15: Covariance

David Varodayan
February 10, 2020
Adapted from slides by Lisa Yan
CS109 roadmap

Multiple events:

Intersection:
\[ P(E \cap F) = P(EF) \]

Conditional probability:
\[ P(E|F) = \frac{P(EF)}{P(F)} \]

Independence:
\[ P(EF) = P(E)P(F) \]

Joint (Multivariate) distributions

Joint PMF/PDF:
\[ p_{X,Y}(x,y) \]
\[ f_{X,Y}(x,y) \]

Conditional distributions:
\[ p_{X|Y}(x|y) \]
\[ f_{X|Y}(x|y) \]

Independent RVs

Sum of independent RVs
Tracking in 2-D space

- You have a **prior belief** about the 2-D location of an object, \((X, Y)\).
- You observe a **noisy distance measurement**, \(D = 4\).
- What is your **updated (posterior) belief** of the 2-D location of the object after observing the measurement?

Recall Bayes terminology:

\[
\text{posterior belief} \\
\frac{f_{X,Y|D}(x, y|d)}{f_D(d)} = \frac{f_{D|X,Y}(d|x, y)f_{X,Y}(x, y)}{\text{normalization constant}}
\]

likelihood (of evidence)  prior belief
Tracking in 2-D space

- You have a **prior belief** about the 2-D location of an object, \((X, Y)\).
- You observe a noisy distance measurement, \(D = 4\).
- What is your updated (posterior) belief of the 2-D location of the object after observing the measurement?

Let \((X, Y) = \) object’s 2-D location. (your satellite is at (0,0))

Suppose the prior distribution is a symmetric bivariate normal distribution:

\[
 f_{X,Y}(x, y) = \frac{1}{2\pi \sigma^2} e^{-\frac{[(x-3)^2+(y-3)^2]}{2(2^2)}} = K_1 \cdot e^{-\frac{[(x-3)^2+(y-3)^2]}{8}}
\]

(normalizing constant)
Tracking in 2-D space

- You have a prior belief about the 2-D location of an object, $(X, Y)$.
- You observe a noisy distance measurement, $D = 4$.
- What is your updated (posterior) belief of the 2-D location of the object after observing the measurement?

Let $D = \text{distance from the satellite (radially)}$.
Suppose you knew your actual position: $(x, y)$.
- $D$ is still noisy! Suppose noise is unit variance: $\sigma^2 = 1$
- On average, $D$ is your actual position: $\mu = \sqrt{x^2 + y^2}$

If you knew your actual location $(x, y)$, you could say **how likely** a measurement $D = 4$ is!!
Tracking in 2-D space

- You have a prior belief about the 2-D location of an object, \((X, Y)\).
- You observe a noisy distance measurement, \(D = 4\).
- What is your updated (posterior) belief of the 2-D location of the object after observing the measurement?

If you knew your actual location \((x, y)\), you could say how likely a measurement \(D = 4\) is!!

If noise is normal: \[ D|X, Y \sim N\left(\mu = \sqrt{x^2 + y^2}, \sigma^2 = 1\right) \]

Distance measurement of a ping is normal with respect to the true location.
Tracking in 2-D space

- You have a prior belief about the 2-D location of an object, \((X, Y)\).
- You observe a noisy distance measurement, \(D = 4\).
- What is your updated (posterior) belief of the 2-D location of the object after observing the measurement?

If you knew your actual location \((x, y)\), you could say how likely a measurement \(D = 4\) is!!

\[
f_{D|X,Y}(D = d | X = x, Y = y) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(d-\mu)^2}{2\sigma^2}}
\]

\[
D|X,Y \sim \mathcal{N}(\mu = \sqrt{x^2 + y^2}, \sigma^2 = 1)
\]

\[
K_2 \cdot e^{-\frac{(d-\sqrt{x^2+y^2})^2}{2}}
\]

normalize constant
Tracking in 2-D space

- You have a prior belief about the 2-D location of an object, \((X, Y)\).
- You observe a noisy distance measurement, \(D = 4\).
- What is your updated (posterior) belief of the 2-D location of the object after observing the measurement?

