Discrete Probability Distributions

Chapter 6

Discrete Distributions
- Uniform Distribution
- Bernoulli Distribution
- Binomial Distribution
- Poisson Distribution
- Hypergeometric Distribution
- Geometric Distribution
- Transformations of Random Variables (optional)

Random Variables

- A random variable is a function or rule that assigns a numerical value to each outcome in the sample space of a random experiment.

  Nomenclature:
  - Capital letters are used to represent random variables (e.g., $X$, $Y$).
  - Lower case letters are used to represent values of the random variable (e.g., $x$, $y$).

- A discrete random variable has a countable number of distinct values.
Discrete Distributions

Random Variables
For example,

<table>
<thead>
<tr>
<th>Decision Problem</th>
<th>Discrete Random Variable (Range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>On the late morning (9 to 12) work shift, L.L. Bean’s order processing center staff can handle up to 5 orders per minute. The mean arrival rate is 3.5 orders per minute. What is the probability that more than 5 orders will arrive in a given minute?</td>
<td>$X = \text{number of phone calls that arrive in a given minute at the L.L. Bean order processing center}$ ($X = 0, 1, 2, ...$)</td>
</tr>
</tbody>
</table>

Discrete Distributions

Probability Distributions

- A *discrete probability distribution* assigns a probability to each value of a discrete random variable $X$.
- To be a valid probability, each probability must be between $0 \leq P(x_i) \leq 1$.
- And the sum of all the probabilities for the values of $X$ must be equal to unity.

\[ \sum_{i=1}^{n} P(x_i) = 1 \]
When you flip a coin three times, the sample space has eight equally likely simple events. They are:

<table>
<thead>
<tr>
<th>1\textsuperscript{st} Toss</th>
<th>2\textsuperscript{nd} Toss</th>
<th>3\textsuperscript{rd} Toss</th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td>H</td>
<td>H</td>
<td>T</td>
</tr>
<tr>
<td>H</td>
<td>T</td>
<td>H</td>
</tr>
<tr>
<td>H</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>T</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td>T</td>
<td>T</td>
<td>H</td>
</tr>
<tr>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
</tbody>
</table>

If $X$ is the number of heads, then $X$ is a random variable whose probability distribution is as follows:

<table>
<thead>
<tr>
<th>Possible Events</th>
<th>$x$</th>
<th>$P(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>TTT</td>
<td>0</td>
<td>1/8</td>
</tr>
<tr>
<td>HTT, THT, TTH</td>
<td>1</td>
<td>3/8</td>
</tr>
<tr>
<td>HHT, HTH, THH</td>
<td>2</td>
<td>3/8</td>
</tr>
<tr>
<td>HHH</td>
<td>3</td>
<td>1/8</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>
Discrete Distributions

Example: Coin Flips

Note that the values of $X$ need not be equally likely. However, they must sum to unity.

Note also that a discrete probability distribution is defined only at specific points on the $X$-axis.

![Probability Distribution for Coin Flips](image)

Discrete Distributions

Expected Value

- The expected value $E(X)$ of a discrete random variable is the sum of all $X$-values weighted by their respective probabilities.

- If there are $n$ distinct values of $X$,

  $$E(X) = \mu = \sum_{i=1}^{n} x_i P(x_i)$$

- The $E(X)$ is a measure of central tendency.
The probability distribution of emergency service calls on Sunday by Ace Appliance Repair is:

<table>
<thead>
<tr>
<th>x</th>
<th>P(x)</th>
<th>xP(x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.05</td>
<td>0.00</td>
</tr>
<tr>
<td>1</td>
<td>0.10</td>
<td>0.10</td>
</tr>
<tr>
<td>2</td>
<td>0.30</td>
<td>0.60</td>
</tr>
<tr>
<td>3</td>
<td>0.25</td>
<td>0.75</td>
</tr>
<tr>
<td>4</td>
<td>0.20</td>
<td>0.80</td>
</tr>
<tr>
<td>5</td>
<td>0.10</td>
<td>0.50</td>
</tr>
<tr>
<td>Total</td>
<td>1.00</td>
<td>2.75</td>
</tr>
</tbody>
</table>

What is the average or expected number of service calls?

First calculate $x_iP(x_i)$:

<table>
<thead>
<tr>
<th>x</th>
<th>P(x)</th>
<th>xP(x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.05</td>
<td>0.00</td>
</tr>
<tr>
<td>1</td>
<td>0.10</td>
<td>0.10</td>
</tr>
<tr>
<td>2</td>
<td>0.30</td>
<td>0.60</td>
</tr>
<tr>
<td>3</td>
<td>0.25</td>
<td>0.75</td>
</tr>
<tr>
<td>4</td>
<td>0.20</td>
<td>0.80</td>
</tr>
<tr>
<td>5</td>
<td>0.10</td>
<td>0.50</td>
</tr>
<tr>
<td>Total</td>
<td>1.00</td>
<td>2.75</td>
</tr>
</tbody>
</table>

The sum of the $xP(x)$ column is the expected value or mean of the discrete distribution.

$$E(X) = \mu = \sum_{i=1}^{5} x_i P(x_i)$$
Discrete Distributions

**Example: Service Calls**

This particular probability distribution is not symmetric around the mean $= 2.75$.

However, the mean is still the balancing point, or fulcrum.

Because $E(X)$ is an average, it does not have to be an observable point.

---

**Application: Life Insurance**

- Expected value is the basis of life insurance.
- For example, what is the probability that a 30-year-old white female will die within the next year?
- Based on mortality statistics, the probability is .00059 and the probability of living another year is $1 - .00059 = .99941$.
- What premium should a life insurance company charge to break even on a $500,000 1$-year term policy?
Let $X$ be the amount paid by the company to settle the policy.

<table>
<thead>
<tr>
<th>Event</th>
<th>$x$</th>
<th>$P(x)$</th>
<th>$xP(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Live</td>
<td>0</td>
<td>.99941</td>
<td>0.00</td>
</tr>
<tr>
<td>Die</td>
<td>500,000</td>
<td>.00059</td>
<td>295.00</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>1.00000</td>
<td>295.00</td>
</tr>
</tbody>
</table>

The total expected payout is 

So, the premium should be $295 plus whatever return the company needs to cover administrative overhead and profit.

---

Application: Life Insurance

Expected value can be applied to raffles and lotteries.

If it costs $2 to buy a ticket in a raffle to win a new car worth $55,000 and 29,346 raffle tickets are sold, what is the expected value of a raffle ticket?

