3. Bivariate Random Variables

Dave Goldsman

H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

3/2/20
In this introductory lesson, we’ll cover …

- What we mean by bivariate (or joint) random variables.
- The discrete case.
- The continuous case.
- Bivariate cdf’s.
In this introductory lesson, we’ll cover …

- What we mean by bivariate (or joint) random variables.
- The discrete case.
- The continuous case.
- Bivariate cdf’s.
Lesson 3.1 — Introduction

In this introductory lesson, we’ll cover . . .

- What we mean by bivariate (or joint) random variables.
- The discrete case.
- The continuous case.
- Bivariate cdf’s.

In this module, we’ll look at what happens when you consider two random variables simultaneously.
In this introductory lesson, we’ll cover …

- What we mean by bivariate (or joint) random variables.
- The discrete case.
- The continuous case.
- Bivariate cdf’s.

In this module, we’ll look at what happens when you consider two random variables simultaneously.

Example: Choose a person at random. Look at their height and weight \((X, Y)\). Obviously, \(X\) and \(Y\) will be related somehow.
Discrete Case
Discrete Case

**Definition**: If $X$ and $Y$ are discrete random variables, then $(X, Y)$ is called a **jointly discrete bivariate random variable**.
Discrete Case

**Definition**: If $X$ and $Y$ are discrete random variables, then $(X, Y)$ is called a jointly discrete bivariate random variable.

The **joint (or bivariate) pmf** is

$$f(x, y) = P(X = x, Y = y), \quad \forall x, y.$$
Discrete Case

**Definition:** If $X$ and $Y$ are discrete random variables, then $(X, Y)$ is called a **jointly discrete bivariate random variable**.

The **joint (or bivariate) pmf** is

$$f(x, y) = P(X = x, Y = y), \quad \forall x, y.$$  

**Properties:**

- $0 \leq f(x, y) \leq 1$. 

Discrete Case

**Definition:** If $X$ and $Y$ are discrete random variables, then $(X, Y)$ is called a jointly discrete bivariate random variable.

The **joint (or bivariate) pmf** is

$$f(x, y) = P(X = x, Y = y), \quad \forall x, y.$$  

**Properties:**

- $0 \leq f(x, y) \leq 1$.
- $\sum_x \sum_y f(x, y) = 1$. 

Discrete Case

**Definition**: If $X$ and $Y$ are discrete random variables, then $(X, Y)$ is called a jointly discrete bivariate random variable.

The **joint (or bivariate) pmf** is

$$f(x, y) = P(X = x, Y = y), \quad \forall x, y.$$  

**Properties**:

- $0 \leq f(x, y) \leq 1$.
- $\sum_x \sum_y f(x, y) = 1$.
- $A \subseteq \mathbb{R}^2 \Rightarrow P((X, Y) \in A) = \sum \sum (x, y) \in A f(x, y).$
Example: 3 sox in a box (numbered 1,2,3). Draw 2 sox at random without replacement. \( X = \# \) of the first sock; \( Y = \# \) of the second sock. The joint pmf \( f(x, y) \) is

\[
\begin{array}{ccc}
X & Y = 1 & Y = 2 \\
1 & 1/6 & 0 \\
2 & 0 & 1/6 \\
3 & 1/6 & 0 \\
\end{array}
\]

\[
\begin{array}{ccc}
X = 1 & X = 2 & X = 3 \\
0 & 1/3 & 0 \\
1/3 & 0 & 1/3 \\
1/3 & 1/3 & 0 \\
\end{array}
\]
**Example**: 3 sox in a box (numbered 1, 2, 3). Draw 2 sox at random without replacement. $X = \#$ of the first sock; $Y = \#$ of the second sock. The joint pmf $f(x, y)$ is

<table>
<thead>
<tr>
<th>$f(x, y)$</th>
<th>$X = 1$</th>
<th>$X = 2$</th>
<th>$X = 3$</th>
<th>$P(Y = y)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X = 1$</td>
<td>1/6</td>
<td>0</td>
<td>1/6</td>
<td>0</td>
</tr>
<tr>
<td>$X = 2$</td>
<td>1/3</td>
<td>1/3</td>
<td>0</td>
<td>1/3</td>
</tr>
<tr>
<td>$X = 3$</td>
<td>0</td>
<td>1/6</td>
<td>1/6</td>
<td>1/3</td>
</tr>
</tbody>
</table>
**Example**: 3 sox in a box (numbered 1,2,3). Draw 2 sox at random without replacement. $X = \# \text{ of the first sock}; \ Y = \# \text{ of the second sock}$. The joint pmf $f(x, y)$ is

<table>
<thead>
<tr>
<th>$f(x, y)$</th>
<th>$X = 1$</th>
<th>$X = 2$</th>
<th>$X = 3$</th>
<th>$P(Y = y)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y = 1$</td>
<td>0</td>
<td>1/6</td>
<td>1/6</td>
<td>1/3</td>
</tr>
</tbody>
</table>

$f_x(x) \equiv P(X = x)$ is the "marginal" pmf of $X$. $f_y(y) \equiv P(Y = y)$ is the "marginal" pmf of $Y$. 
**Example:** 3 sox in a box (numbered 1,2,3). Draw 2 sox at random without replacement. $X = \#$ of the first sock; $Y = \#$ of the second sock. The joint pmf $f(x, y)$ is

<table>
<thead>
<tr>
<th>$f(x, y)$</th>
<th>$X = 1$</th>
<th>$X = 2$</th>
<th>$X = 3$</th>
<th>$P(Y = y)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y = 1$</td>
<td>0</td>
<td>1/6</td>
<td>1/6</td>
<td>1/3</td>
</tr>
<tr>
<td>$Y = 2$</td>
<td>1/6</td>
<td>0</td>
<td>1/6</td>
<td>1/3</td>
</tr>
</tbody>
</table>

$f_X(x) \equiv P(X = x)$ is the “marginal” pmf of $X$. $f_Y(y) \equiv P(Y = y)$ is the “marginal” pmf of $Y$. 
**Example**: 3 sox in a box (numbered 1,2,3). Draw 2 sox at random without replacement. \( X = \# \) of the first sock; \( Y = \# \) of the second sock. The joint pmf \( f(x, y) \) is

<table>
<thead>
<tr>
<th>( f(x, y) )</th>
<th>( X = 1 )</th>
<th>( X = 2 )</th>
<th>( X = 3 )</th>
<th>( P(Y = y) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Y = 1 )</td>
<td>0</td>
<td>1/6</td>
<td>1/6</td>
<td>1/3</td>
</tr>
<tr>
<td>( Y = 2 )</td>
<td>1/6</td>
<td>0</td>
<td>1/6</td>
<td>1/3</td>
</tr>
<tr>
<td>( Y = 3 )</td>
<td>1/6</td>
<td>1/6</td>
<td>0</td>
<td>1/3</td>
</tr>
</tbody>
</table>
**Example:** 3 sox in a box (numbered 1,2,3). Draw 2 sox at random without replacement. \( X = \# \) of the first sock; \( Y = \# \) of the second sock. The joint pmf \( f(x, y) \) is

<table>
<thead>
<tr>
<th>( f(x, y) )</th>
<th>( X = 1 )</th>
<th>( X = 2 )</th>
<th>( X = 3 )</th>
<th>( P(Y = y) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Y = 1 )</td>
<td>0</td>
<td>1/6</td>
<td>1/6</td>
<td>1/3</td>
</tr>
<tr>
<td>( Y = 2 )</td>
<td>1/6</td>
<td>0</td>
<td>1/6</td>
<td>1/3</td>
</tr>
<tr>
<td>( Y = 3 )</td>
<td>1/6</td>
<td>1/6</td>
<td>0</td>
<td>1/3</td>
</tr>
</tbody>
</table>

\( P(X = x) \):

| \( P(X = x) \) | 1/3 | 1/3 | 1/3 | 1   |
**Example**: 3 sox in a box (numbered 1,2,3). Draw 2 sox at random without replacement. $X = \#$ of the first sock; $Y = \#$ of the second sock. The joint pmf $f(x,y)$ is

<table>
<thead>
<tr>
<th>$f(x,y)$</th>
<th>$X = 1$</th>
<th>$X = 2$</th>
<th>$X = 3$</th>
<th>$P(Y = y)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y = 1$</td>
<td>0</td>
<td>1/6</td>
<td>1/6</td>
<td>1/3</td>
</tr>
<tr>
<td>$Y = 2$</td>
<td>1/6</td>
<td>0</td>
<td>1/6</td>
<td>1/3</td>
</tr>
<tr>
<td>$Y = 3$</td>
<td>1/6</td>
<td>1/6</td>
<td>0</td>
<td>1/3</td>
</tr>
<tr>
<td>$P(X = x)$</td>
<td>1/3</td>
<td>1/3</td>
<td>1/3</td>
<td>1</td>
</tr>
</tbody>
</table>

$f_X(x) \equiv P(X = x)$ is the "marginal" pmf of $X$. 
**Example:** 3 sox in a box (numbered 1, 2, 3). Draw 2 sox at random without replacement. $X = \#$ of the first sock; $Y = \#$ of the second sock. The joint pmf $f(x, y)$ is

<table>
<thead>
<tr>
<th>$f(x, y)$</th>
<th>$X = 1$</th>
<th>$X = 2$</th>
<th>$X = 3$</th>
<th>$P(Y = y)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y = 1$</td>
<td>0</td>
<td>1/6</td>
<td>1/6</td>
<td>1/3</td>
</tr>
<tr>
<td>$Y = 2$</td>
<td>1/6</td>
<td>0</td>
<td>1/6</td>
<td>1/3</td>
</tr>
<tr>
<td>$Y = 3$</td>
<td>1/6</td>
<td>1/6</td>
<td>0</td>
<td>1/3</td>
</tr>
</tbody>
</table>

| $P(X = x)$ | 1/3 | 1/3 | 1/3 | 1 |

$f_{X}(x) \equiv P(X = x)$ is the “marginal” pmf of $X$.

$f_{Y}(y) \equiv P(Y = y)$ is the “marginal” pmf of $Y$. 
By the Law of Total Probability,
By the Law of Total Probability,

\[ P(X = 1) \]
By the Law of Total Probability,

\[ P(X = 1) = \sum_{y=1}^{3} P(X = 1, Y = y) \]
By the Law of Total Probability,

\[ P(X = 1) = \sum_{y=1}^{3} P(X = 1, Y = y) = 1/3. \]
By the Law of Total Probability,

\[ P(X = 1) = \sum_{y=1}^{3} P(X = 1, Y = y) = \frac{1}{3}. \]

In addition,

\[ P(X \geq 2, Y \geq 2) = 0 + \frac{1}{6} + \frac{1}{6} + 0 = \frac{1}{3}. \]
By the Law of Total Probability,

\[ P(X = 1) = \sum_{y=1}^{3} P(X = 1, Y = y) = 1/3. \]

In addition,

\[
P(X \geq 2, Y \geq 2) = \sum \sum f(x, y) = 0 + 1/6 + 1/6 + 0 = 1/3.
\]
By the Law of Total Probability,

\[ P(X = 1) = \sum_{y=1}^{3} P(X = 1, Y = y) = \frac{1}{3}. \]

In addition,

\[ P(X \geq 2, Y \geq 2) = \sum_{x \geq 2} \sum_{y \geq 2} f(x, y) = f(2, 2) + f(2, 3) + f(3, 2) + f(3, 3) = 0 + \frac{1}{6} + \frac{1}{6} + 0 = \frac{1}{3}. \]
By the Law of Total Probability,

\[ P(X = 1) = \sum_{y=1}^{3} P(X = 1, Y = y) = \frac{1}{3}. \]

In addition,

\[
P(X \geq 2, Y \geq 2)
= \sum_{x \geq 2} \sum_{y \geq 2} f(x, y)
= f(2, 2) + f(2, 3) + f(3, 2) + f(3, 3)
= 0 + \frac{1}{6} + \frac{1}{6} + 0 = \frac{1}{3}. \]
Continuous Case

Definition: If $X$ and $Y$ are continuous RVs, then $(X,Y)$ is a jointly continuous bivariate RV if there exists a magic function $f(x,y)$ such that $f(x,y) \geq 0$, $\forall x,y$. 

$$\int\int_{\mathbb{R}^2} f(x,y) \, dx\,dy = 1.$$ 

$P(A) = P((X,Y) \in A) = \int\int_{A} f(x,y) \, dx\,dy$. 

In this case, $f(x,y)$ is called the joint pdf. 

If $A \subseteq \mathbb{R}^2$, then $P(A)$ is the volume between $f(x,y)$ and $A$. 

Think of $f(x,y) \, dx\,dy \approx P(x < X < x + dx, y < Y < y + dy)$.

It's easy to see how this generalizes the 1-dimensional pdf, $f(x)$. 

ISYE 6739 — Goldsman
Continuous Case

**Definition:** If $X$ and $Y$ are continuous RVs, then $(X, Y)$ is a *jointly continuous bivariate RV* if there exists a magic function $f(x, y)$ such that

\[
\int \int_{\mathbb{R}^2} f(x, y) \, dx \, dy = 1
\]

**In this case,** $f(x, y)$ is called the **joint pdf**.

If $A \subseteq \mathbb{R}^2$, then $P(A)$ is the volume between $f(x, y)$ and $A$.

Think of $f(x, y) \, dx \, dy \approx P(x < X < x + dx, y < Y < y + dy)$.

It's easy to see how this generalizes the 1-dimensional pdf, $f(x)$. 
Continuous Case

**Definition:** If $X$ and $Y$ are continuous RVs, then $(X, Y)$ is a jointly continuous bivariate RV if there exists a magic function $f(x, y)$ such that

- $f(x, y) \geq 0, \forall x, y.$
Continuous Case

**Definition:** If $X$ and $Y$ are continuous RVs, then $(X, Y)$ is a **jointly continuous bivariate RV** if there exists a magic function $f(x, y)$ such that

- $f(x, y) \geq 0$, $\forall x, y$.
- $\int \int_{\mathbb{R}^2} f(x, y) \, dx \, dy = 1$. 


Continuous Case

**Definition:** If $X$ and $Y$ are continuous RVs, then $(X, Y)$ is a **jointly continuous bivariate RV** if there exists a magic function $f(x, y)$ such that

- $f(x, y) \geq 0, \forall x, y.$
- $\int \int_{\mathbb{R}^2} f(x, y) \, dx \, dy = 1.$
- $P(A) = P((X, Y) \in A) = \int \int_{A} f(x, y) \, dx \, dy.$
Continuous Case

**Definition:** If $X$ and $Y$ are continuous RVs, then $(X, Y)$ is a **jointly continuous bivariate RV** if there exists a magic function $f(x, y)$ such that

- $f(x, y) \geq 0, \forall x, y.$
- $\int \int_{\mathbb{R}^2} f(x, y) \, dx \, dy = 1.$
- $P(A) = P((X, Y) \in A) = \int \int_A f(x, y) \, dx \, dy.$

In this case, $f(x, y)$ is called the **joint pdf**.
Continuous Case

**Definition**: If $X$ and $Y$ are continuous RVs, then $(X, Y)$ is a **jointly continuous bivariate RV** if there exists a magic function $f(x, y)$ such that

- $f(x, y) \geq 0$, $\forall x, y$.
- $\int \int_{\mathbb{R}^2} f(x, y) \, dx \, dy = 1$.
- $P(A) = P((X, Y) \in A) = \int \int_A f(x, y) \, dx \, dy$.

In this case, $f(x, y)$ is called the **joint pdf**.

If $A \subseteq \mathbb{R}^2$, then $P(A)$ is the volume between $f(x, y)$ and $A$. 
Continuous Case

**Definition**: If $X$ and $Y$ are continuous RVs, then $(X, Y)$ is a **jointly continuous bivariate RV** if there exists a magic function $f(x, y)$ such that

1. $f(x, y) \geq 0$, $\forall x, y$.
2. $\int \int_{\mathbb{R}^2} f(x, y) \, dx \, dy = 1$.
3. $P(A) = P((X, Y) \in A) = \int \int_A f(x, y) \, dx \, dy$.

In this case, $f(x, y)$ is called the **joint pdf**.

If $A \subseteq \mathbb{R}^2$, then $P(A)$ is the volume between $f(x, y)$ and $A$.

Think of

$$f(x, y) \, dx \, dy \approx P(x < X < x + dx, y < Y < y + dy).$$
Continuous Case

**Definition:** If $X$ and $Y$ are continuous RVs, then $(X, Y)$ is a **jointly continuous bivariate RV** if there exists a magic function $f(x, y)$ such that

- $f(x, y) \geq 0, \forall x, y.$
- $\int \int_{\mathbb{R}^2} f(x, y) \, dx \, dy = 1.$
- $P(A) = P((X, Y) \in A) = \int \int_{A} f(x, y) \, dx \, dy.$

In this case, $f(x, y)$ is called the **joint pdf**.

If $A \subseteq \mathbb{R}^2$, then $P(A)$ is the volume between $f(x, y)$ and $A$.

Think of

$$f(x, y) \, dx \, dy \approx P(x < X < x + dx, y < Y < y + dy).$$

It’s easy to see how this generalizes the 1-dimensional pdf, $f(x)$. 
**Example**: Choose a point \((X, Y)\) at random in the interior of the circle inscribed in the unit square, e.g., \(C \equiv (x - \frac{1}{2})^2 + (y - \frac{1}{2})^2 \leq \frac{1}{4}\).

Find the pdf of \((X, Y)\).
**Example:** Choose a point \((X, Y)\) at random in the interior of the circle inscribed in the unit square, e.g., \(C \equiv (x - \frac{1}{2})^2 + (y - \frac{1}{2})^2 \leq \frac{1}{4}\).

Find the pdf of \((X, Y)\).

Since the area of the circle is \(\pi/4\),

\[
f(x, y) = \begin{cases} 
4/\pi & \text{if } (x, y) \in C \\
0 & \text{otherwise.}
\end{cases}
\]
**Example:** Choose a point \((X, Y)\) at random in the interior of the circle inscribed in the unit square, e.g., \(C \equiv (x - \frac{1}{2})^2 + (y - \frac{1}{2})^2 \leq \frac{1}{4}\).

Find the pdf of \((X, Y)\).

Since the area of the circle is \(\pi/4\),

\[
f(x, y) = \begin{cases} 
4/\pi & \text{if } (x, y) \in C \\
0 & \text{otherwise.} \end{cases}
\]

**Application:** Toss \(n\) darts randomly into the unit square. The probability that any individual dart will land in the circle is \(\pi/4\). It stands to reason that the proportion of darts, \(\hat{p}_n\), that land in the circle will be approximately \(\pi/4\).
**Example:** Choose a point $(X, Y)$ at random in the interior of the circle inscribed in the unit square, e.g., $C \equiv (x - \frac{1}{2})^2 + (y - \frac{1}{2})^2 \leq \frac{1}{4}$.

Find the pdf of $(X, Y)$.

Since the area of the circle is $\pi/4$,

$$f(x, y) = \begin{cases} 
\frac{4}{\pi} & \text{if } (x, y) \in C \\
0 & \text{otherwise.}
\end{cases}$$

**Application:** Toss $n$ darts randomly into the unit square. The probability that any individual dart will land in the circle is $\pi/4$. It stands to reason that the proportion of darts, $\hat{p}_n$, that land in the circle will be approximately $\pi/4$. So you can use $4\hat{p}_n$ to estimate $\pi$!
**Example**: Suppose that

$$f(x, y) = \begin{cases} 
4xy & \text{if } 0 \leq x \leq 1, 0 \leq y \leq 1 \\
0 & \text{otherwise.}
\end{cases}$$

Find the probability (volume) of the region $0 \leq y \leq 1 - x^2$.

$$V = \int_0^1 \int_0^{\sqrt{1-y}} 4xy \, dx \, dy = \frac{1}{3}.$$**Moral**: Be careful with limits!
**Example**: Suppose that

\[ f(x, y) = \begin{cases} 
4xy & \text{if } 0 \leq x \leq 1, 0 \leq y \leq 1 \\
0 & \text{otherwise.} 
\end{cases} \]

Find the probability (volume) of the region \( 0 \leq y \leq 1 - x^2 \).
**Example:** Suppose that

\[ f(x, y) = \begin{cases} 
4xy & \text{if } 0 \leq x \leq 1, 0 \leq y \leq 1 \\
0 & \text{otherwise.} 
\end{cases} \]

Find the probability (volume) of the region \(0 \leq y \leq 1 - x^2\).

\[
V = \int_0^1 \int_0^{1-x^2} 4xy \, dy \, dx
\]

Moral: Be careful with limits!
**Example:** Suppose that

\[ f(x, y) = \begin{cases} 4xy & \text{if } 0 \leq x \leq 1, \ 0 \leq y \leq 1 \\ 0 & \text{otherwise.} \end{cases} \]

Find the probability (volume) of the region \( 0 \leq y \leq 1 - x^2 \).

\[
V = \int_0^1 \int_0^{1-x^2} 4xy \, dy \, dx
\]

\[
= \int_0^1 \int_0^{\sqrt{1-y}} 4xy \, dx \, dy
\]

\[
= \frac{1}{3}
\]

Moral: Be careful with limits!
**Example:** Suppose that

\[ f(x, y) = \begin{cases} 
4xy & \text{if } 0 \leq x \leq 1, 0 \leq y \leq 1 \\
0 & \text{otherwise.} 
\end{cases} \]

Find the probability (volume) of the region \(0 \leq y \leq 1 - x^2\).

\[
V = \int_0^1 \int_0^{1-x^2} 4xy \, dy \, dx
\]

\[
= \int_0^1 \int_0^{\sqrt{1-y}} 4xy \, dx \, dy
\]

\[
= \frac{1}{3}.
\]
Example: Suppose that

\[ f(x, y) = \begin{cases} 
4xy & \text{if } 0 \leq x \leq 1, 0 \leq y \leq 1 \\
0 & \text{otherwise.} 
\end{cases} \]

Find the probability (volume) of the region \(0 \leq y \leq 1 - x^2\).

\[ V = \int_0^1 \int_0^{1-x^2} 4xy \, dy \, dx \]
\[ = \int_0^1 \int_0^{\sqrt{1-y}} 4xy \, dx \, dy \]
\[ = \frac{1}{3}. \]

Moral: Be careful with limits! \(\Box\)
Bivariate cdf’s
Bivariate cdf’s

**Definition:** The joint (bivariate) cdf of $X$ and $Y$ is $F(x, y) \equiv P(X \leq x, Y \leq y)$, for all $x, y$. 
Bivariate cdf’s

**Definition:** The joint (bivariate) cdf of $X$ and $Y$ is 
\[ F(x, y) \equiv P(X \leq x, Y \leq y), \text{ for all } x, y. \]

\[
F(x, y) = \begin{cases} 
\sum \sum_{s \leq x, t \leq y} f(s, t) & \text{discrete} \\
\int_{-\infty}^{y} \int_{-\infty}^{x} f(s, t) 
\end{cases} \] ds \, dt \text{ continuous.}

\[ = \int_{-\infty}^{x} \int_{-\infty}^{y} f(s, t) \, ds \, dt \text{ continuous.} \]
Bivariate cdf’s

**Definition:** The joint (bivariate) cdf of $X$ and $Y$ is

$$F(x, y) \equiv P(X \leq x, Y \leq y), \text{ for all } x, y.$$ 

$$F(x, y) = \begin{cases} \sum \sum_{s \leq x, t \leq y} f(s, t) & \text{discrete} \\ \int_{-\infty}^{y} \int_{-\infty}^{x} f(s, t) \, ds \, dt & \text{continuous.} \end{cases}$$

Going from cdf’s to pdf’s (continuous case):
**Bivariate cdf’s**

**Definition:** The joint (bivariate) cdf of $X$ and $Y$ is $F(x, y) \equiv P(X \leq x, Y \leq y)$, for all $x, y$.

$$F(x, y) = \begin{cases} \sum \sum_{s \leq x, t \leq y} f(s, t) & \text{discrete} \\ \int_{-\infty}^{y} \int_{-\infty}^{x} f(s, t) \, ds \, dt & \text{continuous.} \end{cases}$$

Going from cdf’s to pdf’s (continuous case):

1-dimensional: $f(x) = F'(x) = \frac{d}{dx} \int_{-\infty}^{x} f(t) \, dt.$
Bivariate cdf’s

**Definition:** The joint (bivariate) cdf of $X$ and $Y$ is $F(x, y) \equiv P(X \leq x, Y \leq y)$, for all $x, y$.

$F(x, y) = \begin{cases} 
\sum \sum_{s \leq x, t \leq y} f(s, t) & \text{discrete} \\
\int_{-\infty}^{y} \int_{-\infty}^{x} f(s, t) \, ds \, dt & \text{continuous}.
\end{cases}$

Going from cdf’s to pdf’s (continuous case):

1-dimension: $f(x) = F'(x) = \frac{d}{dx} \int_{-\infty}^{x} f(t) \, dt$.

2-dimensions: $f(x, y) = \frac{\partial^2}{\partial x \partial y} F(x, y) = \frac{\partial^2}{\partial x \partial y} \int_{-\infty}^{x} \int_{-\infty}^{y} f(s, t) \, dt \, ds.$
Properties:
Properties:

$F(x, y)$ is non-decreasing in both $x$ and $y$. 
Properties:

$F(x, y)$ is non-decreasing in both $x$ and $y$.

$$\lim_{x \to -\infty} F(x, y) = \lim_{y \to -\infty} F(x, y) = 0.$$

$F(x, y)$ is continuous from the right in both $x$ and $y$. 

$$\lim_{x \to \infty} \lim_{y \to \infty} F(x, y) = 1.$$
Properties:

$F(x, y)$ is non-decreasing in both $x$ and $y$.

$$\lim_{x \to -\infty} F(x, y) = \lim_{y \to -\infty} F(x, y) = 0.$$ 

$$\lim_{x \to \infty} F(x, y) = F_Y(y) = P(Y \leq y) \quad (\text{"marginal" cdf of } Y).$$
Properties:

$F(x, y)$ is non-decreasing in both $x$ and $y$.

$$\lim_{x \to -\infty} F(x, y) = \lim_{y \to -\infty} F(x, y) = 0.$$ $$\lim_{x \to \infty} F(x, y) = F_Y(y) = P(Y \leq y) \quad \text{("marginal" cdf of $Y$).}$$ $$\lim_{y \to \infty} F(x, y) = F_X(x) = P(X \leq x) \quad \text{("marginal" cdf of $X$).}$$
**Properties:**

\( F(x, y) \) is non-decreasing in both \( x \) and \( y \).

\[
\lim_{x \to -\infty} F(x, y) = \lim_{y \to -\infty} F(x, y) = 0.
\]

\[
\lim_{x \to \infty} F(x, y) = F_Y(y) = P(Y \leq y) \quad ("\text{marginal" cdf of } Y).
\]

\[
\lim_{y \to \infty} F(x, y) = F_X(x) = P(X \leq x) \quad ("\text{marginal" cdf of } X).
\]

\[
\lim_{x \to \infty} \lim_{y \to \infty} F(x, y) = 1.
\]
Properties:

$F(x, y)$ is non-decreasing in both $x$ and $y$.

$$\lim_{x \to -\infty} F(x, y) = \lim_{y \to -\infty} F(x, y) = 0.$$  

$$\lim_{x \to \infty} F(x, y) = F_Y(y) = P(Y \leq y) \quad \text{("marginal" cdf of } Y).$$  

$$\lim_{y \to \infty} F(x, y) = F_X(x) = P(X \leq x) \quad \text{("marginal" cdf of } X).$$  

$$\lim_{x \to \infty} \lim_{y \to \infty} F(x, y) = 1.$$  

$F(x, y)$ is continuous from the right in both $x$ and $y$.  


Example: Suppose

\[
F(x, y) = \begin{cases} 
1 - e^{-x} - e^{-y} + e^{-(x+y)} & \text{if } x \geq 0, y \geq 0 \\
0 & \text{if } x < 0 \text{ or } y < 0.
\end{cases}
\]
**Example:** Suppose

\[ F(x, y) = \begin{cases} 
1 - e^{-x} - e^{-y} + e^{-(x+y)} & \text{if } x \geq 0, y \geq 0 \\
0 & \text{if } x < 0 \text{ or } y < 0.
\end{cases} \]

The marginal cdf of \( X \) is

\[ F_X(x) = \lim_{y \to \infty} F(x, y) = \begin{cases} 
1 - e^{-x} & \text{if } x \geq 0 \\
0 & \text{if } x < 0.
\end{cases} \]
**Example:** Suppose

\[
F(x, y) = \begin{cases} 
1 - e^{-x} - e^{-y} + e^{-(x+y)} & \text{if } x \geq 0, y \geq 0 \\
0 & \text{if } x < 0 \text{ or } y < 0.
\end{cases}
\]

The marginal cdf of \(X\) is

\[
F_X(x) = \lim_{y \to \infty} F(x, y) = \begin{cases} 
1 - e^{-x} & \text{if } x \geq 0 \\
0 & \text{if } x < 0.
\end{cases}
\]

The joint pdf is

\[
f(x, y) = \frac{\partial^2}{\partial x \partial y} F(x, y)
\]
Example: Suppose

\[ F(x, y) = \begin{cases} 
1 - e^{-x} - e^{-y} + e^{-(x+y)} & \text{if } x \geq 0, y \geq 0 \\
0 & \text{if } x < 0 \text{ or } y < 0.
\end{cases} \]

The marginal cdf of \( X \) is

\[ F_X(x) = \lim_{y \to \infty} F(x, y) = \begin{cases} 
1 - e^{-x} & \text{if } x \geq 0 \\
0 & \text{if } x < 0.
\end{cases} \]

The joint pdf is

\[ f(x, y) = \frac{\partial^2}{\partial x \partial y} F(x, y) = \frac{\partial}{\partial y} (e^{-x} - e^{-y} e^{-x}) \]
Example: Suppose

\[ F(x, y) = \begin{cases} 
1 - e^{-x} - e^{-y} + e^{-(x+y)} & \text{if } x \geq 0, \, y \geq 0 \\
0 & \text{if } x < 0 \text{ or } y < 0.
\end{cases} \]

The marginal cdf of \( X \) is

\[ F_X(x) = \lim_{y \to \infty} F(x, y) = \begin{cases} 
1 - e^{-x} & \text{if } x \geq 0 \\
0 & \text{if } x < 0.
\end{cases} \]

The joint pdf is

\[ f(x, y) = \frac{\partial^2}{\partial x \partial y} F(x, y) = \frac{\partial}{\partial y} (e^{-x} - e^{-y} e^{-x}) = e^{-(x+y)} \quad \text{if } x \geq 0, \, y \geq 0. \]
Marginal Distributions

Introduction

Marginal Distributions

Conditional Distributions

Independent Random Variables

Consequences of Independence

Random Samples

Conditional Expectation

Double Expectation

Honors Class: First-Step Analysis

Honors Class: Random Sums of Random Variables

Honors Class: Standard Conditioning Argument

Covariance and Correlation

Correlation and Causation

A Couple of Worked Correlation Examples

Some Useful Covariance / Correlation Theorems

Moment Generating Functions, Revisited

Honors Bivariate Functions of Random Variables
Lesson 3.2 — Marginal Distributions

We're also interested in the individual (marginal) distributions of $X$ and $Y$.

**Definition**: If $X$ and $Y$ are jointly discrete, then the marginal pmf's of $X$ and $Y$ are, respectively, 

$$f_X(x) = P(X = x) = \sum_y f(x, y)$$

and

$$f_Y(y) = P(Y = y) = \sum_x f(x, y).$$
Lesson 3.2 — Marginal Distributions

We’re also interested in the individual (marginal) distributions of $X$ and $Y$. 
Lesson 3.2 — Marginal Distributions

We’re also interested in the individual (marginal) distributions of $X$ and $Y$.

**Definition**: If $X$ and $Y$ are jointly discrete, then the marginal pmfs of $X$ and $Y$ are, respectively,

$$f_X(x) = P(X = x) = \sum_y f(x, y)$$
We’re also interested in the individual (marginal) distributions of $X$ and $Y$.

**Definition**: If $X$ and $Y$ are jointly *discrete*, then the marginal pmf’s of $X$ and $Y$ are, respectively,

$$f_X(x) = P(X = x) = \sum_y f(x, y)$$

and

$$f_Y(y) = P(Y = y) = \sum_x f(x, y).$$
Example (discrete case): $f(x, y) = P(X = x, Y = y)$. 
Example (discrete case): \( f(x, y) = P(X = x, Y = y) \).

<table>
<thead>
<tr>
<th>( f(x, y) )</th>
<th>( X = 1 )</th>
<th>( X = 2 )</th>
<th>( X = 3 )</th>
<th>( P(Y = y) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Y = 40 )</td>
<td>0.01</td>
<td>0.07</td>
<td>0.12</td>
<td>0.2</td>
</tr>
<tr>
<td>( Y = 60 )</td>
<td>0.29</td>
<td>0.03</td>
<td>0.48</td>
<td>0.8</td>
</tr>
<tr>
<td>( P(X = x) )</td>
<td>0.3</td>
<td>0.1</td>
<td>0.6</td>
<td>1</td>
</tr>
</tbody>
</table>
Example (discrete case): $f(x, y) = P(X = x, Y = y)$.

<table>
<thead>
<tr>
<th>$f(x, y)$</th>
<th>$X = 1$</th>
<th>$X = 2$</th>
<th>$X = 3$</th>
<th>$P(Y = y)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y = 40$</td>
<td>0.01</td>
<td>0.07</td>
<td>0.12</td>
<td>0.2</td>
</tr>
<tr>
<td>$Y = 60$</td>
<td>0.29</td>
<td>0.03</td>
<td>0.48</td>
<td>0.8</td>
</tr>
<tr>
<td>$P(X = x)$</td>
<td>0.3</td>
<td>0.1</td>
<td>0.6</td>
<td>1</td>
</tr>
</tbody>
</table>

By total probability,

$$P(X = 1) = P(X = 1, Y = \text{any #}) = 0.3.$$
Example (discrete case): \( f(x, y) = P(X = x, Y = y) \).
**Example** (discrete case): \( f(x, y) = P(X = x, Y = y) \).