Prior belief

\[
f_{X,Y}(x, y) = K_1 \cdot e^{-\frac{[(x-3)^2 + (y-3)^2]}{8}}
\]

Top-down view

Observation likelihood

\[
f_{D|X,Y}(d|x, y) = K_2 \cdot e^{-\frac{-(d-\sqrt{x^2+y^2})^2}{2}}
\]

Posterior belief

\[
f_{X,Y|D}(x, y|4) = f_{X,Y|D}(X = x, Y = y|D = 4)
\]
Tracking in 2-D space

What is your updated (posterior) belief of the 2-D location of the object after observing the measurement?

\[ f_{X,Y|D}(X = x, Y = y|D = 4) = \frac{f_{D|X,Y}(D = 4|X = x, Y = y)f_{X,Y}(x, y)}{f(D = 4)} \]

\[ = K_2 \cdot e^{-\frac{(4-\sqrt{x^2+y^2})^2}{2}} \cdot K_1 \cdot e^{-\frac{(x-3)^2+(y-3)^2}{8}} \]

Bayes’ Theorem

\[ = \frac{f(D = 4)}{f(D = 4)} \]

\[ = K_3 \cdot e^{-\left[\frac{(4-\sqrt{x^2+y^2})^2}{2} + \frac{(x-3)^2+(y-3)^2}{8}\right]} \]

also constant

\[ = K_4 \cdot e^{-\left[\frac{(4-\sqrt{x^2+y^2})^2}{2} + \frac{(x-3)^2+(y-3)^2}{8}\right]} \]
Tracking in 2-D space: Posterior belief

Prior belief

Top-down view 3-D view

\[ f_{x,y}(x, y) = K_1 \cdot e^{-\frac{(x-3)^2 + (y-3)^2}{8}} \]

Posterior belief

Top-down view 3-D view

\[ f_{x,y|D}(x, y|4) = K_4 \cdot e^{-\frac{\left(4-\sqrt{x^2+y^2}\right)^2}{2} + \frac{(x-3)^2 + (y-3)^2}{8}} \]
Today’s plan

→ Covariance

Variance/covariance of independent RVs

Correlation
A word about today’s diagrams:
Spot the difference

How do the following distributions of two variables differ?

In both distributions: \( E[X] = E[Y], \ Var(X) = Var(Y) \)

\( E[X] \) is the same in both plots
So are \( E[Y], Var(X) \) and \( Var(Y) \)
Covariance

The covariance of two variables $X$ and $Y$ is:

\[
\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])] \\
= E[XY] - E[X]E[Y]
\]

Proof of second part:

\[
\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])] \\
= E[XY] - E[X]E[Y]
\]

(linearity of expectation)

$E[X], E[Y]$ are scalars
Covarying humans

<table>
<thead>
<tr>
<th>Weight (kg)</th>
<th>Height (in)</th>
<th>W · H</th>
</tr>
</thead>
<tbody>
<tr>
<td>64</td>
<td>57</td>
<td>3648</td>
</tr>
<tr>
<td>71</td>
<td>59</td>
<td>4189</td>
</tr>
<tr>
<td>53</td>
<td>49</td>
<td>2597</td>
</tr>
<tr>
<td>67</td>
<td>62</td>
<td>4154</td>
</tr>
<tr>
<td>55</td>
<td>51</td>
<td>2805</td>
</tr>
<tr>
<td>58</td>
<td>50</td>
<td>2900</td>
</tr>
<tr>
<td>77</td>
<td>55</td>
<td>4235</td>
</tr>
<tr>
<td>57</td>
<td>48</td>
<td>2736</td>
</tr>
<tr>
<td>56</td>
<td>42</td>
<td>2352</td>
</tr>
<tr>
<td>51</td>
<td>42</td>
<td>2142</td>
</tr>
<tr>
<td>76</td>
<td>61</td>
<td>4636</td>
</tr>
<tr>
<td>68</td>
<td>57</td>
<td>3876</td>
</tr>
</tbody>
</table>

\[
E[W] = 62.75 \\
E[H] = 52.75 \\
E[WH] = 3355.83
\]

What is the covariance of weight \( W \) and height \( H \)?

\[
= 3355.83 - (62.75)(52.75) \\
= 45.77 \text{ (positive)}
\]

Positive covariance = as one variable increases, so does the second variable.
Covariance reps

Is the covariance positive, negative, or zero?

