(Y) = \lambda.$$

Proof:

Left as an exercise...

Poisson Distribution

Example 2:

Suppose it has been observed that, on average, 180 cars per hour pass a specified point on a particular road in the morning rush hour. Due to impending roadworks it is estimated that congestion will occur closer to the city centre if more than 5 cars pass the point in any one minute. What is the probability of congestion occurring?

Solution:

- Asked: Probability of congestion
- Strategy:
  - What events will bring about congestion? more than 5 cars in any minute
  - Let X be the random variable for the number of cars arriving in any minute.
  - Compute $P(X > 5)$.
    - ★ $P(X > 5) = P(X = 6) + P(X = 7) + \dots$ (can go to infinity since no information on maximum number of cars)
    - ★ Easier to compute: $P(X > 5) = 1 - P(X \leq 5) = 1 - \{P(X = 0) + P(X = 1) + \dots P(X = 5)\}$.

Poisson Distribution

Example 2:

Suppose it has been observed that, on average, 180 cars per hour pass a specified point on a particular road in the morning rush hour. Due to impending roadworks it is estimated that congestion will occur closer to the city centre if more than 5 cars pass the point in any one minute. What is the probability of congestion occurring?

Solution:

- Strategy:
  - Compute $P(X > 5)$.
  - $P(X > 5) = 1 - \{P(X = 0) + P(X = 1) + \ldots P(X = 5)\}$
  - Since X is a random variable for the number of cars arriving in any minute, we can use the pmf of the Poisson distribution with average rate of occurrence parameter $\lambda = 3$ to compute $P(X = 0)$, $P(X = 1)$, $P(X = 2), P(X = 3), P(X = 4), P(X = 5)$.
    - ★ $P(X = 0) = p(0) = \frac{3^0}{0!}e^{-3} =\approx 0.04979$
    - ★ $P(X = 1) = p(1) = \frac{3^1}{1!}e^{-3} =\approx 0.14936$
    - ★ $P(X = 2) = p(2) = \frac{3^2}{2!}e^{-3} =\approx 0.22404$

pmf of Poisson: $p(y) = \frac{\lambda^y}{y!}e^{-\lambda}$

Poisson Distribution

Example 2:

Suppose it has been observed that, on average, 180 cars per hour pass a specified point on a particular road in the morning rush hour. Due to impending roadworks it is estimated that congestion will occur closer to the city centre if more than 5 cars pass the point in any one minute. What is the probability of congestion occurring?

Solution:

- Strategy:
  - Compute $P(X > 5)$.
  - $P(X > 5) = 1 - \{P(X = 0) + P(X = 1) + \ldots P(X = 5)\}$
  - Since X is a random variable for the number of cars arriving in any minute, we can use the pmf of the Poisson distribution with average rate of occurrence parameter $\lambda = 3$ to compute $P(X = 0)$, $P(X = 1)$, $P(X = 2), P(X = 3), P(X = 4), P(X = 5)$.
    - ★ $P(X = 3) = p(3) = \frac{3^3}{3!}e^{-3} =\approx 0.22404$
    - ★ $P(X = 4) = p(4) = \frac{3^4}{4!}e^{-3} =\approx 0.16803$
    - ★ $P(X = 5) = p(5) = \frac{3^5}{5!}e^{-3} =\approx 0.10082$

pmf of Poisson: $p(y) = \frac{\lambda^y}{y!}e^{-\lambda}$

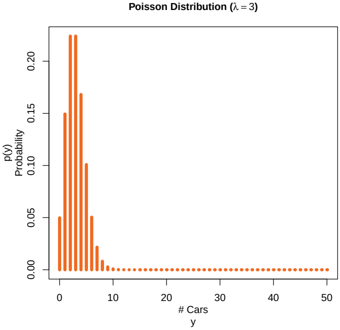
Poisson Distribution

Example 2:

Suppose it has been observed that, on average, 180 cars per hour pass a specified point on a particular road in the morning rush hour. Due to impending roadworks it is estimated that congestion will occur closer to the city centre if more than 5 cars pass the point in any one minute. What is the probability of congestion occurring?

Solution:

- Strategy:
  - Compute $P(X > 5)$.
  - $P(X > 5) = 1 - \{P(X = 0) + P(X = 1) + \dots P(X = 5)\}$
    $= 1 - (0.04979 + 0.14936 + 0.22404 + 0.22404 + 0.16803 + 0.10082)$
    $= 1 - 0.91608 = 0.08392.$
  - The probability of congestion is 0.08392.



Questions?

Homework Exercises: 3.37, 3.43, 3.55, 3.57, 3.65

Solutions will be discussed this Friday by the TA.