<table>
<thead>
<tr>
<th>( f(x, y) )</th>
<th>( X = 1 )</th>
<th>( X = 2 )</th>
<th>( X = 3 )</th>
<th>( P(Y = y) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Y = 40 )</td>
<td>0.06</td>
<td>0.02</td>
<td>0.12</td>
<td>0.2</td>
</tr>
<tr>
<td>( Y = 60 )</td>
<td>0.24</td>
<td>0.08</td>
<td>0.48</td>
<td>0.8</td>
</tr>
<tr>
<td>( P(X = x) )</td>
<td>0.3</td>
<td>0.1</td>
<td>0.6</td>
<td>1</td>
</tr>
</tbody>
</table>
Example (discrete case): \( f(x, y) = P(X = x, Y = y) \).

<table>
<thead>
<tr>
<th>( f(x, y) )</th>
<th>( X = 1 )</th>
<th>( X = 2 )</th>
<th>( X = 3 )</th>
<th>( P(Y = y) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Y = 40 )</td>
<td>0.06</td>
<td>0.02</td>
<td>0.12</td>
<td>0.2</td>
</tr>
<tr>
<td>( Y = 60 )</td>
<td>0.24</td>
<td>0.08</td>
<td>0.48</td>
<td>0.8</td>
</tr>
<tr>
<td>( P(X = x) )</td>
<td>0.3</td>
<td>0.1</td>
<td>0.6</td>
<td>1</td>
</tr>
</tbody>
</table>

Remark: Hmmm... Compared to the last example, this has the same marginals but different joint distribution! That’s because the joint distribution contains much more information than just the marginals.
**Definition**: If $X$ and $Y$ are jointly *continuous*, then the *marginal pdf’s* of $X$ and $Y$ are, respectively,
**Definition:** If $X$ and $Y$ are jointly *continuous*, then the **marginal pdf’s** of $X$ and $Y$ are, respectively,

$$f_X(x) = \int_R f(x, y) \, dy$$
**Definition**: If $X$ and $Y$ are jointly *continuous*, then the *marginal pdf’s* of $X$ and $Y$ are, respectively,

\[
f_X(x) = \int_{\mathbb{R}} f(x, y) \, dy \quad \text{and} \quad f_Y(y) = \int_{\mathbb{R}} f(x, y) \, dx.
\]
**Definition**: If $X$ and $Y$ are jointly *continuous*, then the *marginal pdf’s* of $X$ and $Y$ are, respectively,

$$f_X(x) = \int_{\mathbb{R}} f(x, y) \, dy \quad \text{and} \quad f_Y(y) = \int_{\mathbb{R}} f(x, y) \, dx.$$ 

**Example:**

$$f(x, y) = \begin{cases} 
  e^{-(x+y)} & \text{if } x \geq 0, y \geq 0 \\
  0 & \text{otherwise}.
\end{cases}$$
Definition: If $X$ and $Y$ are jointly continuous, then the marginal pdf’s of $X$ and $Y$ are, respectively,

$$f_X(x) = \int_{\mathbb{R}} f(x, y) \, dy \quad \text{and} \quad f_Y(y) = \int_{\mathbb{R}} f(x, y) \, dx.$$

Example:

$$f(x, y) = \begin{cases} 
  e^{-(x+y)} & \text{if } x \geq 0, y \geq 0 \\
  0 & \text{otherwise.}
\end{cases}$$

Then the marginal pdf of $X$ is

$$f_X(x) = \int_{\mathbb{R}} f(x, y) \, dy.$$
**Definition:** If $X$ and $Y$ are jointly *continuous*, then the *marginal pdf’s* of $X$ and $Y$ are, respectively,

$$f_X(x) = \int_{\mathbb{R}} f(x, y) \, dy \quad \text{and} \quad f_Y(y) = \int_{\mathbb{R}} f(x, y) \, dx.$$ 

**Example:**

$$f(x, y) = \begin{cases} e^{-(x+y)} & \text{if } x \geq 0, y \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

Then the marginal pdf of $X$ is

$$f_X(x) = \int_{\mathbb{R}} f(x, y) \, dy = \int_{0}^{\infty} e^{-(x+y)} \, dy.$$
**Definition:** If $X$ and $Y$ are jointly continuous, then the marginal pdf's of $X$ and $Y$ are, respectively,

\[
f_X(x) = \int_\mathbb{R} f(x, y) \, dy \quad \text{and} \quad f_Y(y) = \int_\mathbb{R} f(x, y) \, dx.
\]

**Example:**

\[
f(x, y) = \begin{cases} 
e^{-(x+y)} & \text{if } x \geq 0, \ y \geq 0 \\ 0 & \text{otherwise.} \end{cases}
\]

Then the marginal pdf of $X$ is

\[
f_X(x) = \int_\mathbb{R} f(x, y) \, dy = \int_0^\infty e^{-(x+y)} \, dy = e^{-x}, \text{ if } x \geq 0. \quad \square
\]
Example:

\[ f(x, y) = \begin{cases} 
\frac{21}{4} x^2 y & \text{if } x^2 \leq y \leq 1 \\
0 & \text{otherwise.} 
\end{cases} \]
Example:

\[ f(x, y) = \begin{cases} \frac{21}{4} x^2 y & \text{if } x^2 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases} \]

Note *funny limits* where the pdf is positive, i.e., \( x^2 \leq y \leq 1 \).
Example:

\[ f(x, y) = \begin{cases} 
\frac{21}{4} x^2 y & \text{if } x^2 \leq y \leq 1 \\
0 & \text{otherwise.}
\end{cases} \]

Note *funny limits* where the pdf is positive, i.e., \( x^2 \leq y \leq 1 \).

\[ f_X(x) = \int_{\mathbb{R}} f(x, y) \, dy \]
Example:

\[ f(x, y) = \begin{cases} \frac{21}{4} x^2 y & \text{if } x^2 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases} \]

Note **funny limits** where the pdf is positive, i.e., \( x^2 \leq y \leq 1 \).

\[ f_X(x) = \int_{\mathbb{R}} f(x, y) \, dy = \int_{x^2}^{1} \frac{21}{4} x^2 y \, dy \]
Example:

\[ f(x, y) = \begin{cases} \frac{21}{4} x^2 y & \text{if } x^2 \leq y \leq 1 \\ 0 & \text{otherwise.} \end{cases} \]

Note \textit{funny limits} where the pdf is positive, i.e., \( x^2 \leq y \leq 1 \).

\[ f_X(x) = \int_{\mathbb{R}} f(x, y) \, dy = \int_{x^2}^{1} \frac{21}{4} x^2 y \, dy = \frac{21}{8} x^2 (1 - x^4), \quad -1 \leq x \leq 1. \]
Example:

\[ f(x, y) = \begin{cases} 
\frac{21}{4} x^2 y & \text{if } x^2 \leq y \leq 1 \\
0 & \text{otherwise.}
\end{cases} \]

Note *funny limits* where the pdf is positive, i.e., \( x^2 \leq y \leq 1 \).

\[ f_X(x) = \int_{\mathbb{R}} f(x, y) \, dy = \int_{x^2}^{1} \frac{21}{4} x^2 y \, dy = \frac{21}{8} x^2 (1 - x^4), \quad -1 \leq x \leq 1. \]

\[ f_Y(y) = \int_{\mathbb{R}} f(x, y) \, dx \]
Example:

\[ f(x, y) = \begin{cases} 
\frac{21}{4} x^2 y & \text{if } x^2 \leq y \leq 1 \\
0 & \text{otherwise.}
\end{cases} \]

Note **funny limits** where the pdf is positive, i.e., \( x^2 \leq y \leq 1 \).

\[ f_X(x) = \int_{\mathbb{R}} f(x, y) \, dy = \int_{x^2}^{1} \frac{21}{4} x^2 y \, dy = \frac{21}{8} x^2 (1 - x^4), \quad -1 \leq x \leq 1. \]

\[ f_Y(y) = \int_{\mathbb{R}} f(x, y) \, dx = \int_{-\sqrt{y}}^{\sqrt{y}} \frac{21}{4} x^2 y \, dx \]
Example:

\[ f(x, y) = \begin{cases} 
\frac{21}{4} x^2 y & \text{if } x^2 \leq y \leq 1 \\
0 & \text{otherwise.}
\end{cases} \]

Note funny limits where the pdf is positive, i.e., \( x^2 \leq y \leq 1 \).

\[ f_X(x) = \int_{\mathbb{R}} f(x, y) \, dy = \int_{x^2}^{1} \frac{21}{4} x^2 y \, dy = \frac{21}{8} x^2 (1 - x^4), \quad -1 \leq x \leq 1. \]

\[ f_Y(y) = \int_{\mathbb{R}} f(x, y) \, dx = \int_{-\sqrt{y}}^{\sqrt{y}} \frac{21}{4} x^2 y \, dx = \frac{7}{2} y^{5/2}, \quad 0 \leq y \leq 1. \]
Conditional Distributions

1 Introduction
2 Marginal Distributions
3 Conditional Distributions
4 Independent Random Variables
5 Consequences of Independence
6 Random Samples
7 Conditional Expectation
8 Double Expectation
9 Honors Class: First-Step Analysis
10 Honors Class: Random Sums of Random Variables
11 Honors Class: Standard Conditioning Argument
12 Covariance and Correlation
13 Correlation and Causation
14 A Couple of Worked Correlation Examples
15 Some Useful Covariance / Correlation Theorems
16 Moment Generating Functions, Revisited
17 Honors Bivariate Functions of Random Variables
Recall conditional probability:

$$P(A | B) = \frac{P(A \cap B)}{P(B)}$$ if $P(B) > 0$.

Suppose that $X$ and $Y$ are jointly discrete RVs. Then if $P(X = x) > 0$,

$$P(Y = y | X = x) = \frac{P(X = x \cap Y = y)}{P(X = x)} = f(x, y) f_X(x).$$

$P(Y = y | X = 2)$ defines the probabilities on $Y$ given that $X = 2$.

**Definition:** If $f_X(x) > 0$, then the conditional pmf/pdf of $Y$ given $X = x$ is $f_{Y|X}(y | x) \equiv f(x, y) f_X(x)$.

**Remark:** We usually just write $f(y | x)$ instead of $f_{Y|X}(y | x)$.

**Remark:** Of course, $f_{X|Y}(x | y) = f(x | y) = f(x, y) f_Y(y)$. 
Recall conditional probability: \( P(A|B) = \frac{P(A \cap B)}{P(B)} \) if \( P(B) > 0 \).
Lesson 3.3 — Conditional Distributions

Recall conditional probability: \( P(A|B) = \frac{P(A \cap B)}{P(B)} \) if \( P(B) > 0 \).

Suppose that \( X \) and \( Y \) are jointly discrete RVs. Then if \( P(X = x) > 0 \),
Recall conditional probability: $P(A|B) = \frac{P(A \cap B)}{P(B)}$ if $P(B) > 0$.

Suppose that $X$ and $Y$ are jointly discrete RVs. Then if $P(X = x) > 0$,

$$P(Y = y|X = x) = \frac{P(X = x \cap Y = y)}{P(X = x)}$$
Lesson 3.3 — Conditional Distributions

Recall conditional probability: \( P(A|B) = \frac{P(A \cap B)}{P(B)} \) if \( P(B) > 0 \).

Suppose that \( X \) and \( Y \) are jointly discrete RVs. Then if \( P(X = x) > 0 \),

\[
P(Y = y|X = x) = \frac{P(X = x \cap Y = y)}{P(X = x)} = \frac{f(x, y)}{f_X(x)}.
\]
Recall conditional probability: \( P(A|B) = \frac{P(A \cap B)}{P(B)} \) if \( P(B) > 0 \).

Suppose that \( X \) and \( Y \) are jointly discrete RVs. Then if \( P(X = x) > 0 \),

\[
P(Y = y|X = x) = \frac{P(X = x \cap Y = y)}{P(X = x)} = \frac{f(x, y)}{f_X(x)}.
\]

\( P(Y = y|X = 2) \) defines the probabilities on \( Y \) given that \( X = 2 \).
Recall conditional probability: \( P(A|B) = \frac{P(A \cap B)}{P(B)} \) if \( P(B) > 0 \).

Suppose that \( X \) and \( Y \) are jointly discrete RVs. Then if \( P(X = x) > 0 \),

\[
P(Y = y|X = x) = \frac{P(X = x \cap Y = y)}{P(X = x)} = \frac{f(x, y)}{f_X(x)}.
\]

\( P(Y = y|X = 2) \) defines the probabilities on \( Y \) given that \( X = 2 \).

**Definition:** If \( f_X(x) > 0 \), then the **conditional pmf/pdf of \( Y \) given \( X = x \)** is

\[
f_{Y|X}(y|x) \equiv \frac{f(x, y)}{f_X(x)}.
\]
Lesson 3.3 — Conditional Distributions

Recall conditional probability: \( P(A|B) = \frac{P(A \cap B)}{P(B)} \) if \( P(B) > 0 \).

Suppose that \( X \) and \( Y \) are jointly discrete RVs. Then if \( P(X = x) > 0 \),

\[
P(Y = y|X = x) = \frac{P(X = x \cap Y = y)}{P(X = x)} = \frac{f(x, y)}{f_X(x)}.
\]

\( P(Y = y|X = 2) \) defines the probabilities on \( Y \) given that \( X = 2 \).

**Definition**: If \( f_X(x) > 0 \), then the **conditional pmf/pdf of \( Y \) given \( X = x \)** is

\[
f_{Y|X}(y|x) \equiv \frac{f(x, y)}{f_X(x)}.
\]

**Remark**: We usually just write \( f(y|x) \) instead of \( f_{Y|X}(y|x) \).
Recall conditional probability: \( P(A|B) = \frac{P(A \cap B)}{P(B)} \) if \( P(B) > 0 \).

Suppose that \( X \) and \( Y \) are jointly discrete RVs. Then if \( P(X = x) > 0 \),

\[
P(Y = y|X = x) = \frac{P(X = x \cap Y = y)}{P(X = x)} = \frac{f(x, y)}{f_X(x)}.
\]

\( P(Y = y|X = 2) \) defines the probabilities on \( Y \) given that \( X = 2 \).

**Definition:** If \( f_X(x) > 0 \), then the **conditional pmf/pdf of \( Y \) given \( X = x \)** is

\[
f_{Y|X}(y|x) \equiv \frac{f(x, y)}{f_X(x)}.
\]

**Remark:** We usually just write \( f(y|x) \) instead of \( f_{Y|X}(y|x) \).

**Remark:** Of course, \( f_{X|Y}(x|y) = f(x|y) = \frac{f(x,y)}{f_Y(y)} \).
**Discrete Example:** \( f(x, y) = P(X = x, Y = y) \).
**Discrete Example:** \( f(x, y) = P(X = x, Y = y) \).

<table>
<thead>
<tr>
<th>( f(x, y) )</th>
<th>( X = 1 )</th>
<th>( X = 2 )</th>
<th>( X = 3 )</th>
<th>( f_Y(y) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Y = 40 )</td>
<td>0.01</td>
<td>0.07</td>
<td>0.12</td>
<td>0.2</td>
</tr>
<tr>
<td>( Y = 60 )</td>
<td>0.29</td>
<td>0.03</td>
<td>0.48</td>
<td>0.8</td>
</tr>
<tr>
<td>( f_X(x) )</td>
<td>0.3</td>
<td>0.1</td>
<td>0.6</td>
<td>1</td>
</tr>
</tbody>
</table>
Conditional Distributions

**Discrete Example:** \( f(x, y) = P(X = x, Y = y). \)

<table>
<thead>
<tr>
<th>( f(x, y) )</th>
<th>( X = 1 )</th>
<th>( X = 2 )</th>
<th>( X = 3 )</th>
<th>( f_Y(y) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Y = 40 )</td>
<td>0.01</td>
<td>0.07</td>
<td>0.12</td>
<td>0.2</td>
</tr>
<tr>
<td>( Y = 60 )</td>
<td><strong>0.29</strong></td>
<td><strong>0.03</strong></td>
<td><strong>0.48</strong></td>
<td><strong>0.8</strong></td>
</tr>
<tr>
<td>( f_X(x) )</td>
<td>0.3</td>
<td>0.1</td>
<td>0.6</td>
<td>1</td>
</tr>
</tbody>
</table>

Then, for example,

\[
f(x | y = 60) = \frac{f(x, 60)}{f_Y(60)}
\]
Discrete Example: \( f(x, y) = P(X = x, Y = y) \).

<table>
<thead>
<tr>
<th>( f(x, y) )</th>
<th>( X = 1 )</th>
<th>( X = 2 )</th>
<th>( X = 3 )</th>
<th>( f_Y(y) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Y = 40 )</td>
<td>0.01</td>
<td>0.07</td>
<td>0.12</td>
<td>0.2</td>
</tr>
<tr>
<td>( Y = 60 )</td>
<td>0.29</td>
<td>0.03</td>
<td>0.48</td>
<td>0.8</td>
</tr>
<tr>
<td>( f_X(x) )</td>
<td>0.3</td>
<td>0.1</td>
<td>0.6</td>
<td>1</td>
</tr>
</tbody>
</table>

Then, for example,

\[
f(x | y = 60) = \frac{f(x, 60)}{f_Y(60)} = \frac{f(x, 60)}{0.8}
\]
Discrete Example: $f(x, y) = P(X = x, Y = y)$.

<table>
<thead>
<tr>
<th>$f(x, y)$</th>
<th>$X = 1$</th>
<th>$X = 2$</th>
<th>$X = 3$</th>
<th>$f_Y(y)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y = 40$</td>
<td>0.01</td>
<td>0.07</td>
<td>0.12</td>
<td>0.2</td>
</tr>
<tr>
<td>$Y = 60$</td>
<td>0.29</td>
<td>0.03</td>
<td>0.48</td>
<td>0.8</td>
</tr>
<tr>
<td>$f_X(x)$</td>
<td>0.3</td>
<td>0.1</td>
<td>0.6</td>
<td>1</td>
</tr>
</tbody>
</table>

Then, for example,

$$f(x | y = 60) = \frac{f(x, 60)}{f_Y(60)} = \frac{f(x, 60)}{0.8} = \begin{cases} \frac{29}{80} & \text{if } x = 1 \\ \frac{3}{80} & \text{if } x = 2 \\ \frac{48}{80} & \text{if } x = 3 \end{cases}$$
Old Continuous Example:

\[ f(x,y) = \begin{cases} 21 & 4x^2y, \text{ if } x^2 \leq y \leq 1 \\ f_X(x) = \begin{cases} 21 & 8x^2(1-x^4), \text{ if } -1 \leq x \leq 1 \\ f_Y(y) = \begin{cases} 7 & 2y^5/2, \text{ if } 0 \leq y \leq 1 \\ \end{cases} \end{cases} \]

Then the conditional pdf of \( Y \) given \( X = x \) is

\[ f(y|x) = \frac{f(x,y)}{f_X(x)} = \begin{cases} 2 & y(1-x^4), \text{ if } x^2 \leq y \leq 1 \end{cases} \]
Old Continuous Example:

\[ f(x, y) = \frac{21}{4} x^2 y, \quad \text{if } x^2 \leq y \leq 1. \]
Old Continuous Example:

\[ f(x, y) = \frac{21}{4} x^2 y, \quad \text{if } x^2 \leq y \leq 1. \]

\[ f_X(x) = \frac{21}{8} x^2 (1 - x^4), \quad \text{if } -1 \leq x \leq 1. \]
Old Continuous Example:

\[ f(x, y) = \frac{21}{4} x^2 y, \quad \text{if } x^2 \leq y \leq 1. \]

\[ f_X(x) = \frac{21}{8} x^2 (1 - x^4), \quad \text{if } -1 \leq x \leq 1. \]

\[ f_Y(y) = \frac{7}{2} y^{5/2}, \quad \text{if } 0 \leq y \leq 1. \]
Old Continuous Example:

\[
f(x, y) = \frac{21}{4}x^2y, \quad \text{if } x^2 \leq y \leq 1.
\]

\[
f_X(x) = \frac{21}{8}x^2(1 - x^4), \quad \text{if } -1 \leq x \leq 1.
\]

\[
f_Y(y) = \frac{7}{2}y^{5/2}, \quad \text{if } 0 \leq y \leq 1.
\]

Then the conditional pdf of \( Y \) given \( X = x \) is

\[
f(y|x) = \frac{f(x, y)}{f_X(x)}
\]
Conditional Distributions

Old Continuous Example:

\[ f(x, y) = \frac{21}{4}x^2y, \quad \text{if } x^2 \leq y \leq 1. \]

\[ f_X(x) = \frac{21}{8}x^2(1 - x^4), \quad \text{if } -1 \leq x \leq 1. \]

\[ f_Y(y) = \frac{7}{2}y^{5/2}, \quad \text{if } 0 \leq y \leq 1. \]

Then the conditional pdf of \( Y \) given \( X = x \) is

\[ f(y|x) = \frac{f(x, y)}{f_X(x)} = \frac{\frac{21}{4}x^2y}{\frac{21}{8}x^2(1 - x^4)} \]

\[ = \frac{21}{8} \frac{x^2}{x^2(1 - x^4)}y \]

\[ = \frac{21}{8} \frac{1}{1 - x^4}y \]

\[ = \frac{21}{8}y \left( \frac{1}{1 - x^4} \right) \]
Old Continuous Example:

\[
f(x, y) = \frac{21}{4} x^2 y, \quad \text{if } x^2 \leq y \leq 1.
\]

\[
f_X(x) = \frac{21}{8} x^2 (1 - x^4), \quad \text{if } -1 \leq x \leq 1.
\]

\[
f_Y(y) = \frac{7}{2} y^{5/2}, \quad \text{if } 0 \leq y \leq 1.
\]

Then the conditional pdf of \(Y\) given \(X = x\) is

\[
f(y|x) = \frac{f(x, y)}{f_X(x)} = \frac{\frac{21}{4} x^2 y}{\frac{21}{8} x^2 (1 - x^4)} = \frac{2y}{1 - x^4}, \quad \text{if } x^2 \leq y \leq 1.
\]
So, for example,

\[
f(y|1/2) = \frac{f\left(\frac{1}{2}, y\right)}{f_X\left(\frac{1}{2}\right)}
\]
So, for example,

\[ f(y|1/2) = \frac{f\left(\frac{1}{2}, y\right)}{f_X\left(\frac{1}{2}\right)} = \frac{\frac{21}{4} \cdot \frac{1}{4} y}{\frac{21}{8} \cdot \frac{1}{4} \cdot \left(1 - \frac{1}{16}\right)} \]
So, for example,

\[
f(y|1/2) = \frac{f(\frac{1}{2}, y)}{f_X(\frac{1}{2})} = \frac{\frac{21}{4} \cdot \frac{1}{4}y}{\frac{21}{8} \cdot \frac{1}{4} \cdot (1 - \frac{1}{16})} = \frac{32}{15}y, \quad \text{if } \frac{1}{4} \leq y \leq 1. \quad \square
\]
So, for example,

\[
f(y|1/2) = \frac{f\left(\frac{1}{2}, y\right)}{f_X\left(\frac{1}{2}\right)} = \frac{\frac{21}{4} \cdot \frac{1}{4} y}{\frac{21}{8} \cdot \frac{1}{4} \cdot (1 - \frac{1}{16})} = \frac{32}{15} y, \quad \text{if} \quad \frac{1}{4} \leq y \leq 1. \quad \square
\]

Note that \(2/(1 - x^4)\) is a constant with respect to \(y\), and we can check to see that \(f(y|x)\) is a legit conditional pdf:

\[
\int_{\mathbb{R}} f(y|x) \, dy = \int_{x^2}^{1} \frac{2y}{1 - x^4} \, dy = 1. \quad \square
\]
**Typical Problem:** Given $f_X(x)$ and $f(y|x)$, find $f_Y(y)$. 
**Typical Problem:** Given $f_X(x)$ and $f(y|x)$, find $f_Y(y)$.

Game Plan: Find $f(x, y) = f_X(x)f(y|x)$ and then $f_Y(y) = \int_{\mathbb{R}} f(x, y) \, dx$. 

Example: Suppose $f_X(x) = 2x$, for $0 < x < 1$. Given $X = x$, suppose that $Y|X \sim \text{Unif}(0, x)$. Now find $f_Y(y)$.

Solution: $Y|X \sim \text{Unif}(0, x)$ implies that $f(y|x) = 1/x$, for $0 < y < x$. So, $f(x, y) = f_X(x)f(y|x) = 2x \cdot 1/x$ for $0 < x < 1$ and $0 < y < x$.

Thus, $f_Y(y) = \int_{\mathbb{R}} f(x, y) \, dx = \int_{y}^{1} 2 \, dx = 2(1 - y)$, $0 < y < 1$. 

2
**Typical Problem:** Given $f_X(x)$ and $f(y|x)$, find $f_Y(y)$.

Game Plan: Find $f(x, y) = f_X(x) f(y|x)$ and then $f_Y(y) = \int_R f(x, y) \, dx$.

**Example:** Suppose $f_X(x) = 2x$, for $0 < x < 1$. Given $X = x$, suppose that $Y|x \sim \text{Unif}(0, x)$. Now find $f_Y(y)$. 
**Typical Problem:** Given $f_X(x)$ and $f(y|x)$, find $f_Y(y)$.

**Game Plan:** Find $f(x,y) = f_X(x)f(y|x)$ and then $f_Y(y) = \int_{\mathbb{R}} f(x,y) \, dx$.

**Example:** Suppose $f_X(x) = 2x$, for $0 < x < 1$. Given $X = x$, suppose that $Y|X \sim \text{Unif}(0, x)$. Now find $f_Y(y)$.

**Solution:** $Y|X \sim \text{Unif}(0, x)$ implies that $f(y|x) = 1/x$, for $0 < y < x$. So,
**Typical Problem:** Given $f_X(x)$ and $f(y|x)$, find $f_Y(y)$.

Game Plan: Find $f(x, y) = f_X(x) f(y|x)$ and then $f_Y(y) = \int_{\mathbb{R}} f(x, y) \, dx$.

**Example:** Suppose $f_X(x) = 2x$, for $0 < x < 1$. Given $X = x$, suppose that $Y|x \sim \text{Unif}(0, x)$. Now find $f_Y(y)$.

Solution: $Y|x \sim \text{Unif}(0, x)$ implies that $f(y|x) = 1/x$, for $0 < y < x$. So,

$$f(x, y) = f_X(x) f(y|x)$$
**Typical Problem:** Given $f_X(x)$ and $f(y|x)$, find $f_Y(y)$.

**Game Plan:** Find $f(x, y) = f_X(x)f(y|x)$ and then $f_Y(y) = \int_{\mathbb{R}} f(x, y) \, dx$.

**Example:** Suppose $f_X(x) = 2x$, for $0 < x < 1$. Given $X = x$, suppose that $Y|X \sim \text{Unif}(0, x)$. Now find $f_Y(y)$.

**Solution:** $Y|X \sim \text{Unif}(0, x)$ implies that $f(y|x) = 1/x$, for $0 < y < x$. So,

\[
 f(x, y) = f_X(x)f(y|x) = 2x \cdot \frac{1}{x} = 2 \quad \text{for} \quad 0 < x < 1 \text{ and } 0 < y < x
\]
**Typical Problem:** Given $f_X(x)$ and $f(y|x)$, find $f_Y(y)$.

Game Plan: Find $f(x, y) = f_X(x) f(y|x)$ and then $f_Y(y) = \int_{\mathbb{R}} f(x, y) \, dx$.

**Example:** Suppose $f_X(x) = 2x$, for $0 < x < 1$. Given $X = x$, suppose that $Y|x \sim \text{Unif}(0, x)$. Now find $f_Y(y)$.

Solution: $Y|x \sim \text{Unif}(0, x)$ implies that $f(y|x) = 1/x$, for $0 < y < x$. So,

$$f(x, y) = f_X(x) f(y|x)$$

$$= 2x \cdot \frac{1}{x} \quad \text{for } 0 < x < 1 \text{ and } 0 < y < x$$

$$= 2 \quad 0 < y < x < 1 \text{ (still have funny limits).}$$
**Typical Problem:** Given \( f_X(x) \) and \( f(y|x) \), find \( f_Y(y) \).

Game Plan: Find \( f(x, y) = f_X(x) f(y|x) \) and then \( f_Y(y) = \int_{\mathbb{R}} f(x, y) \, dx \).

**Example:** Suppose \( f_X(x) = 2x \), for \( 0 < x < 1 \). Given \( X = x \), suppose that \( Y|x \sim \text{Unif}(0, x) \). Now find \( f_Y(y) \).

Solution: \( Y|x \sim \text{Unif}(0, x) \) implies that \( f(y|x) = \frac{1}{x} \), for \( 0 < y < x \). So,

\[
f(x, y) = f_X(x) f(y|x)
= 2x \cdot \frac{1}{x}
= 2 \quad \text{for } 0 < x < 1 \text{ and } 0 < y < x
= 2 \quad 0 < y < x < 1 \text{ (still have funny limits)}.
\]

Thus,

\[
f_Y(y) = \int_{\mathbb{R}} f(x, y) \, dx
\]
**Typical Problem:** Given $f_X(x)$ and $f(y|x)$, find $f_Y(y)$.

Game Plan: Find $f(x, y) = f_X(x)f(y|x)$ and then $f_Y(y) = \int_{\mathbb{R}} f(x, y) \, dx$.

**Example:** Suppose $f_X(x) = 2x$, for $0 < x < 1$. Given $X = x$, suppose that $Y|x \sim \text{Unif}(0, x)$. Now find $f_Y(y)$.

Solution: $Y|x \sim \text{Unif}(0, x)$ implies that $f(y|x) = 1/x$, for $0 < y < x$. So,

$$f(x, y) = f_X(x)f(y|x)$$
$$= 2x \cdot \frac{1}{x} \quad \text{for } 0 < x < 1 \text{ and } 0 < y < x$$
$$= 2 \quad 0 < y < x < 1 \text{ (still have funny limits)}.$$ 

Thus,

$$f_Y(y) = \int_{\mathbb{R}} f(x, y) \, dx = \int_{y}^{1} 2 \, dx$$
**Typical Problem:** Given $f_X(x)$ and $f(y|x)$, find $f_Y(y)$.

Game Plan: Find $f(x, y) = f_X(x) f(y|x)$ and then $f_Y(y) = \int_{\mathbb{R}} f(x, y) \, dx$.

**Example:** Suppose $f_X(x) = 2x$, for $0 < x < 1$. Given $X = x$, suppose that $Y|x \sim \text{Unif}(0, x)$. Now find $f_Y(y)$.

Solution: $Y|x \sim \text{Unif}(0, x)$ implies that $f(y|x) = 1/x$, for $0 < y < x$. So,

$$f(x, y) = f_X(x) f(y|x)$$
$$= 2x \cdot \frac{1}{x} \quad \text{for } 0 < x < 1 \text{ and } 0 < y < x$$
$$= 2 \quad 0 < y < x < 1 \text{ (still have funny limits)}.$$

Thus,

$$f_Y(y) = \int_{\mathbb{R}} f(x, y) \, dx = \int_y^1 2 \, dx = 2(1 - y), \quad 0 < y < 1.$$
4 Independent Random Variables
Recall that two events are independent if
\[ P(A \cap B) = P(A) \cdot P(B) \].

Then
\[ P(A | B) = \frac{P(A \cap B)}{P(B)} = P(A) \frac{P(B)}{P(B)} = P(A) \].

And similarly,
\[ P(B | A) = \frac{P(B \cap A)}{P(A)} = P(B) \frac{P(A)}{P(A)} = P(B) \].

Now we want to define independence for random variables, i.e., the outcome of \( X \) doesn’t influence the outcome of \( Y \)(and vice versa).

**Definition:** \( X \) and \( Y \) are independent RVs if, for all \( x \) and \( y \),
\[ f(x,y) = f_X(x) \cdot f_Y(y) \].
Recall that two events are independent if $P(A \cap B) = P(A)P(B)$.

Definition: $X$ and $Y$ are independent RVs if, for all $x$ and $y$, $f(x,y) = f_X(x)f_Y(y)$. 
Recall that two events are independent if $P(A \cap B) = P(A)P(B)$.

Then

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$
Recall that two events are independent if $P(A \cap B) = P(A)P(B)$. Then

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)}$$
Lesson 3.4 — Independent Random Variables

Recall that two events are independent if $P(A \cap B) = P(A)P(B)$.

Then

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A).$$
Recall that two events are independent if \( P(A \cap B) = P(A)P(B) \).

Then

\[
P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A).
\]

And similarly, \( P(B|A) = P(B) \).
Recall that two events are independent if \( P(A \cap B) = P(A)P(B) \).

Then
\[
P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A).
\]

And similarly, \( P(B|A) = P(B) \).

Now we want to define independence for random variables, i.e., the outcome of \( X \) doesn’t influence the outcome of \( Y \) (and vice versa).
Recall that two events are independent if $P(A \cap B) = P(A)P(B)$.

Then

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A).$$

And similarly, $P(B|A) = P(B)$.

Now we want to define independence for random variables, i.e., the outcome of $X$ doesn’t influence the outcome of $Y$ (and vice versa).

**Definition:** $X$ and $Y$ are independent RVs if, for all $x$ and $y$,

$$f(x, y) = f_X(x)f_Y(y).$$
Equivalent definitions:

If $X$ and $Y$ aren’t independent, then they’re dependent.

Nice, Intuitive Theorem: $X$ and $Y$ are independent if and only if $f(y|x) = f_Y(y) \forall x,y$.

Proof:

$$f(y|x) = f(x,y) = f_X(x) f_Y(y) = f_Y(y).$$

Similarly, $X$ and $Y$ independent implies $f(x|y) = f_X(x)$.
Equivalent definitions:

\[ F(x, y) = F_X(x)F_Y(y), \quad \forall x, y \]
Equivalent definitions:

\[ F(x, y) = F_X(x)F_Y(y), \quad \forall x, y \]

or

\[ P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y), \quad \forall x, y. \]
Equivalent definitions:

\[ F(x, y) = F_X(x)F_Y(y), \quad \forall x, y \]

or

\[ P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y), \quad \forall x, y. \]

If \( X \) and \( Y \) aren’t independent, then they’re \textbf{dependent}.
Equivalent definitions:

\[ F(x, y) = F_X(x)F_Y(y), \quad \forall x, y \]

or

\[ P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y), \quad \forall x, y. \]

If \( X \) and \( Y \) aren’t independent, then they’re dependent.