If you buy 1 ticket, what is the chance you will win = \[
\frac{1}{29,346}
\]

lose = \[
\frac{29,345}{29,346}
\]
Discrete Distributions

Application: Raffle Tickets

• Now, calculate the $E(X)$:

$$E(X) = (\text{value if you win}) \times P(\text{win}) + (\text{value if you lose}) \times P(\text{lose})$$

$$= (55,000) \left( \frac{1}{29,346} \right) + (0) \left( \frac{29,345}{29,346} \right)$$

$$= (55,000)(0.000034076) + (0)(0.999965924) = 1.87$$

• The raffle ticket is actually worth $1.87. Is it worth spending $2.00 for it?

Discrete Distributions

Actuarial Fairness

• An actuarially fair insurance program must collect as much in overall revenue as it pays out in claims.

• Accomplish this by setting the premiums to reflect empirical experience with the insured group.

• If the pool of insured persons is large enough, the total payout is predictable.
Discrete Distributions

**Variance and Standard Deviation**

- If there are \( n \) distinct values of \( X \), then the variance of a discrete random variable is:

\[
V(X) = \sigma^2 = \sum_{i=1}^{n} [x_i - \mu]^2 P(x_i)
\]

- The variance is a weighted average of the dispersion about the mean and is denoted either as \( \sigma^2 \) or \( V(X) \).

- The standard deviation is the square root of the variance and is denoted \( \sigma \).

\[
\sigma = \sqrt{\sigma^2} = \sqrt{V(X)}
\]

---

**Example: Bed and Breakfast**

The Bay Street Inn is a 7-room bed-and-breakfast in Santa Thersea, Ca.

<table>
<thead>
<tr>
<th>( x )</th>
<th>( P(x) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.05</td>
</tr>
<tr>
<td>1</td>
<td>0.05</td>
</tr>
<tr>
<td>2</td>
<td>0.06</td>
</tr>
<tr>
<td>3</td>
<td>0.10</td>
</tr>
<tr>
<td>4</td>
<td>0.13</td>
</tr>
<tr>
<td>5</td>
<td>0.20</td>
</tr>
<tr>
<td>6</td>
<td>0.15</td>
</tr>
<tr>
<td>7</td>
<td>0.26</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1.00</strong></td>
</tr>
</tbody>
</table>

The probability distribution of room rentals during February is:
Discrete Distributions

**Example: Bed and Breakfast**

First find the expected value

\[ E(X) = \mu = \sum_{i=1}^{7} x_i P(x_i) \]

= 4.71 rooms

<table>
<thead>
<tr>
<th>x</th>
<th>P(x)</th>
<th>x P(x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.05</td>
<td>0.00</td>
</tr>
<tr>
<td>1</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>2</td>
<td>0.06</td>
<td>0.12</td>
</tr>
<tr>
<td>3</td>
<td>0.10</td>
<td>0.30</td>
</tr>
<tr>
<td>4</td>
<td>0.13</td>
<td>0.52</td>
</tr>
<tr>
<td>5</td>
<td>0.20</td>
<td>1.00</td>
</tr>
<tr>
<td>6</td>
<td>0.15</td>
<td>0.90</td>
</tr>
<tr>
<td>7</td>
<td>0.26</td>
<td>1.82</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>4.71</td>
</tr>
</tbody>
</table>

The standard deviation is:

\[ \sigma = \sqrt{\sum_{i=1}^{7} [x_i - \mu]^2 P(x_i)} \]

= \sqrt{4.2259} = 2.0577 rooms

\[ \sigma^2 = 4.225900 \]
Discrete Distributions

Example: Bed and Breakfast
The histogram shows that the distribution is skewed to the left and bimodal.

The mode is 7 rooms rented but the average is only 4.71 room rentals.

= 2.06 indicates considerable variation around the average.

Discrete Distributions

What is a PDF or CDF?

- A probability distribution function (PDF) is a mathematical function that shows the probability of each X-value.

- A cumulative distribution function (CDF) is a mathematical function that shows the cumulative sum of probabilities, adding from the smallest to the largest X-value, gradually approaching unity.
Discrete Distributions

What is a PDF or CDF?

Consider the following illustrative histograms:

Illustrative PDF
(Probability Density Function)

Cumulative CDF
(Cumulative Density Function)

The equations for these functions depend on the parameter(s) of the distribution.

Uniform Distribution

Characteristics of the Uniform Distribution

• The uniform distribution describes a random variable with a finite number of integer values from $a$ to $b$ (the only two parameters).

• Each value of the random variable is equally likely to occur.

• Consider the following summary of the uniform distribution:
# Uniform Distribution

| Parameters | $a = \text{lower limit}$  
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$b = \text{upper limit}$</td>
</tr>
<tr>
<td>PDF</td>
<td>$P(x) = \frac{1}{b - a + 1}$</td>
</tr>
<tr>
<td>Range</td>
<td>$a \leq x \leq b$</td>
</tr>
<tr>
<td>Mean</td>
<td>$\frac{a + b}{2}$</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>$\sqrt{\frac{(b-a+1)^2 - 1}{12}}$</td>
</tr>
<tr>
<td>Random data generation in Excel</td>
<td>$a + \text{INT}((b-a+1) \times \text{RAND})$</td>
</tr>
<tr>
<td>Comments</td>
<td>Used as a benchmark, to generate random integers, or to create other distributions.</td>
</tr>
</tbody>
</table>

## Example: Rolling a Die

- The number of dots on the roll of a die form a uniform random variable with six equally likely integer values: 1, 2, 3, 4, 5, 6
- What is the probability of rolling any of these?

![PDF for one die](image1)

![CDF for one die](image2)
Uniform Distribution

Example: Rolling a Die

- The PDF for all \( x \) is: \( P(x) = \frac{1}{b-a+1} = \frac{1}{6-1+1} = \frac{1}{6} \)

- Calculate the mean as:
  \[
  \frac{a+b}{2} = \frac{1+6}{2} = 3.5
  \]

- Calculate the standard deviation as:
  \[
  \sqrt{\frac{(b-a+1)^2 - 1}{12} - \left\lfloor \frac{6-1+1}{12} \right\rfloor} = \sqrt{\frac{12 - 1}{12}} = 1.708
  \]

Uniform Distribution

Application: Pumping Gas (Figure 6.9)

On a gas pump, the last two digits (pennies) displayed will be a uniform random integer (assuming the pump stops automatically).

