Lecture 19:

- Covariance and Correlation
- Auto-Correlation [if time]
Joint Random Variables: Mean and Variance?

The notion of a mean (a single number representing the ”center” of the distribution) can be extended to a

Midpoint = Mean Vector = means of the marginal distributions

This defines the “centroid” or “center of gravity” of the distribution:

Example:

\[
X = [0, 1, 2, 3] \\
Y = [0, 2, 4, 3]
\]

\[
XY = [ (0,0), \\
(1,2), \\
(2,4), \\
(3,3) ]
\]

Midpoint = \( (\mu_X, \mu_Y) \)

= \( \left( \frac{3}{2}, \frac{9}{4} \right) \)
Joint Random Variables: Mean and Variance?

The notion of a mean (a single number representing the ”center” of the distribution) can be extended to a

Midpoint  = Mean Vector = means of the marginal distributions

This defines the “centroid” or “center of gravity” of the distribution:
Joint Random Variables: Mean and Variance?

The notion of a mean (a single number representing the "center" of the distribution) can be extended to a

\[
\text{Midpoint} = \text{Mean Vector} = \text{means of the marginal distributions}
\]

This defines the “centroid” or “center of gravity” of the distribution:

\[
X = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}
Y = \{0, 1, 4, 9, 16, 25, 36, 49, 64, 81, 100\}
\]
Joint Random Variables: Mean and Variance?

The notion of a mean (a single number representing the "center" of the distribution) can be extended to a midpoint.

Midpoint = Mean Vector = means of the marginal distributions

This defines the "centroid" or "center of gravity" of the distribution:
Joint Random Variables: Mean and Variance?

The notion of a mean (a single number representing the "center" of the distribution) can be extended to a

Midpoint = Mean Vector = means of the marginal distributions

This defines the “centroid” or “center of gravity” of the distribution:
Joint Random Variables: Mean and Variance?

The notion of a mean (a single number representing the “center” of the distribution) for a can be extended to a

Midpoint = Mean Vector = means of the marginal distributions

This defines the “centroid” or “center of gravity” of the distribution:
Joint Random Variables: Covariance

Recall: The Variance of $X$ is defined by:

$$Var(X) = E[(X - \mu_X)^2] = E[(X - \mu_X) * (X - \mu_X)]$$

$$= E(X^2) - \mu_X^2 = E(X * X) - \mu_X * \mu_X$$

The Covariance of two JRVs $X$ and $Y$ is defined as follows:

$$Cov(X, Y) = E[(X - \mu_X) * (Y - \mu_Y)]$$

$$= E(X * Y) - \mu_X * \mu_Y$$

The Covariance of two JRVs $X$ and $Y$ has the same defects as the variance of a single RV:

- The units are the product of the units of $X$ and $Y$: if $X = \text{height}$ and $Y = \text{weight}$, then the units might be foot-pounds!

- The scale is hard to work with: What does a covariance of 123.445 foot-pounds mean?
Therefore we standardize the covariance so it is unit-less and in the interval \([-1 .. 1]\).

The Correlation Coefficient of \(X\) and \(Y\) is defined as:

\[
\rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sigma_X \cdot \sigma_Y} = \frac{E[(X - \mu_X) \cdot (Y - \mu_Y)]}{\sigma_X \cdot \sigma_Y} = E\left[\frac{X - \mu_X}{\sigma_X} \cdot \frac{Y - \mu_Y}{\sigma_X}\right]
\]

\[
= E[Z_X \cdot Z_Y]
\]

where \(Z_X\) and \(Z_Y\) are the standardized forms of \(X\) and \(Y\).

To compute, it is best to use:

\[
\rho_{X,Y} = \frac{E(X \cdot Y) - \mu_X \cdot \mu_Y}{\sigma_X \cdot \sigma_Y}
\]
Joint Random Variables: Correlation Coefficient

Motivation for the Correlation Coefficient

$$\rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sigma_X \cdot \sigma_Y} = E[Z_X \cdot Z_Y]$$

Example: Two Sine Waves in Phase (Perfectly Correlated)

For sine waves $Z_X$ and $Z_Y$:

$$E[Z_X \cdot Z_Y] > 0$$

$$\cos(x)^2 = \frac{\cos(2x)}{2} + \frac{1}{2}$$

$1 \cdot 1 = 1$

$-1 \cdot -1 = 1$
Joint Random Variables: Correlation Coefficient

Motivation for the Correlation Coefficient

\[ \rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sigma_X \times \sigma_Y} = E[Z_X \times Z_Y] \]