**Nice, Intuitive Theorem:** \( X \) and \( Y \) are independent if and only if
\[ f(y|x) = f_Y(y) \quad \forall x, y. \]
Equivalent definitions:

\[ F(x, y) = F_X(x) F_Y(y), \quad \forall x, y \]

or

\[ P(X \leq x, Y \leq y) = P(X \leq x) P(Y \leq y), \quad \forall x, y. \]

If \( X \) and \( Y \) aren’t independent, then they’re dependent.

**Nice, Intuitive Theorem:** \( X \) and \( Y \) are independent if and only if
\[ f(y|x) = f_Y(y) \quad \forall x, y. \]

**Proof:**

\[ f(y|x) = \frac{f(x, y)}{f_X(x)} \]
Equivalent definitions:

\[ F(x, y) = F_X(x)F_Y(y), \quad \forall x, y \]

or

\[ P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y), \quad \forall x, y. \]

If \( X \) and \( Y \) aren’t independent, then they’re dependent.

**Nice, Intuitive Theorem:** \( X \) and \( Y \) are independent if and only if 
\[ f(y|x) = f_Y(y) \quad \forall x, y. \]

**Proof:**

\[ f(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{f_X(x)f_Y(y)}{f_X(x)} \]
Equivalent definitions:

\[ F(x, y) = F_X(x)F_Y(y), \quad \forall x, y \]

or

\[ P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y), \quad \forall x, y. \]

If \( X \) and \( Y \) aren’t independent, then they’re dependent.

**Nice, Intuitive Theorem:** \( X \) and \( Y \) are independent if and only if 

\[ f(y|x) = f_Y(y) \quad \forall x, y. \]

**Proof:**

\[ f(y|x) = \frac{f(x, y)}{f_X(x)} = \frac{f_X(x)f_Y(y)}{f_X(x)} = f_Y(y). \quad \square \]
Equivalent definitions:

\[ F(x, y) = F_X(x)F_Y(y), \quad \forall x, y \]

or

\[ P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y), \quad \forall x, y. \]

If \( X \) and \( Y \) aren’t independent, then they’re dependent.

**Nice, Intuitive Theorem:** \( X \) and \( Y \) are independent if and only if \( f(y|x) = f_Y(y) \forall x, y. \)

**Proof:**

\[ f(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{f_X(x)f_Y(y)}{f_X(x)} = f_Y(y). \quad \square \]

Similarly, \( X \) and \( Y \) independent implies \( f(x|y) = f_X(x). \)
Example (discrete): \( f(x, y) = P(X = x, Y = y) \).
Example (discrete): $f(x, y) = P(X = x, Y = y)$.

\[
\begin{array}{c|cc|c}
  f(x, y) & X = 1 & X = 2 & f_Y(y) \\
  \hline
  Y = 2 & 0.12 & 0.28 & 0.4 \\
  Y = 3 & 0.18 & 0.42 & 0.6 \\
  f_X(x) & 0.3 & 0.7 & 1 \\
\end{array}
\]
**Example** (discrete): $f(x, y) = P(X = x, Y = y)$.

<table>
<thead>
<tr>
<th>$f(x, y)$</th>
<th>$X = 1$</th>
<th>$X = 2$</th>
<th>$f_Y(y)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y = 2$</td>
<td>0.12</td>
<td>0.28</td>
<td>0.4</td>
</tr>
<tr>
<td>$Y = 3$</td>
<td>0.18</td>
<td>0.42</td>
<td>0.6</td>
</tr>
<tr>
<td>$f_X(x)$</td>
<td>0.3</td>
<td>0.7</td>
<td>1</td>
</tr>
</tbody>
</table>

$X$ and $Y$ are independent since $f(x, y) = f_X(x)f_Y(y)$, $\forall x, y$. □
Example (continuous): Suppose $f(x, y) = 6xy^2$, $0 \leq x \leq 1$, $0 \leq y \leq 1$. 
Example (continuous): Suppose $f(x, y) = 6xy^2$, $0 \leq x \leq 1$, $0 \leq y \leq 1$. After some work (which can be avoided by the next theorem), we can derive

$$f_X(x) = 2x, \text{ if } 0 \leq x \leq 1 \quad \text{and}$$
Example (continuous): Suppose $f(x, y) = 6xy^2$, $0 \leq x \leq 1$, $0 \leq y \leq 1$. After some work (which can be avoided by the next theorem), we can derive

$$f_X(x) = 2x, \text{ if } 0 \leq x \leq 1 \text{ and}$$
$$f_Y(y) = 3y^2, \text{ if } 0 \leq y \leq 1.$$
Example (continuous): Suppose \( f(x, y) = 6xy^2, \ 0 \leq x \leq 1, \ 0 \leq y \leq 1. \) After some work (which can be avoided by the next theorem), we can derive

\[
\begin{align*}
 f_X(x) &= 2x, \text{ if } 0 \leq x \leq 1 \quad \text{and} \\
 f_Y(y) &= 3y^2, \text{ if } 0 \leq y \leq 1. \\
\end{align*}
\]

\( X \) and \( Y \) are independent since \( f(x, y) = f_X(x)f_Y(y), \ \forall x, y. \) \( \square \)
Easy way to tell if $X$ and $Y$ are independent.

**Theorem**: $X$ and $Y$ are independent iff $f(x,y) = a(x)b(y)$, $\forall x,y$, for some functions $a(x)$ and $b(y)$ (not necessarily pdf's).

So if $f(x,y)$ factors into separate functions of $x$ and $y$, then $X$ and $Y$ are independent.

But if there are funny limits, this messes up the factorization, so in that case, $X$ and $Y$ will be dependent — watch out!

**Example**: $f(x,y) = 6xy^2$, $0 \leq x \leq 1$, $0 \leq y \leq 1$.

Take $a(x) = 6x$, $0 \leq x \leq 1$, and $b(y) = y^2$, $0 \leq y \leq 1$.

Thus, $X$ and $Y$ are independent (as above).
Easy way to tell if $X$ and $Y$ are independent.

**Theorem:** $X$ and $Y$ are independent iff $f(x, y) = a(x)b(y)$, $\forall x, y$, for some functions $a(x)$ and $b(y)$ (not necessarily pdf’s).
Easy way to tell if $X$ and $Y$ are independent.

**Theorem**: $X$ and $Y$ are independent iff $f(x, y) = a(x)b(y)$, $\forall x, y$, for some functions $a(x)$ and $b(y)$ (not necessarily pdf’s).

So if $f(x, y)$ factors into separate functions of $x$ and $y$, then $X$ and $Y$ are independent.
Easy way to tell if $X$ and $Y$ are independent.

**Theorem:** $X$ and $Y$ are independent iff $f(x, y) = a(x)b(y)$, $\forall x, y$, for some functions $a(x)$ and $b(y)$ (not necessarily pdf’s).

So if $f(x, y)$ factors into separate functions of $x$ and $y$, then $X$ and $Y$ are independent.

But if there are *funny limits*, this messes up the factorization, so in that case, $X$ and $Y$ will be dependent — watch out!
Easy way to tell if $X$ and $Y$ are independent.

**Theorem:** $X$ and $Y$ are independent iff $f(x, y) = a(x)b(y)$, $\forall x, y$, for some functions $a(x)$ and $b(y)$ (not necessarily pdf’s).

So if $f(x, y)$ factors into separate functions of $x$ and $y$, then $X$ and $Y$ are independent.

But if there are *funny limits*, this messes up the factorization, so in that case, $X$ and $Y$ will be dependent — watch out!

**Example:** $f(x, y) = 6xy^2$, $0 \leq x \leq 1$, $0 \leq y \leq 1$. 
Easy way to tell if $X$ and $Y$ are independent.

**Theorem:** $X$ and $Y$ are independent iff $f(x, y) = a(x)b(y)$, $\forall x, y$, for some functions $a(x)$ and $b(y)$ (not necessarily pdf’s).

So if $f(x, y)$ factors into separate functions of $x$ and $y$, then $X$ and $Y$ are independent.

But if there are *funny limits*, this messes up the factorization, so in that case, $X$ and $Y$ will be dependent — watch out!

**Example:** $f(x, y) = 6xy^2$, $0 \leq x \leq 1$, $0 \leq y \leq 1$. Take

$$a(x) = 6x, \ 0 \leq x \leq 1, \text{ and } b(y) = y^2, \ 0 \leq y \leq 1.$$
Easy way to tell if $X$ and $Y$ are independent. . . .

**Theorem:** $X$ and $Y$ are independent iff $f(x, y) = a(x)b(y)$, $\forall x, y$, for some functions $a(x)$ and $b(y)$ (not necessarily pdf’s).

So if $f(x, y)$ factors into separate functions of $x$ and $y$, then $X$ and $Y$ are independent.

But if there are *funny limits*, this messes up the factorization, so in that case, $X$ and $Y$ will be dependent — watch out!

**Example:** $f(x, y) = 6xy^2$, $0 \leq x \leq 1$, $0 \leq y \leq 1$. Take

$$a(x) = 6x, \quad 0 \leq x \leq 1,$$
$$b(y) = y^2, \quad 0 \leq y \leq 1.$$

Thus, $X$ and $Y$ are independent (as above). $\Box$
Example: \( f(x, y) = \frac{21}{4} x^2 y, \ x^2 \leq y \leq 1. \)
Example: $f(x, y) = \frac{21}{4} x^2 y, \quad x^2 \leq y \leq 1.$

Funny (non-rectangular) limits make factoring into marginals impossible. Thus, $X$ and $Y$ are not independent. $\Box$
**Example:** $f(x, y) = \frac{21}{4} x^2 y, \quad x^2 \leq y \leq 1.$

Funny (non-rectangular) limits make factoring into marginals impossible. Thus, $X$ and $Y$ are *not* independent. □

**Example:** $f(x, y) = \frac{c}{x+y}, \quad 1 \leq x \leq 2, \quad 1 \leq y \leq 3.$
**Example:** \( f(x, y) = \frac{21}{4} x^2 y, \ x^2 \leq y \leq 1. \)

Funny (non-rectangular) limits make factoring into marginals impossible. Thus, \( X \) and \( Y \) are *not* independent. \( \square \)

**Example:** \( f(x, y) = \frac{c}{x+y}, \ 1 \leq x \leq 2, \ 1 \leq y \leq 3. \)

Can’t factor \( f(x, y) \) into functions of \( x \) and \( y \) separately. Thus, \( X \) and \( Y \) are *not* independent. \( \square \)
Example: \( f(x, y) = \frac{21}{4} x^2 y, \quad x^2 \leq y \leq 1. \)

Funny (non-rectangular) limits make factoring into marginals impossible. Thus, \( X \) and \( Y \) are not independent. \( \square \)

Example: \( f(x, y) = \frac{c}{x+y}, \quad 1 \leq x \leq 2, \quad 1 \leq y \leq 3. \)

Can’t factor \( f(x, y) \) into functions of \( x \) and \( y \) separately. Thus, \( X \) and \( Y \) are not independent. \( \square \)

Now that we can figure out if \( X \) and \( Y \) are independent, what can we do with that knowledge?
Consequences of Independence

1. Introduction
2. Marginal Distributions
3. Conditional Distributions
4. Independent Random Variables
5. Consequences of Independence
6. Random Samples
7. Conditional Expectation
8. Double Expectation
9. Honors Class: First-Step Analysis
10. Honors Class: Random Sums of Random Variables
11. Honors Class: Standard Conditioning Argument
12. Covariance and Correlation
13. Correlation and Causation
14. A Couple of Worked Correlation Examples
15. Some Useful Covariance / Correlation Theorems
16. Moment Generating Functions, Revisited
17. Honors Bivariate Functions of Random Variables
Lesson 3.5 — Consequences of Independence

Definition / Theorem (two-dimensional Unconscious Statistician):

Let $h(X,Y)$ be a function of the RVs $X$ and $Y$.

Then

$$E[h(X,Y)] = \sum_x \sum_y h(x,y) f(x,y) \quad \text{discrete}$$

$$\int \int_R h(x,y) f(x,y) dx dy \quad \text{continuous}.$$

Theorem:

Whether or not $X$ and $Y$ are independent, $E[X+Y] = E[X] + E[Y]$. 
Definition/Theorem (two-dimensional Unconscious Statistician): Let $h(X, Y)$ be a function of the RVs $X$ and $Y$. 
Lesson 3.5 — Consequences of Independence

Definition/Theorem (two-dimensional Unconscious Statistician):
Let $h(X, Y)$ be a function of the RVs $X$ and $Y$. Then

$$E[h(X, Y)] = \begin{cases} 
\sum_x \sum_y h(x, y)f(x, y) & \text{discrete} \\
\int_{\mathbb{R}} \int_{\mathbb{R}} h(x, y)f(x, y) \, dx \, dy & \text{continuous.}
\end{cases}$$
Lesson 3.5 — Consequences of Independence

Definition/Theorem (two-dimensional Unconscious Statistician): Let $h(X, Y)$ be a function of the RVs $X$ and $Y$. Then

$$E[h(X, Y)] = \begin{cases} \sum_x \sum_y h(x, y) f(x, y) & \text{discrete} \\ \int_{\mathbb{R}} \int_{\mathbb{R}} h(x, y) f(x, y) \, dx \, dy & \text{continuous.} \end{cases}$$

Theorem: Whether or not $X$ and $Y$ are independent,

$$E[X + Y] = E[X] + E[Y].$$
**Proof** (continuous case):

\[
E[X + Y] = \int_{\mathbb{R}} \int_{\mathbb{R}} (x + y) f(x,y) \, dx \, dy \quad (2-D \text{ LOTUS})
\]

\[
= \int_{\mathbb{R}} \int_{\mathbb{R}} xf(x,y) \, dx \, dy + \int_{\mathbb{R}} \int_{\mathbb{R}} yf(x,y) \, dx \, dy
\]

\[
= \int_{\mathbb{R}} x \int_{\mathbb{R}} f(x,y) \, dy \, dx + \int_{\mathbb{R}} y \int_{\mathbb{R}} f(x,y) \, dx \, dy
\]

\[
= \int_{\mathbb{R}} xf_X(x) \, dx + \int_{\mathbb{R}} yf_Y(y) \, dy
\]

\[
E[X] + E[Y].
\]
Proof (continuous case):

\[
E[X + Y] = \int_{\mathbb{R}} \int_{\mathbb{R}} (x + y) f(x, y) \, dx \, dy \quad \text{(2-D LOTUS)}
\]
Proof (continuous case):

\[ E[X + Y] = \int_{\mathbb{R}} \int_{\mathbb{R}} (x + y) f(x, y) \, dx \, dy \quad \text{(2-D LOTUS)} \]

\[ = \int_{\mathbb{R}} \int_{\mathbb{R}} x f(x, y) \, dx \, dy + \int_{\mathbb{R}} \int_{\mathbb{R}} y f(x, y) \, dx \, dy \]
Proof (continuous case):

\[
E[X + Y] = \int \int_{\mathbb{R}} (x + y) f(x, y) \, dx \, dy \quad \text{(2-D LOTUS)}
\]

\[
= \int \int_{\mathbb{R}} x f(x, y) \, dx \, dy + \int \int_{\mathbb{R}} y f(x, y) \, dx \, dy
\]

\[
= \int_{\mathbb{R}} x \int_{\mathbb{R}} f(x, y) \, dy \, dx + \int_{\mathbb{R}} y \int_{\mathbb{R}} f(x, y) \, dx \, dy
\]
Proof (continuous case):

\[
E[X + Y] = \int_{\mathbb{R}} \int_{\mathbb{R}} (x + y) f(x, y) \, dx \, dy \quad \text{(2-D LOTUS)}
\]

\[
= \int_{\mathbb{R}} \int_{\mathbb{R}} x f(x, y) \, dx \, dy + \int_{\mathbb{R}} \int_{\mathbb{R}} y f(x, y) \, dx \, dy
\]

\[
= \int_{\mathbb{R}} x \int_{\mathbb{R}} f(x, y) \, dy \, dx + \int_{\mathbb{R}} y \int_{\mathbb{R}} f(x, y) \, dx \, dy
\]

\[
= \int_{\mathbb{R}} x f_X(x) \, dx + \int_{\mathbb{R}} y f_Y(y) \, dy
\]
Proof (continuous case):

\[
E[X + Y] = \int_{\mathbb{R}} \int_{\mathbb{R}} (x + y) f(x, y) \, dx \, dy \quad (2 \text{-D LOTUS})
\]

\[
= \int_{\mathbb{R}} \int_{\mathbb{R}} x f(x, y) \, dx \, dy + \int_{\mathbb{R}} \int_{\mathbb{R}} y f(x, y) \, dx \, dy
\]

\[
= \int_{\mathbb{R}} x \int_{\mathbb{R}} f(x, y) \, dy \, dx + \int_{\mathbb{R}} y \int_{\mathbb{R}} f(x, y) \, dx \, dy
\]

\[
= \int_{\mathbb{R}} x f_X(x) \, dx + \int_{\mathbb{R}} y f_Y(y) \, dy
\]

\[
= E[X] + E[Y]. \quad \square
\]
One can generalize this result to more than two random variables.
One can generalize this result to more than two random variables.

**Corollary**: If $X_1, X_2, \ldots, X_n$ are RVs, then

$$E\left[ \sum_{i=1}^{n} X_i \right] = \sum_{i=1}^{n} E[X_i].$$
One can generalize this result to more than two random variables.

**Corollary**: If $X_1, X_2, \ldots, X_n$ are RVs, then

$$E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i].$$

**Proof**: Induction. □
Theorem: If \( X \) and \( Y \) are independent, then \( \mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y] \).
**Theorem:** If $X$ and $Y$ are independent, then $E[XY] = E[X]E[Y]$.

**Proof** (continuous case):

\[
E[XY] = \int_{\mathbb{R}} \int_{\mathbb{R}} xy f_X(x)f_Y(y) \, dx \, dy
\]  

\[
= \int_{\mathbb{R}} x f_X(x) \, dx \int_{\mathbb{R}} y f_Y(y) \, dy
\]  

\[
= E[X]E[Y].
\]
Theorem: If $X$ and $Y$ are independent, then $\mathbb{E}[XY] = \mathbb{E}[X] \mathbb{E}[Y]$.

Proof (continuous case):

\[
\mathbb{E}[XY] = \int_{\mathbb{R}} \int_{\mathbb{R}} xy f(x, y) \, dx \, dy \quad (\text{2-D LOTUS})
\]
**Theorem:** If $X$ and $Y$ are *independent*, then $E[XY] = E[X]E[Y]$.

**Proof** (continuous case):

\[
E[XY] = \int_{\mathbb{R}} \int_{\mathbb{R}} xyf(x,y) \, dx \, dy \quad \text{(2-D LOTUS)}
\]

\[
= \int_{\mathbb{R}} \int_{\mathbb{R}} xyf_X(x)f_Y(y) \, dx \, dy \quad \text{(X and Y are indep)}
\]
Theorem: If $X$ and $Y$ are independent, then $E[XY] = E[X]E[Y]$.

Proof (continuous case):

\[ E[XY] = \int_{\mathbb{R}} \int_{\mathbb{R}} xyf(x, y)\, dx\, dy \quad (2\text{-D LOTUS}) \]

\[ = \int_{\mathbb{R}} \int_{\mathbb{R}} xyf_X(x)f_Y(y)\, dx\, dy \quad (X \text{ and } Y \text{ are indep}) \]

\[ = \left( \int_{\mathbb{R}} xf_X(x)\, dx \right) \left( \int_{\mathbb{R}} yf_Y(y)\, dy \right) \]
Theorem: If $X$ and $Y$ are independent, then $E[XY] = E[X]E[Y]$.

Proof (continuous case):

$$E[XY] = \int_{\mathbb{R}} \int_{\mathbb{R}} xyf(x, y) \, dx \, dy \quad (2\text{-D LOTUS})$$

$$= \int_{\mathbb{R}} \int_{\mathbb{R}} xyf_X(x)f_Y(y) \, dx \, dy \quad (X \text{ and } Y \text{ are indep})$$

$$= \left( \int_{\mathbb{R}} xf_X(x) \, dx \right) \left( \int_{\mathbb{R}} yf_Y(y) \, dy \right)$$

$$= E[X]E[Y]. \quad \Box$$
**Theorem:** If $X$ and $Y$ are independent, then $E[XY] = E[X]E[Y]$.

**Proof** (continuous case):

\[
E[XY] = \int_{\mathbb{R}} \int_{\mathbb{R}} xyf(x, y) \, dx \, dy \quad (2\text{-D LOTUS})
\]

\[
= \int_{\mathbb{R}} \int_{\mathbb{R}} xyf_X(x)f_Y(y) \, dx \, dy \quad (X \text{ and } Y \text{ are indep})
\]

\[
= \left( \int_{\mathbb{R}} x f_X(x) \, dx \right) \left( \int_{\mathbb{R}} y f_Y(y) \, dy \right)
\]

\[
= E[X]E[Y]. \quad \square
\]

**Remark:** The above theorem is *not* necessarily true if $X$ and $Y$ are dependent. See the upcoming discussion on covariance.
**Theorem:** If $X$ and $Y$ are *independent*, then

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y).$$
Theorem: If \( X \) and \( Y \) are \textit{independent}, then

\[
\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y).
\]
**Theorem:** If $X$ and $Y$ are independent, then

\[
\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y).
\]

**Proof:**

\[
\text{Var}(X + Y) = \mathbb{E}[(X + Y)^2] - (\mathbb{E}[X + Y])^2
\]
**Theorem:** If $X$ and $Y$ are *independent*, then

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y).$$

**Proof:**

\[
\text{Var}(X + Y) = \mathbb{E}[(X + Y)^2] - (\mathbb{E}[X + Y])^2 \\
= \mathbb{E}[X^2 + 2XY + Y^2] - (\mathbb{E}[X] + \mathbb{E}[Y])^2
\]
**Theorem**: If \( X \) and \( Y \) are *independent*, then

\[
\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y).
\]

**Proof**: 

\[
\text{Var}(X + Y) = E[(X + Y)^2] - (E[X + Y])^2 \\
= E[X^2 + 2XY + Y^2] - (E[X] + E[Y])^2 \\
= E[X^2] + 2E[XY] + E[Y^2] - \left\{ (E[X])^2 + 2E[X]E[Y] + (E[Y])^2 \right\}
\]

Remark: The assumption of independence really is important here. If \( X \) and \( Y \) aren't independent, then the result might not hold!
**Theorem:** If \( X \) and \( Y \) are *independent*, then

\[
\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y).
\]

**Proof:**

\[
\begin{align*}
\text{Var}(X + Y) &= \mathbb{E}[(X + Y)^2] - (\mathbb{E}[X + Y])^2 \\
&= \mathbb{E}[X^2 + 2XY + Y^2] - (\mathbb{E}[X] + \mathbb{E}[Y])^2 \\
&= \mathbb{E}[X^2] + 2\mathbb{E}[XY] + \mathbb{E}[Y^2] - \left\{ (\mathbb{E}[X])^2 + 2\mathbb{E}[X]\mathbb{E}[Y] + (\mathbb{E}[Y])^2 \right\} \\
&= \mathbb{E}[X^2] + 2\mathbb{E}[X]\mathbb{E}[Y] + \mathbb{E}[Y^2] - (\mathbb{E}[X])^2 - 2\mathbb{E}[X]\mathbb{E}[Y] - (\mathbb{E}[Y])^2 \\
&= \mathbb{E}[X^2] + 2\mathbb{E}[X]\mathbb{E}[Y] + \mathbb{E}[Y^2] - (\mathbb{E}[X])^2 - 2\mathbb{E}[X]\mathbb{E}[Y] - (\mathbb{E}[Y])^2
\end{align*}
\]

(since \( X \) and \( Y \) are independent)
Theorem: If $X$ and $Y$ are independent, then

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y).$$

Proof:

$$\text{Var}(X + Y) = E[(X + Y)^2] - (E[X + Y])^2$$

$$= E[X^2 + 2XY + Y^2] - (E[X] + E[Y])^2$$

$$= E[X^2] + 2E[XY] + E[Y^2] - \left\{ (E[X])^2 + 2E[X]E[Y] + (E[Y])^2 \right\}$$


(since $X$ and $Y$ are independent)

$$= E[X^2] - (E[X])^2 + E[Y^2] - (E[Y])^2. \quad \square$$
Theorem: If \( X \) and \( Y \) are \textit{independent}, then

\[
\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y).
\]

Proof:

\[
\begin{align*}
\text{Var}(X + Y) &= E[(X + Y)^2] - (E[X + Y])^2 \\
&= E[X^2 + 2XY + Y^2] - (E[X] + E[Y])^2 \\
&= E[X^2] + 2E[XY] + E[Y^2] - \left\{ (E[X])^2 + 2E[X]E[Y] + (E[Y])^2 \right\} \\
&\quad \text{(since } X \text{ and } Y \text{ are independent)} \\
&= E[X^2] - (E[X])^2 + E[Y^2] - (E[Y])^2.
\end{align*}
\]

Remark: The assumption of independence really is important here. If \( X \) and \( Y \) aren’t independent, then the result might not hold!
Can generalize…
Can generalize…

**Corollary**: If $X_1, X_2, \ldots, X_n$ are independent RVs, then

$$\text{Var} \left( \sum_{i=1}^{n} X_i \right) = \sum_{i=1}^{n} \text{Var}(X_i).$$
Can generalize...

**Corollary**: If $X_1, X_2, \ldots, X_n$ are **independent** RVs, then

$$\text{Var} \left( \sum_{i=1}^{n} X_i \right) = \sum_{i=1}^{n} \text{Var}(X_i).$$

**Proof**: Induction. □
Can generalize...

**Corollary**: If $X_1, X_2, \ldots, X_n$ are independent RVs, then

$$\text{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \text{Var}(X_i).$$

**Proof**: Induction. □

**Corollary**: If $X_1, X_2, \ldots, X_n$ are independent RVs, then

$$\text{Var}\left(\sum_{i=1}^{n} a_i X_i + b\right) = \sum_{i=1}^{n} a_i^2 \text{Var}(X_i).$$
<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Introduction</td>
</tr>
<tr>
<td>2</td>
<td>Marginal Distributions</td>
</tr>
<tr>
<td>3</td>
<td>Conditional Distributions</td>
</tr>
<tr>
<td>4</td>
<td>Independent Random Variables</td>
</tr>
<tr>
<td>5</td>
<td>Consequences of Independence</td>
</tr>
<tr>
<td>6</td>
<td>Random Samples</td>
</tr>
<tr>
<td>7</td>
<td>Conditional Expectation</td>
</tr>
<tr>
<td>8</td>
<td>Double Expectation</td>
</tr>
<tr>
<td>9</td>
<td>Honors Class: First-Step Analysis</td>
</tr>
<tr>
<td>10</td>
<td>Honors Class: Random Sums of Random Variables</td>
</tr>
<tr>
<td>11</td>
<td>Honors Class: Standard Conditioning Argument</td>
</tr>
<tr>
<td>12</td>
<td>Covariance and Correlation</td>
</tr>
<tr>
<td>13</td>
<td>Correlation and Causation</td>
</tr>
<tr>
<td>14</td>
<td>A Couple of Worked Correlation Examples</td>
</tr>
<tr>
<td>15</td>
<td>Some Useful Covariance / Correlation Theorems</td>
</tr>
<tr>
<td>16</td>
<td>Moment Generating Functions, Revisited</td>
</tr>
<tr>
<td>17</td>
<td>Honors Bivariate Functions of Random Variables</td>
</tr>
</tbody>
</table>
Definition: $X_1, X_2, ..., X_n$ form a random sample if $X_i$'s are all independent. Each $X_i$ has the same pmf/pdf $f(x)$.

Notation: $X_1, ..., X_n \ iid \sim f(x)$ ("independent and identically distributed").

Example/Theorem: Suppose $X_1, ..., X_n \ iid \sim f(x)$, with $E[X_i] = \mu$, and $\text{Var}(X_i) = \sigma^2$. Define the sample mean as $\bar{X} \equiv \frac{1}{n} \sum_{i=1}^{n} X_i$. Then $E[\bar{X}] = \frac{1}{n} \sum_{i=1}^{n} E[X_i] = \frac{1}{n} \sum_{i=1}^{n} \mu = \mu$. So the mean of $\bar{X}$ is the same as the mean of $X_i$.
Lesson 3.6 — Random Samples

Definition: \( X_1, X_2, \ldots, X_n \) form a random sample if

\[ X_i \]’s are all independent.
Each \( X_i \) has the same pmf/pdf \( f(x) \).

Notation: \( X_1, \ldots, X_n \text{ iid} \sim f(x) \) ("independent and identically distributed").

Example/Theorem: Suppose \( X_1, \ldots, X_n \text{ iid} \sim f(x) \), with \( E[X_i] = \mu \), and \( \text{Var}(X_i) = \sigma^2 \). Define the sample mean as \( \bar{X} \equiv \frac{\sum_{i=1}^{n} X_i}{n} \). Then

\[ E[\bar{X}] = E[\frac{1}{n} \sum_{i=1}^{n} X_i] = \frac{1}{n} \sum_{i=1}^{n} E[X_i] = \frac{1}{n} \sum_{i=1}^{n} \mu = \mu. \]
So the mean of \( \bar{X} \) is the same as the mean of \( X_i \).
Lesson 3.6 — Random Samples

**Definition:** $X_1, X_2, \ldots, X_n$ form a **random sample** if

- $X_i$’s are all **independent**.
Lesson 3.6 — Random Samples

**Definition:** $X_1, X_2, \ldots, X_n$ form a **random sample** if

- $X_i$’s are all **independent**.
- Each $X_i$ has the same pmf/pdf $f(x)$. 
**Definition:** $X_1, X_2, \ldots, X_n$ form a **random sample** if

- $X_i$’s are all **independent**.
- Each $X_i$ has the same pmf/pdf $f(x)$.

**Notation:** $X_1, \ldots, X_n \overset{iid}{\sim} f(x)$ (“**independent and identically distributed**”).
Lesson 3.6 — Random Samples

**Definition:** $X_1, X_2, \ldots, X_n$ form a **random sample** if
- $X_i$’s are all *independent*.
- Each $X_i$ has the same pmf/pdf $f(x)$.

**Notation:** $X_1, \ldots, X_n \overset{iid}{\sim} f(x)$ (“independent and identically distributed”).

**Example/Theorem:** Suppose $X_1, \ldots, X_n \overset{iid}{\sim} f(x)$, with $E[X_i] = \mu$, and $\text{Var}(X_i) = \sigma^2$. Define the **sample mean** as $\bar{X} \equiv \sum_{i=1}^{n} X_i/n$. Then
Random Samples

Lesson 3.6 — Random Samples

Definition: $X_1, X_2, \ldots, X_n$ form a random sample if

- $X_i$’s are all independent.
- Each $X_i$ has the same pmf/pdf $f(x)$.

Notation: $X_1, \ldots, X_n \overset{iid}{\sim} f(x)$ ("independent and identically distributed").

Example/Theorem: Suppose $X_1, \ldots, X_n \overset{iid}{\sim} f(x)$, with $E[X_i] = \mu$, and $\text{Var}(X_i) = \sigma^2$. Define the sample mean as $\bar{X} \equiv \sum_{i=1}^{n} X_i/n$. Then

$$E[\bar{X}]$$
Lesson 3.6 — Random Samples

**Definition:** $X_1, X_2, \ldots, X_n$ form a **random sample** if

- $X_i$’s are all *independent*.
- Each $X_i$ has the same pmf/pdf $f(x)$.

**Notation:** $X_1, \ldots, X_n \text{ iid } \sim f(x)$ ("independent and identically distributed").

**Example/Theorem:** Suppose $X_1, \ldots, X_n \text{ iid } \sim f(x)$, with $E[X_i] = \mu$, and $\text{Var}(X_i) = \sigma^2$. Define the **sample mean** as $\bar{X} \equiv \frac{1}{n} \sum_{i=1}^{n} X_i$. Then

$$
E[\bar{X}] = E \left[ \frac{1}{n} \sum_{i=1}^{n} X_i \right]
$$
Lesson 3.6 — Random Samples

Definition: $X_1, X_2, \ldots, X_n$ form a **random sample** if

- $X_i$’s are all **independent**.
- Each $X_i$ has the same pmf/pdf $f(x)$.

Notation: $X_1, \ldots, X_n \overset{iid}{\sim} f(x)$ ("independent and identically distributed").

Example/Theorem: Suppose $X_1, \ldots, X_n \overset{iid}{\sim} f(x)$, with $E[X_i] = \mu$, and $\text{Var}(X_i) = \sigma^2$. Define the **sample mean** as $\bar{X} \equiv \frac{1}{n} \sum_{i=1}^{n} X_i$. Then

$$E[\bar{X}] = E\left[\frac{1}{n} \sum_{i=1}^{n} X_i\right] = \frac{1}{n} \sum_{i=1}^{n} E[X_i]$$
**Definition**: $X_1, X_2, \ldots, X_n$ form a random sample if

- $X_i$’s are all independent.
- Each $X_i$ has the same pmf/pdf $f(x)$.

**Notation**: $X_1, \ldots, X_n \overset{iid}{\sim} f(x)$ (“independent and identically distributed”).

**Example/Theorem**: Suppose $X_1, \ldots, X_n \overset{iid}{\sim} f(x)$, with $E[X_i] = \mu$, and $\text{Var}(X_i) = \sigma^2$. Define the sample mean as $\bar{X} \equiv \sum_{i=1}^{n} X_i / n$. Then

$$E[\bar{X}] = E\left[\frac{1}{n} \sum_{i=1}^{n} X_i\right] = \frac{1}{n} \sum_{i=1}^{n} E[X_i] = \frac{1}{n} \sum_{i=1}^{n} \mu$$
Lesson 3.6 — Random Samples

Definition: $X_1, X_2, \ldots, X_n$ form a random sample if
- $X_i$’s are all independent.
- Each $X_i$ has the same pmf/pdf $f(x)$.