The parameters are: \( a = 00 \) and \( b = 99 \)
Uniform Distribution

Application: Pumping Gas

- The PDF for all $x$ is:

$$P(x) = \frac{1}{b-a+1} = \frac{1}{99-0+1} = \frac{1}{100} = .010$$

- Calculate the mean as:

$$\frac{a+b}{2} = \frac{0+99}{2} = 49.5$$

- Calculate the standard deviation as:

$$\sqrt{\frac{(b-a+1)^2-1}{12}} = \sqrt{\frac{(99-0+1)^2-1}{12}} = 28.87$$

Uniform Distribution

Uniform Random Integers

- To generate random integers from a discrete uniform distribution, use Excel function

$$=a+\text{INT}((b-a+1\times\text{RAND}()))$$

- To create integers 1 through $N$, set $a = 1$ and $b = N$ and use Excel function

$$=1+\text{INT}(N\times\text{RAND}())$$

- To obtain $n$ distinct random integers, generate a few extra numbers and then delete the duplicate values.
The finance department at Zymurgy, Inc., has a new digital copier that requires a unique user ID code for each of the 37 users.

- Generate unique 4-digit uniform random integers from 1000 to 9999 using the function =1000+INT(9000*RAND()) in an Excel spreadsheet.

After entering the formula, drag it down to fill 50 cells with randomly generated numbers following the uniform distribution.
After highlighting and copying the cells to the clipboard, paste only the values (not the formulas) to another column using *Paste Special – Values*. Now these values can be sorted.

**Application: Copier Codes**

Sort the random numbers using *Data – Sort*.

Use the first 37 random numbers as copier codes for the current employees and save the remaining codes for future employees.
Uniform Distribution

Uniform Model in LearningStats

Here is the uniform distribution for one die from LearningStats.

Bernoulli Distribution

Bernoulli Experiments

- A random experiment with only 2 outcomes is a Bernoulli experiment.
- One outcome is arbitrarily labeled a “success” (denoted $X = 1$) and the other a “failure” (denoted $X = 0$).
  
  \[
  \text{is the } P(\text{success}), \quad 1 - \text{is the } P(\text{failure}).
  \]
- “Success” is usually defined as the less likely outcome so that $< .5$ for convenience.
- Note that $P(0) + P(1) = (1 - ) + = 1$ and $0 \leq 1.$
Bernoulli Distribution

Bernoulli Experiments

Consider the following Bernoulli experiments:

<table>
<thead>
<tr>
<th>Bernoulli Experiment</th>
<th>Possible Outcomes</th>
<th>Probability of “Success”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flip a coin</td>
<td>1 = heads</td>
<td>( \pi = .50 )</td>
</tr>
<tr>
<td></td>
<td>0 = tails</td>
<td></td>
</tr>
<tr>
<td>Inspect a jet turbine blade</td>
<td>1 = crack found</td>
<td>( \pi = .001 )</td>
</tr>
<tr>
<td></td>
<td>0 = no crack found</td>
<td></td>
</tr>
<tr>
<td>Purchase a tank of gas</td>
<td>1 = pay by credit card</td>
<td>( \pi = .78 )</td>
</tr>
<tr>
<td></td>
<td>0 = do not pay by credit</td>
<td></td>
</tr>
<tr>
<td></td>
<td>card</td>
<td></td>
</tr>
<tr>
<td>Do a mammogram test</td>
<td>1 = positive test</td>
<td>( \pi = .0004 )</td>
</tr>
<tr>
<td></td>
<td>0 = negative test</td>
<td></td>
</tr>
</tbody>
</table>

Bernoulli Experiments

- The expected value (mean) of a Bernoulli experiment is calculated as:

\[
E(X) = \sum_{i=1}^{2} x_i P(x_i) = (0)(1-\pi) + (1)(\pi) = \pi
\]

- The variance of a Bernoulli experiment is calculated as:

\[
V(X) = \sum_{i=1}^{2} \left[ x_i - E(X) \right]^2 P(x_i) = (0-\pi)^2(1-\pi) + (1-\pi)^2(\pi) = \pi(1-\pi)
\]

- The mean and variance are useful in developing the next model.
Binomial Distribution

Characteristics of the Binomial Distribution

• The binomial distribution arises when a Bernoulli experiment is repeated $n$ times.

• Each Bernoulli trial is independent so the probability of success remains constant on each trial.

• In a binomial experiment, we are interested in $X = \text{number of successes in } n \text{ trials}$. So,

$$X = X_1 + X_2 + \ldots + X_n$$

• The probability of a particular number of successes $P(X)$ is determined by parameters $n$ and $p$.

Binomial Distribution

Characteristics of the Binomial Distribution

• The mean of a binomial distribution is found by adding the means for each of the $n$ Bernoulli independent events:

$$+ \ldots + = n$$

• The variance of a binomial distribution is found by adding the variances for each of the $n$ Bernoulli independent events:

$$(1- )+ (1- ) + \ldots + (1- ) = n (1- )$$

• The standard deviation is

$$\sqrt{n (1- )}$$
Binomial Distribution

| Parameters | $n =$ number of trials  
| $\pi =$ probability of success  
| PDF | $P(x) = \frac{n!}{x!(n-x)!}\pi^x(1-\pi)^{n-x}$  
| Excel function | =BINOMDIST($k$,$n$,$\pi$,0)  
| Range | $X = 0, 1, 2, \ldots, n$  
| Mean | $n\pi$  
| Std. Dev. | $\sqrt{n\pi(1-\pi)}$  
| Random data generation in Excel | Sum $n$ values of =1+INT(2*RAND()) or use Excel’s Tools | Data Analysis  
| Comments | Skewed right if $\pi < .50$, skewed left if $\pi > .50$, and symmetric if $\pi = .50$.  

Example: Quick Oil Change Shop

- It is important to quick oil change shops to ensure that a car’s service time is not considered “late” by the customer.
- Service times are defined as either late or not late.
- $X$ is the number of cars that are late out of the total number of cars serviced.
- Assumptions:
  - cars are independent of each other  
  - probability of a late car is consistent
Binomial Distribution

Example: Quick Oil Change Shop

- What is the probability that exactly 2 of the next \( n = 10 \) cars serviced are late (\( P(X = 2) \))?
- \( P(\text{car is late}) = .10 \)
- \( P(\text{car not late}) = 1 - .10 = .90 \)

\[
P(x) = \frac{n!}{x!(n-x)!} \pi^x (1-\pi)^{n-x}
\]

\[
P(X = 2) = \frac{10!}{2!(10-2)!} (.1)^2(1-.10)^{10-2} = .1937
\]

Binomial Distribution

Example: Quick Oil Change Shop

- Alternatively, we could find \( P(X = 2) \) using the Excel function \( \text{=BINOMDIST}(k,n, ,0) \) where

\[
k = \text{the number of “successes” in } n \text{ trials}
\]
\[
n = \text{the number of independent trials}
\]
\[
= \text{probability of a “success”}
\]
\[
0 \text{ means that we want to calculate } P(X = 2) \text{ rather than } P(X \leq 2)
\]
A binomial distribution is skewed right if $p < .50$, skewed left if $p > .50$, and symmetric if $p = .50$.

