S&DS 241 Lecture 21
Covariance and correlation
B-H: 7.3, 10.1
Given a random variable $X$

Recall mean and variance

- $EX$: mean value
- $\text{Var}(X) = E((X - EX)^2)$: mean-square deviation around the mean
Given two random variables $X$ and $Y$

Suppose

- $X$ has mean $EX$ and $\text{Var}(X) = \sigma_X^2$
- $Y$ has mean $EY$ and $\text{Var}(Y) = \sigma_Y^2$
Given two random variables $X$ and $Y$

Suppose

- $X$ has mean $EX$ and $\text{Var}(X) = \sigma_X^2$
- $Y$ has mean $EY$ and $\text{Var}(Y) = \sigma_Y^2$

Two important quantities:

- Covariance:

  $$\text{Cov}(X, Y) = E((X - EX)(Y - EY))$$

  summary statistic for the “tendency” of $X$ and $Y$ to move in the same direction
Given two random variables $X$ and $Y$

Suppose

- $X$ has mean $E_X$ and $\text{Var}(X) = \sigma^2_X$
- $Y$ has mean $E_Y$ and $\text{Var}(Y) = \sigma^2_Y$

Two important quantities:

- **Covariance:**
  \[
  \text{Cov}(X, Y) = E((X - E_X)(Y - E_Y))
  \]
  summary statistic for the “tendency” of $X$ and $Y$ to move in the same direction

- **Correlation coefficient:**
  \[
  \rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}
  \]
  (Denoted by $\text{Corr}(X, Y)$ in B-H)
Remarks

Covariance:

- \( \text{Cov}(X, Y) = E(XY) - (EX)(EY) \).

Correlation coefficient:

- \(-1 \leq \rho(X, Y) \leq 1\): to be justified later. So \(\rho\) is a normalized version of \(\text{Cov}\).

- Positively correlated: \(\rho > 0\)
  - e.g. \(X = \text{midterm grade}, Y = \text{total grade}\)
  - e.g. \(X = \text{gas price}, Y = \text{stock price of TSLA}\)

- Negatively correlated: \(\rho < 0\)
  - e.g. \(X = \text{yield of crop}, Y = \text{market price}\)
  - e.g. roll a die for 100 times, \(X = \text{number of 1's}, Y = \text{number of 6's}\)
Remarks

Covariance:

- $\text{Cov}(X, Y) = E(XY) - (EX)(EY)$. 
- $\text{Cov}(X, X) = \text{Var}(X)$

Correlation coefficient:

- $-1 \leq \rho(X, Y) \leq 1$: to be justified later. So $\rho$ is a normalized version of $\text{Cov}$.

- Positively correlated: $\rho > 0$ 
  - e.g. $X =$ midterm grade, $Y =$ total grade
- Negatively correlated: $\rho < 0$ 
  - e.g. $X =$ yield of crop, $Y =$ market price
  - e.g. roll a die for 100 times, $X =$ number of 1's, $Y =$ number of 6's
Remarks

Covariance:

- \( \text{Cov}(X, Y) = E(XY) - (EX)(EY) \).
- \( \text{Cov}(X, X) = \text{Var}(X) \)
- \( \text{Cov}(X, Y) = \text{Cov}(Y, X) \)

Correlation coefficient:

- \(-1 \leq \rho(X, Y) \leq 1\): to be justified later. So \(\rho\) is a normalized version of \(\text{Cov}\).

- Positively correlated: \(\rho > 0\)
  - e.g. \(X = \) midterm grade, \(Y = \) total grade
  - e.g. \(X = \) gas price, \(Y = \) stock price of TSLA

- Negatively correlated: \(\rho < 0\)
  - e.g. \(X = \) yield of crop, \(Y = \) market price
  - e.g. roll a die for 100 times, \(X = \) number of 1's, \(Y = \) number of 6's
Remarks

Covariance:

- $\text{Cov}(X, Y) = E(XY) - (E(X))(E(Y))$.
- $\text{Cov}(X, X) = \text{Var}(X)$
- $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
Remarks