Notation: $X_1, \ldots, X_n \sim^\text{iid} f(x)$ (“independent and identically distributed”).

Example/Theorem: Suppose $X_1, \ldots, X_n \sim f(x)$, with $E[X_i] = \mu$, and $\text{Var}(X_i) = \sigma^2$. Define the sample mean as $\bar{X} \equiv \frac{1}{n} \sum_{i=1}^n X_i$. Then

$$E[\bar{X}] = E\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \frac{1}{n} \sum_{i=1}^n E[X_i] = \frac{1}{n} \sum_{i=1}^n \mu = \mu.$$
Lesson 3.6 — Random Samples

Definition: $X_1, X_2, \ldots, X_n$ form a random sample if
- $X_i$’s are all independent.
- Each $X_i$ has the same pmf/pdf $f(x)$.

Notation: $X_1, \ldots, X_n \sim f(x)$ (“independent and identically distributed”).

Example/Theorem: Suppose $X_1, \ldots, X_n \sim f(x)$, with $\mathbb{E}[X_i] = \mu$, and $\text{Var}(X_i) = \sigma^2$. Define the sample mean as $\bar{X} \equiv \sum_{i=1}^{n} X_i / n$. Then

$$
\mathbb{E}[\bar{X}] = \mathbb{E} \left[ \frac{1}{n} \sum_{i=1}^{n} X_i \right] = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[X_i] = \frac{1}{n} \sum_{i=1}^{n} \mu = \mu.
$$

So the mean of $\bar{X}$ is the same as the mean of $X_i$. \qed
Meanwhile, how about the *variance* of the sample mean?
Meanwhile, how about the variance of the sample mean?

\[
\text{Var}(\bar{X}) = \text{Var}\left( \frac{1}{n} \sum_{i=1}^{n} X_i \right)
\]
Meanwhile, how about the variance of the sample mean?

\[
\text{Var}(\bar{X}) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^{n} X_i\right)
\]

\[
= \frac{1}{n^2} \text{Var}\left(\sum_{i=1}^{n} X_i\right)
\]

So the mean of \(\bar{X}\) is the same as the mean of \(X_i\), but the variance decreases! This makes \(\bar{X}\) a great estimator for \(\mu\) (which is usually unknown in practice); the result is referred to as the Law of Large Numbers. Stay tuned.
Meanwhile, how about the variance of the sample mean?

\[
\text{Var}(\bar{X}) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^{n} X_i\right)
\]

\[
= \frac{1}{n^2} \text{Var}\left(\sum_{i=1}^{n} X_i\right)
\]

\[
= \frac{1}{n^2} \sum_{i=1}^{n} \text{Var}(X_i) \quad (X_i's \ text{indep})
\]
Meanwhile, how about the variance of the sample mean?

\[
\text{Var}(\bar{X}) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^{n} X_i \right)
\]

\[
= \frac{1}{n^2} \text{Var}\left(\sum_{i=1}^{n} X_i \right)
\]

\[
= \frac{1}{n^2} \sum_{i=1}^{n} \text{Var}(X_i) \quad (X_i's \text{ indep})
\]

\[
= \frac{1}{n^2} \sum_{i=1}^{n} \sigma^2 = \sigma^2/n.
\]
Meanwhile, how about the variance of the sample mean?

\[
\text{Var}(\bar{X}) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^{n} X_i \right) = \frac{1}{n^2} \text{Var}\left(\sum_{i=1}^{n} X_i \right) = \frac{1}{n^2} \sum_{i=1}^{n} \text{Var}(X_i) \quad (X_i's \text{ indep}) = \frac{1}{n^2} \sum_{i=1}^{n} \sigma^2 = \frac{\sigma^2}{n}.
\]

So the mean of \( \bar{X} \) is the same as the mean of \( X_i \), but the variance decreases! This makes \( \bar{X} \) a great estimator for \( \mu \) (which is usually unknown in practice); the result is referred to as the **Law of Large Numbers**. Stay tuned.
Conditional Expectation

Introduction
Marginal Distributions
Conditional Distributions
Independent Random Variables
Consequences of Independence
Random Samples

Conditional Expectation

Double Expectation
Honors Class: First-Step Analysis
Honors Class: Random Sums of Random Variables
Honors Class: Standard Conditioning Argument
Covariance and Correlation
Correlation and Causation
A Couple of Worked Correlation Examples
Some Useful Covariance / Correlation Theorems
Moment Generating Functions, Revisited
Honors Bivariate Functions of Random Variables
The Next Few Lessons:

- Conditional expectation — definition and examples.
- “Double” expectation — a very cool theorem.
- Honors Class: First-step analysis.
- Honors Class: Random sums of random variables.
- Honors Class: The standard conditioning argument and its applications.
Consider the usual definition of expectation. (E.g., what’s the average weight of a male?)
Consider the usual definition of expectation. (E.g., what’s the average weight of a male?)

\[
E[Y] = \begin{cases} 
\sum_y y f(y) & \text{discrete} \\
\int_{\mathbb{R}} y f(y) \, dy & \text{continuous.}
\end{cases}
\]
Consider the usual definition of expectation. (E.g., what’s the average weight of a male?)

$$E[Y] = \begin{cases} \sum_{y} y f(y) & \text{discrete} \\ \int_{\mathbb{R}} y f(y) \, dy & \text{continuous}. \end{cases}$$

Now suppose we’re interested in the average weight of a 6' tall male.
Consider the usual definition of expectation. (E.g., what’s the average weight of a male?)

\[
E[Y] = \begin{cases} 
\sum_y y f(y) & \text{discrete} \\
\int_{\mathbb{R}} y f(y) \, dy & \text{continuous.}
\end{cases}
\]

Now suppose we’re interested in the average weight of a 6' tall male.

\(f(y|x)\) is the conditional pmf/pdf of \(Y\) given \(X = x\).
Consider the usual definition of expectation. (E.g., what’s the average weight of a male?)

\[ E[Y] = \begin{cases} 
\sum_y y f(y) & \text{discrete} \\ 
\int_{\mathbb{R}} y f(y) \, dy & \text{continuous.} 
\end{cases} \]

Now suppose we’re interested in the average weight of a 6' tall male.

\( f(y|x) \) is the conditional pmf/pdf of \( Y \) given \( X = x \).

**Definition:** The conditional expectation of \( Y \) given \( X = x \) is
Consider the usual definition of expectation. (E.g., what’s the average weight of a male?)

$$E[Y] = \begin{cases} \sum_y y f(y) & \text{discrete} \\ \int_{\mathbb{R}} y f(y) \, dy & \text{continuous.} \end{cases}$$

Now suppose we’re interested in the average weight of a 6' tall male.

$f(y|x)$ is the conditional pmf/pdf of $Y$ given $X = x$.

**Definition:** The **conditional expectation** of $Y$ given $X = x$ is

$$E[Y|X = x] \equiv \begin{cases} \sum_y y f(y|x) & \text{discrete} \\ \int_{\mathbb{R}} y f(y|x) \, dy & \text{continuous.} \end{cases}$$
Consider the usual definition of expectation. (E.g., what’s the average weight of a male?)

$$E[Y] = \begin{cases} \sum_y y f(y) & \text{discrete} \\ \int_{\mathbb{R}} y f(y) \, dy & \text{continuous.} \end{cases}$$

Now suppose we’re interested in the average weight of a 6' tall male.

$f(y|x)$ is the conditional pmf/pdf of $Y$ given $X = x$.

**Definition**: The **conditional expectation** of $Y$ given $X = x$ is

$$E[Y|X = x] \equiv \begin{cases} \sum_y y f(y|x) & \text{discrete} \\ \int_{\mathbb{R}} y f(y|x) \, dy & \text{continuous.} \end{cases}$$

Note that $E[Y|X = x]$ is a function of $x$. 
Discrete Example:
## Conditional Expectation

### Discrete Example:

<table>
<thead>
<tr>
<th></th>
<th>( f(x, y) )</th>
<th>( X = 0 )</th>
<th>( X = 3 )</th>
<th>( f_Y(y) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Y = 2 )</td>
<td>0.11</td>
<td><strong>0.34</strong></td>
<td></td>
<td>0.45</td>
</tr>
<tr>
<td>( Y = 5 )</td>
<td>0.00</td>
<td><strong>0.05</strong></td>
<td></td>
<td>0.05</td>
</tr>
<tr>
<td>( Y = 10 )</td>
<td>0.29</td>
<td><strong>0.21</strong></td>
<td></td>
<td>0.50</td>
</tr>
<tr>
<td>( f_X(x) )</td>
<td>0.40</td>
<td><strong>0.60</strong></td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

The unconditional expectation is \( \mathbb{E}[Y] = \sum y f_Y(y) = 2(0.45) + 5(0.05) + 10(0.50) = 6.15 \).
### Discrete Example:

<table>
<thead>
<tr>
<th>( f(x, y) )</th>
<th>( X = 0 )</th>
<th>( X = 3 )</th>
<th>( f_Y(y) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Y = 2 )</td>
<td>0.11</td>
<td>0.34</td>
<td>0.45</td>
</tr>
<tr>
<td>( Y = 5 )</td>
<td>0.00</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>( Y = 10 )</td>
<td>0.29</td>
<td>0.21</td>
<td>0.50</td>
</tr>
<tr>
<td>( f_X(x) )</td>
<td>0.40</td>
<td>0.60</td>
<td>1</td>
</tr>
</tbody>
</table>

The *unconditional* expectation is

\[
E[Y] = \sum_y y f_Y(y)
\]
Discrete Example:

<table>
<thead>
<tr>
<th>$f(x, y)$</th>
<th>$X = 0$</th>
<th>$X = 3$</th>
<th>$f_Y(y)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y = 2$</td>
<td>0.11</td>
<td>0.34</td>
<td>0.45</td>
</tr>
<tr>
<td>$Y = 5$</td>
<td>0.00</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>$Y = 10$</td>
<td>0.29</td>
<td>0.21</td>
<td>0.50</td>
</tr>
<tr>
<td>$f_X(x)$</td>
<td>0.40</td>
<td>0.60</td>
<td>1</td>
</tr>
</tbody>
</table>

The *unconditional* expectation is

$$E[Y] = \sum_y y f_Y(y) = 2(0.45) + 5(0.05) + 10(0.50) = 6.15.$$
But conditional on $X = 3$, we have
But conditional on $X = 3$, we have

$$f(y|x = 3) = \frac{f(3, y)}{f_X(3)}$$
But conditional on $X = 3$, we have

$$f(y | x = 3) = \frac{f(3, y)}{f_X(3)} = \frac{f(3, y)}{0.60}$$
But conditional on $X = 3$, we have

$$f(y|x = 3) = \frac{f(3, y)}{f_X(3)} = \frac{f(3, y)}{0.60} = \begin{cases} 
\frac{34}{60} & \text{if } y = 2 \\
\frac{5}{60} & \text{if } y = 5 \\
\frac{21}{60} & \text{if } y = 10.
\end{cases}$$
But conditional on $X = 3$, we have

$$f(y|x = 3) = f(3, y) / f_X(3) = f(3, y) / 0.60 = \begin{cases} \frac{34}{60} & \text{if } y = 2 \\ \frac{5}{60} & \text{if } y = 5 \\ \frac{21}{60} & \text{if } y = 10. \end{cases}$$

So the expectation conditional on $X = 3$ is
But conditional on \( X = 3 \), we have

\[
f(y|x = 3) = \frac{f(3, y)}{f_X(3)} = \frac{f(3, y)}{0.60} = \begin{cases} 
\frac{34}{60} & \text{if } y = 2 \\
\frac{5}{60} & \text{if } y = 5 \\
\frac{21}{60} & \text{if } y = 10.
\end{cases}
\]

So the expectation conditional on \( X = 3 \) is

\[
E[Y|X = 3] = \sum_y y f(y|3)
\]
But conditional on $X = 3$, we have

$$f(y|x = 3) = \frac{f(3,y)}{f_X(3)} = \frac{f(3,y)}{0.60} = \begin{cases} 
\frac{34}{60} & \text{if } y = 2 \\
\frac{5}{60} & \text{if } y = 5 \\
\frac{21}{60} & \text{if } y = 10. 
\end{cases}$$

So the expectation conditional on $X = 3$ is

$$E[Y|X = 3] = \sum_y yf(y|3) = 2(34/60) + 5(5/60) + 10(21/60) = 5.05.$$
But conditional on $X = 3$, we have

$$f(y|x = 3) = \frac{f(3, y)}{f_X(3)} = \frac{f(3, y)}{0.60} = \begin{cases} \frac{34}{60} & \text{if } y = 2 \\ \frac{5}{60} & \text{if } y = 5 \\ \frac{21}{60} & \text{if } y = 10. \end{cases}$$

So the expectation conditional on $X = 3$ is

$$E[Y|X = 3] = \sum_y y f(y|3)$$

$$= 2(34/60) + 5(5/60) + 10(21/60)$$

$$= 5.05.$$ 

This compares to the unconditional expectation $E[Y] = 6.15$. So information that $X = 3$ pushes the conditional expected value of $Y$ down to 5.05. □
Old Continuous Example:

\[ f(x, y) = \begin{cases} 4x^2y, & x^2 \leq y \leq 1. \end{cases} \]

Recall that \( f(y|x) = \begin{cases} y - x^4, & x^2 \leq y \leq 1. \end{cases} \)

Thus, \( \mathbb{E}[Y|x] = \int_{\mathbb{R}} y f(y|x) \, dy = \frac{2}{3} \cdot \frac{6^3}{6^4} = \frac{1}{15} \)
Old Continuous Example:

\[ f(x, y) = \frac{21}{4} x^2 y, \quad \text{if } x^2 \leq y \leq 1. \]
Old Continuous Example:

\[ f(x, y) = \frac{21}{4} x^2 y, \quad \text{if } x^2 \leq y \leq 1. \]

Recall that

\[ f(y|x) = \frac{2y}{1 - x^4} \quad \text{if } x^2 \leq y \leq 1. \]
Old Continuous Example:

\[ f(x, y) = \frac{21}{4} x^2 y, \quad \text{if } x^2 \leq y \leq 1. \]

Recall that

\[ f(y|x) = \frac{2y}{1 - x^4}, \quad \text{if } x^2 \leq y \leq 1. \]

Thus,

\[ \mathbb{E}[Y|x] \]
Conditional Expectation

Old Continuous Example:

\[ f(x, y) = \frac{21}{4} x^2 y, \quad \text{if } x^2 \leq y \leq 1. \]

Recall that

\[ f(y|x) = \frac{2y}{1 - x^4} \quad \text{if } x^2 \leq y \leq 1. \]

Thus,

\[ E[Y|x] = \int_{\mathbb{R}} y f(y|x) \, dy \]
Old Continuous Example:

\[ f(x, y) = \frac{21}{4} x^2 y, \quad \text{if } x^2 \leq y \leq 1. \]

Recall that

\[ f(y|x) = \frac{2y}{1 - x^4} \quad \text{if } x^2 \leq y \leq 1. \]

Thus,

\[
E[Y|x] = \int_{\mathbb{R}} y f(y|x) \, dy = \frac{2}{1 - x^4} \int_{x^2}^{1} y^2 \, dy
\]
Old Continuous Example:

\[ f(x, y) = \frac{21}{4} x^2 y, \quad \text{if} \ x^2 \leq y \leq 1. \]

Recall that

\[ f(y|x) = \frac{2y}{1 - x^4} \quad \text{if} \ x^2 \leq y \leq 1. \]

Thus,

\[ E[Y|x] = \int_{\mathbb{R}} yf(y|x) \, dy = \frac{2}{1 - x^4} \int_{x^2}^{1} y^2 \, dy = \frac{2}{3} \cdot \frac{1 - x^6}{1 - x^4}. \]
Old Continuous Example:

\[ f(x, y) = \frac{21}{4} x^2 y, \quad \text{if } x^2 \leq y \leq 1. \]

Recall that

\[ f(y|x) = \frac{2y}{1 - x^4}, \quad \text{if } x^2 \leq y \leq 1. \]

Thus,

\[
E[Y|x] = \int_{\mathbb{R}} y f(y|x) \, dy = \frac{2}{1 - x^4} \int_{x^2}^{1} y^2 \, dy = \frac{2}{3} \cdot \frac{1 - x^6}{1 - x^4}.
\]

So, e.g., \( E[Y|X = 0.5] = \frac{2}{3} \cdot \frac{63}{64} / \frac{15}{16} = 0.70. \)
Double Expectation

1. Introduction
2. Marginal Distributions
3. Conditional Distributions
4. Independent Random Variables
5. Consequences of Independence
6. Random Samples
7. Conditional Expectation
8. Double Expectation
9. Honors Class: First-Step Analysis
10. Honors Class: Random Sums of Random Variables
11. Honors Class: Standard Conditioning Argument
12. Covariance and Correlation
13. Correlation and Causation
14. A Couple of Worked Correlation Examples
15. Some Useful Covariance / Correlation Theorems
16. Moment Generating Functions, Revisited
17. Honors Bivariate Functions of Random Variables
Theorem (double expectation):

\[ E[E(Y|X)] = E[Y] \]

Remarks: Yikes, what the heck is this!?

The expected value (averaged over all \( X \)’s) of the conditional expected value (of \( Y | X \)) is the plain old expected value (of \( Y \)).

Think of the outside expected value as the expected value of \( h(X) = E(Y|X) \). Then LOTUS miraculously gives us \( E[Y] \).

Believe it or not, sometimes it's easier to calculate \( E[Y] \) indirectly by using our double expectation trick.
Lesson 3.8 — Double Expectation

Theorem (double expectation):

\[ E[E(Y|X)] = E[Y]. \]
Lesson 3.8 — Double Expectation

Theorem (double expectation):

\[ E[E(Y|X)] = E[Y]. \]

Remarks: Yikes, what the heck is this!?
**Theorem (double expectation):**

\[ E[E(Y|X)] = E[Y]. \]

**Remarks:** Yikes, what the heck is this!?

The expected value (averaged over all \( X \)'s) of the conditional expected value (of \( Y|X \)) is the plain old expected value (of \( Y \)).
Lesson 3.8 — Double Expectation

Theorem (double expectation):

\[ E[E(Y|X)] = E[Y]. \]

Remarks: Yikes, what the heck is this!?

The expected value (averaged over all \( X \)'s) of the conditional expected value (of \( Y|X \)) is the plain old expected value (of \( Y \)).

Think of the outside expected value as the expected value of \( h(X) = E(Y|X) \). Then LOTUS miraculously gives us \( E[Y] \).
Lesson 3.8 — Double Expectation

Theorem (double expectation):

\[ E[E(Y|X)] = E[Y]. \]

Remarks: Yikes, what the heck is this!?

The expected value (averaged over all \( X \)'s) of the conditional expected value (of \( Y|X \)) is the plain old expected value (of \( Y \)).

Think of the outside expected value as the expected value of \( h(X) = E(Y|X) \). Then LOTUS miraculously gives us \( E[Y] \).

Believe it or not, sometimes it’s easier to calculate \( E[Y] \) indirectly by using our double expectation trick.
**Proof** (continuous case): By the Unconscious Statistician,
**Proof** (continuous case): By the Unconscious Statistician,

\[
E[E(Y|X)] = \int_{\mathbb{R}} E(Y|x) f_X(x) \, dx
\]
**Proof** (continuous case): By the Unconscious Statistician,

\[
E[E(Y|X)] = \int_R E(Y|x) f_X(x) \, dx
\]

\[
= \int_R \left( \int_R y f(y|x) \, dy \right) f_X(x) \, dx
\]

= \int_R y f_Y(y) \, dy = E[Y].
**Proof** (continuous case): By the Unconscious Statistician,

\[
E[E(Y \mid X)] = \int_{\mathbb{R}} E(Y \mid x) f_X(x) \, dx
\]

\[
= \int_{\mathbb{R}} \left( \int_{\mathbb{R}} y f(y \mid x) \, dy \right) f_X(x) \, dx
\]

\[
= \int_{\mathbb{R}} \int_{\mathbb{R}} y f(y \mid x) f_X(x) \, dx \, dy
\]

\[
= E[Y]
\]
Proof (continuous case): By the Unconscious Statistician,

\[
E[E(Y|X)] = \int_{\mathbb{R}} E(Y|x) f_X(x) \, dx
\]

\[
= \int_{\mathbb{R}} \left( \int_{\mathbb{R}} y f(y|x) \, dy \right) f_X(x) \, dx
\]

\[
= \int_{\mathbb{R}} \int_{\mathbb{R}} y f(y|x) f_X(x) \, dx \, dy
\]

\[
= \int_{\mathbb{R}} y \int_{\mathbb{R}} f(x,y) \, dx \, dy
\]
**Proof** (continuous case): By the Unconscious Statistician,

\[
E[E(Y|X)] = \int_{\mathbb{R}} E(Y|x) f_X(x) \, dx
\]

\[
= \int_{\mathbb{R}} \left( \int_{\mathbb{R}} y f(y|x) \, dy \right) f_X(x) \, dx
\]

\[
= \int_{\mathbb{R}} \int_{\mathbb{R}} y f(y|x) f_X(x) \, dx \, dy
\]

\[
= \int_{\mathbb{R}} y \int_{\mathbb{R}} f(x,y) \, dx \, dy
\]

\[
= \int_{\mathbb{R}} y f_Y(y) \, dy
\]
**Proof** (continuous case): By the Unconscious Statistician,

\[
\begin{align*}
\mathbb{E}[\mathbb{E}(Y \mid X)] &= \int_{\mathbb{R}} \mathbb{E}(Y \mid x) f_X(x) \, dx \\
&= \int_{\mathbb{R}} \left( \int_{\mathbb{R}} y f(y \mid x) \, dy \right) f_X(x) \, dx \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} y f(y \mid x) f_X(x) \, dx \, dy \\
&= \int_{\mathbb{R}} y \int_{\mathbb{R}} f(x, y) \, dx \, dy \\
&= \int_{\mathbb{R}} y f_Y(y) \, dy \\
&= \mathbb{E}[Y]. \quad \square
\end{align*}
\]
Old Example: Suppose $f(x, y) = \frac{21}{4}x^2y$, if $x^2 \leq y \leq 1$.

Find $E[Y]$ two ways.
**Old Example:** Suppose \( f(x, y) = \frac{21}{4} x^2 y, \) if \( x^2 \leq y \leq 1. \)

Find \( E[Y] \) *two ways.*

By previous examples, we know that
**Old Example:** Suppose $f(x, y) = \frac{21}{4} x^2 y$, if $x^2 \leq y \leq 1$.

Find $E[Y]$ *two ways*.

By previous examples, we know that

$$f_X(x) = \frac{21}{8} x^2 (1 - x^4), \text{ if } -1 \leq x \leq 1$$
Old Example: Suppose $f(x, y) = \frac{21}{4} x^2 y$, if $x^2 \leq y \leq 1$.

Find $E[Y]$ two ways.

By previous examples, we know that

\[
  f_X(x) = \frac{21}{8} x^2 (1 - x^4), \quad \text{if } -1 \leq x \leq 1
\]

\[
  f_Y(y) = \frac{7}{2} y^{5/2}, \quad \text{if } 0 \leq y \leq 1
\]
Old Example: Suppose \( f(x, y) = \frac{21}{4} x^2 y \), if \( x^2 \leq y \leq 1 \).

Find \( E[Y] \) two ways.

By previous examples, we know that

\[
    f_X(x) = \frac{21}{8} x^2 (1 - x^4), \quad \text{if } -1 \leq x \leq 1
\]

\[
    f_Y(y) = \frac{7}{2} y^{5/2}, \quad \text{if } 0 \leq y \leq 1
\]

\[
    E[Y|X] = \frac{2}{3} \cdot \frac{1 - x^6}{1 - x^4}.
\]
Solution #1 (old, boring way):

\[
E[Y] = \int_{\mathbb{R}} y f_Y(y) \, dy = \int_{1}^{7} \frac{y}{2} \, dy = \frac{7}{9}.
\]
Double Expectation

Solution #1 (old, boring way):

\[
E[Y] = \int_{\mathbb{R}} y f_Y(y) \, dy = \int_{0}^{1} \frac{7}{2}y^{7/2} \, dy = \frac{7}{9}.
\]
Solution #1 (old, boring way):

$$E[Y] = \int_{\mathbb{R}} y f_Y(y) \, dy = \int_{0}^{1} \frac{7}{2} y^{7/2} \, dy = \frac{7}{9}.$$ 

Solution #2 (new, exciting way):

$$E[Y] = E[E(Y|X)] = \int_{\mathbb{R}} E(Y|x) f_X(x) \, dx.$$
Solution #1 (old, boring way):

\[ E[Y] = \int_{\mathbb{R}} y f_Y(y) \, dy = \int_{0}^{1} \frac{7}{2} y^{7/2} \, dy = \frac{7}{9}. \]

Solution #2 (new, exciting way):

\[ E[Y] = E[E(Y|X)] \]
Solution #1 (old, boring way):

\[
E[Y] = \int_{\mathbb{R}} y f_Y(y) \, dy = \int_{0}^{1} \frac{7}{2} y^{7/2} \, dy = \frac{7}{9}.
\]

Solution #2 (new, exciting way):

\[
E[Y] = E[E(Y|X)] = \int_{\mathbb{R}} E(Y|x) f_X(x) \, dx
\]
Solution #1 (old, boring way):

\[ E[Y] = \int_{\mathbb{R}} y f_Y(y) \, dy = \int_0^1 \frac{7}{2} y^{7/2} \, dy = \frac{7}{9}. \]

Solution #2 (new, exciting way):

\[
E[Y] = E[E(Y|X)] = \int_{\mathbb{R}} E(Y|x) f_X(x) \, dx
= \int_{-1}^1 \left( \frac{2}{3} \cdot \frac{1-x^6}{1-x^4} \right) \left( \frac{21}{8} x^2 (1-x^4) \right) dx
\]
Solution #1 (old, boring way):

\[ E[Y] = \int_{\mathbb{R}} y f_Y(y) \, dy = \int_0^1 \frac{7}{2} y^{7/2} \, dy = \frac{7}{9}. \]

Solution #2 (new, exciting way):

\[ E[Y] = E[E(Y \mid X)] = \int_{\mathbb{R}} E(Y \mid x) f_X(x) \, dx \]
\[ = \int_{-1}^1 \left( \frac{2}{3} \cdot \frac{1 - x^6}{1 - x^4} \right) \left( \frac{21}{8} x^2 (1 - x^4) \right) \, dx = \frac{7}{9}. \]
Solution #1 (old, boring way):

$$E[Y] = \int_{\mathbb{R}} y f_Y(y) \, dy = \int_0^1 \frac{7}{2} y^{7/2} \, dy = \frac{7}{9}.$$ 

Solution #2 (new, exciting way):

$$E[Y] = \mathbb{E}[\mathbb{E}(Y|X)]$$
$$= \int_{\mathbb{R}} \mathbb{E}(Y|x) f_X(x) \, dx$$
$$= \int_{-1}^1 \left( \frac{2}{3} \cdot \frac{1 - x^6}{1 - x^4} \right) \left( \frac{21}{8} x^2 (1 - x^4) \right) \, dx$$
$$= \frac{7}{9}.$$ 

Notice that both answers are the same (good)!  □
Example:

"First-step" method to find the mean of $Y \sim \text{Geom}(p)$. Think of $Y$ as the number of coin flips before $H$ appears, where $P(H) = p$.

Furthermore, consider the first step of the coin flip process, and let $X = H$ or $T$ denote the outcome of the first toss. Based on the result $X$ of this first step, we have


$= (1 + E[Y])(1 - p) + (1)(p) \text{ (start from scratch if } X = T)$.

Solving, we get $E[Y] = 1/p$ (which is the correct answer)!
Example: “First-step” method to find the mean of $Y \sim \text{Geom}(p)$. Think of $Y$ as the number of coin flips before H appears, where $P(H) = p$. 

Furthermore, consider the first step of the coin flip process, and let $X = H$ or $T$ denote the outcome of the first toss. Based on the result $X$ of this first step, we have 


Solving, we get $E[Y] = 1/p$ (which is the correct answer)!
Example: “First-step” method to find the mean of \( Y \sim \text{Geom}(p) \). Think of \( Y \) as the number of coin flips before \( H \) appears, where \( P(H) = p \).

Furthermore, consider the first step of the coin flip process, and let \( X = H \) or \( T \) denote the outcome of the first toss.
Lesson 3.9 — Honors Class: First-Step Analysis

Example: “First-step” method to find the mean of $Y \sim \text{Geom}(p)$. Think of $Y$ as the number of coin flips before H appears, where $P(H) = p$.

Furthermore, consider the first step of the coin flip process, and let $X = H$ or T denote the outcome of the first toss. Based on the result $X$ of this first step, we have

$$E[Y] = E[E(Y|X)]$$
Example: “First-step” method to find the mean of $Y \sim \text{Geom}(p)$. Think of $Y$ as the number of coin flips before H appears, where $P(H) = p$.

Furthermore, consider the first step of the coin flip process, and let $X = H$ or $T$ denote the outcome of the first toss. Based on the result $X$ of this first step, we have

$$E[Y] = E[E(Y|X)]$$
$$= \sum_x E[Y|x] f_X(x)$$
Example: “First-step” method to find the mean of $Y \sim \text{Geom}(p)$. Think of $Y$ as the number of coin flips before H appears, where $P(H) = p$.

Furthermore, consider the first step of the coin flip process, and let $X = \text{H or T}$ denote the outcome of the first toss. Based on the result $X$ of this first step, we have

$$E[Y] = E[E(Y|X)] = \sum_x E[Y|x]f_X(x) = E[Y|X = \text{T}]P(X = \text{T}) + E[Y|X = \text{H}]P(X = \text{H})$$
Example: “First-step” method to find the mean of $Y \sim \text{Geom}(p)$. Think of $Y$ as the number of coin flips before H appears, where $P(H) = p$.

Furthermore, consider the first step of the coin flip process, and let $X = H$ or $T$ denote the outcome of the first toss. Based on the result $X$ of this first step, we have

**Example:** “First-step” method to find the mean of $Y \sim \text{Geom}(p)$. Think of $Y$ as the number of coin flips before H appears, where $P(H) = p$.

Furthermore, consider the first step of the coin flip process, and let $X = H$ or $T$ denote the outcome of the first toss. Based on the result $X$ of this first step, we have

\[
E[Y] = E[E(Y|X)] \\
= \sum_x E[Y|x] f_X(x) \\
= E[Y|X = T] P(X = T) + E[Y|X = H] P(X = H) \\
= (1 + E[Y]) (1 - p) + (1)(p) \quad \text{(start from scratch if } X = T). 
\]

Solving, we get $E[Y] = 1/p$ (which is the correct answer)! $\square$
**Example**: Consider a sequence of coin flips. What is the expected number of flips $Y$ until “HT” appears for the first time?
**Example:** Consider a sequence of coin flips. What is the expected number of flips $Y$ until “HT” appears for the first time?

Clearly, $Y = A + B$, where $A$ is the number of flips until the first “H” appears, and $B$ is the number of subsequent flips until “T” appears for the first time after the sequence of H’s begins.
**Example:** Consider a sequence of coin flips. What is the expected number of flips $Y$ until “HT” appears for the first time?

Clearly, $Y = A + B$, where $A$ is the number of flips until the first “H” appears, and $B$ is the number of subsequent flips until “T” appears for the first time after the sequence of H’s begins.

For instance, the sequence TTTHHT corresponds to $Y = A + B = 4 + 2 = 6$. 
**Example**: Consider a sequence of coin flips. What is the expected number of flips $Y$ until “HT” appears for the first time?

Clearly, $Y = A + B$, where $A$ is the number of flips until the first “H” appears, and $B$ is the number of subsequent flips until “T” appears for the first time after the sequence of H’s begins.

For instance, the sequence TTTHTHT corresponds to $Y = A + B = 4 + 2 = 6$.

In any case, it’s obvious that $A$ and $B$ are iid Geom($p = 1/2$), so by the previous example, $E[Y] = E[A] + E[B] = (1/p) + (1/p) = 4$.  \qed
**Example:** Consider a sequence of coin flips. What is the expected number of flips $Y$ until “HT” appears for the first time?

Clearly, $Y = A + B$, where $A$ is the number of flips until the first “H” appears, and $B$ is the number of subsequent flips until “T” appears for the first time after the sequence of H’s begins.

For instance, the sequence TTTHHT corresponds to $Y = A + B = 4 + 2 = 6$.

In any case, it’s obvious that $A$ and $B$ are iid Geom($p = 1/2$), so by the previous example, $\mathbb{E}[Y] = \mathbb{E}[A] + \mathbb{E}[B] = (1/p) + (1/p) = 4$. □

This example didn’t involve first-step analysis (besides using the expected value of a geometric RV). But the next related example will....
**Example**: Again consider a sequence of coin flips. What is the expected number of flips $Y$ until “HH” appears for the first time?
**Example**: Again consider a sequence of coin flips. What is the expected number of flips $Y$ until “HH” appears for the first time?

For instance, the sequence TTHTTHH corresponds to $Y = 7$ tries.
**Example:** Again consider a sequence of coin flips. What is the expected number of flips $Y$ until “HH” appears for the first time?

For instance, the sequence TTHTTHH corresponds to $Y = 7$ tries.

Using an enhanced first-step analysis, we see that

**Example**: Again consider a sequence of coin flips. What is the expected number of flips $Y$ until “HH” appears for the first time?

For instance, the sequence TTHTTHH corresponds to $Y = 7$ tries.

Using an enhanced first-step analysis, we see that

$$
$$

$$
= E[Y|T]P(T) + \left\{ E[Y|HH]P(HH|H) + E[Y|HT]P(HT|H) \right\} P(H)
$$
Example: Again consider a sequence of coin flips. What is the expected number of flips $Y$ until “HH” appears for the first time?

For instance, the sequence TTHTTHH corresponds to $Y = 7$ tries.