Skewness decreases as $n$ increases, regardless of the value of $p$.

To illustrate, consider the following graphs:
Binomial Distribution

Binomial Shape

<table>
<thead>
<tr>
<th>π = 0.20</th>
<th>π = 0.50</th>
<th>π = 0.80</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skewed right</td>
<td>Symmetric</td>
<td>Skewed left</td>
</tr>
</tbody>
</table>

$n = 10$

$\begin{array}{c}
\text{Number of Successes} \\
\hline
0 & 5 & 10 & 15 & 20 \\
\end{array}$

$\begin{array}{c}
\text{Number of Successes} \\
\hline
0 & 5 & 10 & 15 & 20 \\
\end{array}$

$n = 20$

$\begin{array}{c}
\text{Number of Successes} \\
\hline
0 & 5 & 10 & 15 & 20 \\
\end{array}$

$\begin{array}{c}
\text{Number of Successes} \\
\hline
0 & 5 & 10 & 15 & 20 \\
\end{array}$
Binomial Distribution

Application: Uninsured Patients

- On average, 20% of the emergency room patients at Greenwood General Hospital lack health insurance.
- In a random sample of 4 patients, what is the probability that at least 2 will be uninsured?
- \( X \) = number of uninsured patients (“success”)
- \( P(\text{uninsured}) = 0.20 \) or .20
- \( P(\text{insured}) = 1 - 0.20 = 0.80 \)
- \( n = 4 \) patients
- The range is \( X = 0, 1, 2, 3, 4 \) patients.
What is the mean and standard deviation of this binomial distribution?

Mean = \( n = (4)(.20) = 0.8 \) patients

Standard deviation = \( \sqrt{np(1-p)} \)

= \( \sqrt{4(0.20)(1-0.20)} \)

= 0.8 patients

Here is the binomial distribution for \( n = 4, p = .20 \)

<table>
<thead>
<tr>
<th>( x )</th>
<th>PDF</th>
<th>CDF</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>.4096 = ( P(X=0) )</td>
<td>.4096 = ( P(X\leq0) = P(0) )</td>
</tr>
<tr>
<td>1</td>
<td>.4096 = ( P(X=1) )</td>
<td>.8192 = ( P(X\leq1) = P(0)+P(1) )</td>
</tr>
<tr>
<td>2</td>
<td>.1536 = ( P(X=2) )</td>
<td>.9728 = ( P(X\leq2) = P(0)+P(1)+P(2) )</td>
</tr>
<tr>
<td>3</td>
<td>.0256 = ( P(X=3) )</td>
<td>.9984 = ( P(X\leq3) = P(0)+P(1)+P(2)+P(3) )</td>
</tr>
<tr>
<td>4</td>
<td>.0016 = ( P(X=4) )</td>
<td>1.0000 = ( P(X\leq4) = P(0)+P(1)+P(2)+P(3)+P(4) )</td>
</tr>
</tbody>
</table>

These probabilities can be calculated using a calculator or Excel’s function

=BINOMDIST(\( x \),n,p,cumulative) where

\( cumulative = 0 \) for a PDF or \( = 1 \) for a CDF
Binomial Distribution

Application: Uninsured Patients

PDF formula calculations:

\[ P(0) = \frac{4!}{0!(4-0)!} (0.2)^0 (1-0.2)^{4-0} = 1 \times 0.2^0 \times 0.8^4 \]
\[ P(1) = \frac{4!}{1!(4-1)!} (0.2)^1 (1-0.2)^{4-1} = 4 \times 0.2^1 \times 0.8^3 \]
\[ P(2) = \frac{4!}{2!(4-2)!} (0.2)^2 (1-0.2)^{4-2} = 4 \times 0.2^2 \times 0.8^2 \]
\[ P(3) = \frac{4!}{3!(4-3)!} (0.2)^3 (1-0.2)^{4-3} = 4 \times 0.2^3 \times 0.8^1 \]
\[ P(4) = \frac{4!}{4!(4-4)!} (0.2)^4 (1-0.2)^{4-4} = 1 \times 0.2^4 \times 0.8^0 \]

Excel formula:

= BINOMDIST(0,4,0.2,0) = .4096
= BINOMDIST(1,4,0.2,0) = .4096
= BINOMDIST(2,4,0.2,0) = .1536
= BINOMDIST(3,4,0.2,0) = .0256
= BINOMDIST(4,4,0.2,0) = .0016

Binomial probabilities can also be determined by looking them up in a table (Appendix A) for selected values of \( n \) (row) and \( p \) (column).

<table>
<thead>
<tr>
<th>( n )</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.9801</td>
<td>0.9604</td>
<td>0.9025</td>
<td>0.7200</td>
<td>0.0000</td>
</tr>
<tr>
<td>1</td>
<td>0.0198</td>
<td>0.0392</td>
<td>0.0950</td>
<td>0.1800</td>
<td>0.0000</td>
</tr>
<tr>
<td>2</td>
<td>0.0000</td>
<td>0.0012</td>
<td>0.0071</td>
<td>0.0270</td>
<td>0.0000</td>
</tr>
<tr>
<td>3</td>
<td>0.0000</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0000</td>
</tr>
<tr>
<td>4</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Application: Uninsured Patients

Binomial probabilities can also be determined by looking them up in a table (Appendix A) for selected values of \( n \) (row) and \( p \) (column).
Individual probabilities can be added to obtain any desired event probability. For example, the probability that the sample of 4 patients will contain at least 2 uninsured patients is

\[ P(X \geq 2) = P(2) + P(3) + P(4) \]

\[ = .1536 + .0256 + .0016 = .1808 \]

HINT: What inequality means “at least?”

What is the probability that fewer than 2 patients have insurance?

\[ P(X < 2) = P(0) + P(1) \]

\[ = .4096 + .4096 = .8192 \]

HINT: What inequality means “fewer than?”