Covariance:
- $\text{Cov}(X, Y) = E(XY) - (EX)(EY)$.
- $\text{Cov}(X, X) = \text{Var}(X)$
- $\text{Cov}(X, Y) = \text{Cov}(Y, X)$

Correlation coefficient:
- $-1 \leq \rho(X, Y) \leq 1$: to be justified later. So $\rho$ is a normalized version of $\text{Cov}$.
Remarks

Covariance:

- \( \text{Cov}(X, Y) = E(XY) - (EX)(EY) \).
- \( \text{Cov}(X, X) = \text{Var}(X) \)
- \( \text{Cov}(X, Y) = \text{Cov}(Y, X) \)

Correlation coefficient:

- \(-1 \leq \rho(X, Y) \leq 1\): to be justified later. So \( \rho \) is a normalized version of \( \text{Cov} \)
- Positively correlated: \( \rho > 0 \)
Remarks

Covariance:

- \( \text{Cov}(X, Y) = E(XY) - (EX)(EY) \).
- \( \text{Cov}(X, X) = \text{Var}(X) \)
- \( \text{Cov}(X, Y) = \text{Cov}(Y, X) \)

Correlation coefficient:

- \( -1 \leq \rho(X, Y) \leq 1 \): to be justified later. So \( \rho \) is a normalized version of \( \text{Cov} \)
- Positively correlated: \( \rho > 0 \)
  - e.g. \( X=\)midterm grade, \( Y=\)total grade
- Negatively correlated: \( \rho < 0 \)
  - e.g. \( X=\)yield of crop, \( Y=\)market price
  - e.g. roll a die for 100 times, \( X=\)number of 1’s, \( Y=\)number of 6’s
Remarks

Covariance:

- \( \text{Cov}(X, Y) = E(XY) - (EX)(EY). \)
- \( \text{Cov}(X, X) = \text{Var}(X) \)
- \( \text{Cov}(X, Y) = \text{Cov}(Y, X) \)

Correlation coefficient:

- \(-1 \leq \rho(X, Y) \leq 1\): to be justified later. So \( \rho \) is a normalized version of \( \text{Cov} \)
- Positively correlated: \( \rho > 0 \)
  - e.g. \( X=\)midterm grade, \( Y=\)total grade
  - e.g. \( X=\)gas price, \( Y=\)stock price of TSLA
Remarks

Covariance:

- \( \text{Cov}(X, Y) = E(XY) - (EX)(EY) \).
- \( \text{Cov}(X, X) = \text{Var}(X) \)
- \( \text{Cov}(X, Y) = \text{Cov}(Y, X) \)

Correlation coefficient:

- \(-1 \leq \rho(X, Y) \leq 1\): to be justified later. So \( \rho \) is a normalized version of \( \text{Cov} \)
- **Positively correlated**: \( \rho > 0 \)
  - e.g. \( X=\text{midterm grade}, Y=\text{total grade} \)
  - e.g. \( X=\text{gas price}, Y=\text{stock price of TSLA} \)
- **Negatively correlated**: \( \rho < 0 \)
Remarks

Covariance:

• $\text{Cov}(X, Y) = E(XY) - (EX)(EY)$.  
• $\text{Cov}(X, X) = \text{Var}(X)$  
• $\text{Cov}(X, Y) = \text{Cov}(Y, X)$

Correlation coefficient:

• $-1 \leq \rho(X, Y) \leq 1$: to be justified later. So $\rho$ is a normalized version of $\text{Cov}$

• **Positively correlated**: $\rho > 0$
  ▶ e.g. $X=\text{midterm grade}, Y=\text{total grade}$
  ▶ e.g. $X=\text{gas price}, Y=\text{stock price of TSLA}$

• **Negatively correlated**: $\rho < 0$
  ▶ e.g. $X=\text{yield of crop}, Y=\text{market price}$
Remarks

Covariance:

- \( \text{Cov}(X, Y) = E(XY) - (EX)(EY) \).
- \( \text{Cov}(X, X) = \text{Var}(X) \)
- \( \text{Cov}(X, Y) = \text{Cov}(Y, X) \)