Using an enhanced first-step analysis, we see that

$$
= E[Y|T]P(T)
+ \left\{ E[Y|HH]P(HH|H) + E[Y|HT]P(HT|H) \right\} P(H)
= (1 + E[Y])(0.5) + \left\{ (2)(0.5) + (2 + E[Y])(0.5) \right\}(0.5)
$$
Example: Again consider a sequence of coin flips. What is the expected number of flips $Y$ until “HH” appears for the first time?

For instance, the sequence TTHTTHH corresponds to $Y = 7$ tries.

Using an enhanced first-step analysis, we see that

\[
\]

\[
= E[Y|T]P(T)
\]

\[
+ \left\{ E[Y|HH]P(HH|H) + E[Y|HT]P(HT|H) \right\} P(H)
\]

\[
= (1 + E[Y])(0.5) + \left\{ (2)(0.5) + (2 + E[Y])(0.5) \right\}(0.5)
\]

(since we have to start over once we see a T)
**Example**: Again consider a sequence of coin flips. What is the expected number of flips $Y$ until “HH” appears for the first time?

For instance, the sequence TTHTTTHH corresponds to $Y = 7$ tries.

Using an enhanced first-step analysis, we see that

$$
$$

$$
$$

$$
= (1 + E[Y])(0.5) + \{(2)(0.5) + (2 + E[Y])(0.5)\}(0.5)
$$

(since we have to start over once we see a T)

$$
= 1.5 + 0.75 E[Y].
$$
**Example:** Again consider a sequence of coin flips. What is the expected number of flips \( Y \) until “HH” appears for the first time?

For instance, the sequence TTHTTHH corresponds to \( Y = 7 \) tries.

Using an enhanced first-step analysis, we see that

\[
= E[Y|T]P(T) \\
\quad + \{E[Y|HH]P(HH|H) + E[Y|HT]P(HT|H)\}P(H) \\
= (1 + E[Y])(0.5) + \{(2)(0.5) + (2 + E[Y])(0.5)\}(0.5) \\
\quad \text{(since we have to start over once we see a T)} \\
= 1.5 + 0.75 E[Y].
\]

Solving, we obtain \( E[Y] = 6 \), which is perhaps surprising given the result from the previous example. \( \square \)
Bonuss Theorem (expectation of sum of a random number of RVs): Suppose that $X_1, X_2, \ldots$ are independent RVs, all with the same mean. Also suppose that $N$ is a nonnegative, integer-valued RV that's independent of the $X_i$'s. Then
\[
E\left[\sum_{i=1}^{\sum N_i} X_i\right] = E[N]E[X_1].
\]
Remark: You have to be very careful here. In particular, note that $E\left[\sum_{i=1}^{\sum N_i} X_i\right] \neq N E[X_1]$, since the LHS is a number and the RHS is random.
Bonus Theorem (expectation of sum of a random number of RVs):

Suppose that $X_1, X_2, \ldots$ are independent RVs, all with the same mean. Also suppose that $N$ is a nonnegative, integer-valued RV that’s independent of the $X_i$’s. Then

$$E\left[\sum_{i=1}^{N} X_i\right] = E[N]E[X_1].$$

Remark: You have to be very careful here. In particular, note that $E\left[\sum_{i=1}^{N} X_i\right] \neq N E[X_1]$, since the LHS is a number and the RHS is random.
Bonus Theorem (expectation of sum of a random number of RVs):

Suppose that $X_1, X_2, \ldots$ are independent RVs, all with the same mean.
**Bonus Theorem (expectation of sum of a random number of RVs):**

Suppose that $X_1, X_2, \ldots$ are independent RVs, all with the same mean.

Also suppose that $N$ is a nonnegative, integer-valued RV that’s independent of the $X_i$’s. Then
Bonus Theorem (expectation of sum of a random number of RVs):

Suppose that $X_1, X_2, \ldots$ are independent RVs, all with the same mean.

Also suppose that $N$ is a nonnegative, integer-valued RV that’s independent of the $X_i$’s. Then

$$E\left[\sum_{i=1}^{N} X_i\right] = E[N]E[X_1].$$
Bonus Theorem (expectation of sum of a random number of RVs):

Suppose that $X_1, X_2, \ldots$ are independent RVs, all with the same mean.

Also suppose that $N$ is a nonnegative, integer-valued RV that’s independent of the $X_i$’s. Then

$$
E\left[\sum_{i=1}^{N} X_i\right] = E[N]E[X_1].
$$

Remark: You have to be very careful here. In particular, note that $E\left[\sum_{i=1}^{N} X_i\right] \neq NE[X_1]$, since the LHS is a number and the RHS is random.
Proof (cf. Ross): By double expectation,
**Proof** (cf. Ross): By double expectation,

\[
E\left( \sum_{i=1}^{N} X_i \right) = E \left[ E\left( \sum_{i=1}^{N} X_i \mid N \right) \right]
\]
**Proof** (cf. Ross): By double expectation,

\[
E\left(\sum_{i=1}^{N} X_i\right) = E\left[E\left(\sum_{i=1}^{N} X_i \mid N\right)\right]
\]

\[
= \sum_{n=1}^{\infty} E\left(\sum_{i=1}^{N} X_i \mid N = n\right)P(N = n)
\]

\[
= E\left[X_1\right] \sum_{n=1}^{\infty} nP(N = n)
\]

\[
= E\left[X_1\right] \sum_{n=1}^{\infty} nP(N = n)
\]
**Proof** (cf. Ross): By double expectation,

\[
E\left(\sum_{i=1}^{N} X_i\right) = E\left[E\left(\sum_{i=1}^{N} X_i \mid N\right)\right]
\]

\[
= \sum_{n=1}^{\infty} E\left(\sum_{i=1}^{N} X_i \mid N = n\right) P(N = n)
\]

\[
= \sum_{n=1}^{\infty} E\left(\sum_{i=1}^{n} X_i \mid N = n\right) P(N = n)
\]
Proof (cf. Ross): By double expectation,

\[
E\left( \sum_{i=1}^{N} X_i \right) = E \left[ E\left( \sum_{i=1}^{N} X_i \mid N \right) \right]
\]

\[
= \sum_{n=1}^{\infty} E \left( \sum_{i=1}^{n} X_i \mid N = n \right) P(N = n)
\]

\[
= \sum_{n=1}^{\infty} \sum_{i=1}^{n} E\left( X_i \mid N = n \right) P(N = n)
\]

\[
= \sum_{n=1}^{\infty} \sum_{i=1}^{n} E\left( X_i \right) P(N = n) \quad (N \text{ and } X_i\text{'s indep})
\]
Proof (cf. Ross): By double expectation,

\[
E\left(\sum_{i=1}^{N} X_i\right) = E\left[E\left(\sum_{i=1}^{N} X_i \mid N\right)\right]
\]

\[
= \sum_{n=1}^{\infty} E\left(\sum_{i=1}^{N} X_i \mid N = n\right) P(N = n)
\]

\[
= \sum_{n=1}^{\infty} E\left(\sum_{i=1}^{n} X_i \mid N = n\right) P(N = n)
\]

\[
= \sum_{n=1}^{\infty} E\left(\sum_{i=1}^{n} X_i\right) P(N = n) \quad (N \text{ and } X_i \text{'s indep})
\]

\[
= \sum_{n=1}^{\infty} nE[X_1] P(N = n)
\]
**Proof** (cf. Ross): By double expectation,

\[
E\left(\sum_{i=1}^{N} X_i\right) = E\left[E\left(\sum_{i=1}^{N} X_i \mid N\right)\right]
\]

\[
= \sum_{n=1}^{\infty} E\left(\sum_{i=1}^{n} X_i \mid N = n\right) P(N = n)
\]

\[
= \sum_{n=1}^{\infty} E\left(\sum_{i=1}^{n} X_i \right) P(N = n) \quad (N \text{ and } X_i's \text{ indep})
\]

\[
= \sum_{n=1}^{\infty} nE[X_1] P(N = n)
\]

\[
= E[X_1] \sum_{n=1}^{\infty} nP(N = n). \quad \square
\]
**Example:** Suppose the number of times we roll a die is $N \sim \text{Pois}(10)$. If $X_i$ denotes the value of the $i$th toss, then the expected total of all of the rolls is
**Example:** Suppose the number of times we roll a die is $N \sim \text{Pois}(10)$. If $X_i$ denotes the value of the $i$th toss, then the expected total of all of the rolls is

$$E\left( \sum_{i=1}^{N} X_i \right) = E[N]E[X_1] = 10(3.5) = 35. \quad \square$$
**Example:** Suppose the number of times we roll a die is $N \sim \text{Pois}(10)$. If $X_i$ denotes the value of the $i$th toss, then the expected total of all of the rolls is

$$E\left(\sum_{i=1}^{N} X_i\right) = E[N]E[X_1] = 10(3.5) = 35.$$ 

**Theorem:** Under the same conditions as before,

$$\text{Var}\left(\sum_{i=1}^{N} X_i\right) = E[N]\text{Var}(X_1) + (E[X_1])^2\text{Var}(N).$$
Example: Suppose the number of times we roll a die is \( N \sim \text{Pois}(10) \). If \( X_i \) denotes the value of the \( i \)th toss, then the expected total of all of the rolls is

\[
\mathbb{E}\left( \sum_{i=1}^{N} X_i \right) = \mathbb{E}[N]\mathbb{E}[X_1] = 10(3.5) = 35. \quad \square
\]

Theorem: Under the same conditions as before,

\[
\text{Var}\left( \sum_{i=1}^{N} X_i \right) = \mathbb{E}[N]\text{Var}(X_1) + (\mathbb{E}[X_1])^2\text{Var}(N).
\]

Proof: See, for instance, Ross. \( \square \)
| 1 | Introduction |
| 2 | Marginal Distributions |
| 3 | Conditional Distributions |
| 4 | Independent Random Variables |
| 5 | Consequences of Independence |
| 6 | Random Samples |
| 7 | Conditional Expectation |
| 8 | Double Expectation |
| 9 | Honors Class: First-Step Analysis |
| 10 | Honors Class: Random Sums of Random Variables |
| 11 | Honors Class: Standard Conditioning Argument |
| 12 | Covariance and Correlation |
| 13 | Correlation and Causation |
| 14 | A Couple of Worked Correlation Examples |
| 15 | Some Useful Covariance / Correlation Theorems |
| 16 | Moment Generating Functions, Revisited |
| 17 | Honors Bivariate Functions of Random Variables |
Lesson 3.11 — Honors Class: Standard Conditioning Argument

Let $A$ be some event, and define the RV $Y$ as the following indicator function:

$$Y = \begin{cases} 1 & \text{if } A \text{ occurs} \\ 0 & \text{otherwise} \end{cases}.$$

Then

$$E[Y] = \sum_{y} y f_Y(y) = P(Y = 1) = P(A).$$

Similarly, for any RV $X$, we have

$$E[Y | X = x] = \sum_{y} y f_Y(y | x) = P(Y = 1 | X = x) = P(A | X = x).$$

These results suggest an alternative way of calculating $P(A)$...
Lesson 3.11 — Honors Class: Standard Conditioning Argument

Bonus Theorem/Proof (computing probabilities by conditioning):

Let $A$ be some event, and define the RV $Y$ as the following indicator function:

$$Y = \begin{cases} 
1 & \text{if } A \text{ occurs} \\
0 & \text{otherwise} 
\end{cases}.$$

Then

$$E[Y] = \sum y f_Y(y) = P(Y = 1) = P(A).$$

Similarly, for any RV $X$, we have

$$E[Y | X = x] = \sum y f_Y(y | x) = P(Y = 1 | X = x) = P(A | X = x).$$

These results suggest an alternative way of calculating $P(A)$...
Lesson 3.11 — Honors Class: Standard Conditioning Argument

Bonus Theorem/Proof (computing probabilities by conditioning):

Let $A$ be some event, and define the RV $Y$ as the following indicator function:

$$Y = 1_{A} = \begin{cases} 
1 & \text{if } A \text{ occurs} \\
0 & \text{otherwise}
\end{cases}$$

Then

$$\mathbb{E}[Y] = \sum y y f_{Y}(y) = P(Y = 1) = P(A).$$

Similarly, for any RV $X$, we have

$$\mathbb{E}[Y | X = x] = \sum y y f_{Y | X = x}(y) = P(Y = 1 | X = x) = P(A | X = x).$$

These results suggest an alternative way of calculating $P(A)$.

...
Bonus Theorem/Proof (computing probabilities by conditioning):

Let $A$ be some event, and define the RV $Y$ as the following indicator function:

$$Y = 1_A \equiv \begin{cases} 
1 & \text{if } A \text{ occurs} \\
0 & \text{otherwise.}
\end{cases}$$

Then

$$E[Y] = \sum_y y f_Y(y) = P(Y = 1) = P(A).$$

Similarly, for any RV $X$, we have

$$E[Y | X = x] = \sum_y y f_Y(y | x) = P(Y = 1 | X = x) = P(A | X = x).$$

These results suggest an alternative way of calculating $P(A)$.
Lesson 3.11 — Honors Class: Standard Conditioning Argument

**Bonus Theorem/Proof (computing probabilities by conditioning):**

Let $A$ be some event, and define the RV $Y$ as the following indicator function:

\[
Y = 1_A \equiv \begin{cases} 
1 & \text{if } A \text{ occurs} \\
0 & \text{otherwise.}
\end{cases}
\]

Then

\[
E[Y] = \sum_y y f_Y(y)
\]
**Bonus Theorem/Proof (computing probabilities by conditioning):**

Let \( A \) be some event, and define the RV \( Y \) as the following indicator function:

\[
Y = 1_A \equiv \begin{cases} 
1 & \text{if } A \text{ occurs} \\
0 & \text{otherwise.}
\end{cases}
\]

Then

\[
E[Y] = \sum_y yf_Y(y) = P(Y = 1)
\]
Lesson 3.11 — Honors Class: Standard Conditioning Argument

Bonus Theorem/Proof (computing probabilities by conditioning):

Let $A$ be some event, and define the RV $Y$ as the following indicator function:

$$Y = 1_A \equiv \begin{cases} 
1 & \text{if } A \text{ occurs} \\
0 & \text{otherwise}.
\end{cases}$$

Then

$$E[Y] = \sum_y y f_Y(y) = P(Y = 1) = P(A).$$
Lesson 3.11 — Honors Class: Standard Conditioning Argument

Bonus Theorem/Proof *(computing probabilities by conditioning)*:

Let $A$ be some event, and define the RV $Y$ as the following indicator function:

$$Y = 1_A \equiv \begin{cases} 1 & \text{if } A \text{ occurs} \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$E[Y] = \sum_y y f_Y(y) = P(Y = 1) = P(A).$$

Similarly, for any RV $X$, we have
Bonus Theorem/Proof \textit{(computing probabilities by conditioning)}:

Let $A$ be some event, and define the RV $Y$ as the following indicator function:

$$Y = 1_A \equiv \begin{cases} 1 & \text{if } A \text{ occurs} \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$E[Y] = \sum_y y f_Y(y) = P(Y = 1) = P(A).$$

Similarly, for any RV $X$, we have

$$E[Y|X = x] = \sum_y y f(y|x).$$
Lesson 3.11 — Honors Class: Standard Conditioning Argument

Bonus Theorem/Proof (computing probabilities by conditioning):

Let $A$ be some event, and define the RV $Y$ as the following indicator function:

$$Y = 1_A \equiv \begin{cases} 1 & \text{if } A \text{ occurs} \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$E[Y] = \sum_y y f_Y(y) = P(Y = 1) = P(A).$$

Similarly, for any RV $X$, we have

$$E[Y|X = x] = \sum_y y f(y|x) = P(Y = 1|X = x).$$
**Lesson 3.11 — Honors Class: Standard Conditioning Argument**

**Bonus Theorem/Proof (computing probabilities by conditioning):**

Let $A$ be some event, and define the RV $Y$ as the following indicator function:

$$Y = 1_A \equiv \begin{cases} 1 & \text{if } A \text{ occurs} \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$\mathbb{E}[Y] = \sum_y y f_Y(y) = P(Y = 1) = P(A).$$

Similarly, for any RV $X$, we have

$$\mathbb{E}[Y|X = x] = \sum_y y f(y|x) = P(Y = 1|X = x) = P(A|X = x).$$
Lesson 3.11 — Honors Class: Standard Conditioning Argument

**Bonus Theorem/Proof (computing probabilities by conditioning):**

Let $A$ be some event, and define the RV $Y$ as the following indicator function:

$$Y = 1_A \equiv \begin{cases} 
  1 & \text{if } A \text{ occurs} \\
  0 & \text{otherwise}.
\end{cases}$$

Then

$$E[Y] = \sum_y y f_Y(y) = P(Y = 1) = P(A).$$

Similarly, for any RV $X$, we have

$$E[Y|X = x] = \sum_y y f(y|x) = P(Y = 1|X = x) = P(A|X = x).$$

These results suggest an alternative way of calculating $P(A)$...
**Theorem:** If $X$ is a continuous RV (similar result if $X$ is discrete), then

$$P(A) = \int_{\mathbb{R}} P(A \mid X = x) f_X(x) \, dx.$$
**Theorem:** If $X$ is a continuous RV (similar result if $X$ is discrete), then

$$P(A) = \int_{\mathbb{R}} P(A|X = x) f_X(x) \, dx.$$ 

**Proof:**

$$P(A) = E[Y] \quad \text{(where we take } Y = 1_A \text{)}$$
**Theorem:** If $X$ is a continuous RV (similar result if $X$ is discrete), then

\[ P(A) = \int_{\mathbb{R}} P(A|X = x) f_X(x) \, dx. \]

**Proof:**

\[
P(A) = E[Y] \quad \text{(where we take } Y = 1_A) \\
= E[E(Y|X)] \quad \text{(double expectation)}
\]
Theorem: If $X$ is a continuous RV (similar result if $X$ is discrete), then

$$P(A) = \int_{\mathbb{R}} P(A|X = x) f_X(x) \, dx.$$ 

Proof:

$$P(A) = \mathbb{E}[Y] \quad (\text{where we take } Y = 1_A)$$

$$= \mathbb{E}[\mathbb{E}(Y|X)] \quad (\text{double expectation})$$

$$= \int_{\mathbb{R}} \mathbb{E}[Y|x] f_X(x) \, dx \quad (\text{LOTUS})$$
**Theorem:** If $X$ is a continuous RV (similar result if $X$ is discrete), then

$$P(A) = \int_{\mathbb{R}} P(A|X = x) f_X(x) \, dx.$$  

**Proof:**

$$P(A) = E[Y] \quad \text{(where we take } Y = 1_A)$$

$$= E[E(Y|X)] \quad \text{(double expectation)}$$

$$= \int_{\mathbb{R}} E[Y|x] f_X(x) \, dx \quad \text{(LOTUS)}$$

$$= \int_{\mathbb{R}} P(A|X = x) f_X(x) \, dx \quad \text{(since } Y = 1_A). \quad \Box$$
**Theorem:** If $X$ is a continuous RV (similar result if $X$ is discrete), then

$$P(A) = \int_{\mathbb{R}} P(A|X = x) f_X(x) \, dx.$$  

**Proof:**

$$P(A) = \mathbb{E}[Y] \quad \text{(where we take } Y = 1_A)$$

$$= \mathbb{E}[\mathbb{E}(Y|X)] \quad \text{(double expectation)}$$

$$= \int_{\mathbb{R}} \mathbb{E}[Y|x] f_X(x) \, dx \quad \text{(LOTUS)}$$

$$= \int_{\mathbb{R}} P(A|X = x) f_X(x) \, dx \quad \text{(since } Y = 1_A). \quad \Box$$

**Remark:** We call this the “standard conditioning argument.”
**Theorem:** If $X$ is a continuous RV (similar result if $X$ is discrete), then

\[
P(A) = \int_{\mathbb{R}} P(A|X = x) f_X(x) \, dx.
\]

**Proof:**

\[
P(A) = E[Y] \quad \text{(where we take } Y = 1_A) \\
= E[E(Y|X)] \quad \text{(double expectation)} \\
= \int_{\mathbb{R}} E[Y|X = x] f_X(x) \, dx \quad \text{(LOTUS)} \\
= \int_{\mathbb{R}} P(A|X = x) f_X(x) \, dx \quad \text{(since } Y = 1_A). \quad \square
\]

**Remark:** We call this the “standard conditioning argument.” Yes, it looks complicated. But sometimes you need to take a step backward to go two steps forward!
Example/Theorem: If $X$ and $Y$ are independent continuous RVs, with pdf $f_X(\cdot)$ and cdf $F_Y(\cdot)$, respectively. Then
**Example/Theorem:** If $X$ and $Y$ are independent continuous RVs, with pdf $f_X(\cdot)$ and cdf $F_Y(\cdot)$, respectively. Then

$$P(Y \leq X) = \int_{\mathbb{R}} F_Y(x) f_X(x) \, dx.$$
Example/Theorem: If $X$ and $Y$ are independent continuous RVs, with pdf $f_X(\cdot)$ and cdf $F_Y(\cdot)$, respectively. Then

$$P(Y \leq X) = \int_{\mathbb{R}} F_Y(x) f_X(x) \, dx.$$ 

Proof: (Actually, there are many proofs.) Let the event $A = \{Y \leq X\}$. 


**Example/Theorem:** If $X$ and $Y$ are independent continuous RVs, with pdf $f_X(\cdot)$ and cdf $F_Y(\cdot)$, respectively. Then

$$P(Y \leq X) = \int_{\mathbb{R}} F_Y(x) f_X(x) \, dx.$$ 

**Proof:** (Actually, there are many proofs.) Let the event $A = \{Y \leq X\}$. Then

$$P(Y \leq X) = \int_{\mathbb{R}} P(Y \leq X | X = x) f_X(x) \, dx.$$
**Example/Theorem:** If $X$ and $Y$ are independent continuous RVs, with pdf $f_X(\cdot)$ and cdf $F_Y(\cdot)$, respectively. Then

\[
P(Y \leq X) = \int_{\mathbb{R}} F_Y(x) f_X(x) \, dx.
\]

**Proof:** (Actually, there are many proofs.) Let the event $A = \{Y \leq X\}$. Then

\[
P(Y \leq X) = \int_{\mathbb{R}} P(Y \leq X|X = x) f_X(x) \, dx
\]

\[
= \int_{\mathbb{R}} P(Y \leq x|X = x) f_X(x) \, dx
\]
Example/Theorem: If $X$ and $Y$ are independent continuous RVs, with pdf $f_X(\cdot)$ and cdf $F_Y(\cdot)$, respectively. Then

$$P(Y \leq X) = \int_{\mathbb{R}} F_Y(x) f_X(x) \, dx.$$ 

Proof: (Actually, there are many proofs.) Let the event $A = \{Y \leq X\}$. Then

$$P(Y \leq X) = \int_{\mathbb{R}} P(Y \leq X|X = x) f_X(x) \, dx$$

$$= \int_{\mathbb{R}} P(Y \leq x|X = x) f_X(x) \, dx$$

$$= \int_{\mathbb{R}} P(Y \leq x) f_X(x) \, dx \quad (X, Y \text{ are independent}).$$
**Example**: If $X \sim \text{Exp}(\alpha)$ and $Y \sim \text{Exp}(\beta)$ are independent RVs, then
**Example:** If $X \sim \text{Exp}(\alpha)$ and $Y \sim \text{Exp}(\beta)$ are independent RVs, then

$$P(Y \leq X) = \int_{\mathbb{R}} F_Y(x) f_X(x) \, dx$$
Example: If $X \sim \text{Exp}(\alpha)$ and $Y \sim \text{Exp}(\beta)$ are independent RVs, then

\[
P(Y \leq X) = \int_{\mathbb{R}} F_Y(x) f_X(x) \, dx
\]

\[
= \int_{0}^{\infty} (1 - e^{-\beta x}) \alpha e^{-\alpha x} \, dx
\]
Example: If $X \sim \text{Exp}(\alpha)$ and $Y \sim \text{Exp}(\beta)$ are independent RVs, then

\[
P(Y \leq X) = \int_{\mathbb{R}} F_Y(x) f_X(x) \, dx
\]

\[
= \int_{0}^{\infty} (1 - e^{-\beta x}) \alpha e^{-\alpha x} \, dx
\]

\[
= \frac{\beta}{\alpha + \beta}.
\]

Remark: Think of $X$ as the time until the next male driver shows up at a parking lot (at rate $\alpha$/hour) and $Y$ as the time for the next female driver (at rate $\beta$/hour). Then $P(Y \leq X) = \frac{\beta}{\alpha + \beta}$ is the intuitively reasonable probability that the next driver to arrive will be female.
**Example:** If $X \sim \text{Exp}(\alpha)$ and $Y \sim \text{Exp}(\beta)$ are independent RVs, then

$$P(Y \leq X) = \int_{\mathbb{R}} F_Y(x) f_X(x) \, dx$$

$$= \int_{0}^{\infty} (1 - e^{-\beta x}) \alpha e^{-\alpha x} \, dx$$

$$= \frac{\beta}{\alpha + \beta}.$$ 

**Remark:** Think of $X$ as the time until the next male driver shows up at a parking lot (at rate $\alpha$/hour) and $Y$ as the time for the next female driver (at rate $\beta$/hour). Then $P(Y \leq X) = \beta/(\alpha + \beta)$ is the intuitively reasonable probability that the next driver to arrive will be female.
Example/Theorem: Suppose $X$ and $Y$ are independent continuous RVs, with pdf $f_X(\cdot)$ and cdf $F_Y(\cdot)$, respectively. Define the sum $Z = X + Y$. 
Example/Theorem: Suppose $X$ and $Y$ are independent continuous RVs, with pdf $f_X(\cdot)$ and cdf $F_Y(\cdot)$, respectively. Define the sum $Z = X + Y$. Then

$$P(Z \leq z) = \int_{\mathbb{R}} F_Y(z - x) f_X(x) \, dx.$$
**Example/Theorem:** Suppose $X$ and $Y$ are independent continuous RVs, with pdf $f_X(\cdot)$ and cdf $F_Y(\cdot)$, respectively. Define the sum $Z = X + Y$. Then

$$P(Z \leq z) = \int_{\mathbb{R}} F_Y(z - x) f_X(x) \, dx.$$ 

As expression such as the above for $P(Z \leq z)$ is often called a *convolution*. 


**Example/Theorem:** Suppose $X$ and $Y$ are independent continuous RVs, with pdf $f_X(\cdot)$ and cdf $F_Y(\cdot)$, respectively. Define the sum $Z = X + Y$. Then

$$P(Z \leq z) = \int_{\mathbb{R}} F_Y(z - x) f_X(x) \, dx.$$ 

As expression such as the above for $P(Z \leq z)$ is often called a *convolution*.

**Proof:**

$$P(Z \leq z) = \int_{\mathbb{R}} P(X + Y \leq z | X = x) f_X(x) \, dx$$
**Example/Theorem:** Suppose $X$ and $Y$ are independent continuous RVs, with pdf $f_X(\cdot)$ and cdf $F_Y(\cdot)$, respectively. Define the sum $Z = X + Y$. Then

$$P(Z \leq z) = \int_{\mathbb{R}} F_Y(z-x) f_X(x) \, dx.$$ 

As expression such as the above for $P(Z \leq z)$ is often called a *convolution*.

**Proof:**

$$P(Z \leq z) = \int_{\mathbb{R}} P(X + Y \leq z | X = x) f_X(x) \, dx$$

$$= \int_{\mathbb{R}} P(Y \leq z-x | X = x) f_X(x) \, dx$$
**Example/Theorem:** Suppose $X$ and $Y$ are independent continuous RVs, with pdf $f_X(\cdot)$ and cdf $F_Y(\cdot)$, respectively. Define the sum $Z = X + Y$. Then

$$P(Z \leq z) = \int_{\mathbb{R}} F_Y(z-x) f_X(x) \, dx.$$ 

As expression such as the above for $P(Z \leq z)$ is often called a *convolution*.

**Proof:**

$$P(Z \leq z) = \int_{\mathbb{R}} P(X + Y \leq z | X = x) f_X(x) \, dx$$

$$= \int_{\mathbb{R}} P(Y \leq z - x | X = x) f_X(x) \, dx$$

$$= \int_{\mathbb{R}} P(Y \leq z - x) f_X(x) \, dx \quad (X, Y \text{ are indep}).$$
**Example**: Suppose $X, Y \overset{iid}{\sim} \text{Exp}(\lambda)$, and let $Z = X + Y$. 

$$
\begin{align*}
P(Z \leq z) &= \int_{\mathbb{R}} F_Y(z-x)f_X(x) \, dx \\
&= \int_0^z (1 - e^{-\lambda(z-x)}) \lambda e^{-\lambda x} \, dx \\
&= 1 - e^{-\lambda z} - \lambda z e^{-\lambda z} + \lambda z e^{-\lambda z} - \lambda z e^{-\lambda z} \\
&= 1 - e^{-\lambda z}.
\end{align*}
$$

Thus, the pdf of $Z$ is

$$
d_Z P(Z \leq z) = \frac{\lambda}{2} z e^{-\lambda z}, \quad z \geq 0. $$

This turns out to mean that $Z \sim \text{Gamma}(2, \lambda)$, aka $\text{Erlang}_2(\lambda)$. 

2

3/2/20 67 / 110
**Example:** Suppose $X, Y \overset{iid}{\sim} \text{Exp}(\lambda)$, and let $Z = X + Y$. Then

\[
P(Z \leq z) = \int_{\mathbb{R}} F_Y(z - x) f_X(x) \, dx
\]
**Example:** Suppose $X, Y \overset{iid}{\sim} \text{Exp}(\lambda)$, and let $Z = X + Y$. Then

$$P(Z \leq z) = \int_{\mathbb{R}} F_Y(z - x) f_X(x) \, dx$$

$$= \int_{0}^{z} (1 - e^{-\lambda(z-x)}) \lambda e^{-\lambda x} \, dx$$

(must have $x \geq 0$ and $z - x \geq 0$)
Example: Suppose $X, Y \overset{iid}{\sim} \text{Exp}(\lambda)$, and let $Z = X + Y$. Then

$$P(Z \leq z) = \int_{\mathbb{R}} F_Y(z - x) f_X(x) \, dx$$

$$= \int_{0}^{z} (1 - e^{-\lambda(z-x)}) \lambda e^{-\lambda x} \, dx$$

(must have $x \geq 0$ and $z - x \geq 0$)

$$= 1 - e^{-\lambda z} - \lambda z e^{-\lambda z}, \quad \text{if } z \geq 0.$$
**Example:** Suppose $X, Y \overset{\text{iid}}{\sim} \text{Exp}(\lambda)$, and let $Z = X + Y$. Then

$$P(Z \leq z) = \int_{\mathbb{R}} F_Y(z - x)f_X(x)\,dx$$

$$= \int_0^z (1 - e^{-\lambda(z-x)})\lambda e^{-\lambda x}\,dx$$

(must have $x \geq 0$ and $z - x \geq 0$)

$$= 1 - e^{-\lambda z} - \lambda ze^{-\lambda z}, \text{ if } z \geq 0.$$

Thus, the pdf of $Z$ is

$$\frac{d}{dz} P(Z \leq z) = \lambda^2 ze^{-\lambda z}, \text{ if } z \geq 0.$$
**Example:** Suppose $X, Y \overset{iid}{\sim} \text{Exp}(\lambda)$, and let $Z = X + Y$. Then

\[
P(Z \leq z) = \int_{\mathbb{R}} F_Y(z - x) f_X(x) \, dx
\]

\[
= \int_{0}^{z} (1 - e^{-\lambda(z-x)}) \lambda e^{-\lambda x} \, dx
\]

(must have $x \geq 0$ and $z - x \geq 0$)

\[
= 1 - e^{-\lambda z} - \lambda z e^{-\lambda z}, \quad \text{if } z \geq 0.
\]

Thus, the pdf of $Z$ is

\[
\frac{d}{dz} P(Z \leq z) = \lambda^2 z e^{-\lambda z}, \quad z \geq 0.
\]

This turns out to mean that $Z \sim \text{Gamma}(2, \lambda)$, aka $\text{Erlang}_2(\lambda)$. \qed
You can do the similar kinds of convolutions with discrete RVs. We state the following result without proof (which is straightforward).
You can do the similar kinds of convolutions with discrete RVs. We state the following result without proof (which is straightforward).