What is the probability that no more than 2 patients have insurance?

\[ P(X \leq 2) = P(0) + P(1) + P(2) \]

\[ = .4096 + .4096 + .1536 = .9728 \]

HINT: What inequality means “no more than?”

Binomial Distribution
Binomial Distribution

**Compound Events**

It is helpful to sketch a diagram:

- "At least two" events
- "Fewer than two" events
- "Fewer than 2 or more than 2" events

Using Software: Excel

- Use Excel’s Insert | Function menu to calculate the probability of $x = 67$ successes in $n = 1,024$ trials with probability $p = .048$.
- Or use =BINOMDIST(67,1024,0.048,0)
Binomial Distribution

Using Software: MegaStat

- Compute an entire binomial PDF for any \( n \) and (e.g., \( n = 10, \ p = 0.5 \)) in MegaStat.

Using Software: MegaStat

MegaStat also gives you the option to create a graph of the PDF:
### Binomial Distribution

#### Using Software: Visual Statistics

- Using *Visual Statistics* Module 4, here is a binomial distribution for $n = 10$ and $p = 0.50$:

Copy and paste graph as a bitmap. Copy and paste probabilities into Excel.

“Spin” $n$ and $\pi$ and superimpose a normal curve on the binomial distribution.

#### Using Software: LearningStats

Here, $n = 50$ and $p = 0.095$.

Spin buttons let you vary $n$ and $\pi$. 
Binomial Distribution

Binomial Random Data

- Generate a single binomial random number in Excel by summing \( n \) Bernoulli random variables (0 or 1) using the function \( = 0 + \text{INT}(1 \times \text{RAND}) \).
- Alternatively, use Excel's Tools | Data Analysis to get binomial random data.
- This will generate 20 binomial random data values using \( n = 4 \) and \( p = .20 \).

Binomial Distribution

Recognizing Binomial Applications

- Can you recognize a binomial situation? Look for \( n \) independent Bernoulli trials with constant probability of success.

In a sample of 20 friends:

- How many are left-handed?
- How many have ever worked on a factory floor?
- How many own a motorcycle?
Binomial Distribution

Recognizing Binomial Applications
- Can you recognize a binomial situation?
  In a sample of 50 cars in a parking lot:
  - How many are parked end-first?
  - How many are blue?
  - How many have hybrid engines?
- In a sample of 10 emergency patients with chest pain:
  - How many will be admitted?
  - How many will need bypass surgery?
  - How many will be uninsured?

Poisson Distribution

Poisson Processes
- The *Poisson distribution* was named for French mathematician Siméon Poisson (1781-1840).
- The Poisson distribution describes the number of occurrences within a randomly chosen unit of time or space.
- For example, within a minute, hour, day, square foot, or linear mile.
Poisson Distribution

- Called the *model of arrivals*, most Poisson applications model *arrivals per unit of time*.
- The events occur randomly and independently over a continuum of time or space:

```
One Unit     of Time     One Unit
\|\|\|\|\|\|\|\|\|
\|\|\|\|\|\|\|\|\|
•   •  ••      • •         •      ••••   •       • • •• •   • •••   • •

Flow of Time →
```

Each dot (•) is an occurrence of the event of interest.

Poisson Distribution

- Let $X =$ the number of events per unit of time.
- $X$ is a random variable that depends on when the unit of time is observed.
- For example, we could get $X = 3$ or $X = 1$ or $X = 5$ events, depending on where the randomly chosen unit of time happens to fall.

```
One Unit     of Time     One Unit
\|\|\|\|\|\|\|\|\|
\|\|\|\|\|\|\|\|\|
•   •   ••••   •       • • •• •   • •••   • •

Flow of Time →
```
Arrivals (e.g., customers, defects, accidents) must be independent of each other.

Some examples of Poisson models in which assumptions are sufficiently met are:

- $X$ = number of customers arriving at a bank ATM in a given minute.
- $X$ = number of file server virus infections at a data center during a 24-hour period.
- $X$ = number of blemishes per sheet of white bond paper.

The Poisson model’s only parameter is (Greek letter “lambda”).

represents the mean number of events per unit of time or space.

The unit of time should be short enough that the mean arrival rate is not large ($< 20$).

To make smaller, convert to a smaller time unit (e.g., convert hours to minutes).
The number of events that can occur in a given unit of time is not bounded, therefore $X$ has no obvious limit. However, Poisson probabilities taper off toward zero as $X$ increases.

The Poisson distribution is sometimes called the model of rare events.

### Poisson Distribution

#### Poisson Processes

- $\lambda$ = mean arrivals per unit of time or space

### Parameters

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDF</td>
<td>$P(x) = \frac{\lambda^x e^{-\lambda}}{x!}$</td>
</tr>
<tr>
<td>Range</td>
<td>$X = 0, 1, 2, \ldots$ (no obvious upper limit)</td>
</tr>
<tr>
<td>Mean</td>
<td>$\lambda$</td>
</tr>
<tr>
<td>St. Dev.</td>
<td>$\sqrt{\lambda}$</td>
</tr>
<tr>
<td>Random data</td>
<td>Use Excel’s Tools</td>
</tr>
<tr>
<td>Comments</td>
<td>Always right-skewed, but less so for larger $\lambda$.</td>
</tr>
</tbody>
</table>
Here are some Poisson PDFs.

<table>
<thead>
<tr>
<th>x</th>
<th>λ = 0.1</th>
<th>λ = 0.5</th>
<th>λ = 0.8</th>
<th>λ = 1.6</th>
<th>λ = 2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>.9048</td>
<td>.6065</td>
<td>.4493</td>
<td>.2019</td>
<td>.1353</td>
</tr>
<tr>
<td>1</td>
<td>.0905</td>
<td>.3033</td>
<td>.3595</td>
<td>.3230</td>
<td>.2707</td>
</tr>
<tr>
<td>2</td>
<td>.0045</td>
<td>.0758</td>
<td>.1438</td>
<td>.2584</td>
<td>.2707</td>
</tr>
<tr>
<td>3</td>
<td>.0002</td>
<td>.0126</td>
<td>.0383</td>
<td>.1378</td>
<td>.1804</td>
</tr>
<tr>
<td>4</td>
<td>--</td>
<td>.0016</td>
<td>.0077</td>
<td>.0551</td>
<td>.0902</td>
</tr>
<tr>
<td>5</td>
<td>--</td>
<td>.0002</td>
<td>.0012</td>
<td>.0176</td>
<td>.0361</td>
</tr>
<tr>
<td>6</td>
<td>--</td>
<td>--</td>
<td>.0002</td>
<td>.0047</td>
<td>.0120</td>
</tr>
<tr>
<td>7</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>.0011</td>
<td>.0034</td>
</tr>
<tr>
<td>8</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>.0002</td>
<td>.0009</td>
</tr>
<tr>
<td>9</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>.0002</td>
</tr>
<tr>
<td>Sum</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

Poisson processes are always right-skewed but become less skewed and more bell-shaped as \( \lambda \) increases.
Example: Credit Union Customers

- On Thursday morning between 9 A.M. and 10 A.M., customers arrive and enter the queue at the Oxnard University Credit Union at a mean rate of 1.7 customers per minute.