Correlation coefficient:

- \(-1 \leq \rho(X, Y) \leq 1\): to be justified later. So \( \rho \) is a normalized version of \( \text{Cov} \)
- \underline{Positively correlated}: \( \rho > 0 \)
  - e.g. \( X=\text{midterm grade}, Y=\text{total grade} \)
  - e.g. \( X=\text{gas price}, Y=\text{stock price of TSLA} \)
- \underline{Negatively correlated}: \( \rho < 0 \)
  - e.g. \( X=\text{yield of crop}, Y=\text{market price} \)
  - e.g. roll a die for 100 times, \( X=\text{number of 1's}, Y=\text{number of 6's} \)
Uncorrelated

We say $X$ and $Y$ are **uncorrelated** if

$$\text{Cov}(X, Y) = 0$$

$$\iff \rho(X, Y) = 0$$

$$\iff E(XY) = (EX)(EY)$$
Independent versus uncorrelated

- Independent $\Rightarrow$ uncorrelated: $E(XY) = (EX)(EY)$.
- Uncorrelated $\not\Rightarrow$ independent:

Example (HW3): Choose a point uniformly at random to be one of the four vertices of the diamond below. Let $(X, Y)$ denote its coordinate.

Then $E(XY) = 0 = EX = EY$, but $X$ and $Y$ are dependent.
Properties

1. Shift-invariance: \( \text{Cov}(X + b, Y + d) = \text{Cov}(X, Y) \)

   ▶ Next: WLOG assume all random variables have zero mean

2. Bilinearity:
   \[
   \text{Cov}(X + Y, W + Z) = \text{Cov}(X, W) + \text{Cov}(Y, W) + \text{Cov}(X, Z) + \text{Cov}(Y, Z)
   \]
Properties

1. **Shift-invariance:** \( \text{Cov}(X + b, Y + d) = \text{Cov}(X, Y) \)
   - Next: WLOG assume all random variables have zero mean

2. \( \text{Cov}(aX, cY) = ac \text{Cov}(X, Y) \).
Properties

1. **Shift-invariance:** $\text{Cov}(X + b, Y + d) = \text{Cov}(X, Y)$
   
   ▶ Next: WLOG assume all random variables have zero mean

2. $\text{Cov}(aX, cY) = ac \text{Cov}(X, Y)$. More generally

   $$\text{Cov}(aX + b, cY + d) = ac\text{Cov}(X, Y)$$

   and

   $$\rho(aX + b, cY + d) = \rho(X, Y), \quad \text{provided that } a, c > 0$$

Interpretation: the corr coeff between temperature in New York and that in New Haven is unchanged when expressed in either °F or °C
Properties

1. Shift-invariance: \( \text{Cov}(X + b, Y + d) = \text{Cov}(X, Y) \)
   
   Next: WLOG assume all random variables have zero mean

2. \( \text{Cov}(aX, cY) = ac \text{Cov}(X, Y) \). More generally

\[
\text{Cov}(aX + b, cY + d) = ac \text{Cov}(X, Y)
\]

and

\[
\rho(aX + b, cY + d) = \rho(X, Y), \quad \text{provided that } a, c > 0
\]

Interpretation: the corr coeff between temperature in New York and that in New Haven is unchanged when expressed in either °F or °C

3. Bilinearity:

\[
\text{Cov}(X + Y, W + Z) = \text{Cov}(X, W) + \text{Cov}(Y, W) + \text{Cov}(X, Z) + \text{Cov}(Y, Z)
\]
Properties

4 Var(\(X + Y\)) = Var(X) + Var(Y) + 2Cov(X, Y), or equivalently

\[
\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y + 2\rho(X, Y)\sigma_X\sigma_Y
\]

Therefore

\[
\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) \iff X \text{ and } Y \text{ are uncorrelated}
\]
Properties

4. \( \text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y) \), or equivalently

\[
\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2 + 2\rho(X, Y)\sigma_X\sigma_Y
\]

Therefore

\( \text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) \iff X \text{ and } Y \text{ are uncorrelated} \)

