1 Covariance

1.1 Motivation

- While the joint PDF of two random variables fully describes the relationship between two random variables, we would like to have a measurements to summarizes how closely related two random variables are
- We have already touched on this with the concept of Independence
- When two random varibles are independent, that means that there is no relationship between two random variables
- Covariance is how we will numerically summarize how closely related two random variables are

1.2 Definitions

- First we define the covariance of two random variables

  Definition 1. *Let X_1 and X_2 be random variables with finite first moments (i.e. $E[X_1]$ and $E[X_2]$ are real valued). Then the **Covariance** between X_1 and X_2 is*

  $$Cov(X_1, X_2) = E[(X_1 - E[X_1])(X_2 - E[X_2])]$$

- And we will also define a related concept, the correlation of two random variables

  Definition 2. *Let X_1 and X_2 be random variables with finite second moments (i.e. $E[X_1^2]$ and $E[X_2^2]$ are real valued). Then the **Correlation** between X_1 and X_2 is*

  $$Corr(X_1, X_2) = \frac{Cov(X_1, X_2)}{\sqrt{V[X_1]V[X_2]}}$$

1.3 Theorems

- We have a few theorems to help make finding the Covariance easier
- Our first theorem is akin to our Variance Formula

  Theorem 1. *Let X_1 and X_2 be random variables with finite first moments (i.e. $E[X_1]$ and $E[X_2]$ are real valued). Then*

  $$Cov(X_1, X_2) = E[X_1 X_2] - E[X_1]E[X_2]$$

Proof.

$$
\begin{aligned}
Cov(X_1, X_2) &= E[(X_1 - E[X_1])(X_2 - E[X_2])] \\
&= E[X_1X_2 - E[X_1]X_2 - E[X_2]X_1 + E[X_1]E[X_2]] \\
&= E[X_1X_2] - E[E[X_1]X_2] - E[E[X_2]X_1] + E[E[X_1]E[X_2]] \\
&= E[X_1X_2] - E[X_1]E[X_2] - E[X_2]E[X_1] + E[X_1]E[X_2] \\
&= E[X_1X_2] - E[X_1]E[X_2]
\end{aligned}
$$

□

- Our Second theorem Relates to the relationship bewteen independence and covariance

  Theorem 2. *Let X_1 and X_2 be random variables with finite first moments (i.e. $E[X_1]$ and $E[X_2]$ are real valued). If X_1 and X_2 are independent, then*

  $$Cov(X_1, X_2) = 0$$

  Proof.

  $$
  \begin{aligned}
  Cov(X_1, X_2) &= E[X_1X_2] - E[X_1]E[X_2] \leftarrow \text{By previous theorem} \\
  &= E[X_1]E[X_2] - E[X_1]E[X_2] \leftarrow \text{Because } X_1 \text{ and } X_2 \text{ are independent} \\
  &= 0
  \end{aligned}
  $$

  □

2 Expectation of Linear Combinations

2.1 Motivation

- Next we will cover the Exectation, Variance and Covariance of Random variables that are actually the linear combination of other random variables
- This is useful because we often combine measurements into a summary measurement (for example, the sample mean)

2.2 Theorems

- All of these theorems will be based on

  $$U_x = \sum_{i=1}^{n} a_i X_i \text{ and } U_y = \sum_{j=1}^{m} a_j Y_j$$

  where $X_1, ..., X_n$ and $Y_1, ..., Y_m$ are random variables and $a_1, ..., a_n, b_1, ..., b_m$ are all real valued constants

- Expectation

Theorem 3.

$$E[U_x] = \sum_{i=1}^{n} a_i E[X_i]$$

Proof.

$$\begin{aligned} E[U_x] &= E[\sum_{i=1}^{n} a_i X_i] \\ &= \sum_{i=1}^{n} E[a_i X_i] \\ &= \sum_{i=1}^{n} a_i E[X_i] \end{aligned}$$

□

- Variance

Theorem 4.

$$V[U_x] = \sum_{i=1}^{n} a_i^2 V[X_i] + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} a_i a_j Cov[X_i, X_j]$$