- Find the PDF, mean and standard deviation:

  \[ P(x) = \frac{\lambda^x e^{-\lambda}}{x!} = \frac{(1.7)^x e^{-1.7}}{x!} \]

  Mean = \( \lambda = 1.7 \) customers per minute.

  Standard deviation = \( \sqrt{\lambda} = 1.304 \) cust/min

Here is the Poisson probability distribution for \( \lambda = 1.7 \) customers per minute on average.

- Note that \( x \) represents the number of customers.

  For example, \( P(X=4) \) is the probability that there are exactly 4 customers in the bank.
### Poisson Distribution

#### Using the Poisson Formula

<table>
<thead>
<tr>
<th>Formula</th>
<th>Excel function</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(0) = \frac{1.7^0 e^{-1.7}}{0!} = .1827$</td>
<td>=POISSON(0,1.7,0)</td>
</tr>
<tr>
<td>$P(1) = \frac{1.7^1 e^{-1.7}}{1!} = .3106$</td>
<td>=POISSON(1,1.7,0)</td>
</tr>
<tr>
<td>$P(2) = \frac{1.7^2 e^{-1.7}}{2!} = .2640$</td>
<td>=POISSON(2,1.7,0)</td>
</tr>
<tr>
<td>$P(3) = \frac{1.7^3 e^{-1.7}}{3!} = .1496$</td>
<td>=POISSON(3,1.7,0)</td>
</tr>
<tr>
<td>$P(4) = \frac{1.7^4 e^{-1.7}}{4!} = .0636$</td>
<td>=POISSON(4,1.7,0)</td>
</tr>
</tbody>
</table>

These probabilities can be calculated using a calculator or Excel.

---

### Poisson Distribution

#### Using the Poisson Formula

<table>
<thead>
<tr>
<th>Formula</th>
<th>Excel function</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(5) = \frac{1.7^5 e^{-1.7}}{5!} = .0216$</td>
<td>=POISSON(5,1.7,0)</td>
</tr>
<tr>
<td>$P(6) = \frac{1.7^6 e^{-1.7}}{6!} = .0061$</td>
<td>=POISSON(6,1.7,0)</td>
</tr>
<tr>
<td>$P(7) = \frac{1.7^7 e^{-1.7}}{7!} = .0015$</td>
<td>=POISSON(7,1.7,0)</td>
</tr>
<tr>
<td>$P(8) = \frac{1.7^8 e^{-1.7}}{8!} = .0003$</td>
<td>=POISSON(8,1.7,0)</td>
</tr>
<tr>
<td>$P(9) = \frac{1.7^9 e^{-1.7}}{9!} = .0001$</td>
<td>=POISSON(9,1.7,0)</td>
</tr>
</tbody>
</table>

Beyond $X = 9$, the probabilities are below .0001.
Here are the graphs of the distributions:

- Poisson PDF for $\lambda = 1.7$
- Poisson CDF for $\lambda = 1.7$

- The most likely event is 1 arrival ($P(1) = .3106$ or 31.1% chance).
- This will help the credit union schedule tellers.

Compound Events

- Cumulative probabilities can be evaluated by summing individual $X$ probabilities.

What is the probability that two or fewer customers will arrive in a given minute?

$$P(X \leq 2) = P(0) + P(1) + P(2)$$

$$= .1827 + .3106 + .2640 = .7573$$
What is the probability of at least three customers (the complimentary event)?

\[ P(X \geq 3) = 1 - P(X \leq 2) = 1 - 0.7573 = 0.2427 \]

You can also use Excel’s function =POISSON(2,1.7,1) to obtain this probability.

Poisson Distribution

Compound Events

- You can also use Excel’s function =POISSON(2,1.7,1) to obtain this probability.
- What is the probability of at least three customers (the complimentary event)?

\[ P(X \geq 3) = 1 - P(X \leq 2) = 1 - 0.7573 = 0.2427 \]

Poisson Distribution

Poisson Probabilities: Tables (Appendix B)

<table>
<thead>
<tr>
<th>X</th>
<th>1.6</th>
<th>1.7</th>
<th>1.8</th>
<th>1.9</th>
<th>2.0</th>
<th>2.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.2019</td>
<td>0.1827</td>
<td>0.1653</td>
<td>0.1496</td>
<td>0.1353</td>
<td>0.1225</td>
</tr>
<tr>
<td>1</td>
<td>0.3230</td>
<td>0.3106</td>
<td>0.2975</td>
<td>0.2842</td>
<td>0.2707</td>
<td>0.2572</td>
</tr>
<tr>
<td>2</td>
<td>0.2584</td>
<td>0.2640</td>
<td>0.2678</td>
<td>0.2700</td>
<td>0.2707</td>
<td>0.2700</td>
</tr>
<tr>
<td>3</td>
<td>0.1378</td>
<td>0.1496</td>
<td>0.1607</td>
<td>0.1710</td>
<td>0.1804</td>
<td>0.1890</td>
</tr>
<tr>
<td>4</td>
<td>0.0551</td>
<td>0.0636</td>
<td>0.0723</td>
<td>0.0812</td>
<td>0.0902</td>
<td>0.0992</td>
</tr>
<tr>
<td>5</td>
<td>0.0176</td>
<td>0.0216</td>
<td>0.0260</td>
<td>0.0309</td>
<td>0.0361</td>
<td>0.0417</td>
</tr>
<tr>
<td>6</td>
<td>0.0047</td>
<td>0.0061</td>
<td>0.0078</td>
<td>0.0098</td>
<td>0.0120</td>
<td>0.0146</td>
</tr>
<tr>
<td>7</td>
<td>0.0011</td>
<td>0.0015</td>
<td>0.0020</td>
<td>0.0027</td>
<td>0.0034</td>
<td>0.0044</td>
</tr>
<tr>
<td>8</td>
<td>0.0002</td>
<td>0.0003</td>
<td>0.0005</td>
<td>0.0006</td>
<td>0.0009</td>
<td>0.0011</td>
</tr>
<tr>
<td>9</td>
<td>--</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0002</td>
<td>0.0003</td>
</tr>
<tr>
<td>10</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>0.0001</td>
</tr>
<tr>
<td>11</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

Appendix B facilitates Poisson calculations but doesn't go beyond \( \lambda = 20 \).
Poisson Distribution

**Using Software: Excel**

Excel's menus for calculating Poisson probabilities

The resulting probabilities are more accurate than those from Appendix B.