5. \[
\text{Cov} \left( \sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j \right) = \sum_{i=1}^{n} \sum_{j=1}^{m} \text{Cov}(X_i, Y_j)
\]
Properties

4. \( \text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y) \), or equivalently

\[
\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y + 2\rho(X, Y)\sigma_X\sigma_Y
\]

Therefore

\[
\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) \iff X \text{ and } Y \text{ are uncorrelated}
\]

5. \[
\text{Cov}\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j\right) = \sum_{i=1}^{n} \sum_{j=1}^{m} \text{Cov}(X_i, Y_j)
\]

6. \[
\text{Var}(X_1 + \cdots + X_n) = \text{Var}(X_1) + \cdots + \text{Var}(X_n) + \sum_{i \neq j} \text{Cov}(X_i, X_j)
\]
Variance of a sum

Corollary: When do variances add up?

• Suppose \(X_1, \ldots, X_n\) are uncorrelated, that is, \(\text{Cov}(X_i, X_j) = 0\) whenever \(i \neq j\). Then

\[
\text{Var}(X_1 + \cdots + X_n) = \text{Var}(X_1) + \cdots + \text{Var}(X_n)
\]

Special case: \(X_1, \ldots, X_n\) are independent
Example

Let $X$ be a random variable (signal) with unit variance. Let $Y$ be a noisy observation of $X$

$$Y = X + Z$$

where the noise $Z$ is independent of $X$ and has variance $\sigma^2$. 

$X$ and $Y$ are highly (positively) correlated when $\sigma$ is small (less noisy). 

$X$ and $Y$ are almost uncorrelated when $\sigma$ is large (very noisy).
Example

Let $X$ be a random variable (signal) with unit variance. Let $Y$ be a noisy observation of $X$

$$Y = X + Z$$

where the noise $Z$ is independent of $X$ and has variance $\sigma^2$.

- **Covariance**

  $$\text{Cov}(X, Y) = \text{Cov}(X, X + Z) = \text{Cov}(X, X) + \text{Cov}(X, Z) = 1$$

  $$\text{Var}(X) + 0 = 1$$

- **Correlation coefficient**

  $$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = \frac{1}{\sqrt{1 + \sigma^2}}$$

- $X$ and $Y$ are highly (positively) correlated when $\sigma$ is small (less noisy).
- $X$ and $Y$ are almost uncorrelated when $\sigma$ is large (very noisy).
Example

Let $X$ be a random variable (signal) with unit variance. Let $Y$ be a noisy observation of $X$

$$Y = X + Z$$

where the noise $Z$ is independent of $X$ and has variance $\sigma^2$.

- **Covariance**

  $$\text{Cov}(X, Y) = \text{Cov}(X, X + Z) = \text{Cov}(X, X) + \text{Cov}(X, Z) = 1$$

  $$\text{Var}(X)$$

  $$0$$

- **Correlation coefficient**

  $$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = \frac{1}{\sqrt{1 + \sigma^2}}$$

- $X$ and $Y$ are highly (positively) correlated when $\sigma$ is small (less noisy)

- $X$ and $Y$ are almost uncorrelated when $\sigma$ is large (very noisy)
Example

Scatter plot of 500 independent samples of $(X, Y)$:

$\sigma = 0.1, \rho = 0.995$
Example

Scatter plot of 500 independent samples of \((X, Y)\):

\[ \sigma = 0.3, \rho = 0.96 \]
Example

Scatter plot of 500 independent samples of \((X, Y)\):

\[
\sigma = 0.5, \, \rho = 0.89
\]
Example

Scatter plot of 500 independent samples of $(X, Y)$:

\[ \sigma = 1, \rho = 0.71 \]
Example

Scatter plot of 500 independent samples of \((X, Y)\):

\[
\sigma = 2, \rho = 0.45
\]
Example

Scatter plot of 500 independent samples of \((X, Y)\):

\[ \sigma = 5, \rho = 0.2 \]
Example

Similarly, if $Y = -X + Z$, then

$$\rho(X, Y) = -\frac{1}{\sqrt{1 + \sigma^2}}$$

Scatter plot of 500 independent samples of $(X, Y)$:

$$\sigma = 0.3, \rho = -0.96$$
Intuition

- If $X$ and $Y$ are highly correlated (i.e., $\rho(X, Y) \approx \pm 1$), then they are approximately related using a straight line. Thus we can predict one by the other by a linear equation. This is the idea of linear regression (in its most basic form)
Caveat