Or, Equivalently

$$V[U_x] = \sum_{i=1}^{n} a_i^2 V[X_i] + 2\sum\sum_{1\leq i<j\leq n} a_i a_j Cov[X_i, X_j]$$

Or, Equivalently

$$V[U_x] = \sum_{i=1}^{n} a_i^2 V[X_i] + \sum_{\substack{i=1\\ i\neq j}}^{n}\sum_{j=1}^{n} a_i a_j Cov[X_i, X_j]$$

Proof.

$$\begin{aligned} V[U_x] &= E[(U_x - E[U_x])^2] \\ &= E[(\sum_{i=1}^{n} a_i X_i - E[\sum_{i=1}^{n} a_i X_i])^2] \\ &= E[(\sum_{i=1}^{n} a_i X_i - \sum_{i=1}^{n} a_i E[X_i])^2] \end{aligned}$$

$$
\begin{aligned}
&= E[(\sum_{i=1}^{n} a_i[X_i - E[X_i]])^2] \\
&= E[\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j [X_i - E[X_i]][X_j - E[X_j]]] \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j E[[X_i - E[X_i]][X_j - E[X_j]]] \\
&= \sum_{i=1}^{n} a_i^2 E[[X_i - E[X_i]]^2] + \sum_{\substack{i=1 \\ i \neq j}}^{n}\sum_{j=1}^{n} a_i a_j E[[X_i - E[X_i]][X_j - E[X_j]]] \\
&= \sum_{i=1}^{n} a_i^2 V[X_i] + \sum_{\substack{i=1 \\ i \neq j}}^{n}\sum_{j=1}^{n} a_i a_j Cov[X_1, X_i] \\
&= \sum_{i=1}^{n} a_i^2 V[X_i] + 2\sum\sum_{1 \leq i < j \leq n} a_i a_j Cov[X_i, X_j] \\
&= \sum_{i=1}^{n} a_i^2 V[X_i] + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} a_i a_j Cov[X_i, X_j]
\end{aligned}
$$

□

- Covariance

  Theorem 5.

$$
Cov[U_x, U_y] = \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j Cov[x_i, Y_j]
$$

Proof.

$$
\begin{aligned}
Cov[U_x, U_y] &= E[(U_x - E[U_x])(U_y - E[U_y])] \\
&= E[(\sum_{i=1}^{n} a_i X_i - E[\sum_{i=1}^{n} a_i X_i])(\sum_{j=1}^{m} b_j Y_j - E[\sum_{j=1}^{m} b_j Y_j])] \\
&= E[(\sum_{i=1}^{n} a_i (X_i - E[X_i]))(\sum_{j=1}^{m} b_j (Y_j - E[Y_j]))] \\
&= E[(\sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j (X_i - E[X_i])(Y_j - E[Y_j])] \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} E[a_i b_j (X_i - E[X_i])(Y_j - E[Y_j])]
\end{aligned}
$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j E[(X_i - E[X_i])(Y_j - E[Y_j])]$$
$$= \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j Cov[X_i, Y_i]$$

□

3 Conditional Expectation

3.1 Motivation

- Here we formally introduce the concept of conditional Expectation, and some related theorems
- The concept is really just recognizing that the conditional distribution is itself a distribution

3.2 Definitions

- The definition of the Conditional expectation for discrete random variables is as follows:

  Definition 3. *Let X_1 and X_2 be discrete random variables where the conditional distribution of $X_1|X_2 = x_2$ is $p_{1|2}(x_1) = p(x_1|x_2)$, and let g be a real valued function defined for all possible values of $X_1|X_2 = x_2$. Then the conditional expectation of $g(X_1)|X_2 = x_2$ is*

  $$E[g(X_1)|X_2 = x_2] = E_{1|2}[g(X_1)|X_2 = x_2] = \sum_{x_1 \in S_1} g(x_1) p_{1|2}(x_1)$$

- And the definition of the Conditional expectation for continuous random variables is as follows:

  Definition 4. *Let X_1 and X_2 be continuous random variables where the conditional distribution of $X_1|X_2 = x_2$ is $f_{1|2}(x_1) = f(x_1|x_2)$, and let g be a real valued function defined for all possible values of $X_1|X_2 = x_2$. Then the conditional expectation of $g(X_1)|X_2 = x_2$ is*

  $$E[g(X_1)|X_2 = x_2] = E_{1|2}[g(X_1)|X_2 = x_2] = \int_{-\infty}^{\infty} g(x_1) f_{1|2}(x_1)\mathrm{d}x_1$$

3.3 Theorems

- Here we introduce some theorems related to conditional expectations

Theorem 6. *Let X_1 and X_2 be random variales and let g be a real valued function defined for all possible values of X_1, then*

$$E[g(X_1)] = E[E[g(X_1)|X_2 = x_2]]$$

Proof.
Here we present the proof when X_1 and X_2 are continuous. The proof is similar when X_1 and X_2 are discrete. Let f be the joint pdf of X_1 an X_2, $f_{1|2}$ be the conditional pdf of $X_1|X_2 = x_2$, and f_2 be the marginal pdf of X_2. Then we have

$$\begin{aligned}
E[g(X_1)] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x_1)f(x_1,x_2)\mathrm{d}x_1\mathrm{d}x_2 \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x_1)f_{1|2}(x_1|X_2 = x_2)f_2(x_2)\mathrm{d}x_1\mathrm{d}x_2 \\
&= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} g(x_1)f_{1|2}(x_1|X_2 = x_2)\mathrm{d}x_1\right] f_2(x_2)\mathrm{d}x_2 \\
&= \int_{-\infty}^{\infty} E[g(X_1)|X_2 = x_2]f_2(x_2)\mathrm{d}x_2 \\
&= E[E[g(X_1)|X_2 = x_2]]
\end{aligned}$$

□

Theorem 7. *Let X_1 and X_2 be random variales and let g be a real valued function defined for all possible values of X_1, then*

$$V[g(X_1)] = E[V[g(X_1)|X_2 = x_2]] + V[E[g(X_1)|X_2 = x_2]]$$

where

$$\begin{aligned}
V[g(X_1)|X_2 = x_2] &= E[(g(X_1) - E[g(X_1)|X_2 = x_2])^2|X_2 = x_2] \\
&= E[g(X_1)^2|X_2 = x_2] - (E[g(X_1)|X_2 = x_2])^2
\end{aligned}$$

Proof.

$$\begin{aligned}
V[g(X_1)] &= E[g(X_1)^2] - E^2[g(X_1)] \\
&= E\left[E[g(X_1)^2|X_2 = x_2]\right] - (E\left[E[g(X_1)|X_2 = x_2]\right])^2 \\
&= E\left[E[g(X_1)^2|X_2 = x_2]\right] + \left(-E\left[(E[g(X_1)|X_2 = x_2])^2\right] + E\left[(E[g(X_1)|X_2 = x_2])^2\right]\right) \\
&\quad -(E\left[E[g(X_1)|X_2 = x_2]\right])^2
\end{aligned}$$

$$
\begin{aligned}
&= \left(E\left[E[g(X_1)^2|X_2=x_2]\right] - E\left[(E[g(X_1)|X_2=x_2])^2\right]\right) \\
&\quad + \left(E\left[(E[g(X_1)|X_2=x_2])^2\right] - (E\left[E[g(X_1)|X_2=x_2]\right])^2\right) \\
&= \left(E\left[E[g(X_1)^2|X_2=x_2] - (E[g(X_1)|X_2=x_2])^2\right]\right) \\
&\quad + \left(E\left[(E[g(X_1)|X_2=x_2])^2\right] - (E\left[E[g(X_1)|X_2=x_2]\right])^2\right) \\
&= (E\left[V[g(X_1)|X_2=x_2]\right]) \\
&\quad + \left(E\left[(E[g(X_1)|X_2=x_2])^2\right] - (E\left[E[g(X_1)|X_2=x_2]\right])^2\right) \\
&= (E\left[V[g(X_1)|X_2=x_2]\right]) + (V\left[E[g(X_1)|X_2=x_2]\right])
\end{aligned}
$$

$\square$