---

Poisson Distribution

**Using Software: Visual Statistics**

Module 4\( (\lambda = 1.7)\)

Copy and paste the graph as a bitmap; copy and paste the probabilities into Excel.

“Spin” \(\lambda\) and overlay a normal curve.
Poisson Distribution

**Recognizing Poisson Applications**

- Can you recognize a Poisson situation?
- Look for arrivals of “rare” independent events with *no obvious upper limit*.
  - In the last week, how many credit card applications did you receive by mail?
  - In the last week, how many checks did you write?
  - In the last week, how many e-mail viruses did your firewall detect?

**Poisson Approximation to Binomial**

- The Poisson distribution may be used to approximate a binomial by setting \( l = n \).
- This approximation is helpful when \( n \) is large and Excel is not available.
- For example, suppose \( n = 1,000 \) women are screened for a rare type of cancer.
- This cancer has a nationwide incidence of 6 cases per 10,000. What is \( p \)?
  - \( p = 6/10,000 = .0006 \)
- This is a binomial distribution with \( n = 1,000 \) and \( p = .0006 \).
Poisson Distribution

Poisson Approximation to Binomial

- Since the binomial formula involves factorials (which are cumbersome as \( n \) increases), use the Poisson distribution as an approximation:

- Set \( \lambda = n p \) = (1000)(.0006) = .6

- Now use Appendix B or the Poisson PDF to calculate the probability of \( x \) successes. For example:

\[
P(X \leq 2) = P(0) + P(1) + P(2) = .5488 + .3293 + .0988 = .9769
\]

Here is a comparison of Binomial probabilities and the respective Poisson approximations.

<table>
<thead>
<tr>
<th>Poisson approximation:</th>
<th>Actual Binomial probability:</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P(0) = .6^0 e^{-0.6} / 0! = .5488 )</td>
<td>( P(2) = \frac{1000!}{2!(1000-2)!} \cdot 0.0006^2 (1-.0006)^{1000-2} = .5487 )</td>
</tr>
<tr>
<td>( P(1) = .6^1 e^{-0.6} / 1! = .3293 )</td>
<td>( P(1) = \frac{1000!}{1!(1000-1)!} \cdot 0.0006 (1-.0006)^{1000-1} = .3294 )</td>
</tr>
<tr>
<td>( P(2) = .6^2 e^{-0.6} / 2! = .0988 )</td>
<td>( P(0) = \frac{1000!}{0!(1000-0)!} \cdot 0.0006^0 (1-.0006)^{1000-0} = .0988 )</td>
</tr>
</tbody>
</table>

Rule of thumb: the approximation is adequate if \( n \geq 20 \) and \( p \leq .05 \).
The hypergeometric distribution is similar to the binomial distribution.

However, unlike the binomial, sampling is **without replacement** from a finite population of \( N \) items.

The hypergeometric distribution may be skewed right or left and is symmetric only if the proportion of successes in the population is 50%.

### Hypergeometric Distribution

**Characteristics of the Hypergeometric Dist.**

- The *hypergeometric distribution* is similar to the binomial distribution.
- However, unlike the binomial, sampling is **without replacement** from a finite population of \( N \) items.
- The hypergeometric distribution may be skewed right or left and is symmetric only if the proportion of successes in the population is 50%.

### Parameters

- \( N \): number of items in the population
- \( n \): sample size
- \( s \): number of "successes" in population

### PDF

\[
P(x) = \frac{s \binom{N-s}{n-x} \binom{s}{x}}{\binom{N}{n}}
\]

### Range

\( X = \max(0, n-N+s) \leq x \leq \min(s, n) \)

### Mean

\( m = np \) where \( p = s/N \)

### St. Dev.

\[
\sqrt{n \pi (1-\pi) \left( \frac{N-n}{N-1} \right)}
\]

### Comments

Similar to binomial, but sampling is without replacement from a finite population. Can be approximated by binomial with \( \pi = s/N \) if \( n/N < 0.05 \) (i.e., less than 5% sample).
Hypergeometric Distribution

Characteristics of the Hypergeometric Dist.
The hypergeometric PDF uses the formula for combinations:

\[ P(x) = \frac{s \binom{x}{s} \cdot \left(\frac{N-s}{N-\binom{x}{s}}\right)}{\binom{N}{n}} \]

where

- \( s \binom{x}{s} = \) the number of ways to choose \( x \) successes from \( s \) successes in the population
- \( \binom{N-s}{n-x} = \) the number of ways to choose \( n-x \) failures from \( N-s \) failures in the population
- \( \binom{N}{n} = \) the number of ways to choose \( n \) items from \( N \) items in the population

Example: Damaged iPods

- In a shipment of 10 iPods, 2 were damaged and 8 are good.
- The receiving department at Best Buy tests a sample of 3 iPods at random to see if they are defective.
- Let the random variable \( X \) be the number of damaged iPods in the sample.
Hypergeometric Distribution

Example: Damaged iPods

Now describe the problem:

- \( N = 10 \) (number of iPods in the shipment)
- \( n = 3 \) (sample size drawn from the shipment)
- \( s = 2 \) (number of damaged iPods in the shipment (“successes” in population))
- \( N-s = 8 \) (number of non-damaged iPods in the shipment)
- \( x \) = number of damaged iPods in the sample (“successes” in sample)
- \( n-x \) = number of non-damaged iPods in the sample

Hypergeometric Distribution

Example: Damaged iPods

- This is not a binomial problem because is not constant.
- What is the probability of getting a damaged iPod on the first draw from the sample?

\[
p_1 = \frac{2}{10}
\]

- Now, what is the probability of getting a damaged iPod on the second draw?

\[
p_2 = \frac{1}{9} \text{ (if the first iPod was damaged)} \quad \text{or} \quad \frac{2}{9} \text{ (if the first iPod was undamaged)}
\]
Hypergeometric Distribution

Example: Damaged iPods

- What about on the third draw?
  \[ p_3 = \frac{0}{8} \text{ or } \frac{1}{8} \text{ or } \frac{2}{8} \]
  depending on what happened in the first two draws.