Example: Two Sine Waves 180° out of Phase (Perfectly Anti-Correlated)

\[
E[Z_X \times Z_Y] < 0
\]
Joint Random Variables: Correlation Coefficient

Motivation for the Correlation Coefficient

\[ \rho_{X,Y} = \frac{\text{Cov}(X,Y)}{\sigma_X \cdot \sigma_Y} = E[Z_X \cdot Z_Y] \]

Example: Two Random Sine Waves (No Correlation)

\[ E[Z_X \cdot Z_Y] = 0 \]
The range of the Correlation Coefficient of X and Y is from -1 to 1:

\[
\rho_{X,Y} = \frac{E(X \ast Y) - \mu_X \ast \mu_Y}{\sigma_X \ast \sigma_Y}
\]

-1.0  0.0  1.0

Inversely correlated  Uncorrelated  Correlated
Joint Random Variables

Example 1: Toss 2 coins;  \(X = \# \text{ heads on first}, \ Y = \# \text{ heads on second}\)

<table>
<thead>
<tr>
<th>X ( \times ) Y</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>1</td>
<td>0.25</td>
<td>0.25</td>
</tr>
</tbody>
</table>

\[
\begin{align*}
E(X * Y) &= 0 * 0.25 + 1 * 0.25 + 0 * 0.25 + 0 * 0.25 = 0.25 \\
\rho_{X,Y} &= \frac{Cov(X,Y)}{\sigma_X * \sigma_Y} = \frac{E(X * Y) - \mu_X * \mu_Y}{\sigma_X * \sigma_Y} = \frac{0.25 - 0.5 * 0.5}{0.5 * 0.5} = 0.0
\end{align*}
\]
Joint Random Variables

Calculating the Covariance and Correlation Coefficient is best done with either a spreadsheet or Python:

**Example: Toss 1 coin; \( X = \# \text{ heads}, Y = \# \text{ heads} \)**

\[
\begin{array}{cccccccc}
X & 0 & 1 \\
\hline
0 & 0.125 & 0.5 \\
1 & 0.5 & 0.5 \\
\end{array}
\]

\[
\begin{array}{cccccccc}
Y & 0 & 1 \\
\hline
0 & 0.125 & 0.25 \\
1 & 0.5 & 0.5 \\
\end{array}
\]

\[
\sigma_X = 0.5, \quad \mu_X = 0.5, \quad \sigma_Y = 0.5, \quad \mu_Y = 0.5
\]

\[
\sigma_{XY} = 0.25
\]

\[
\rho_{X,Y} = \frac{\text{Cov}(X,Y)}{\sigma_X \cdot \sigma_Y} = \frac{E(X \cdot Y) - \mu_X \cdot \mu_Y}{\sigma_X \cdot \sigma_Y} = \frac{0.5 - 0.5 \cdot 0.5}{0.5 \cdot 0.5} = \frac{0.25}{0.25} = 1.0
\]

\[
E(X \cdot Y) = 0 \cdot 0.0 + 1 \cdot 0.5 + 0 \cdot 0.5 + 0 \cdot 0.0 = 0.5
\]
Joint Random Variables

Calculating the Covariance and Correlation Coefficient is best done with either a spreadsheet or Python:

**Example: Toss 2 coins; X = # heads on first coin, Y = total # of heads**

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>p(x)</th>
<th>p(y)</th>
<th>p(x,y) * p(x)</th>
<th>cov(X,Y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0.125</td>
<td>0.250</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0.125</td>
<td>0.500</td>
<td>0.250</td>
<td>0.250</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>0.125</td>
<td>0.250</td>
<td>0.500</td>
<td>0.750</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0.500</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.500</td>
<td>0.250</td>
<td>0.250</td>
<td>0.250</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>0.500</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0.250</td>
<td>0.250</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0.250</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>0.250</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

\[
\rho_{X,Y} = \frac{Cov(X, Y)}{\sigma_X \cdot \sigma_Y} = \frac{E(X \cdot Y) - \mu_X \cdot \mu_Y}{\sigma_X \cdot \sigma_Y} = \frac{0.75 - 0.5 \times 1.0}{0.5 \times 0.707} = \frac{0.25}{0.354} = 0.707
\]
Joint Random Variables

Calculating the Covariance and Correlation Coefficient is best done with either a spreadsheet or Python:

Example: Toss 2 coins; $X =$ # heads on first coin, $Y =$ total # of heads

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td>H</td>
<td>=&gt; (1,2)</td>
</tr>
<tr>
<td>H</td>
<td>T</td>
<td>=&gt; (1,1)</td>
</tr>
<tr>
<td>T</td>
<td>H</td>
<td>=&gt; (0,1)</td>
</tr>
<tr>
<td>T</td>
<td>T</td>
<td>=&gt; (0,0)</td>
</tr>
</tbody>
</table>

$X = [1, 1, 0, 0]$  
$Y = [2, 1, 1, 0]$