**Example/Theorem**: Suppose $X$ and $Y$ are two independent integer-valued RVs with pmf’s $f_X(x)$ and $f_Y(y)$. Then the pmf of $Z = X + Y$ is

$$f_Z(z) = P(Z = z) = \sum_{x=-\infty}^{\infty} f_X(x) f_Y(z - x).$$
Example Suppose $X$ and $Y$ are iid Bern$(p)$. Then the pmf of $Z = X + Y$ is
**Example** Suppose $X$ and $Y$ are iid Bern($p$). Then the pmf of $Z = X + Y$ is

$$f_Z(z) = \sum_{x=-\infty}^{\infty} f_X(x) f_Y(z-x)$$
**Example** Suppose $X$ and $Y$ are iid Bern($p$). Then the pmf of $Z = X + Y$ is

$$f_Z(z) = \sum_{x=-\infty}^{\infty} f_X(x) f_Y(z-x)$$

$$= f_X(0) f_Y(z) + f_X(1) f_Y(z-1) \quad (X \text{ can only } = 0 \text{ or } 1)$$

Thus, $Z \sim \text{Bin}(2,p)$, a fond blast from the past!
Example Suppose $X$ and $Y$ are iid Bern($p$). Then the pmf of $Z = X + Y$ is

$$f_Z(z) = \sum_{x=-\infty}^{\infty} f_X(x) f_Y(z-x)$$

$$= f_X(0) f_Y(z) + f_X(1) f_Y(z-1) \quad (X \text{ can only } = 0 \text{ or } 1)$$

$$= f_X(0) f_Y(z) 1_{\{0,1\}}(z) + f_X(1) f_Y(z-1) 1_{\{1,2\}}(z)$$

(1_{\{\cdot\}}(z) \text{ functions indicate nonzero } f_Y(\cdot)'s)
**Example** Suppose $X$ and $Y$ are iid Bern($p$). Then the pmf of $Z = X + Y$ is

\[
f_Z(z) = \sum_{x=-\infty}^{\infty} f_X(x) f_Y(z - x)
\]

\[
= f_X(0) f_Y(z) + f_X(1) f_Y(z - 1) \quad (X \text{ can only } = 0 \text{ or } 1)
\]

\[
= f_X(0) f_Y(z) 1_{\{0,1\}}(z) + f_X(1) f_Y(z - 1) 1_{\{1,2\}}(z)
\]

(\(1_{\{.\}}(z)\) functions indicate nonzero \(f_Y(\cdot)\)'s)

\[
= p^0 q^{1-0}
\]
Example Suppose $X$ and $Y$ are iid $\text{Bern}(p)$. Then the pmf of $Z = X + Y$ is

\[
f_Z(z) = \sum_{x=-\infty}^{\infty} f_X(x) f_Y(z-x) = f_X(0) f_Y(z) + f_X(1) f_Y(z-1) \quad (X \text{ can only } = 0 \text{ or } 1)
\]

\[
= f_X(0) f_Y(z) 1_{\{0,1\}}(z) + f_X(1) f_Y(z-1) 1_{\{1,2\}}(z) \\
(1_{\{\cdot\}}(z) \text{ functions indicate nonzero } f_Y(\cdot)'s)
\]

\[
= p^0 q^{1-0} p^z q^{1-z}
\]
**Example** Suppose $X$ and $Y$ are iid Bern$(p)$. Then the pmf of $Z = X + Y$ is

\[
f_Z(z) = \sum_{x=-\infty}^{\infty} f_X(x) f_Y(z-x)
= f_X(0) f_Y(z) + f_X(1) f_Y(z-1) \quad (X \text{ can only } = 0 \text{ or } 1)
= f_X(0) f_Y(z) 1_{\{0,1\}}(z) + f_X(1) f_Y(z-1) 1_{\{1,2\}}(z)
\quad (1_{\{.\}}(z) \text{ functions indicate nonzero } f_Y(\cdot)'s)
= p^0 q^{1-0} p^z q^{1-z} 1_{\{0,1\}}(z)
\]
Example Suppose $X$ and $Y$ are iid Bern($p$). Then the pmf of $Z = X + Y$ is

$$f_Z(z) = \sum_{x=-\infty}^{\infty} f_X(x) f_Y(z-x)$$

$$= f_X(0) f_Y(z) + f_X(1) f_Y(z-1) \quad (X \text{ can only } = 0 \text{ or } 1)$$

$$= f_X(0) f_Y(z) 1_{\{0,1\}}(z) + f_X(1) f_Y(z-1) 1_{\{1,2\}}(z)$$

$$\quad (1_{\cdot}(z) \text{ functions indicate nonzero } f_Y(\cdot)\text{’s})$$

$$= p^0 q^{1-0} p^z q^{1-z} 1_{\{0,1\}}(z) + p^1 q^{1-1} p^{z-1} q^{2-z} 1_{\{1,2\}}(z)$$
Example Suppose $X$ and $Y$ are iid Bern($p$). Then the pmf of $Z = X + Y$ is

$$f_Z(z) = \sum_{x=-\infty}^{\infty} f_X(x) f_Y(z-x)$$

$$= f_X(0) f_Y(z) + f_X(1) f_Y(z-1) \quad (X \text{ can only } = 0 \text{ or } 1)$$

$$= f_X(0) f_Y(z) 1_{\{0,1\}}(z) + f_X(1) f_Y(z-1) 1_{\{1,2\}}(z)$$

$(1_{\{\cdot\}}(z)$ functions indicate nonzero $f_Y(\cdot)$’s)

$$= p^0 q^{1-z} p^1 q^{1-z} 1_{\{0,1\}}(z) + p^1 q^{1-z} p^{z-1} q^{2-z} 1_{\{1,2\}}(z)$$

$$= p^z q^{2-z}$$
Example Suppose $X$ and $Y$ are iid Bern($p$). Then the pmf of $Z = X + Y$ is

$$f_Z(z) = \sum_{x=-\infty}^{\infty} f_X(x) f_Y(z - x)$$

$$= f_X(0) f_Y(z) + f_X(1) f_Y(z - 1) \quad (X \text{ can only } = 0 \text{ or } 1)$$

$$= f_X(0) f_Y(z) 1_{\{0,1\}}(z) + f_X(1) f_Y(z - 1) 1_{\{1,2\}}(z)$$

$(1_{\{\cdot\}}(z)$ functions indicate nonzero $f_Y(\cdot)$’s)

$$= p^0 q^{1-z} p^z q^{1-z} 1_{\{0,1\}}(z) + p^1 q^{1-1} p^{z-1} q^{2-z} 1_{\{1,2\}}(z)$$

$$= p^z q^{2-z} \left[ 1_{\{0,1\}}(z) + 1_{\{1,2\}}(z) \right]$$

Thus, $Z \sim \text{Bin}(2,p)$, a fond blast from the past!
**Example** Suppose \( X \) and \( Y \) are iid Bern(\( p \)). Then the pmf of \( Z = X + Y \) is

\[
f_Z(z) = \sum_{x=-\infty}^{\infty} f_X(x) f_Y(z-x)
\]

\[
= f_X(0) f_Y(z) + f_X(1) f_Y(z-1) \quad (X \text{ can only } = 0 \text{ or } 1)
\]

\[
= f_X(0) f_Y(z) 1_{\{0,1\}}(z) + f_X(1) f_Y(z-1) 1_{\{1,2\}}(z)
\]

\( (1_{\{\cdot\}}(z) \) functions indicate nonzero \( f_Y(\cdot) \)'s)

\[
= p^0 q^{1-0} p^z q^{1-z} 1_{\{0,1\}}(z) + p^1 q^{1-1} p^{z-1} q^{2-z} 1_{\{1,2\}}(z)
\]

\[
= p^z q^{2-z} \left[ 1_{\{0,1\}}(z) + 1_{\{1,2\}}(z) \right]
\]

\[
= \binom{2}{z} p^z q^{2-z}, \quad z = 0, 1, 2.
\]
**Example** Suppose $X$ and $Y$ are iid Bern($p$). Then the pmf of $Z = X + Y$ is

$$f_Z(z) = \sum_{x=-\infty}^{\infty} f_X(x)f_Y(z-x)$$

$$= f_X(0)f_Y(z) + f_X(1)f_Y(z-1) \quad (X \text{ can only } = 0 \text{ or } 1)$$

$$= f_X(0)f_Y(z)1_{\{0,1\}}(z) + f_X(1)f_Y(z-1)1_{\{1,2\}}(z)$$

(1$_{\{\cdot\}}$(z) functions indicate nonzero $f_Y(\cdot)$’s)

$$= p^0 q^{1-0} p^z q^{1-z} 1_{\{0,1\}}(z) + p^1 q^{1-1} p^{z-1} q^{2-z} 1_{\{1,2\}}(z)$$

$$= p^z q^{2-z} \left[ 1_{\{0,1\}}(z) + 1_{\{1,2\}}(z) \right]$$

$$= \binom{z}{2} p^z q^{2-z}, \quad z = 0, 1, 2.$$  

Thus, $Z \sim \text{Bin}(2, p)$, a fond blast from the past!  \(\Box\)
12 Covariance and Correlation
Lesson 3.12 — Covariance and Correlation

Covariance and correlation are measures used to define the degree of association between $X$ and $Y$ if they don’t happen to be independent.

**Definition:** The covariance between $X$ and $Y$ is $\text{Cov}(X,Y) \equiv \sigma_{XY} \equiv E[(X - E[X])(Y - E[Y])]$.

**Remark:** $\text{Cov}(X,X) = E[(X - E[X])^2] = \text{Var}(X)$. 


Lesson 3.12 — Covariance and Correlation

In the next few lessons we’ll cover:
Lesson 3.12 — Covariance and Correlation

In the next few lessons we’ll cover:

- Basic Concepts of Covariance and Correlation
- Causation
- A Couple of Worked Examples
- Some Useful Theorems
Lesson 3.12 — Covariance and Correlation

In the next few lessons we’ll cover:

- Basic Concepts of Covariance and Correlation
- Causation
- A Couple of Worked Examples
- Some Useful Theorems

Covariance and correlation are measures used to define the degree of association between $X$ and $Y$ if they don’t happen to be independent.
Lesson 3.12 — Covariance and Correlation

In the next few lessons we’ll cover:

- Basic Concepts of Covariance and Correlation
- Causation
- A Couple of Worked Examples
- Some Useful Theorems

Covariance and correlation are measures used to define the degree of association between $X$ and $Y$ if they don’t happen to be independent.

**Definition:** The covariance between $X$ and $Y$ is

$$\text{Cov}(X, Y) \equiv \sigma_{XY} \equiv \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])].$$
Lesson 3.12 — Covariance and Correlation

In the next few lessons we’ll cover:

- Basic Concepts of Covariance and Correlation
- Causation
- A Couple of Worked Examples
- Some Useful Theorems

Covariance and correlation are measures used to define the degree of *association* between $X$ and $Y$ if they don’t happen to be independent.

**Definition:** The covariance between $X$ and $Y$ is

$$\text{Cov}(X, Y) \equiv \sigma_{XY} \equiv E[(X - E[X])(Y - E[Y])].$$

**Remark:** $\text{Cov}(X, X) = E[(X - E[X])^2] = \text{Var}(X)$. 
**Remark**: If $X$ and $Y$ have positive covariance, then $X$ and $Y$ move “in the same direction.” Think height and weight.
**Remark**: If $X$ and $Y$ have positive covariance, then $X$ and $Y$ move “in the same direction.” Think height and weight.
If $X$ and $Y$ have negative covariance, then $X$ and $Y$ move “in opposite directions.” Think snowfall and temperature.
If $X$ and $Y$ have negative covariance, then $X$ and $Y$ move “in opposite directions.” Think snowfall and temperature.
If $X$ and $Y$ are independent, then of course they have no association with each other. In fact, we’ll prove below that independence implies that the covariance is 0 (but not the other way around).
If $X$ and $Y$ are independent, then of course they have no association with each other. In fact, we’ll prove below that independence implies that the covariance is 0 (but not the other way around).

**Example**: IBM stock price vs. temperature on Mars are independent — at least that’s what they want you to believe!
If $X$ and $Y$ are independent, then of course they have no association with each other. In fact, we’ll prove below that independence implies that the covariance is 0 (but not the other way around).

**Example:** IBM stock price vs. temperature on Mars are independent — at least that’s what they want you to believe!
Theorem (easier way to calculate covariance):

\[ \text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]. \]
Theorem (easier way to calculate covariance):

\[ \text{Cov}(X, Y) = E[XY] - E[X]E[Y]. \]

Proof:

\[ \text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])]. \]
Theorem (easier way to calculate covariance):

\[
\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].
\]

Proof:

\[
\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]
= \mathbb{E}
\left[
XY - X\mathbb{E}[Y] - Y\mathbb{E}[X] + \mathbb{E}[X]\mathbb{E}[Y]
\right].
\]
Theorem (easier way to calculate covariance):

\[ \text{Cov}(X, Y) = E[XY] - E[X]E[Y]. \]

Proof:

\[
\begin{align*}
\text{Cov}(X, Y) &= E[(X - E[X])(Y - E[Y])] \\
\end{align*}
\]
**Theorem** (easier way to calculate covariance):

\[ \text{Cov}(X, Y) = E[XY] - E[X]E[Y]. \]

**Proof:**

\[
\begin{align*}
\text{Cov}(X, Y) &= E[(X - E[X])(Y - E[Y])] \\
\end{align*}
\]

**Theorem**: $X$ and $Y$ independent implies $\text{Cov}(X, Y) = 0$. 
**Theorem** (easier way to calculate covariance):

\[ \text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]. \]

**Proof:**

\begin{align*}
\text{Cov}(X, Y) & = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \\
 & = \mathbb{E}[XY - X\mathbb{E}[Y] - Y\mathbb{E}[X] + \mathbb{E}[X]\mathbb{E}[Y]] \\
 & = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] - \mathbb{E}[Y]\mathbb{E}[X] + \mathbb{E}[X]\mathbb{E}[Y]. \quad \square
\end{align*}

**Theorem:** If \( X \) and \( Y \) are independent, then \( \text{Cov}(X, Y) = 0. \)

**Proof:** By a previous theorem, if \( X \) and \( Y \) are independent, then \( \mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]. \) Then
**Theorem** (easier way to calculate covariance):

\[ \text{Cov}(X, Y) = E[XY] - E[X]E[Y]. \]

**Proof:**

\[
\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])] \\
\]

**Theorem:** \( X \) and \( Y \) independent implies \( \text{Cov}(X, Y) = 0. \)

**Proof:** By a previous theorem, \( X \) and \( Y \) independent implies \( E[XY] = E[X]E[Y]. \) Then

\[
\]
**Danger Will Robinson!**  \( \text{Cov}(X, Y) = 0 \) **does not imply** that \( X \) and \( Y \) are independent!!
Danger Will Robinson! $\text{Cov}(X, Y) = 0$ does not imply that $X$ and $Y$ are independent!!

Example: Suppose $X \sim \text{Unif}(-1, 1)$ and $Y = X^2$ (so $X$ and $Y$ are clearly dependent).
**Danger Will Robinson!** \( \text{Cov}(X, Y) = 0 \) does not imply that \( X \) and \( Y \) are independent!!

**Example:** Suppose \( X \sim \text{Unif}(-1, 1) \) and \( Y = X^2 \) (so \( X \) and \( Y \) are clearly dependent).

But

\[
\mathbb{E}[X] = \int_{-1}^{1} x \cdot \frac{1}{2} \, dx = 0 \quad \text{and}
\]
Danger Will Robinson! Cov\( (X, Y) = 0 \) does not imply that \( X \) and \( Y \) are independent!!

**Example**: Suppose \( X \sim \text{Unif}(-1, 1) \) and \( Y = X^2 \) (so \( X \) and \( Y \) are clearly dependent).

But

\[
E[X] = \int_{-1}^{1} x \cdot \frac{1}{2} \, dx = 0 \quad \text{and} \quad E[XY] = E[X^3] = \int_{-1}^{1} x^3 \cdot \frac{1}{2} \, dx = 0,
\]
Danger Will Robinson! Cov(X, Y) = 0 does not imply that X and Y are independent!!

Example: Suppose X ∼ Unif(−1, 1) and Y = X^2 (so X and Y are clearly dependent).

But

\[ E[X] = \int_{-1}^{1} x \cdot \frac{1}{2} \, dx = 0 \] and

\[ E[XY] = E[X^3] = \int_{-1}^{1} x^3 \cdot \frac{1}{2} \, dx = 0, \]

so

\[ \text{Cov}(X, Y) = E[XY] - E[X]E[Y] = 0. \]
In fact, here’s a graphical illustration of this zero-correlation dependence phenomenon, where we’ve actually added some normal noise to $Y$ to make it look prettier.
In fact, here’s a graphical illustration of this zero-correlation dependence phenomenon, where we’ve actually added some normal noise to $Y$ to make it look prettier.
**Definition:** The correlation between $X$ and $Y$ is

\[ \rho = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \]

**Remark:** Covariance has "square" units; correlation is unitless.

**Corollary:** $X, Y$ independent implies $\rho = 0$.

**Theorem:** It can be shown that $-1 \leq \rho \leq 1$.

$\rho \approx 1$ is "high" correlation.

$\rho \approx 0$ is "low" correlation.

$\rho \approx -1$ is "high" negative correlation.

**Example:** Height is highly correlated with weight.

Temperature on Mars has low correlation with IBM stock price.
**Definition:** The **correlation** between $X$ and $Y$ is

$$\rho = \text{Corr}(X, Y) \equiv \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}.$$

**Remark:** Covariance has "square" units; correlation is unitless.

**Corollary:** $X$, $Y$ independent implies $\rho = 0$.

**Theorem:** It can be shown that $-1 \leq \rho \leq 1$.

$\rho \approx 1$ is "high" correlation.

$\rho \approx 0$ is "low" correlation.

$\rho \approx -1$ is "high" negative correlation.

**Example:** Height is highly correlated with weight.

Temperature on Mars has low correlation with IBM stock price.
**Definition:** The correlation between $X$ and $Y$ is

$$
\rho = \text{Corr}(X, Y) \equiv \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}.
$$

**Remark:** Covariance has “square” units; correlation is unitless.
**Definition:** The correlation between $X$ and $Y$ is

$$
\rho = \text{Corr}(X, Y) \equiv \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}.
$$

**Remark:** Covariance has “square” units; correlation is unitless.

**Corollary:** $X, Y$ independent implies $\rho = 0$. 

Example: Height is highly correlated with weight. Temperature on Mars has low correlation with IBM stock price.
**Definition:** The correlation between $X$ and $Y$ is

$$
\rho = \text{Corr}(X, Y) \equiv \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y}.
$$

**Remark:** Covariance has “square” units; correlation is unitless.

**Corollary:** $X, Y$ independent implies $\rho = 0$.

**Theorem:** It can be shown that $-1 \leq \rho \leq 1$. 

$\rho \approx 1$ is “high” correlation. 

$\rho \approx 0$ is “low” correlation. 

$\rho \approx -1$ is “high” negative correlation.

**Example:** Height is highly correlated with weight. 

Temperature on Mars has low correlation with IBM stock price.
**Definition:** The **correlation** between $X$ and $Y$ is

$$
\rho = \text{Corr}(X, Y) \equiv \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}.
$$

**Remark:** Covariance has “square” units; correlation is unitless.

**Corollary:** $X, Y$ independent implies $\rho = 0$.

**Theorem:** It can be shown that $-1 \leq \rho \leq 1$.

$\rho \approx 1$ is “high” correlation.

$\rho \approx 0$ is “low” correlation.

$\rho \approx -1$ is “high” negative correlation.

**Example:** Height is highly correlated with weight.

Temperature on Mars has low correlation with IBM stock price.
**Definition:** The **correlation** between $X$ and $Y$ is

$$\rho = \text{Corr}(X, Y) \equiv \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}.$$ 

**Remark:** Covariance has “square” units; correlation is unitless.

**Corollary:** $X, Y$ independent implies $\rho = 0$.

**Theorem:** It can be shown that $-1 \leq \rho \leq 1$.

$\rho \approx 1$ is “high” correlation.

$\rho \approx 0$ is “low” correlation.
**Definition:** The correlation between $X$ and $Y$ is

$$
\rho = \text{Corr}(X, Y) \equiv \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y}.
$$

**Remark:** Covariance has “square” units; correlation is unitless.

**Corollary:** $X, Y$ independent implies $\rho = 0$.

**Theorem:** It can be shown that $-1 \leq \rho \leq 1$.

- $\rho \approx 1$ is “high” correlation.
- $\rho \approx 0$ is “low” correlation.
- $\rho \approx -1$ is “high” negative correlation.

Example: Height is highly correlated with weight. Temperature on Mars has low correlation with IBM stock price.
**Definition:** The correlation between $X$ and $Y$ is

$$\rho = \text{Corr}(X, Y) \equiv \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}.$$ 

**Remark:** Covariance has “square” units; correlation is unitless.

**Corollary:** $X, Y$ independent implies $\rho = 0$.

**Theorem:** It can be shown that $-1 \leq \rho \leq 1$.

$\rho \approx 1$ is “high” correlation.
$\rho \approx 0$ is “low” correlation.
$\rho \approx -1$ is “high” negative correlation.

**Example:** Height is *highly* correlated with weight.
**Definition:** The correlation between $X$ and $Y$ is

\[ \rho = \text{Corr}(X, Y) \equiv \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}. \]

**Remark:** Covariance has “square” units; correlation is unitless.

**Corollary:** $X, Y$ independent implies $\rho = 0$.

**Theorem:** It can be shown that $-1 \leq \rho \leq 1$.

- $\rho \approx 1$ is “high” correlation.
- $\rho \approx 0$ is “low” correlation.
- $\rho \approx -1$ is “high” negative correlation.

**Example:** Height is *highly* correlated with weight. Temperature on Mars has *low* correlation with IBM stock price.
Correlation and Causation

1 Introduction
2 Marginal Distributions
3 Conditional Distributions
4 Independent Random Variables
5 Consequences of Independence
6 Random Samples
7 Conditional Expectation
8 Double Expectation
9 Honors Class: First-Step Analysis
10 Honors Class: Random Sums of Random Variables
11 Honors Class: Standard Conditioning Argument
12 Covariance and Correlation
13 **Correlation and Causation**
14 A Couple of Worked Correlation Examples
15 Some Useful Covariance / Correlation Theorems
16 Moment Generating Functions, Revisited
17 Honors Bivariate Functions of Random Variables
Correlation does not necessarily imply causality! This is a very common pitfall in many areas of data analysis and public discourse.

Example in which correlation does imply causality: Height and weight are positively correlated, and larger height does indeed tend to cause greater weight.

Example in which correlation does not imply causality: Temperature and lemonade sales have positive correlation, and temperature has a causal influence on lemonade sales. Similarly, temperature and overheating cars are positively correlated with a causal relationship. It’s also likely that lemonade sales and overheating cars are positively correlated, but there’s no causal relationship there.

Example of a zero correlation relationship with causality: We’ve seen that it’s possible for two dependent RVs to be uncorrelated.
NOTE! Correlation does not necessarily imply causality! This is a very common pitfall in many areas of data analysis and public discourse.
NOTE! Correlation does not necessarily imply causality! This is a very common pitfall in many areas of data analysis and public discourse.

Example in which correlation does imply causality:

Height and weight are positively correlated, and larger height does indeed tend to cause greater weight.

Example in which correlation does not imply causality:

Temperature and lemonade sales have positive correlation, and temperature has causal influence on lemonade sales. Similarly, temperature and overheating cars are positively correlated with a causal relationship. It’s also likely that lemonade sales and overheating cars are positively correlated, but there’s no causal relationship there.

Example of a zero correlation relationship with causality:

We’ve seen that it’s possible for two dependent RVs to be uncorrelated.
NOTE! Correlation does not necessarily imply causality! This is a very common pitfall in many areas of data analysis and public discourse.

Example in which correlation does imply causality: Height and weight are positively correlated, and larger height does indeed tend to cause greater weight.
NOTE! Correlation does not necessarily imply causality! This is a very common pitfall in many areas of data analysis and public discourse.

Example in which correlation does imply causality: Height and weight are positively correlated, and larger height does indeed tend to cause greater weight.

Example in which correlation does not imply causality:
Lesson 3.13 — Correlation and Causation

**NOTE! Correlation does not necessarily imply causality!** This is a very common pitfall in many areas of data analysis and public discourse.

**Example in which correlation does imply causality:** Height and weight are positively correlated, and larger height does indeed tend to cause greater weight.

**Example in which correlation does not imply causality:** Temperature and lemonade sales have positive cor, and temp has causal influence on lemonade sales. Similarly, temp and overheating cars are positively correlated with a causal relationship. It’s also likely that lemonade sales and overheating cars are positively correlated, but there’s no causal relationship there.
NOTE! Correlation does not necessarily imply causality! This is a very common pitfall in many areas of data analysis and public discourse.

Example in which correlation does imply causality: Height and weight are positively correlated, and larger height does indeed tend to cause greater weight.

Example in which correlation does not imply causality: Temperature and lemonade sales have positive correlation, and temp has causal influence on lemonade sales. Similarly, temp and overheating cars are positively correlated with a causal relationship. It’s also likely that lemonade sales and overheating cars are positively correlated, but there’s no causal relationship there.

Example of a zero correlation relationship with causality!
Lesson 3.13 — Correlation and Causation

NOTE! Correlation does not necessarily imply causality! This is a very common pitfall in many areas of data analysis and public discourse.

Example in which correlation does imply causality: Height and weight are positively correlated, and larger height does indeed tend to cause greater weight.

Example in which correlation does not imply causality: Temperature and lemonade sales have positive corr, and temp has causal influence on lemonade sales. Similarly, temp and overheating cars are positively correlated with a causal relationship. It’s also likely that lemonade sales and overheating cars are positively correlated, but there’s no causal relationship there.

Example of a zero correlation relationship with causality! We’ve seen that it’s possible for two dependent RVs to be uncorrelated.
To prove that $X$ causes $Y$, one must establish that:

1. $X$ occurred before $Y$;
2. The relationship between $X$ and $Y$ is not completely due to random chance; and
3. Nothing else accounts for the relationship (which is violated in the lemonade sales / overheating cars example above).

These items can be often be established via mathematical analysis, statistical analysis of appropriate data, or consultation with appropriate experts.
To prove that $X$ causes $Y$, one must establish that:

- $X$ occurred before $Y$;
To prove that $X$ causes $Y$, one must establish that:

- $X$ occurred before $Y$;
- The relationship between $X$ and $Y$ is not completely due to random chance; and
To prove that $X$ causes $Y$, one must establish that:

- $X$ occurred before $Y$;
- The relationship between $X$ and $Y$ is not completely due to random chance; and
- Nothing else accounts for the relationship (which is violated in the lemonade sales / overheating cars example above).
To prove that $X$ causes $Y$, one must establish that:

- $X$ occurred before $Y$;
- The relationship between $X$ and $Y$ is not completely due to random chance; and
- Nothing else accounts for the relationship (which is violated in the lemonade sales / overheating cars example above).

These items can be often be established via mathematical analysis, statistical analysis of appropriate data, or consultation with appropriate experts.
The three examples above seem to give conflicting guidance with respect to the relationship between correlation and causality. How can we interpret these findings in a meaningful way? Here are the takeaways:

1. If the correlation between \( X \) and \( Y \) is (significantly) nonzero, there is some type of relationship between the two items, which may or may not be causal; but this should raise our curiosity.

2. If the correlation between \( X \) and \( Y \) is 0, we are not quite out of the woods with respect to dependence and causality. In order to definitively rule out a relationship between \( X \) and \( Y \), it is always highly recommended protocol to, at the very least, plot data from \( X \) and \( Y \) against each other to see if there is a nonlinear relationship, as in the uncorrelated-yet-dependent example. Consult with appropriate experts.
The three examples above seem to give conflicting guidance with respect to the relationship between correlation and causality. How can we interpret these findings in a meaningful way? Here are the takeaways:

- If the correlation between $X$ and $Y$ is (significantly) nonzero, there is some type of relationship between the two items, which may or may not be causal; but this should raise our curiosity.
- If the correlation between $X$ and $Y$ is 0, we are not quite out of the woods with respect to dependence and causality. In order to definitively rule out a relationship between $X$ and $Y$, it is always highly recommended protocol to, at the very least, plot data from $X$ and $Y$ against each other to see if there is a nonlinear relationship, as in the uncorrelated-yet-dependent example. Consult with appropriate experts.
The three examples above seem to give conflicting guidance with respect to the relationship between correlation and causality. How can we interpret these findings in a meaningful way? Here are the takeaways:

- If the correlation between $X$ and $Y$ is (significantly) nonzero, there is some type of relationship between the two items, which may or may not be causal; but this should raise our curiosity.

- If the correlation between $X$ and $Y$ is 0, we are not quite out of the woods with respect to dependence and causality. In order to definitively rule out a relationship between $X$ and $Y$, it is always highly recommended protocol to, at the very least,
The three examples above seem to give conflicting guidance with respect to the relationship between correlation and causality. How can we interpret these findings in a meaningful way? Here are the takeaways:

- If the correlation between $X$ and $Y$ is (significantly) nonzero, there is some type of relationship between the two items, which may or may not be causal; but this should raise our curiosity.
- If the correlation between $X$ and $Y$ is 0, we are not quite out of the woods with respect to dependence and causality. In order to definitively rule out a relationship between $X$ and $Y$, it is always highly recommended protocol to, at the very least,
  - Plot data from $X$ and $Y$ against each other to see if there is a nonlinear relationship, as in the uncorrelated-yet-dependent example.
The three examples above seem to give conflicting guidance with respect to the relationship between correlation and causality. How can we interpret these findings in a meaningful way? Here are the takeaways:

- If the correlation between $X$ and $Y$ is (significantly) nonzero, there is some type of relationship between the two items, which may or may not be causal; but this should raise our curiosity.

- If the correlation between $X$ and $Y$ is 0, we are not quite out of the woods with respect to dependence and causality. In order to definitively rule out a relationship between $X$ and $Y$, it is always highly recommended protocol to, at the very least,
  - Plot data from $X$ and $Y$ against each other to see if there is a nonlinear relationship, as in the uncorrelated-yet-yet-dependent example.
  - Consult with appropriate experts.
<table>
<thead>
<tr>
<th>Number</th>
<th>Section</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Introduction</td>
</tr>
<tr>
<td>2</td>
<td>Marginal Distributions</td>
</tr>
<tr>
<td>3</td>
<td>Conditional Distributions</td>
</tr>
<tr>
<td>4</td>
<td>Independent Random Variables</td>
</tr>
<tr>
<td>5</td>
<td>Consequences of Independence</td>
</tr>
<tr>
<td>6</td>
<td>Random Samples</td>
</tr>
<tr>
<td>7</td>
<td>Conditional Expectation</td>
</tr>
<tr>
<td>8</td>
<td>Double Expectation</td>
</tr>
<tr>
<td>9</td>
<td>Honors Class: First-Step Analysis</td>
</tr>
<tr>
<td>10</td>
<td>Honors Class: Random Sums of Random Variables</td>
</tr>
<tr>
<td>11</td>
<td>Honors Class: Standard Conditioning Argument</td>
</tr>
<tr>
<td>12</td>
<td>Covariance and Correlation</td>
</tr>
<tr>
<td>13</td>
<td>Correlation and Causation</td>
</tr>
<tr>
<td>14</td>
<td>A Couple of Worked Correlation Examples</td>
</tr>
<tr>
<td>15</td>
<td>Some Useful Covariance / Correlation Theorems</td>
</tr>
<tr>
<td>16</td>
<td>Moment Generating Functions, Revisited</td>
</tr>
<tr>
<td>17</td>
<td>Honors Bivariate Functions of Random Variables</td>
</tr>
</tbody>
</table>
Lesson 3.14 — A Couple of Worked Correlation Examples

Discrete Example:
Suppose \( X \) is the GPA of a UGA student, and \( Y \) is their IQ. Here's the joint pmf.

\[
\begin{array}{c|c|c|c}
X & Y = 40 & Y = 50 & Y = 60 \\
\hline
2 & 0.0 & 0.2 & 0.2 \\
3 & 0.4 & & \\
4 & & & \\
\end{array}
\]

\[
\begin{array}{c|c|c|c}
X & Y = 40 & Y = 50 & Y = 60 \\
\hline
0.5 & 0.3 & 0.2 & \\
1 & & & \\
\end{array}
\]

We'll spare the details, but here are the relevant calculations.
Discrete Example: Suppose $X$ is the GPA of a UGA student, and $Y$ is their IQ. Here’s the joint pmf.
**Discrete Example:** Suppose $X$ is the GPA of a UGA student, and $Y$ is their IQ. Here’s the joint pmf.

<table>
<thead>
<tr>
<th>$f(x, y)$</th>
<th>$X = 2$</th>
<th>$X = 3$</th>
<th>$X = 4$</th>
<th>$f_Y(y)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y = 40$</td>
<td>0.0</td>
<td>0.2</td>
<td>0.2</td>
<td>0.4</td>
</tr>
<tr>
<td>$Y = 50$</td>
<td>0.1</td>
<td>0.1</td>
<td>0.0</td>
<td>0.2</td>
</tr>
<tr>
<td>$Y = 60$</td>
<td>0.4</td>
<td>0.0</td>
<td>0.0</td>
<td>0.4</td>
</tr>
<tr>
<td>$f_X(x)$</td>
<td>0.5</td>
<td>0.3</td>
<td>0.2</td>
<td>1</td>
</tr>
</tbody>
</table>
Discrete Example: Suppose $X$ is the GPA of a UGA student, and $Y$ is their IQ. Here’s the joint pmf.

<table>
<thead>
<tr>
<th>$f(x, y)$</th>
<th>$X = 2$</th>
<th>$X = 3$</th>
<th>$X = 4$</th>
<th>$f_Y(y)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y = 40$</td>
<td>0.0</td>
<td>0.2</td>
<td>0.2</td>
<td>0.4</td>
</tr>
<tr>
<td>$Y = 50$</td>
<td>0.1</td>
<td>0.1</td>
<td>0.0</td>
<td>0.2</td>
</tr>
<tr>
<td>$Y = 60$</td>
<td>0.4</td>
<td>0.0</td>
<td>0.0</td>
<td>0.4</td>
</tr>
</tbody>
</table>

$\sum f_X(x) = 1$, $\sum f_Y(y) = 1$.