1. \( Y = Y \)  
   \( X = x \)  
   \( E[X] \)  
   \( E[Y] \)  
   positive

2. \( Y = Y \)  
   \( X = x \)  
   \( E[X] \)  
   \( E[Y] \)  
   negative

3. \( Y = Y \)  
   \( X = x \)  
   \( E[X] \)  
   \( E[Y] \)  
   zero

\[
\]
Properties of Covariance

The **covariance** of two variables $X$ and $Y$ is:

\[
\]

**True/False:**

1. $\text{Cov}(X, Y) = \text{Cov}(Y, X)$ \(\text{T}\)
2. $\text{Cov}(X, X) = E[X \cdot X] - E[X]E[X] = \text{Var}(X)$ \(\text{T}\)
3. $\text{Cov}(aX + b, Y) = a\text{Cov}(X, Y)$ \(\text{T}\) (not linear)
4. $\text{Cov}\left(\sum_i X_i, \sum_j Y_j\right) = \sum_i \sum_j \text{Cov}(X_i, Y_j)$ \(\text{T}\) (proof left to you)
Announcements

Midquarter feedback (optional but appreciated)
Link posted in announcement on CS109 webpage
https://forms.gle/6JC6a4oyrH5hEGTy7
Closes: Wednesday February 12, 11:59pm
Today’s plan

Covariance

Variance/covariance of sum of RVs

Correlation
Variance of sum of RVs

If $X$ and $Y$ are random variables, then

$$\text{Var}(X + Y) = \text{Var}(X) + 2 \cdot \text{Cov}(X, Y) + \text{Var}(Y)$$

Proof: \[\text{Var}(X + Y) = \text{Cov}(X + Y, X + Y)\]
\[= \text{Cov}(X, X) + \text{Cov}(X, Y) + \text{Cov}(Y, X) + \text{Cov}(Y, Y)\]
\[= \text{Var}(X) + 2 \cdot \text{Cov}(X, Y) + \text{Var}(Y)\]

More generally:
\[\text{Var}\left( \sum_{i=1}^{n} X_i \right) = \sum_{i=1}^{n} \text{Var}(X_i) + 2 \sum_{i=1}^{n} \sum_{j=i+1}^{n} \text{Cov}(X_i, X_j)\]

(proof in extra slides)
Variance of sum of independent random variables

If $X$ and $Y$ are independent, then:

$$E[XY] = E[X]E[Y]$$

Therefore for independent $X$ and $Y$:

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

Proof of covariance:

$$\text{Cov}(X, Y) = E[XY] - E[X]E[Y]$$
$$= 0$$

NOT bidirectional: Cov$(X, Y) = 0$ does NOT imply independence of $X$ and $Y$!
Zero covariance does not imply independence

Let $X$ take on values $\{-1,0,1\}$ with equal probability $1/3$.

Define $Y = \begin{cases} 
1 & \text{if } X = 0 \\
0 & \text{otherwise}
\end{cases}$

What is the joint PMF of $X$ and $Y$?

A. $\begin{array}{c|ccc}
Y & -1 & 0 & 1 \\
0 & 1/6 & 1/6 & 1/6 \\
1 & 1/6 & 1/6 & 1/6 \\
\end{array}$

B. $\begin{array}{c|ccc}
X & -1 & 0 & 1 \\
0 & 1/3 & 0 & 1/3 \\
1 & 0 & 1/3 & 0 \\
\end{array}$

C. $\begin{array}{c|ccc}
X & -1 & 0 & 1 \\
0 & 0 & 1/3 & 0 \\
1 & 1/3 & 0 & 1/3 \\
\end{array}$
Zero covariance does not imply independence

Let $X$ take on values $\{-1,0,1\}$ with equal probability $1/3$.