- Correlation only capture “linear” dependence.
  - Example: $X \sim \text{Unif}(-1, 1)$, $Y = X^2$. $\text{Cov}(X, Y) = 0$ (Exercise)
  - Cannot predict $Y$ using linear instrument, but $X$ completely determines $Y$ (non-linearly)
Example: exam score

Assume $X$ (Pset score), $Y$ (midterm score) and $Z$ (final score) are independent and $\text{Unif}(0, 100)$. Total score:

$$S = 0.3X + 0.3Y + 0.4Z$$

Then

- $\text{Var}(X) = \text{Var}(Y) = \text{Var}(Z) = \sigma^2$ and
  $\text{Var}(S) = (0.3^2 + 0.3^2 + 0.4^2)\sigma^2$

- Correlation between Pset and total

$$\rho(X, S) = \frac{\text{Cov}(X, 0.3X + 0.3Y + 0.4Z)}{\sqrt{\text{Var}(X)\text{Var}(S)}} = \frac{0.3}{\sqrt{0.3^2 + 0.3^2 + 0.4^2}}$$

$$\approx 0.51$$

- Similarly, $\rho(Y, S) \approx 0.51$, $\rho(Z, S) \approx 0.69$. 
Example: dice

Roll a fair die for \( n \) times, \( X \) = number of \( \square \), \( Y \) = number of \( \Box \). Find \( \rho(X, Y) \)

- \( X \sim \text{Bin}(n, 1/6) \), \( Y \sim \text{Bin}(n, 1/6) \)
Example: dice

Roll a fair die for $n$ times, $X =$ number of \( \square \), $Y =$ number of \( \blacksquare \). Find $\rho(X, Y)$

- $X \sim \text{Bin}(n, 1/6)$, $Y \sim \text{Bin}(n, 1/6)$
- Are they independent?
Example: dice

Roll a fair die for \( n \) times, \( X = \) number of \( \bullet \), \( Y = \) number of \( \blacksquare \). Find \( \rho(X, Y) \)

- \( X \sim \text{Bin}(n, 1/6) \), \( Y \sim \text{Bin}(n, 1/6) \)
- Are they independent? No.
Example: dice

Roll a fair die for \( n \) times, \( X = \) number of \( \Box \), \( Y = \) number of \( \blacksquare \). Find \( \rho(X, Y) \)

- \( X \sim \text{Bin}(n, 1/6), \ Y \sim \text{Bin}(n, 1/6) \)
- Are they independent? No.
- Decompose binomial as sum of independent Bernoullis:

\[
X = \sum_{i=1}^{n} X_i, \quad X_i = \begin{cases} 1 & \text{ith toss is } 1 \\ 0 & \text{else} \end{cases}
\]

\[
Y = \sum_{i=1}^{n} Y_i, \quad Y_i = \begin{cases} 1 & \text{ith toss is } 6 \\ 0 & \text{else} \end{cases}
\]

- Each \( X_i \) and \( Y_i \) are dependent; for \( i \neq j \), \( X_i \) and \( Y_j \) are independent
Example: dice

\[ \text{Cov}(X, Y) = \text{Cov}(X_1 + \ldots + X_n, Y_1 + \ldots + Y_n) = \sum_{i=1}^{n} \sum_{j=1}^{n} \text{Cov}(X_i, Y_j) \]
Example: dice

\[
\text{Cov}(X, Y) = \text{Cov}(X_1 + \ldots + X_n, Y_1 + \ldots + Y_n) = \sum_{i=1}^{n} \sum_{j=1}^{n} \text{Cov}(X_i, Y_j)
\]