We’ll spare the details, but here are the relevant calculations...
A Couple of Worked Correlation Examples

\[ E[X] = \sum x f_X(x) = 2.7, \]
A Couple of Worked Correlation Examples

\[ E[X] = \sum_x x f_X(x) = 2.7, \]

\[ E[X^2] = \sum_x x^2 f_X(x) = 7.9, \text{ and} \]
A Couple of Worked Correlation Examples

\[
E[X] = \sum_x x f_X(x) = 2.7,
\]

\[
E[X^2] = \sum_x x^2 f_X(x) = 7.9, \quad \text{and}
\]

\[
\text{Var}(X) = E[X^2] - (E[X])^2 = 0.61.
\]
$$E[X] = \sum_x x f_X(x) = 2.7,$$

$$E[X^2] = \sum_x x^2 f_X(x) = 7.9,$$ and

$$\text{Var}(X) = E[X^2] - (E[X])^2 = 0.61.$$

Similarly, $E[Y] = 50$, $E[Y^2] = 2580$, and $\text{Var}(Y) = 80$. Finally,
A Couple of Worked Correlation Examples

\[ E[X] = \sum_x x f_X(x) = 2.7, \]

\[ E[X^2] = \sum_x x^2 f_X(x) = 7.9, \quad \text{and} \]

\[ \text{Var}(X) = E[X^2] - (E[X])^2 = 0.61. \]

Similarly, \( E[Y] = 50, E[Y^2] = 2580, \) and \( \text{Var}(Y) = 80. \) Finally,

\[ E[XY] = \sum_x \sum_y xy f(x, y) \]
\[ E[X] = \sum_x x f_X(x) = 2.7, \]
\[ E[X^2] = \sum_x x^2 f_X(x) = 7.9, \quad \text{and} \]
\[ \text{Var}(X) = E[X^2] - (E[X])^2 = 0.61. \]

Similarly, \( E[Y] = 50, E[Y^2] = 2580, \) and \( \text{Var}(Y) = 80. \) Finally,
\[
E[XY] = \sum_x \sum_y xy f(x, y) \\
= 2(40)(0.0) + 3(40)(0.2) + \cdots + 4(60)(0.0) = 129,
\]
A Couple of Worked Correlation Examples

\[ E[X] = \sum_x x f_X(x) = 2.7, \]
\[ E[X^2] = \sum_x x^2 f_X(x) = 7.9, \text{ and} \]
\[ \text{Var}(X) = E[X^2] - (E[X])^2 = 0.61. \]

Similarly, \( E[Y] = 50, E[Y^2] = 2580, \) and \( \text{Var}(Y) = 80. \) Finally,
\[ E[XY] = \sum_x \sum_y xy f(x, y) \]
\[ = 2(40)(0.0) + 3(40)(0.2) + \cdots + 4(60)(0.0) = 129, \]
\[ \text{Cov}(X,Y) = E[XY] - E[X]E[Y] = -6.0, \text{ and} \]
A Couple of Worked Correlation Examples

\[ E[X] = \sum_{x} x f_X(x) = 2.7, \]
\[ E[X^2] = \sum_{x} x^2 f_X(x) = 7.9, \quad \text{and} \]
\[ \text{Var}(X) = E[X^2] - (E[X])^2 = 0.61. \]

Similarly, \( E[Y] = 50, E[Y^2] = 2580, \) and \( \text{Var}(Y) = 80. \) Finally,

\[ E[XY] = \sum_{x} \sum_{y} xy f(x, y) \]
\[ = 2(40)(0.0) + 3(40)(0.2) + \cdots + 4(60)(0.0) = 129, \]
\[ \text{Cov}(X, Y) = E[XY] - E[X]E[Y] = -6.0, \quad \text{and} \]
\[ \rho = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = -0.859. \quad \square \]
Continuous Example: Suppose $f(x, y) = 10x^2y$, $0 \leq y \leq x \leq 1$. 
Continuous Example: Suppose \( f(x, y) = 10x^2y, \ 0 \leq y \leq x \leq 1. \)

\[
f_X(x) = \int_0^x 10x^2y \, dy = 5x^4, \quad 0 \leq x \leq 1,
\]
Continuous Example: Suppose \( f(x, y) = 10x^2y, 0 \leq y \leq x \leq 1. \)

\[
f_X(x) = \int_0^x 10x^2y \, dy = 5x^4, \quad 0 \leq x \leq 1,
\]

\[
E[X] = \int_0^1 5x^5 \, dx = 5/6,
\]
**Continuous Example:** Suppose \( f(x, y) = 10x^2y, \ 0 \leq y \leq x \leq 1. \)

\[
f_X(x) = \int_0^x 10x^2y \, dy = 5x^4, \quad 0 \leq x \leq 1,
\]

\[
E[X] = \int_0^1 5x^5 \, dx = 5/6,
\]

\[
E[X^2] = \int_0^1 5x^6 \, dx = 5/7,
\]
Continuous Example: Suppose $f(x, y) = 10x^2y$, $0 \leq y \leq x \leq 1$.

\[
f_X(x) = \int_0^x 10x^2y \, dy = 5x^4, \quad 0 \leq x \leq 1,
\]

\[
E[X] = \int_0^1 5x^5 \, dx = 5/6,
\]

\[
E[X^2] = \int_0^1 5x^6 \, dx = 5/7,
\]

\[
\text{Var}(X) = E[X^2] - (E[X])^2 = 0.01984.
\]
Similarly,

\[ f_Y(y) = \int_y^1 10x^2y \, dx = \frac{10}{3} y(1 - y^3), \quad 0 \leq y \leq 1, \]
Similarly,

\[ f_Y(y) = \int_y^1 10x^2y \, dx = \frac{10}{3} y(1 - y^3), \quad 0 \leq y \leq 1, \]

\[ E[Y] = \frac{5}{9}, \quad \text{Var}(Y) = 0.04850, \]
Similarly,

\[ f_Y(y) = \int_y^1 10x^2 y \, dx = \frac{10}{3} y(1 - y^3), \quad 0 \leq y \leq 1, \]

\[ E[Y] = \frac{5}{9}, \quad \text{Var}(Y) = 0.04850, \]

\[ E[XY] = \int_0^1 \int_0^x 10x^3 y^2 \, dy \, dx = \frac{10}{21}, \]
Similarly,

\[ f_Y(y) = \int_y^1 10x^2y \, dx = \frac{10}{3} y(1 - y^3), \quad 0 \leq y \leq 1, \]

\[ E[Y] = \frac{5}{9}, \quad \text{Var}(Y) = 0.04850, \]

\[ E[XY] = \int_0^1 \int_0^x 10x^3y^2 \, dy \, dx = \frac{10}{21}, \]

\[ \text{Cov}(X, Y) = E[XY] - E[X]E[Y] = 0.01323, \]
Similarly,

\[
\begin{align*}
  f_Y(y) &= \int_y^1 10x^2y \, dx = \frac{10}{3}y(1 - y^3), \quad 0 \leq y \leq 1, \\
  E[Y] &= \frac{5}{9}, \quad \text{Var}(Y) = 0.04850, \\
  E[XY] &= \int_0^1 \int_0^x 10x^3y^2 \, dy \, dx = \frac{10}{21}, \\
  \text{Cov}(X,Y) &= E[XY] - E[X]E[Y] = 0.01323, \\
  \rho &= \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = 0.4265. \quad \Box
\end{align*}
\]
| 1  | Introduction                        |
| 2  | Marginal Distributions              |
| 3  | Conditional Distributions           |
| 4  | Independent Random Variables        |
| 5  | Consequences of Independence        |
| 6  | Random Samples                      |
| 7  | Conditional Expectation             |
| 8  | Double Expectation                  |
| 9  | Honors Class: First-Step Analysis  |
| 10 | Honors Class: Random Sums of Random Variables |
| 11 | Honors Class: Standard Conditioning Argument |
| 12 | Covariance and Correlation          |
| 13 | Correlation and Causation           |
| 14 | A Couple of Worked Correlation Examples |
| 15 | Some Useful Covariance / Correlation Theorems |
| 16 | Moment Generating Functions, Revisited |
| 17 | Honors Bivariate Functions of Random Variables |
Some Useful Covariance / Correlation Theorems

Lesson 3.15 — Some Useful Covariance / Correlation Theorems

Theorem:

\[ \text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X,Y) \]

whether or not \(X\) and \(Y\) are independent.

Remark: If \(X\), \(Y\) are independent, the covariance term goes away.

Proof: By the work we did on a previous proof,

\[ \text{Var}(X + Y) = E[X^2] - (E[X])^2 + E[Y^2] - (E[Y])^2 + 2(E[XY] - E[X]E[Y]) \]

\[ = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X,Y) \]
Theorem: \( \text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2 \text{Cov}(X, Y) \), whether or not \( X \) and \( Y \) are independent.
Theorem: $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$, whether or not $X$ and $Y$ are independent.

Remark: If $X, Y$ are independent, the covariance term goes away.
Lesson 3.15 — Some Useful Covariance / Correlation Theorems

**Theorem:** \( \text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y) \), whether or not \( X \) and \( Y \) are independent.

**Remark:** If \( X, Y \) are independent, the covariance term goes away.

**Proof:** By the work we did on a previous proof,
Lesson 3.15 — Some Useful Covariance / Correlation Theorems

**Theorem:** \( \text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y) \), whether or not \( X \) and \( Y \) are independent.

**Remark:** If \( X, Y \) are independent, the covariance term goes away.

**Proof:** By the work we did on a previous proof,

\[
\begin{align*}
\text{Var}(X + Y) &= E[X^2] - (E[X])^2 + E[Y^2] - (E[Y])^2 \\
&+ 2(E[XY] - E[X]E[Y])
\end{align*}
\]
Lesson 3.15 — Some Useful Covariance / Correlation Theorems

**Theorem:** \( \text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y) \), *whether or not* \( X \) and \( Y \) are independent.

**Remark:** If \( X, Y \) are independent, the covariance term goes away.

**Proof:** By the work we did on a previous proof,

\[
\text{Var}(X + Y) = E[X^2] - (E[X])^2 + E[Y^2] - (E[Y])^2 + 2(E[XY] - E[X]E[Y])
\]

\[
= \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y). \quad \square
\]
Theorem:

\[ \text{Var}\left( \sum_{i=1}^{n} X_i \right) = \sum_{i=1}^{n} \text{Var}(X_i) + 2 \sum_{i=1}^{n} \sum_{i<j} \text{Cov}(X_i, X_j). \]
Theorem:

\[
\text{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \text{Var}(X_i) + 2 \sum_{i<j} \text{Cov}(X_i, X_j).
\]

Proof: Induction.
Some Useful Covariance / Correlation Theorems

Theorem:

\[ \text{Var} \left( \sum_{i=1}^{n} X_i \right) = \sum_{i=1}^{n} \text{Var}(X_i) + 2 \sum_{i<j} \text{Cov}(X_i, X_j). \]

Proof: Induction.

Corollary: If all \( X_i \)'s are independent, then

\[ \text{Var} \left( \sum_{i=1}^{n} X_i \right) = \sum_{i=1}^{n} \text{Var}(X_i). \]
**Theorem:** $\text{Cov}(aX, bY + c) = ab \text{Cov}(X, Y)$. 

**Proof:** Put the above two results together.
**Theorem:** \( \text{Cov}(aX, bY + c) = ab \text{Cov}(X, Y). \)

**Proof:**

\[
\text{Cov}(aX, bY + c) = E[aX \cdot (bY + c)] - E[aX]E[bY + c]
\]
Theorem: \( \text{Cov}(aX, bY + c) = ab \text{Cov}(X, Y) \). 

Proof:

\[
\text{Cov}(aX, bY + c) = E[aX \cdot (bY + c)] - E[aX]E[bY + c] \\
\]
**Theorem:** $\text{Cov}(aX, bY + c) = ab \text{Cov}(X, Y)$.

**Proof:**

\[
\begin{align*}
\text{Cov}(aX, bY + c) &= E[aX \cdot (bY + c)] - E[aX]E[bY + c] \\
&= abE[XY] - abE[X]E[Y] + acE[X] - acE[X]
\end{align*}
\]
Theorem: $\text{Cov}(aX, bY + c) = ab \text{Cov}(X, Y)$.

Proof:

\[
\text{Cov}(aX, bY + c) = \mathbb{E}[aX \cdot (bY + c)] - \mathbb{E}[aX]\mathbb{E}[bY + c]
\]
\[
= \mathbb{E}[abXY] + \mathbb{E}[acX] - \mathbb{E}[aX]\mathbb{E}[bY] - \mathbb{E}[aX]\mathbb{E}[c]
\]
\[
= ab \mathbb{E}[XY] - ab \mathbb{E}[X]\mathbb{E}[Y] + ac\mathbb{E}[X] - ac\mathbb{E}[X]
\]
\[
= ab \text{Cov}(X, Y). \quad \square
\]
**Theorem:** $\text{Cov}(aX, bY + c) = ab \text{Cov}(X, Y)$.

**Proof:**

$$\text{Cov}(aX, bY + c) = E[aX \cdot (bY + c)] - E[aX]E[bY + c]$$


$$= ab E[XY] - ab E[X]E[Y] + ac E[X] - ac E[X]$$

$$= ab \text{Cov}(X, Y). \quad \Box$$

**Theorem:**

$$\text{Var}\left(\sum_{i=1}^{n} a_i X_i + c\right) = \sum_{i=1}^{n} a_i^2 \text{Var}(X_i) + 2 \sum_{i=1}^{n} \sum_{j<i} a_i a_j \text{Cov}(X_i, X_j).$$
Some Useful Covariance / Correlation Theorems

**Theorem:** $\text{Cov}(aX, bY + c) = ab \text{Cov}(X, Y)$.

**Proof:**

$$\text{Cov}(aX, bY + c) = E[aX \cdot (bY + c)] - E[aX]E[bY + c]$$


$$= ab E[XY] - ab E[X]E[Y] + acE[X] - acE[X]$$

$$= ab \text{Cov}(X, Y). \quad \Box$$

**Theorem:**

$$\text{Var}\left(\sum_{i=1}^{n} a_iX_i + c\right) = \sum_{i=1}^{n} a_i^2 \text{Var}(X_i) + 2 \sum \sum_{i<j} a_ia_j \text{Cov}(X_i, X_j).$$

**Proof:** Put the above two results together. \quad \Box
Example: \( \text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y) - 2\text{Cov}(X, Y) \).
Some Useful Covariance / Correlation Theorems

Example: \( \text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y) - 2\text{Cov}(X, Y). \)

Example: Suppose \( \text{Var}(X) = \text{Var}(Y) = \text{Var}(Z) = 10, \) \( \text{Cov}(X, Y) = 3, \text{Cov}(X, Z) = -2, \) and \( \text{Cov}(Y, Z) = 0. \) Then

\[ \text{Var}(X - 2Y + 3Z) = \text{Var}(X) + 4\text{Var}(Y) + 9\text{Var}(Z) - 4\text{Cov}(X, Y) + 6\text{Cov}(X, Z) - 12\text{Cov}(Y, Z) = 14(10) - 4(3) + 6(-2) - 12(0) = 116. \]
Example: $\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y) - 2\text{Cov}(X, Y)$.

Example: Suppose $\text{Var}(X) = \text{Var}(Y) = \text{Var}(Z) = 10$, $\text{Cov}(X, Y) = 3$, $\text{Cov}(X, Z) = -2$, and $\text{Cov}(Y, Z) = 0$. Then

$$\text{Var}(X - 2Y + 3Z)$$
Example: \( \text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y) - 2\text{Cov}(X, Y) \). 

Example: Suppose \( \text{Var}(X) = \text{Var}(Y) = \text{Var}(Z) = 10 \), \( \text{Cov}(X, Y) = 3 \), \( \text{Cov}(X, Z) = -2 \), and \( \text{Cov}(Y, Z) = 0 \). Then

\[
\begin{align*}
\text{Var}(X - 2Y + 3Z) &= \text{Var}(X) + 4\text{Var}(Y) + 9\text{Var}(Z) \\
&\quad - 4\text{Cov}(X, Y) + 6\text{Cov}(X, Z) - 12\text{Cov}(Y, Z)
\end{align*}
\]
**Example:** \( \text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y) - 2\text{Cov}(X, Y) \).

**Example:** Suppose \( \text{Var}(X) = \text{Var}(Y) = \text{Var}(Z) = 10 \), \( \text{Cov}(X, Y) = 3 \), \( \text{Cov}(X, Z) = -2 \), and \( \text{Cov}(Y, Z) = 0 \). Then

\[
\text{Var}(X - 2Y + 3Z) \\
= \text{Var}(X) + 4\text{Var}(Y) + 9\text{Var}(Z) \\
- 4\text{Cov}(X, Y) + 6\text{Cov}(X, Z) - 12\text{Cov}(Y, Z) \\
= 14(10) - 4(3) + 6(-2) - 12(0) = 116. \quad \square
\]
Moment Generating Functions, Revisited

Introduction
Marginal Distributions
Conditional Distributions
Independent Random Variables
Consequences of Independence
Random Samples
Conditional Expectation
Double Expectation
Honors Class: First-Step Analysis
Honors Class: Random Sums of Random Variables
Honors Class: Standard Conditioning Argument
Covariance and Correlation
Correlation and Causation
A Couple of Worked Correlation Examples
Some Useful Covariance / Correlation Theorems
Moment Generating Functions, Revisited
Honors Bivariate Functions of Random Variables
Moment Generating Functions, Revisited

Lesson 3.16 — Moment Generating Functions, Revisited

Old Definition:
$M_X(t) \equiv E[e^{tX}]$ is the moment generating function (mgf) of the RV $X$.

Old Example:
If $X \sim \text{Bern}(p)$, then
$$M_X(t) = E[e^{tX}] = \sum_x e^{tx} f(x) = e^{t} \cdot 1^p + e^{t} \cdot 0^q = e^{t}$$

Old Example:
If $X \sim \text{Exp}(\lambda)$, then
$$M_X(t) = E[e^{tX}] = \int_{\mathbb{R}} e^{tx} f(x) \, dx = \lambda \left( \frac{1}{\lambda} - t \right) \text{if } \lambda > t.$$
Old Definition: $M_X(t) \equiv E[e^{tX}]$ is the moment generating function (mgf) of the RV $X$. 
Lesson 3.16 — Moment Generating Functions, Revisited

**Old Definition:** \( M_X(t) \equiv E[e^{tX}] \) is the **moment generating function** (mgf) of the RV \( X \).

**Old Example:** If \( X \sim \text{Bern}(p) \), then

\[
M_X(t) = E[e^{tX}] = \sum_x e^{tx} f(x)
\]
Lesson 3.16 — Moment Generating Functions, Revisited

Old Definition: $M_X(t) \equiv \mathbb{E}[e^{tX}]$ is the moment generating function (mgf) of the RV $X$.

Old Example: If $X \sim \text{Bern}(p)$, then

$$M_X(t) = \mathbb{E}[e^{tX}] = \sum_x e^{tx} f(x) = e^{t \cdot 1} p + e^{t \cdot 0} q$$
Old Definition: \( M_X(t) \equiv \mathbb{E}[e^{tX}] \) is the \textbf{moment generating function} (mgf) of the RV \( X \).

Old Example: If \( X \sim \text{Bern}(p) \), then

\[
M_X(t) = \mathbb{E}[e^{tX}] = \sum_x e^{tx} f(x) = e^{t\cdot1} p + e^{t\cdot0} q = pe^t + q. \quad \square
\]
Lesson 3.16 — Moment Generating Functions, Revisited

Old Definition: \( M_X(t) \equiv E[e^{tX}] \) is the **moment generating function** (mgf) of the RV \( X \).

Old Example: If \( X \sim \text{Bern}(p) \), then

\[
M_X(t) = E[e^{tX}] = \sum_x e^{tx} f(x) = e^{t \cdot 1}p + e^{t \cdot 0}q = pe^t + q. \]

\( \blacksquare \)

Old Example: If \( X \sim \text{Exp}(\lambda) \), then

\[
M_X(t) = E[e^{tX}] = \int_{\mathbb{R}} e^{tx} f(x) \, dx
\]
Lesson 3.16 — Moment Generating Functions, Revisited

**Old Definition:** $M_X(t) \equiv \mathbb{E}[e^{tX}]$ is the moment generating function (mgf) of the RV $X$.

**Old Example:** If $X \sim \text{Bern}(p)$, then

$$M_X(t) = \mathbb{E}[e^{tX}] = \sum_x e^{tx} f(x) = e^{t \cdot 1} p + e^{t \cdot 0} q = pe^t + q. \quad \square$$

**Old Example:** If $X \sim \text{Exp}(\lambda)$, then

$$M_X(t) = \mathbb{E}[e^{tX}] = \int_{\mathbb{R}} e^{tx} f(x) \, dx = \frac{\lambda}{\lambda - t} \quad \text{if } \lambda > t. \quad \square$$
Lesson 3.16 — Moment Generating Functions, Revisited

**Old Definition:** $M_X(t) \equiv E[e^{tX}]$ is the **moment generating function** (mgf) of the RV $X$.

**Old Example:** If $X \sim \text{Bern}(p)$, then

$$M_X(t) = E[e^{tX}] = \sum_x e^{tx}f(x) = e^{t \cdot 1}p + e^{t \cdot 0}q = pe^t + q. \quad \square$$

**Old Example:** If $X \sim \text{Exp}(\lambda)$, then

$$M_X(t) = E[e^{tX}] = \int_{\mathbb{R}} e^{tx}f(x) \, dx = \frac{\lambda}{\lambda - t} \quad \text{if } \lambda > t. \quad \square$$

**Old Theorem** (why it’s called the mgf): Under certain technical conditions,

$$E[X^k]$$
**Lesson 3.16 — Moment Generating Functions, Revisited**

**Old Definition:** \( M_X(t) \equiv \mathbb{E}[e^{tX}] \) is the **moment generating function** (mgf) of the RV \( X \).

**Old Example:** If \( X \sim \text{Bern}(p) \), then

\[
M_X(t) = \mathbb{E}[e^{tX}] = \sum_x e^{tx} f(x) = e^{t1}p + e^{t0}q = pe^t + q. \quad \square
\]

**Old Example:** If \( X \sim \text{Exp}(\lambda) \), then

\[
M_X(t) = \mathbb{E}[e^{tX}] = \int_{\mathbb{R}} e^{tx} f(x) \, dx = \frac{\lambda}{\lambda - t} \quad \text{if } \lambda > t. \quad \square
\]

**Old Theorem** (why it’s called the mgf): Under certain technical conditions,

\[
\mathbb{E}[X^k] = \left. \frac{d^k}{dt^k} M_X(t) \right|_{t=0}, \quad k = 1, 2, \ldots
\]
New Theorem (mgf of the sum of independent RVs): Suppose $X_1, \ldots, X_n$ are independent. Let $Y = \sum_{i=1}^{n} X_i$. Then
New Theorem (mgf of the sum of independent RVs): Suppose $X_1, \ldots, X_n$ are independent. Let $Y = \sum_{i=1}^n X_i$. Then

$$M_Y(t) = \prod_{i=1}^n M_{X_i}(t).$$
**New Theorem** (mgf of the sum of independent RVs): Suppose $X_1, \ldots, X_n$ are *independent*. Let $Y = \sum_{i=1}^{n} X_i$. Then

$$M_Y(t) = \prod_{i=1}^{n} M_{X_i}(t).$$

**Proof:**

$$M_Y(t) = E[e^{tY}]$$
New Theorem (mgf of the sum of independent RVs): Suppose $X_1, \ldots, X_n$ are independent. Let $Y = \sum_{i=1}^{n} X_i$. Then

$$M_Y(t) = \prod_{i=1}^{n} M_{X_i}(t).$$

Proof:

$$M_Y(t) = E[e^{tY}] = E[e^{t\sum X_i}]$$
**New Theorem** (mgf of the sum of independent RVs): Suppose $X_1, \ldots, X_n$ are independent. Let $Y = \sum_{i=1}^{n} X_i$. Then

$$M_Y(t) = \prod_{i=1}^{n} M_{X_i}(t).$$

**Proof:**

$$M_Y(t) = \mathbb{E}[e^{tY}]$$

$$= \mathbb{E}[e^{t \sum X_i}]$$

$$= \mathbb{E} \left[ \prod_{i=1}^{n} e^{tX_i} \right]$$
New Theorem (mgf of the sum of independent RVs): Suppose $X_1, \ldots, X_n$ are independent. Let $Y = \sum_{i=1}^{n} X_i$. Then

$$M_Y(t) = \prod_{i=1}^{n} M_{X_i}(t).$$

Proof:

$$M_Y(t) = E[e^{tY}]$$

$$= E[e^{t\sum X_i}]$$

$$= E\left[\prod_{i=1}^{n} e^{tX_i}\right]$$

$$= \prod_{i=1}^{n} E[e^{tX_i}] \quad (X_i's \text{ independent})$$
New Theorem (mgf of the sum of independent RVs): Suppose $X_1, \ldots, X_n$ are independent. Let $Y = \sum_{i=1}^n X_i$. Then

$$M_Y(t) = \prod_{i=1}^n M_{X_i}(t).$$

Proof:

$$M_Y(t) = E[e^{tY}]$$

$$= E[e^{t\sum X_i}]$$

$$= E\left[ \prod_{i=1}^n e^{tX_i} \right]$$

$$= \prod_{i=1}^n E[e^{tX_i}] \quad (X_i's \text{ independent})$$

$$= \prod_{i=1}^n M_{X_i}(t). \qed$$
**Corollary**: If $X_1, \ldots, X_n$ are iid and $Y = \sum_{i=1}^{n} X_i$, then

$$M_Y(t) = [M_{X_1}(t)]^n.$$
Corollary: If $X_1, \ldots, X_n$ are iid and $Y = \sum_{i=1}^{n} X_i$, then

$$M_Y(t) = [M_{X_1}(t)]^n.$$
Corollary: If $X_1, \ldots, X_n$ are iid and $Y = \sum_{i=1}^{n} X_i$, then

$$M_Y(t) = [M_{X_1}(t)]^n.$$ 

Example: Suppose $X_1, \ldots, X_n \overset{iid}{\sim} \text{Bern}(p)$. Then by a previous example,
Corollary: If $X_1, \ldots, X_n$ are iid and $Y = \sum_{i=1}^{n} X_i$, then

$$M_Y(t) = [M_{X_1}(t)]^n.$$ 

Example: Suppose $X_1, \ldots, X_n \sim \text{Bern}(p)$. Then by a previous example,

$$M_Y(t) = [M_{X_1}(t)]^n = (pe^t + q)^n.$$ 

**Corollary:** If $X_1, \ldots, X_n$ are iid and $Y = \sum_{i=1}^{n} X_i$, then

$$M_Y(t) = [M_{X_1}(t)]^n.$$ 

**Example:** Suppose $X_1, \ldots, X_n \overset{iid}{\sim} \text{Bern}(p)$. Then by a previous example,

$$M_Y(t) = [M_{X_1}(t)]^n = (pe^t + q)^n.$$ 

So what use is a result like this? We can use results such as this with our old friend...
Corollary: If $X_1, \ldots, X_n$ are iid and $Y = \sum_{i=1}^{n} X_i$, then

$$M_Y(t) = [M_{X_1}(t)]^n.$$  

Example: Suppose $X_1, \ldots, X_n \sim \text{Bern}(p)$. Then by a previous example,

$$M_Y(t) = [M_{X_1}(t)]^n = (pe^t + q)^n.$$  

So what use is a result like this? We can use results such as this with our old friend.

Old Theorem (identifying distributions): In this class, each distribution has a unique mgf.
Example/Theorem: The sum $Y$ of $n$ iid Bern$(p)$ RVs is the same as a Bin$(n, p)$ RV.
**Example/Theorem:** The sum $Y$ of $n$ iid $\text{Bern}(p)$ RVs is the same as a $\text{Bin}(n, p)$ RV.

By the previous example and uniqueness, all we need to show is that the mgf of $Z \sim \text{Bin}(n, p)$ matches $M_Y(t) = (pe^t + q)^n$. 
Example/Theorem: The sum $Y$ of $n$ iid Bern($p$) RVs is the same as a Bin($n, p$) RV.

By the previous example and uniqueness, all we need to show is that the mgf of $Z \sim \text{Bin}(n, p)$ matches $M_Y(t) = (pe^t + q)^n$. To this end, we have

$$M_Z(t) = \mathbb{E}[e^{tZ}]$$
**Example/Theorem:** The sum $Y$ of $n$ iid Bern($p$) RVs is the same as a Bin($n$, $p$) RV.

By the previous example and uniqueness, all we need to show is that the mgf of $Z \sim \text{Bin}(n, p)$ matches $M_Y(t) = (pe^t + q)^n$. To this end, we have

$$M_Z(t) = \mathbb{E}[e^{tZ}]$$

$$= \sum_{z} e^{tz} P(Z = z)$$
**Example/Theorem:** The sum $Y$ of $n$ iid Bern($p$) RVs is the same as a Bin($n, p$) RV.

By the previous example and uniqueness, all we need to show is that the mgf of $Z \sim \text{Bin}(n, p)$ matches $M_Y(t) = (pe^t + q)^n$. To this end, we have

$$
M_Z(t) = \mathbb{E}[e^{tZ}]
$$

$$
= \sum_{z} e^{tz} P(Z = z)
$$

$$
= \sum_{z=0}^{n} e^{tz} \binom{n}{z} p^z q^{n-z}
$$
**Example/Theorem**: The sum $Y$ of $n$ iid Bern($p$) RVs is the same as a Bin($n, p$) RV.

By the previous example and uniqueness, all we need to show is that the mgf of $Z \sim$ Bin($n, p$) matches $M_Y(t) = (pe^t + q)^n$. To this end, we have

$$M_Z(t) = \mathbb{E}[e^{tZ}] = \sum_z e^{tz} P(Z = z) = \sum_{z=0}^{n} \binom{n}{z} p^z q^{n-z} \left(pe^t\right)^z q^{n-z} = \left(pe^t + q\right)^n.$$
**Example/Theorem:** The sum $Y$ of $n$ iid Bern($p$) RVs is the same as a Bin($n, p$) RV.

By the previous example and uniqueness, all we need to show is that the mgf of $Z \sim \text{Bin}(n, p)$ matches $M_Y(t) = (pe^t + q)^n$. To this end, we have

$$M_Z(t) = E[e^{tZ}]$$

$$= \sum_{z} e^{tz} P(Z = z)$$

$$= \sum_{z=0}^{n} e^{tz} \binom{n}{z} p^z q^{n-z}$$

$$= \sum_{z=0}^{n} \binom{n}{z} (pe^t)^z q^{n-z}$$

$$= (pe^t + q)^n \quad \text{(by the Binomial Theorem).} \quad \square$$
**Example**: You can identify a distribution by its mgf.
Example: You can identify a distribution by its mgf.

\[ M_X(t) = \left( \frac{3}{4}e^t + \frac{1}{4} \right)^{15} \]
Example: You can identify a distribution by its mgf.

\[ M_X(t) = \left( \frac{3}{4} e^t + \frac{1}{4} \right)^{15} \]

implies that \( X \sim \text{Bin}(15, 0.75) \).  \( \square \)
Example: You can identify a distribution by its mgf.

\[ M_X(t) = \left( \frac{3}{4}e^t + \frac{1}{4} \right)^{15} \]

implies that \( X \sim \text{Bin}(15, 0.75) \). \( \square \)

Old Theorem (mgf of a linear function of \( X \)): Suppose \( X \) has mgf \( M_X(t) \) and let \( Y = aX + b \). Then \( M_Y(t) = e^{tb}M_X(at) \).
Example: You can identify a distribution by its mgf.

\[ M_X(t) = \left( \frac{3}{4} e^t + \frac{1}{4} \right)^{15} \]

implies that \( X \sim \text{Bin}(15, 0.75) \).  \( \square \)

Old Theorem (mgf of a linear function of \( X \)): Suppose \( X \) has mgf \( M_X(t) \) and let \( Y = aX + b \). Then \( M_Y(t) = e^{tb} M_X(at) \).

Example:

\[ M_Y(t) = e^{-2t} \left( \frac{3}{4} e^{3t} + \frac{1}{4} \right)^{15} \]
Example: You can identify a distribution by its mgf.

\[ M_X(t) = \left( \frac{3}{4} e^t + \frac{1}{4} \right)^{15} \]

implies that \( X \sim \text{Bin}(15, 0.75) \). \( \square \)

Old Theorem (mgf of a linear function of \( X \)): Suppose \( X \) has mgf \( M_X(t) \) and let \( Y = aX + b \). Then \( M_Y(t) = e^{tb} M_X(at) \).

Example:

\[ M_Y(t) = e^{-2t} \left( \frac{3}{4} e^{3t} + \frac{1}{4} \right)^{15} = e^{bt} (pe^{at} + q)^n \]
**Example:** You can identify a distribution by its mgf.

\[ M_X(t) = \left( \frac{3}{4} e^t + \frac{1}{4} \right)^{15} \]

implies that \( X \sim \text{Bin}(15, 0.75) \). \[ \square \]

**Old Theorem** (mgf of a linear function of \( X \)): Suppose \( X \) has mgf \( M_X(t) \) and let \( Y = aX + b \). Then \( M_Y(t) = e^{tb} M_X(at) \).

**Example:**

\[ M_Y(t) = e^{-2t} \left( \frac{3}{4} e^{3t} + \frac{1}{4} \right)^{15} = e^{bt} (pe^{at} + q)^n = e^{bt} M_X(at), \]
**Example**: You can identify a distribution by its mgf.

\[
M_X(t) = \left(\frac{3}{4}e^t + \frac{1}{4}\right)^{15}
\]

implies that \(X \sim \text{Bin}(15, 0.75).\)  \(\square\)

**Old Theorem** (mgf of a linear function of \(X\)): Suppose \(X\) has mgf \(M_X(t)\) and let \(Y = aX + b\). Then \(M_Y(t) = e^{tb}M_X(at)\).