Define $Y = \begin{cases} 
1 & \text{if } X = 0 \\
0 & \text{otherwise}
\end{cases}$

<table>
<thead>
<tr>
<th>$Y$</th>
<th>$-1$</th>
<th>$0$</th>
<th>$1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0$</td>
<td>$1/3$</td>
<td>$0$</td>
<td>$1/3$</td>
</tr>
<tr>
<td>$1$</td>
<td>$0$</td>
<td>$1/3$</td>
<td>$0$</td>
</tr>
<tr>
<td>$\frac{1}{3}$</td>
<td>$\frac{1}{3}$</td>
<td>$\frac{1}{3}$</td>
<td>$\frac{1}{3}$</td>
</tr>
</tbody>
</table>

1. $E[X] = 0$  
   $E[Y] = \frac{1}{3}$

2. $E[XY] = 0$  
   ($xy$ is always 0)


4. Are $X$ and $Y$ independent?  
   \[ \text{No} \]  
   \[ P(Y=0) = \frac{2}{3} \neq P(Y=0|X=0) = 0 \]
Variance of sum of independent random variables

If $X$ and $Y$ are independent, then:

$$E[XY] = E[X]E[Y]$$

(proof in extra slides)

Therefore for independent $X$ and $Y$:

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

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

Proof of variance:

$$\text{Var}(X + Y) = \text{Var}(X) + 2 \cdot \text{Cov}(X, Y) + \text{Var}(Y)$$

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

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

Proof (proved earlier)

$X$ and $Y$ are independent

1. Also not bidirectional
2. Does not apply to dependent $X$ and $Y$
Variance of the Binomial

\[ X \sim \text{Bin}(n, p) \quad \text{Var}(X) = np(1 - p) \]

Let 
\[ X = \sum_{i=1}^{n} X_i \]

Let \( X_i = \text{ith trial is heads} \)
\( X_i \sim \text{Ber}(p) \)
\( \text{Var}(X_i) = p(1 - p) \)

\( X_i \) are independent (by definition)

\[ \text{Var}(X) = \text{Var} \left( \sum_{i=1}^{n} X_i \right) \]
\[ = \sum_{i=1}^{n} \text{Var}(X_i) \quad \text{X}_i \text{ are independent, therefore variance of sum} \]
\[ = \sum_{i=1}^{n} p(1 - p) \quad \text{sum of variance} \]
\[ = np(1 - p) \quad \text{Variance of Bernoulli} \]
Today’s plan

- Covariance
- Variance/covariance of sum of independent RVs
- Correlation
Covarying humans

What is the covariance of weight $W$ and height $H$?


= 3355.83 - (62.75)(52.75)

= 45.77 (positive)

What about weight (lb) and height (cm)?

Cov(2.20$W$, 2.54$H$)

= $E[2.20W \cdot 2.54H] - E[2.20W]E[2.54H]$

= 18752.38 - (138.05)(133.99)

= 255.06 (positive)

Covariance depends on units!

For covariance, the sign (+/−) is more meaningful than the value.
Correlation

The correlation of two variables $X$ and $Y$ is:

$$\rho(X,Y) = \frac{\text{Cov}(X,Y)}{\sigma_X \sigma_Y}$$

\[ \sigma_X^2 = \text{Var}(X), \quad \sigma_Y^2 = \text{Var}(Y) \]

- Note: $-1 \leq \rho(X,Y) \leq 1$
- Correlation measures the linear relationship between $X$ and $Y$:

  $$\rho(X,Y) = 1 \quad \Rightarrow \quad Y = aX + b, \text{where } a = \frac{\sigma_Y}{\sigma_X}$$
  $$\rho(X,Y) = -1 \quad \Rightarrow \quad Y = aX + b, \text{where } a = -\frac{\sigma_Y}{\sigma_X}$$
  $$\rho(X,Y) = 0 \quad \Rightarrow \quad \text{“uncorrelated” (absence of linear relationship)}$$
Correlation reps

What is the correlation coefficient $\rho(X, Y)$?