\[
= \sum_{i=1}^{n} \text{Cov}(X_i, Y_i) + \sum_{i \neq j} \text{Cov}(X_i, Y_j)
\]

\[
0, \text{ by independence}
\]

\[\frac{\rho(X, Y)}{\text{Var}(X) \text{Var}(Y)} = -\frac{\frac{\text{Cov}(X, Y)}{n}}{\sqrt{\frac{\text{Var}(X)}{n}} \sqrt{\frac{\text{Var}(Y)}{n}}} = -\frac{1}{\sqrt{6}}\times\frac{\sqrt{6}}{\sqrt{6}} = -\frac{1}{6}\]
Example: dice

\[
\text{Cov}(X, Y) = \text{Cov}(X_1 + \ldots + X_n, Y_1 + \ldots + Y_n) = \sum_{i=1}^{n} \sum_{j=1}^{n} \text{Cov}(X_i, Y_j)
\]

\[
= \sum_{i=1}^{n} \text{Cov}(X_i, Y_i) + \sum_{i \neq j} \text{Cov}(X_i, Y_j)
\]

\[
0, \text{ by independence}
\]

\[
= \sum_{i=1}^{n} E(X_i Y_i) - E(X_i)E(Y_i) = -\frac{n}{36}
\]

\[
0 - \frac{1}{6} \times \frac{1}{6}
\]
Example: dice

\[
\text{Cov}(X, Y) = \text{Cov}(X_1 + \ldots + X_n, Y_1 + \ldots + Y_n) = \sum_{i=1}^{n} \sum_{j=1}^{n} \text{Cov}(X_i, Y_j)
\]

\[
= \sum_{i=1}^{n} \text{Cov}(X_i, Y_i) + \sum_{i \neq j} \text{Cov}(X_i, Y_j)
\]

\[
= \sum_{i=1}^{n} \left( E(X_i Y_i) - E(X_i)E(Y_i) \right) = -\frac{n}{36}
\]

and \(\text{Var}(X) = \text{Var}(Y) = n \times \frac{1}{6} \times \frac{5}{6} \). Thus

\[
\rho(X, Y) = \frac{-\frac{n}{36}}{n \frac{1}{6} \frac{5}{6}} = -\frac{1}{5}
\]
Why $\rho \in [-1, 1]$?
Cauchy-Schwarz inequality (B-H 10.1)

**Theorem**

For any random variables $U, V$:

$$(E(UV))^2 \leq E(U^2)E(V^2),$$

with equality if and only if $U = cV$ for some constant $c$. 

**Corollary**

For any random variables $X, Y$:

$$|\text{Cov}(X, Y)| \leq \sigma_X \sigma_Y,$$

i.e. $$|\rho(X, Y)| \leq 1$$

Furthermore,

- $\rho(X, Y) = 1 \iff Y = aX + b$ for some $a > 0$ (perfectly positively correlated): e.g. $X$ is temperature in $\degree C$ and $Y$ in $\degree F$
- $\rho(X, Y) = -1 \iff Y = aX + b$ for some $a < 0$ (perfectly negatively correlated): e.g. $X$ = number of Heads, $Y$ = number of Tails
Cauchy-Schwarz inequality (B-H 10.1)

Theorem

For any random variables $U, V$:

$$(E(UV))^2 \leq E(U^2)E(V^2),$$

with equality if and only if $U = cV$ for some constant $c$.

Corollary

For any random variables $X, Y$:

$$|\text{Cov}(X, Y)| \leq \sigma_X\sigma_Y,$$

i.e. $|\rho(X, Y)| \leq 1$

Furthermore,

- $\rho(X, Y) = 1 \iff Y = aX + b$ for some $a > 0$ (perfectly positively correlated): e.g. $X$ is temperature in °C and $Y$ in °F
- $\rho(X, Y) = -1 \iff Y = aX + b$ for some $a < 0$ (perfectly negatively correlated): e.g. $X$=number of Heads, $Y$ = number of Tails
Cauchy-Schwarz inequality: proof (optional)

Consider the function:

\[ f(\lambda) = E(U - \lambda V)^2 \]
\[ = E(U^2) + \lambda^2 E(V^2) - 2\lambda E(UV) \]

- By definition, \( f(\lambda) \geq 0 \) for all \( \lambda \). The minimum of the parabola is achieved at \( \lambda_0 = \frac{E(UV)}{E(V^2)} \) to be

\[ E(U^2) - \left( \frac{E(UV)}{E(V^2)} \right)^2, \]

which must be non-negative.