**Example**:

\[
M_Y(t) = e^{-2t}\left(\frac{3}{4}e^{3t} + \frac{1}{4}\right)^{15} = e^{bt}(pe^{at} + q)^n = e^{bt}M_X(at),
\]

which implies that \(Y\) has the same distribution as \(3X - 2\), where \(X \sim \text{Bin}(15, 0.75)\).  \(\square\)
**Theorem** (Additive property of Binomials): If $X_1, \ldots, X_k$ are independent, with $X_i \sim \text{Bin}(n_i, p)$ (where $p$ is the same for all $X_i$'s), then
**Theorem** (Additive property of Binomials): If $X_1, \ldots, X_k$ are independent, with $X_i \sim \text{Bin}(n_i, p)$ (where $p$ is the same for all $X_i$'s), then

$$Y \equiv \sum_{i=1}^{k} X_i \sim \text{Bin}\left(\sum_{i=1}^{k} n_i, p\right).$$
Theorem (Additive property of Binomials): If $X_1, \ldots, X_k$ are independent, with $X_i \sim \text{Bin}(n_i, p)$ (where $p$ is the same for all $X_i$’s), then

$$Y \equiv \sum_{i=1}^{k} X_i \sim \text{Bin}\left( \sum_{i=1}^{k} n_i, p \right).$$

Proof:

$$M_Y(t) = \prod_{i=1}^{k} M_{X_i}(t) \quad \text{(mgf of independent sum)}$$
**Theorem** (Additive property of Binomials): If $X_1, \ldots, X_k$ are independent, with $X_i \sim \text{Bin}(n_i, p)$ (where $p$ is the same for all $X_i$’s), then

$$Y \equiv \sum_{i=1}^{k} X_i \sim \text{Bin}\left(\sum_{i=1}^{k} n_i, p\right).$$

**Proof:**

$$M_Y(t) = \prod_{i=1}^{k} M_{X_i}(t) \quad \text{(mgf of independent sum)}$$

$$= \prod_{i=1}^{k} (pe^t + q)^{n_i} \quad \text{(Binomial}(n_i, p) \text{ mgf})$$
Theorem (Additive property of Binomials): If $X_1, \ldots, X_k$ are independent, with $X_i \sim \text{Bin}(n_i, p)$ (where $p$ is the same for all $X_i$'s), then

$$Y \equiv \sum_{i=1}^{k} X_i \sim \text{Bin}\left(\sum_{i=1}^{k} n_i, p\right).$$

Proof:

$$M_Y(t) = \prod_{i=1}^{k} M_{X_i}(t) \quad \text{(mgf of independent sum)}$$

$$= \prod_{i=1}^{k} (pe^t + q)^{n_i} \quad \text{(Binomial}(n_i, p) \text{ mgf)}$$

$$= (pe^t + q)^{\sum_{i=1}^{k} n_i}.$$
**Theorem** (Additive property of Binomials): If $X_1, \ldots, X_k$ are independent, with $X_i \sim \text{Bin}(n_i, p)$ (where $p$ is the same for all $X_i$'s), then

$$Y \equiv \sum_{i=1}^{k} X_i \sim \text{Bin}\left(\sum_{i=1}^{k} n_i, p\right).$$

**Proof:**

$$M_Y(t) = \prod_{i=1}^{k} M_{X_i}(t) \quad (\text{mgf of independent sum})$$

$$= \prod_{i=1}^{k} (pe^t + q)^{n_i} \quad (\text{Binomial}(n_i, p) \text{ mgf})$$

$$= (pe^t + q)^{\sum_{i=1}^{k} n_i}.$$

This is the mgf of the $\text{Bin}\left(\sum_{i=1}^{k} n_i, p\right)$, so we’re done. □
Honors Bivariate Functions of Random Variables
In earlier work, we looked at functions of a single variable, e.g., what is the expected value of $h(X)$? (LOTUS, from Module 2)

What is the distribution of $h(X)$? (functions of RVs, from Module 2)

And sometimes even functions of two (or more) variables. For example, if the $X_i$'s are independent, what's $\text{Var}\left(\sum_{i=1}^{n} X_i\right)$? (earlier in Module 3)

Use a standard conditioning argument to get the distribution of $X+Y$. (earlier in Module 3)

Goal: Now let's give a general result on the distribution of functions of two random variables, the proof of which is beyond the scope of our class.
In earlier work, we looked at . . .

- Functions of a single variable, e.g., what is the expected value of $h(X)$? (LOTUS, from Module 2)
In earlier work, we looked at...

- Functions of a single variable, e.g., what is the expected value of $h(X)$? (LOTUS, from Module 2)
- What is the distribution of $h(X)$? (functions of RVs, from Module 2)
Lesson 3.17 — Honors Bivariate Functions of Random Variables

In earlier work, we looked at...

- Functions of a single variable, e.g., what is the expected value of $h(X)$? (LOTUS, from Module 2)
- What is the distribution of $h(X)$? (functions of RVs, from Module 2)
- And sometimes even functions of two (or more) variables. For example, if the $X_i$’s are independent, what’s $\text{Var}(\sum_{i=1}^{n} X_i)$? (earlier in Module 3)
In earlier work, we looked at…

- Functions of a single variable, e.g., what is the expected value of $h(X)$? (LOTUS, from Module 2)
- What is the distribution of $h(X)$? (functions of RVs, from Module 2)
- And sometimes even functions of two (or more) variables. For example, if the $X_i$’s are independent, what’s $\text{Var}(\sum_{i=1}^{n} X_i)$? (earlier in Module 3)
- Use a standard conditioning argument to get the distribution of $X + Y$. (earlier in Module 3)
In earlier work, we looked at.

- Functions of a single variable, e.g., what is the expected value of $h(X)$? (LOTUS, from Module 2)
- What is the distribution of $h(X)$? (functions of RVs, from Module 2)
- And sometimes even functions of two (or more) variables. For example, if the $X_i$’s are independent, what’s $\text{Var}(\sum_{i=1}^{n} X_i)$? (earlier in Module 3)
- Use a standard conditioning argument to get the distribution of $X + Y$. (earlier in Module 3)

**Goal:** Now let’s give a general result on the distribution of functions of two random variables, the proof of which is beyond the scope of our class.
**Honors Theorem:** Suppose $X$ and $Y$ are continuous RVs with joint pdf $f(x, y)$, and $V = h_1(X, Y)$ and $W = h_2(X, Y)$ are functions of $X$ and $Y$, and
**Honors Theorem:** Suppose $X$ and $Y$ are continuous RVs with joint pdf $f(x, y)$, and $V = h_1(X, Y)$ and $W = h_2(X, Y)$ are functions of $X$ and $Y$, and

$$X = k_1(V, W) \quad \text{and} \quad Y = k_2(V, W),$$

for suitably chosen inverse functions $k_1$ and $k_2$. 
**Honors Theorem:** Suppose $X$ and $Y$ are continuous RVs with joint pdf $f(x, y)$, and $V = h_1(X, Y)$ and $W = h_2(X, Y)$ are functions of $X$ and $Y$, and

$$X = k_1(V, W) \quad \text{and} \quad Y = k_2(V, W),$$

for suitably chosen inverse functions $k_1$ and $k_2$.

Then the joint pdf of $V$ and $W$ is

$$g(v, w) = f(k_1(v, w), k_2(v, w)) |J(v, w)|,$$
**Honors Theorem:** Suppose $X$ and $Y$ are continuous RVs with joint pdf $f(x, y)$, and $V = h_1(X, Y)$ and $W = h_2(X, Y)$ are functions of $X$ and $Y$, and

$$X = k_1(V, W) \quad \text{and} \quad Y = k_2(V, W),$$

for suitably chosen inverse functions $k_1$ and $k_2$.

Then the joint pdf of $V$ and $W$ is

$$g(v, w) = f(k_1(v, w), k_2(v, w)) \left| J(v, w) \right|,$$

where $|J|$ is the absolute value of the *Jacobian* (determinant) of the transformation, i.e.,
**Honors Theorem:** Suppose $X$ and $Y$ are continuous RVs with joint pdf $f(x, y)$, and $V = h_1(X, Y)$ and $W = h_2(X, Y)$ are functions of $X$ and $Y$, and

$$X = k_1(V, W) \quad \text{and} \quad Y = k_2(V, W),$$

for suitably chosen inverse functions $k_1$ and $k_2$.

Then the joint pdf of $V$ and $W$ is

$$g(v, w) = f(k_1(v, w), k_2(v, w)) \mid J(v, w)\mid,$$

where $\mid J\mid$ is the absolute value of the *Jacobian* (determinant) of the transformation, i.e.,

$$J(v, w) = \begin{vmatrix} \frac{\partial x}{\partial v} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial v} & \frac{\partial y}{\partial w} \end{vmatrix}.$$
Honors Theorem: Suppose $X$ and $Y$ are continuous RVs with joint pdf $f(x, y)$, and $V = h_1(X, Y)$ and $W = h_2(X, Y)$ are functions of $X$ and $Y$, and

\[ X = k_1(V, W) \quad \text{and} \quad Y = k_2(V, W), \]

for suitably chosen inverse functions $k_1$ and $k_2$.

Then the joint pdf of $V$ and $W$ is

\[ g(v, w) = f(k_1(v, w), k_2(v, w)) |J(v, w)|, \]

where $|J|$ is the absolute value of the Jacobian (determinant) of the transformation, i.e.,

\[
J(v, w) = \left| \begin{array}{cc}
\frac{\partial x}{\partial v} & \frac{\partial x}{\partial w} \\
\frac{\partial y}{\partial v} & \frac{\partial y}{\partial w}
\end{array} \right| = \frac{\partial x}{\partial v} \frac{\partial y}{\partial w} - \frac{\partial y}{\partial v} \frac{\partial x}{\partial w}.
\]
**Corollary**: If $X$ and $Y$ are *independent*, then the joint pdf of $V$ and $W$ is

$$g(v, w) = f_X(k_1(v, w)) f_Y(k_2(v, w)) |J(v, w)|.$$
**Corollary**: If $X$ and $Y$ are independent, then the joint pdf of $V$ and $W$ is

$$g(v, w) = f_X(k_1(v, w)) f_Y(k_2(v, w)) |J(v, w)|.$$ 

**Remark**: These results generalize the 1-D method from Module 2.
Corollary: If $X$ and $Y$ are independent, then the joint pdf of $V$ and $W$ is

$$g(v, w) = f_X(k_1(v, w)) f_Y(k_2(v, w)) |J(v, w)|.$$  

Remark: These results generalize the 1-D method from Module 2.

You can use this method to find all sorts of cool stuff, e.g., the distribution of $X + Y$, $X/Y$, etc., as well as the joint pdf of any functions of $X$ and $Y$. 
Corollary: If $X$ and $Y$ are independent, then the joint pdf of $V$ and $W$ is

$$g(v, w) = f_X(k_1(v, w)) f_Y(k_2(v, w)) |J(v, w)|.$$ 

Remark: These results generalize the 1-D method from Module 2.

You can use this method to find all sorts of cool stuff, e.g., the distribution of $X + Y$, $X/Y$, etc., as well as the joint pdf of any functions of $X$ and $Y$.

Remark: Although the notation is nasty, the application isn’t really so bad.
**Example**: Suppose $X$ and $Y$ are iid Exp($\lambda$). Find the pdf of $X + Y$. 

We'll set $V = X + Y$ along with the dummy RV $W = X$.

This yields $X = W = k_1(V,W)$ and $Y = V - W = k_2(V,W)$.

To get the Jacobian term, we calculate $\frac{\partial x}{\partial v} = 0$, $\frac{\partial x}{\partial w} = 1$, $\frac{\partial y}{\partial v} = 1$, and $\frac{\partial y}{\partial w} = -1$, so that $|J| = \left| \begin{array}{cc} 0 & -1 \\ 1 & 1 \end{array} \right| = 1$. 

**Honors Bivariate Functions of Random Variables**
Example: Suppose $X$ and $Y$ are iid Exp($\lambda$). Find the pdf of $X + Y$.

We’ll set $V = X + Y$ along with the dummy RV $W = X$. 
**Example:** Suppose $X$ and $Y$ are iid Exp($\lambda$). Find the pdf of $X + Y$.

We’ll set $V = X + Y$ along with the dummy RV $W = X$.

This yields

$$X = W = k_1(V, W)$$
**Example:** Suppose $X$ and $Y$ are iid Exp($\lambda$). Find the pdf of $X + Y$.

We’ll set $V = X + Y$ along with the dummy RV $W = X$.

This yields

$$X = W = k_1(V, W) \quad \text{and} \quad Y = V - W = k_2(V, W).$$
**Example:** Suppose $X$ and $Y$ are iid Exp($\lambda$). Find the pdf of $X + Y$.

We’ll set $V = X + Y$ along with the dummy RV $W = X$.

This yields

$$X = W = k_1(V, W) \quad \text{and} \quad Y = V - W = k_2(V, W).$$

To get the Jacobian term, we calculate

$$\frac{\partial x}{\partial v} = 0, \quad \frac{\partial x}{\partial w} = 1, \quad \frac{\partial y}{\partial v} = 1, \quad \text{and} \quad \frac{\partial y}{\partial w} = -1,$$
**Example**: Suppose $X$ and $Y$ are iid $\text{Exp}(\lambda)$. Find the pdf of $X + Y$.

We’ll set $V = X + Y$ along with the dummy RV $W = X$.

This yields

$$X = W = k_1(V, W) \quad \text{and} \quad Y = V - W = k_2(V, W).$$

To get the Jacobian term, we calculate

$$\frac{\partial x}{\partial v} = 0, \quad \frac{\partial x}{\partial w} = 1, \quad \frac{\partial y}{\partial v} = 1, \quad \text{and} \quad \frac{\partial y}{\partial w} = -1,$$

so that

$$|J| = \left| \frac{\partial x}{\partial v} \frac{\partial y}{\partial w} - \frac{\partial y}{\partial v} \frac{\partial x}{\partial w} \right| = |0(-1) - 1(1)| = 1.$$
This implies that the joint pdf of $V$ and $W$ is
This implies that the joint pdf of $V$ and $W$ is

$$g(v, w) = f(k_1(v, w), k_2(v, w)) |J(v, w)|$$
This implies that the joint pdf of $V$ and $W$ is

$$g(v, w) = f(k_1(v, w), k_2(v, w)) |J(v, w)|$$

$$= f(w, v - w) \cdot 1$$
This implies that the joint pdf of \( V \) and \( W \) is

\[
g(v, w) = f(k_1(v, w), k_2(v, w)) |J(v, w)| \\
= f(w, v - w) \cdot 1 \\
= f_X(w)f_Y(v - w) \quad (X \text{ and } Y \text{ independent})
\]
This implies that the joint pdf of $V$ and $W$ is

$$g(v, w) = f(k_1(v, w), k_2(v, w)) |J(v, w)|$$

$$= f(w, v - w) \cdot 1$$

$$= f_X(w) f_Y(v - w) \quad (X \text{ and } Y \text{ independent})$$

$$= \lambda e^{-\lambda w} \cdot \lambda e^{-\lambda(v-w)}, \quad \text{for } w > 0 \text{ and } v - w > 0$$
This implies that the joint pdf of $V$ and $W$ is

$$g(v, w) = f(k_1(v, w), k_2(v, w)) |J(v, w)|$$

$$= f(w, v - w) \cdot 1$$

$$= f_X(w)f_Y(v - w) \quad (X \text{ and } Y \text{ independent})$$

$$= \lambda e^{-\lambda w} \cdot \lambda e^{-\lambda(v-w)}, \quad \text{for } w > 0 \text{ and } v - w > 0$$

$$= \lambda^2 e^{-\lambda v}, \quad \text{for } 0 < w < v.$$
This implies that the joint pdf of $V$ and $W$ is

$$g(v, w) = f(k_1(v, w), k_2(v, w)) |J(v, w)|$$

$$= f(w, v - w) \cdot 1$$

$$= f_X(w)f_Y(v - w) \quad (X \text{ and } Y \text{ independent})$$

$$= \lambda e^{-\lambda w} \cdot \lambda e^{-\lambda (v-w)}, \quad \text{for } w > 0 \text{ and } v - w > 0$$

$$= \lambda^2 e^{-\lambda v}, \quad \text{for } 0 < w < v.$$ 

And, finally, we obtain the desired pdf of the sum $V$ (after carefully noting the region of integration),

$$g_V(v) = \int_{-\infty}^{\infty} g(v, w) dw = \lambda^2 e^{-\lambda v}, \quad \text{for } 0 < w < v.$$
This implies that the joint pdf of $V$ and $W$ is

\[
g(v, w) = f(k_1(v, w), k_2(v, w)) |J(v, w)|
\]
\[
= f(w, v - w) \cdot 1 \\
= f_X(w) f_Y(v - w) \quad (X \text{ and } Y \text{ independent})
\]
\[
= \lambda e^{-\lambda w} \cdot \lambda e^{-\lambda(v-w)}, \quad \text{for } w > 0 \text{ and } v - w > 0
\]
\[
= \lambda^2 e^{-\lambda v}, \quad \text{for } 0 < w < v.
\]

And, finally, we obtain the desired pdf of the sum $V$ (after carefully noting the region of integration),

\[
g_V(v) = \int_{\mathbb{R}} g(v, w) \, dw
\]
This implies that the joint pdf of $V$ and $W$ is

$$g(v, w) = f(k_1(v, w), k_2(v, w)) |J(v, w)|$$

$$= f(w, v - w) \cdot 1$$

$$= f_X(w) f_Y(v - w) \quad (X \text{ and } Y \text{ independent})$$

$$= \lambda e^{-\lambda w} \cdot \lambda e^{-\lambda(v-w)}, \quad \text{for } w > 0 \text{ and } v - w > 0$$

$$= \lambda^2 e^{-\lambda v}, \quad \text{for } 0 < w < v.$$

And, finally, we obtain the desired pdf of the sum $V$ (after carefully noting the region of integration),

$$g_V(v) = \int_R g(v, w) \, dw = \int_0^v \lambda^2 e^{-\lambda v} \, dw$$
This implies that the joint pdf of $V$ and $W$ is

$$g(v, w) = f(k_1(v, w), k_2(v, w)) |J(v, w)| = f(w, v - w) \cdot 1 = f_X(w)f_Y(v - w) \quad (X \text{ and } Y \text{ independent}) = \lambda e^{-\lambda w} \cdot \lambda e^{-\lambda(v - w)}, \quad \text{for } w > 0 \text{ and } v - w > 0 = \lambda^2 e^{-\lambda v}, \quad \text{for } 0 < w < v.$$

And, finally, we obtain the desired pdf of the sum $V$ (after carefully noting the region of integration),

$$g_V(v) = \int_{\mathbb{R}} g(v, w) \, dw = \int_0^v \lambda^2 e^{-\lambda v} \, dw = \lambda^2 ve^{-\lambda v}, \quad \text{for } v > 0.$$
This implies that the joint pdf of $V$ and $W$ is

$$
g(v, w) = f(k_1(v, w), k_2(v, w)) |J(v, w)|
= f(w, v - w) \cdot 1
= f_X(w)f_Y(v - w) \quad (X \text{ and } Y \text{ independent})
= \lambda e^{-\lambda w} \cdot \lambda e^{-\lambda(v-w)}, \quad \text{for } w > 0 \text{ and } v - w > 0
= \lambda^2 e^{-\lambda v}, \quad \text{for } 0 < w < v.
$$

And, finally, we obtain the desired pdf of the sum $V$ (after carefully noting the region of integration),

$$
g_V(v) = \int_{\mathbb{R}} g(v, w) \, dw = \int_{0}^{v} \lambda^2 e^{-\lambda v} \, dw = \lambda^2 v e^{-\lambda v}, \quad \text{for } v > 0.
$$

This is the Gamma$(2, \lambda)$ pdf, which matches our answer from earlier in the current module. $\square$
Honors Example: Suppose $X$ and $Y$ are iid Unif(0,1). Find the joint pdf of $V = X + Y$ and $W = X/Y$.
**Honors Example:** Suppose $X$ and $Y$ are iid Unif(0,1). Find the joint pdf of $V = X + Y$ and $W = X/Y$.

After some algebra, we obtain

$$X = \frac{VW}{W + 1} = k_1(V, W) \quad \text{and}$$
**Honors Example:** Suppose $X$ and $Y$ are iid Unif(0,1). Find the joint pdf of $V = X + Y$ and $W = X/Y$.

After some algebra, we obtain

\[ X = \frac{VW}{W+1} = k_1(V, W) \quad \text{and} \quad Y = \frac{V}{W+1} = k_2(V, W). \]
**Honors Example:** Suppose $X$ and $Y$ are iid Unif$(0,1)$. Find the joint pdf of $V = X + Y$ and $W = X/Y$.

After some algebra, we obtain

$$X = \frac{VW}{W + 1} = k_1(V, W) \quad \text{and} \quad Y = \frac{V}{W + 1} = k_2(V, W).$$

After more algebra, we calculate

$$\frac{\partial x}{\partial v} = \frac{w}{w + 1}, \quad \frac{\partial x}{\partial w} = \frac{v}{(w + 1)^2}, \quad \frac{\partial y}{\partial v} = \frac{1}{w + 1}, \quad \frac{\partial y}{\partial w} = \frac{-v}{(w + 1)^2},$$

so that after still more algebra,

$$|J| = \left| \begin{vmatrix} \frac{\partial x}{\partial v} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial v} & \frac{\partial y}{\partial w} \end{vmatrix} \right| = \frac{v}{(w + 1)^2}.$$
**Honors Example:** Suppose $X$ and $Y$ are iid Unif(0,1). Find the joint pdf of $V = X + Y$ and $W = X/Y$.

After some algebra, we obtain

$$X = \frac{VW}{W + 1} = k_1(V, W) \quad \text{and} \quad Y = \frac{V}{W + 1} = k_2(V, W).$$

After more algebra, we calculate

$$\frac{\partial x}{\partial v} = \frac{w}{w + 1}, \quad \frac{\partial x}{\partial w} = \frac{u}{(w + 1)^2}, \quad \frac{\partial y}{\partial v} = \frac{1}{w + 1}, \quad \frac{\partial y}{\partial w} = \frac{-v}{(w + 1)^2},$$

so that after still more algebra,

$$|J| = \left| \frac{\partial x}{\partial v} \frac{\partial y}{\partial w} - \frac{\partial y}{\partial v} \frac{\partial x}{\partial w} \right| = \frac{v}{(w + 1)^2}.$$
This implies that the joint pdf of $V$ and $W$ is
This implies that the joint pdf of $V$ and $W$ is

$$g(v, w) = f(k_1(v, w), k_2(v, w)) |J(v, w)|$$
This implies that the joint pdf of $V$ and $W$ is

$$g(v, w) = f(k_1(v, w), k_2(v, w)) |J(v, w)|$$

$$= f\left(\frac{vw}{v+1}, \frac{v}{w+1}\right) \cdot \frac{v}{(w+1)^2}$$
This implies that the joint pdf of $V$ and $W$ is

\[
g(v, w) = f(k_1(v, w), k_2(v, w)) |J(v, w)|
\]

\[
= f\left(\frac{vw}{w+1}, \frac{v}{w+1}\right) \cdot \frac{v}{(w+1)^2}
\]

\[
= f_X\left(\frac{vw}{w+1}\right) f_Y\left(\frac{v}{w+1}\right) \frac{v}{(w+1)^2} \quad (X \text{ and } Y \text{ indep})
\]
This implies that the joint pdf of $V$ and $W$ is

\[
g(v, w) = f(k_1(v, w), k_2(v, w)) |J(v, w)|
\]

\[
= f\left(\frac{vw}{w+1}, \frac{v}{w+1}\right) \cdot \frac{v}{(w+1)^2}
\]

\[
= f_X\left(\frac{vw}{w+1}\right) f_Y\left(\frac{v}{w+1}\right) \frac{v}{(w+1)^2} \quad (X \text{ and } Y \text{ indep})
\]

\[
= 1 \cdot 1 \cdot \frac{v}{(w+1)^2}, \text{ for } 0 < x, y < 1 \text{ (since } X, Y \sim \text{ Unif}(0,1))
\]
This implies that the joint pdf of $V$ and $W$ is

$$g(v, w) = f(k_1(v, w), k_2(v, w)) |J(v, w)|$$

$$= f\left(\frac{vw}{w+1}, \frac{v}{w+1}\right) \cdot \frac{v}{(w+1)^2}$$

$$= f_X\left(\frac{vw}{w+1}\right) f_Y\left(\frac{v}{w+1}\right) \frac{v}{(w+1)^2} \quad (X \text{ and } Y \text{ indep})$$

$$= 1 \cdot 1 \cdot \frac{v}{(w+1)^2}, \text{ for } 0 < x, y < 1 \text{ (since } X, Y \sim \text{Unif}(0,1))$$

$$= \frac{v}{(w+1)^2}, \text{ for } 0 < x = \frac{vw}{w+1} < 1 \text{ and } 0 < y = \frac{v}{w+1} < 1.$$
This implies that the joint pdf of $V$ and $W$ is

\[
g(v, w) = f\left(k_1(v, w), k_2(v, w)\right) |J(v, w)|
\]

\[
= f\left(\frac{vw}{w+1}, \frac{v}{w+1}\right) \cdot \frac{v}{(w+1)^2}
\]

\[
= f_X\left(\frac{vw}{w+1}\right) f_Y\left(\frac{v}{w+1}\right) \frac{v}{(w+1)^2} \quad (X \text{ and } Y \text{ indep})
\]

\[
= 1 \cdot 1 \cdot \frac{v}{(w+1)^2}, \text{ for } 0 < x, y < 1 \text{ (since } X, Y \sim \text{Unif}(0,1))
\]

\[
= \frac{v}{(w+1)^2}, \text{ for } 0 < x = \frac{vw}{w+1} < 1 \text{ and } 0 < y = \frac{v}{w+1} < 1.
\]

\[
= \frac{v}{(w+1)^2}, \text{ for } 0 < v < 1 + \min\{\frac{1}{w}, w\} \text{ and } w > 0 \text{ (after algebra)}.
\]
This implies that the joint pdf of $V$ and $W$ is

$$g(v, w) = f \left( k_1(v, w), k_2(v, w) \right) |J(v, w)|$$

$$= f \left( \frac{vw}{w+1}, \frac{v}{w+1} \right) \cdot \frac{v}{(w+1)^2}$$

$$= f_X \left( \frac{vw}{w+1} \right) f_Y \left( \frac{v}{w+1} \right) \frac{v}{(w+1)^2} \quad (X \text{ and } Y \text{ indep})$$

$$= 1 \cdot 1 \cdot \frac{v}{(w+1)^2}, \text{ for } 0 < x, y < 1 \text{ (since } X, Y \sim \text{Unif}(0,1))$$

$$= \frac{v}{(w+1)^2}, \text{ for } 0 < x = \frac{vw}{w+1} < 1 \text{ and } 0 < y = \frac{v}{w+1} < 1.$$  

$$= \frac{v}{(w+1)^2}, \text{ for } 0 < v < 1 + \min\{\frac{1}{w}, w\} \text{ and } w > 0 \text{ (after algebra).}$$

Note that you have to be careful about the limits of $v$ and $w$, but this thing really does double integrate to 1! □
We can also get the marginal pdf’s. First of all, for the ratio of the uniforms, we get
We can also get the marginal pdf’s. First of all, for the ratio of the uniforms, we get

\[ g_W(w) = \int_{\mathbb{R}} g(v, w) \, dv \]
We can also get the marginal pdf’s. First of all, for the ratio of the uniforms, we get

\[ g_W(w) = \int \frac{v}{(w + 1)^2} \, dv \]

\[ = \int_0^{1+\min\{1/w,w\}} \frac{v}{(w + 1)^2} \, dv \]
We can also get the marginal pdf’s. First of all, for the ratio of the uniforms, we get

\[
g_W(w) = \int_{\mathbb{R}} g(v, w) \, dv
\]

\[
= \int_{0}^{1 + \min\{1/w, w\}} \frac{v}{(w + 1)^2} \, dv
\]

\[
= \frac{(1 + \min\{1/w, w\})^2}{2(w + 1)^2}
\]
We can also get the marginal pdf’s. First of all, for the ratio of the uniforms, we get

\[ g_W(w) = \int_{\mathbb{R}} g(v, w) \, dv \]

\[ = \int_{0}^{1 + \min\{1/w, w\}} \frac{v}{(w + 1)^2} \, dv \]

\[ = \frac{\left(1 + \min\{1/w, w\}\right)^2}{2(w + 1)^2} \]

\[ = \begin{cases} 
\frac{1}{2}, & \text{if } w \leq 1 \\
\frac{1}{2w^2}, & \text{if } w > 1,
\end{cases} \]

which is a little weird-looking and unexpected to me (it’s flat for \(w \leq 1\), and then decreases to 0 pretty quickly for \(w > 1\)).
We can also get the marginal pdf’s. First of all, for the ratio of the uniforms, we get

\[ g_W(w) = \int_{-\infty}^{\infty} g(v,w) \, dv \]

\[ = \int_0 ^{1 + \min\{1/w, w\}} \frac{v}{(w + 1)^2} \, dv \]

\[ = \frac{(1 + \min\{1/w, w\})^2}{2(w + 1)^2} \]

\[ = \begin{cases} 
\frac{1}{2}, & \text{if } w \leq 1 \\
\frac{1}{2w^2}, & \text{if } w > 1, 
\end{cases} \]

which is a little weird-looking and unexpected to me (it’s flat for \( w \leq 1 \), and then decreases to 0 pretty quickly for \( w > 1 \)). \( \square \)
For the pdf of the sum of the uniforms, we have to calculate

\[ g_V(v) = \int_{\mathbb{R}} g(v, w) \, dw. \]

But first we need to deal with some inequality constraints so that we can integrate over the proper region, namely,

\[ 0 \leq v \leq 1 + \min\{1/w, w\}, \quad 0 \leq v \leq 2, \quad w \geq 0. \]

With a little thought, we see that if \( 0 \leq v \leq 1 \), then there is no constraint on \( w \) except for it being positive. On the other hand, if \( 1 < v \leq 2 \), then you can show (it takes a little work) that

\[ v - 1 \leq w \leq \frac{v - 1}{v - 1}. \]

Thus, we have

\[ g_V(v) = \begin{cases} \int_{\mathbb{R}} g(v, w) \, dw, & \text{if } 0 \leq v \leq 1 \\ \int_{1/(v - 1)}^{v} g(v, w) \, dw, & \text{if } 1 < v \leq 2 \end{cases} \]

which is a \( \text{Triangle}(0, 1, 2) \) pdf. Can you see why? Is there an intuitive explanation for this pdf?
For the pdf of the sum of the uniforms, we have to calculate 
\[ g_V(v) = \int_{\mathbb{R}} g(v, w) \, dw. \] But first we need to deal with some inequality constraints so that we can integrate over the proper region, namely,

\[
0 \leq v \leq 1 + \min\{1/w, w\}, \quad 0 \leq v \leq 2, \quad \text{and} \quad w \geq 0.
\]
For the pdf of the sum of the uniforms, we have to calculate
\[ g_V(v) = \int_{\mathbb{R}} g(v, w) \, dw. \]
But first we need to deal with some inequality constraints so that we can integrate over the proper region, namely,

\[ 0 \leq v \leq 1 + \min\{1/w, w\}, \quad 0 \leq v \leq 2, \quad \text{and} \quad w \geq 0. \]

With a little thought, we see that if \(0 \leq v \leq 1\), then there is no constraint on \(w\) except for it being positive. On the other hand, if \(1 < v \leq 2\), then you can show (it takes a little work) that \(v - 1 \leq w \leq \frac{1}{v-1}\). Thus, we have
Honors Bivariate Functions of Random Variables

For the pdf of the sum of the uniforms, we have to calculate

\[ g_V(v) = \int_{\mathbb{R}} g(v, w) \, dw. \]

But first we need to deal with some inequality constraints so that we can integrate over the proper region, namely,

\[ 0 \leq v \leq 1 + \min\{1/w, w\}, \quad 0 \leq v \leq 2, \quad \text{and} \quad w \geq 0. \]

With a little thought, we see that if \( 0 \leq v \leq 1 \), then there is no constraint on \( w \) except for it being positive. On the other hand, if \( 1 < v \leq 2 \), then you can show (it takes a little work) that \( v - 1 \leq w \leq \frac{1}{v-1} \). Thus, we have

\[ g_V(v) = \begin{cases} \int_0^\infty g(v, w) \, dw, & \text{if } 0 \leq v \leq 1 \\ \int_{v-1}^{1/(v-1)} g(v, w) \, dw, & \text{if } 1 < v \leq 2 \end{cases} \]
For the pdf of the sum of the uniforms, we have to calculate
\[ g_V(v) = \int_{\mathbb{R}} g(v, w) \, dw. \]
But first we need to deal with some inequality constraints so that we can integrate over the proper region, namely,

\[ 0 \leq v \leq 1 + \min\{1/w, w\}, \quad 0 \leq v \leq 2, \quad \text{and} \quad w \geq 0. \]

With a little thought, we see that if \( 0 \leq v \leq 1 \), then there is no constraint on \( w \) except for it being positive. On the other hand, if \( 1 < v \leq 2 \), then you can show (it takes a little work) that \( v - 1 \leq w \leq \frac{1}{v-1} \). Thus, we have

\[
g_V(v) = \begin{cases} 
\int_0^\infty g(v, w) \, dw, & \text{if } 0 \leq v \leq 1 \\
\int_{v-1}^{1/(v-1)} g(v, w) \, dw, & \text{if } 1 < v \leq 2 
\end{cases}
\]

\[
= \begin{cases} 
v, & \text{if } 0 \leq v \leq 1 \\
2 - v, & \text{if } 1 < v \leq 2 
\end{cases}
\]
(after algebra).

This is a Triangle(0,1,2) pdf. Can you see why? Is there an intuitive explanation for this pdf?
For the pdf of the sum of the uniforms, we have to calculate
\[ g_V(v) = \int_{\mathbb{R}} g(v, w) \, dw. \]
But first we need to deal with some inequality constraints so that we can integrate over the proper region, namely,
\[
0 \leq v \leq 1 + \min\{1/w, w\}, \quad 0 \leq v \leq 2, \quad \text{and} \quad w \geq 0.
\]

With a little thought, we see that if \( 0 \leq v \leq 1 \), then there is no constraint on \( w \) except for it being positive. On the other hand, if \( 1 < v \leq 2 \), then you can show (it takes a little work) that \( v - 1 \leq w \leq \frac{1}{v-1} \). Thus, we have

\[
g_V(v) = \begin{cases} 
\int_{0}^{\infty} g(v, w) \, dw, & \text{if } 0 \leq v \leq 1 \\
\int_{v-1}^{1/(v-1)} g(v, w) \, dw, & \text{if } 1 < v \leq 2 
\end{cases}
\]

\[
= \begin{cases} 
v, & \text{if } 0 \leq v \leq 1 \\
2 - v, & \text{if } 1 < v \leq 2 
\end{cases}
\]

This is a \textbf{Triangle}(0,1,2) pdf. Can you see why? Is there an intuitive explanation for this pdf? \)[ }
And Now a Word From Our Sponsor…
And Now a Word From Our Sponsor…

We are finally done with the most-difficult module of the course. Congratulations and Felicitations!!!
And Now a Word From Our Sponsor…

We are finally done with the most-difficult module of the course. Congratulations and Felicitations!!!

Things will get easier from now on!
And Now a Word From Our Sponsor…

We are finally done with the most-difficult module of the course. Congratulations and Felicitations!!!

Things will get easier from now on! Happy days are here again!
And Now a Word From Our Sponsor…

We are finally done with the most-difficult module of the course. Congratulations and Felicitations!!!

Things will get easier from now on! Happy days are here again! 😊😊😊😊