1. [Image of a line with a negative slope]
   -1

2. [Image of a line with a positive slope]
   +1

3. [Image of a scatter plot with no apparent trend]

4. [Image of a scatter plot with a non-linear trend]

A. $\rho(X, Y) = 1$
B. $\rho(X, Y) = -1$
C. $\rho(X, Y) = 0$
D. Other
Correlation reps

What is the correlation coefficient $\rho(X, Y)$?

1. $\rho(X, Y) = -1$
   
   $Y = -\frac{\sigma_Y}{\sigma_X} X + b$

2. $\rho(X, Y) = 1$
   
   $Y = \frac{\sigma_Y}{\sigma_X} X + b$

3. $\rho(X, Y) = 0$
   
   “uncorrelated”

4. $\rho(X, Y) = 0$
   
   $Y = X^2 + \text{noise}$

Correlation measures linearity. $X$ and $Y$ can be nonlinearly related even if $\text{Cov}(X, Y) = 0$.
Spurious Correlations

"Correlation does not imply causation"

\( \rho(X, Y) \) is used a lot to statistically quantify the relationship b/t X and Y.

Correlation: 0.947091

https://www.tylervigen.com/spurious-correlations

Stanford University
$\rho(X, Y)$ is used a lot to statistically quantify the relationship between $X$ and $Y$.

**Correlation:**
0.947091

**Per capita cheese consumption** correlates with **Number of people who died by becoming tangled in their bedsheets**

![Graph showing the relationship between cheese consumption and deaths from tangled bedsheets](image)
Arcade revenue vs. CS PhDs

“Correlation does not imply causation”

Correlation: 0.947091

Total revenue generated by arcades correlates with Computer science doctorates awarded in the US

Data sources: U.S. Census Bureau and National Science Foundation

https://www.tylervigen.com/spurious-correlations

Stanford University
Extra slides

Expectation of a product of independent RVs

Variance of sums of variables
Expectation of product of independent RVs

If $X$ and $Y$ are independent, then:

$$E[XY] = E[X]E[Y]$$
$$E[g(X)h(Y)] = E[g(X)]E[h(Y)]$$

More generally,

Proof:

$$E[g(X)h(Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x)h(y)f_{X,Y}(x,y)\,dx\,dy$$
$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x)h(y)f_X(x)f_Y(y)\,dx\,dy$$
$$= \int_{-\infty}^{\infty} h(y)f_Y(y)\,dy \int_{-\infty}^{\infty} g(x)f_X(x)\,dx$$
$$= \left(\int_{-\infty}^{\infty} g(x)f_X(x)\,dx\right)\left(\int_{-\infty}^{\infty} h(y)f_Y(y)\,dy\right)$$
$$= E[g(X)]E[h(Y)]$$

(for discrete proof, replace integrals with summations)

$X$ and $Y$ are independent

Terms dependent on $y$ are constant in integral of $x$

Integrals separate
Variance of Sums of Variables

\[
\text{Var}\left( \sum_{i=1}^{n} X_i \right) = \sum_{i=1}^{n} \text{Var}(X_i) + 2 \sum_{i=1}^{n} \sum_{j=i+1}^{n} \text{Cov}(X_i, X_j)
\]

For 2 variables:

\[
\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)
\]

Proof:

\[
\begin{align*}
\text{Var}\left( \sum_{i=1}^{n} X_i \right) &= \text{Cov}\left( \sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i \right) \\
&= \sum_{i=1}^{n} \text{Var}(X_i) + \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} \text{Cov}(X_i, X_j) \\
&= \sum_{i=1}^{n} \text{Var}(X_i) + 2 \sum_{i=1}^{n} \sum_{j=i+1}^{n} \text{Cov}(X_i, X_j)
\end{align*}
\]

Symmetry of covariance:

\[
\text{Cov}(X, X) = \text{Var}(X)
\]

Adjust summation bounds:

Stanford University