### Using the Hypergeometric Formula

Since there are only 2 damaged iPods in the population, the only possible values of \( x \) are 0, 1, and 2. Here are the probabilities:

<table>
<thead>
<tr>
<th>PDF Formula</th>
<th>Excel Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P(0) = \frac{2C_0 \cdot 8C_3}{10C_3} = \left( \frac{21}{10} \right) \left( \frac{81}{370} \right) = \frac{56}{120} \approx 0.4667 )</td>
<td>=HYPGEOMDIST(0,3,2,10)</td>
</tr>
<tr>
<td>( P(1) = \frac{2C_1 \cdot 8C_2}{10C_3} = \left( \frac{21}{10} \right) \left( \frac{81}{370} \right) = \frac{56}{120} \approx 0.4667 )</td>
<td>=HYPGEOMDIST(1,3,2,10)</td>
</tr>
<tr>
<td>( P(2) = \frac{2C_2 \cdot 8C_1}{10C_3} = \left( \frac{21}{10} \right) \left( \frac{81}{370} \right) = \frac{8}{120} \approx 0.0667 )</td>
<td>=HYPGEOMDIST(2,3,2,10)</td>
</tr>
</tbody>
</table>
Hypergeometric Distribution

Using Software: Excel

Since the hypergeometric formula and tables are tedious and impractical, use Excel’s hypergeometric function to find probabilities.

Using Software: Visual Statistics

Module 4, the probabilities are given below the graph.

Copy and paste graph as a bitmap; copy and paste probabilities into Excel.

“Spin” \( N \) and \( n \); overlay a normal or binomial curve.
Hypergeometric Distribution

Using Software: LearningStats

LearningStats allows you to spin the values of $N$, $n$, and $s$ to get the desired probability.

Figure 6.29

Hypergeometric Distribution

Using Software: LearningStats

- Look for a finite population $(N)$ containing a known number of successes $(s)$ and sampling without replacement $(n$ items in the sample).

- Out of 40 cars are inspected for California emissions compliance, 32 are compliant but 8 are not. A sample of 7 cars is chosen at random. What is the probability that all are compliant? At least 5?
Hypergeometric Distribution

Using Software: LearningStats

- Out of 500 background checks for firearms purchasers, 50 applicants are convicted felons. Through a computer error, 10 applicants are approved without a background check. What is the probability that none is a felon? At least 2?
- Out of 40 blood specimens checked for HIV, 8 actually contain HIV. A worker is accidentally exposed to 5 specimens. What is the probability that none contained HIV?

Hypergeometric Distribution

Binomial Approximation to the Hypergeometric

- Both the binomial and hypergeometric involve samples of size $n$ and treat $X$ as the number of successes.
- The binomial samples with replacement while the hypergeometric samples without replacement.

Rule of Thumb

If $n/N < 0.05$ it is safe to use the binomial approximation to the hypergeometric, using sample size $n$ and success probability $\pi = s/N$. 
For example, suppose we want $P(X=6)$ for a hypergeometric with $N = 400$, $n = 10$, $s = 200$.

$n/N = 10/400 = 0.025 < .05$ so the binomial approximation is acceptable.

Set $s/N = 200/400 = 0.50$ and use Appendix A to obtain the probability.

$P(X=6) = 0.2051$

The geometric distribution describes the number of Bernoulli trials until the first success. $X$ is the number of trials until the first success. $X$ ranges from $\{1, 2, \ldots\}$ since we must have at least one trial to obtain the first success. However, the number of trials is not fixed. $p$ is the constant probability of a success on each trial.
The geometric distribution is always skewed to the right.

Parameters

<table>
<thead>
<tr>
<th>Parameters</th>
<th>= probability of success</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDF</td>
<td>$P(x) = \pi(1-\pi)^{x-1}$</td>
</tr>
<tr>
<td>Range</td>
<td>$X = 1, 2, ...$</td>
</tr>
<tr>
<td>Mean</td>
<td>$1/\pi$</td>
</tr>
<tr>
<td>St. Dev.</td>
<td>$\sqrt{\frac{1-\pi}{\pi^2}}$</td>
</tr>
<tr>
<td>Comments</td>
<td>Describes the number of trials before the first success. Highly skewed.</td>
</tr>
</tbody>
</table>

The mean and standard deviation are nearly the same when $\pi$ is small.

---

**Example: Telefund Calling**

- At Faber University, 15% of the alumni make a donation or pledge during the annual telefund.
- What is the probability that the first donation will not come until the 7th call?
  - What is $\pi$ = .15
  - The PDF is: $P(x) = (1-\pi)^{x-1}$

  $P(7) = .15(1-.15)^{7-1} = .15(.85)^6 = .0566$
What are the mean and standard deviation of this distribution?

\[ \mu = \frac{1}{p} = \frac{1}{0.15} = 6.67 \text{ calls} \]

So, we would expect to call between 6 and 7 alumni until the first donation.

\[ \sigma = \sqrt{\frac{1 - \pi}{\pi^2}} = \sqrt{\frac{1 - 0.15}{(0.15)^2}} = 6.15 \]

The large standard deviation indicates that we should not regard the mean as a good prediction of how many trials are needed.

**Using Software: LearningStats**

*LearningStats* gives both the graph and numeric probabilities of the distribution.
Discrete Distributions Compared

- Bernoulli: $n > 1$, $X = 0, 1$
- Geometric: $n = 20$, $X = 1, 2, ...$
- Binomial: $n, \pi$, $X = 0, 1, ..., n$
- Poisson: $\lambda$, $X = 0, 1, ...$
- Hypergeometric: $N, n, s$, $\text{max}(0, n - N + s) \leq X \leq \text{min}(s, n)$

Transformations of Random Variables

Linear Transformations

- A linear transformation of a random variable $X$ is performed by adding a constant or multiplying by a constant.

**Rule 1:** $aX + b = a \bar{X} + b$ (mean of a transformed variable)

**Rule 2:** $aX + b = a \sigma$ (standard deviation of a transformed variable)
Transformations of Random Variables

Linear Transformations

**Rule 3:** \( x + y = \mu_x + \mu_y \) (mean of a two random variables X and Y)

**Rule 4:** \( s_{x+y} = \sqrt{\sigma_x^2 + \sigma_y^2} \) (standard deviation of sum if X and Y are independent)

Applied Statistics in Business & Economics

End of Chapter 6