- (The case of equality) Suppose the minimum is zero. Then \( f(\lambda_0) = E(U - \lambda_0 V)^2 = 0 \), which means that \( U - \lambda_0 V \) is always zero, i.e., \( U = \lambda_0 V \).
Application: portfolio optimization

- Two assets:
  - Stock return $S$: $\mu_S = 10\%$, $\sigma_S = 10\%$
  - Bond return $B$: $\mu_B = 5\%$, $\sigma_B = 5\%$
  - Correlation coefficient: $\rho(S, B) = -0.5$
  - Incentive to invest in bond: hedge the risk of stock!
Application: portfolio optimization

- Two assets:
  - Stock return $S$: $\mu_S = 10\%$, $\sigma_S = 10\%$
  - Bond return $B$: $\mu_B = 5\%$, $\sigma_B = 5\%$
  - Correlation coefficient: $\rho(S, B) = -0.5$
  - Incentive to invest in bond: hedge the risk of stock!

- Portfolio: invest $\lambda$ fraction of funds in stock and $1 - \lambda$ in bond.
Application: portfolio optimization

- Two assets:
  - Stock return \( S \): \( \mu_S = 10\% \), \( \sigma_S = 10\% \)
  - Bond return \( B \): \( \mu_B = 5\% \), \( \sigma_B = 5\% \)
  - Correlation coefficient: \( \rho(S, B) = -0.5 \)
  - Incentive to invest in bond: hedge the risk of stock!

- Portfolio: invest \( \lambda \) fraction of funds in stock and \( 1 - \lambda \) in bond.

- Return: \( D = \lambda S + (1 - \lambda)B \)
  - Expected return:
    \[
    \mu_D = \lambda \mu_S + (1 - \lambda) \mu_B
    \]
  - Variance:
    \[
    \sigma_D^2 = \lambda^2 \sigma_S^2 + (1 - \lambda)^2 \sigma_B^2 + 2\lambda(1 - \lambda)\rho(S, B)\sigma_S\sigma_B
    \]
Performance of portfolio

- Minimal risk ($\sigma_D = 4\%$) occurred at $\mu_D = 6.25\%$: $\frac{1}{4}S + \frac{3}{4}B$
- As $\lambda \uparrow$, portfolio shifts to stock
  - return $\uparrow$; volatility first $\downarrow$ then $\uparrow$
Performance of portfolio

- Minimal risk ($\sigma_D = 4\%$) occurred at $\mu_D = 6.25\%$: $\frac{1}{4} S + \frac{3}{4} B$
- As $\lambda \uparrow$, portfolio shifts to stock
  - return $\uparrow$; volatility first $\downarrow$ then $\uparrow$
- A reasonable investor would not operate on the left of the minimum
Performance of portfolio

- Minimal risk ($\sigma_D = 4\%$) occurred at $\mu_D = 6.25\%$: $\frac{1}{4} S + \frac{3}{4} B$
- As $\lambda \uparrow$, portfolio shifts to stock
  - return ↑; volatility first ↓ then ↑
- A reasonable investor would not operate on the left of the minimum
- Things become more complicated with multiple assets (Markowitz portfolio optimization)
Performance of portfolio

- Minimal risk ($\sigma_D = 4\%$) occurred at $\mu_D = 6.25\%$: $\frac{1}{4}S + \frac{3}{4}B$
- As $\lambda \uparrow$, portfolio shifts to stock
  - return $\uparrow$; volatility first $\downarrow$ then $\uparrow$
- A reasonable investor would not operate on the left of the minimum
- Things become more complicated with multiple assets (Markowitz portfolio optimization)
- Can you go make a fortune after class? What’s the catch?