Covariance, LLSE, Conditional Expectation, Markov Chains 10

Computer Science Mentors 70

November 14 to November 18, 2016

1 Covariance

1.1 Introduction

The **covariance** of two random variables X and Y is defined as:

$$\text{Cov}(X, Y) := \text{E}((X - \text{E}(X)) \cdot (Y - \text{E}(Y)))$$

1.2 Warm Up

1. Prove that $\text{Cov}(X, X) = \text{Var}(X)$:

Solution:

$$\text{Cov}(X, X) = \text{E}(X \cdot X) - \text{E}(X) \cdot \text{E}(X) = \text{E}(X^2) - \text{E}(X)^2$$

2. Prove that if X and Y are independent, then $\text{Cov}(X, Y) = 0$:

Solution:

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

Remember that a property of expectation is that if X and Y are independent, then $\mathrm{E}(XY) = \mathrm{E}(X) \cdot \mathrm{E}(Y)$, so we get 0 when we subtract

3. Prove that $\mathrm{Cov}(X + Y, Z) = \mathrm{Cov}(X, Z) + \mathrm{Cov}(Y, Z)$:

Solution:

$$\begin{aligned}\mathrm{Cov}(X+Y,Z) &= \mathrm{E}((X+Y)\cdot Z) - \mathrm{E}(X+Y)\cdot \mathrm{E}(Z)\\ &= \mathrm{E}(X\cdot Z) + \mathrm{E}(Y\cdot Z) - (\mathrm{E}(X)\cdot \mathrm{E}(Z) + \mathrm{E}(Y)\cdot \mathrm{E}(Z))\\ &= \mathrm{E}(X\cdot Z) - \mathrm{E}(X)\cdot \mathrm{E}(Z) + \mathrm{E}(Y\cdot Z) - \mathrm{E}(Y)\cdot \mathrm{E}(Z)\\ &= \mathrm{Cov}(X,Z) + \mathrm{Cov}(Y,Z)\end{aligned}$$

1.3 Questions

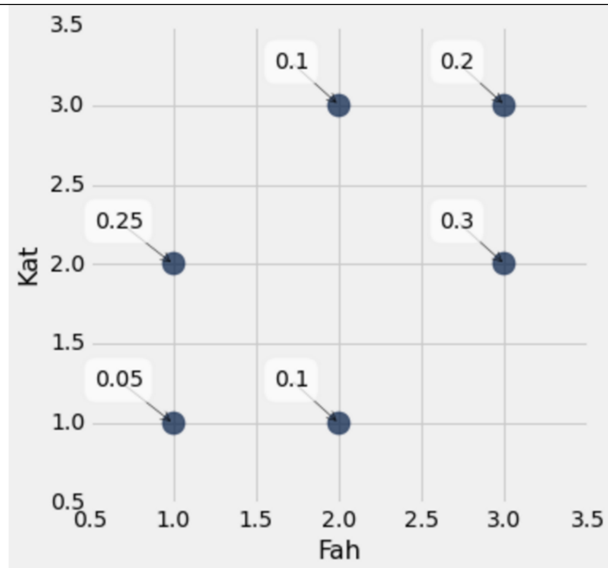
1. Roll 2 dice. Let A be the number of 6's you get, and B be the number of 5's, find $\mathrm{Cov}(A, B)$

Solution: $\mathrm{E}(A) = \frac{1}{6}$ for one die, by linearity of expectation, two dice make $\frac{1}{3}$, same for $\mathrm{E}(B)$ $\mathrm{E}(A) = \frac{1}{3}, \mathrm{E}(B) = \frac{1}{3}$
AB can be either 0 (if no 5's or 6's show up) or 1 (get a 5 and a 6).

$$\begin{aligned}\mathrm{E}(AB) &= 1 \cdot \mathrm{P}[\text{get a 5 and a 6}]\\ &= \mathrm{P}[\text{first die = 5 and second die = 6}] + \mathrm{P}[\text{first die = 6 and second die = 5}]\\ &= \frac{1}{36} + \frac{1}{36}\end{aligned}$$

$$\begin{aligned}\mathrm{Cov}(AB) &= \mathrm{E}(AB) - \mathrm{E}(A)\cdot \mathrm{E}(B)\\ &= \frac{1}{18} - \frac{1}{9}\\ &= -\frac{1}{18}\end{aligned}$$

2. Consider the following distribution with random variables Fah and Kat:



Find the covariance of Fah and Kat.

Solution: E(Fah) $= 1 \cdot .3 + 2 \cdot .2 + 3 \cdot .5 = \mathbf{2.2}$
E(Kat) = $1 \cdot .15 + 2 \cdot .55 + 3 \cdot .3 = \mathbf{2.15}$
E(KatFah) = $1 \cdot 1 \cdot .05 + 1 \cdot 2 \cdot .25 + 2 \cdot 1 \cdot .1 + 2 \cdot 3 \cdot .1 + 3 \cdot 2 \cdot .3 + 3 \cdot 3 \cdot .2 = \mathbf{4.95}$
cov(Kat, Fah) = $4.95 - 2.2 \cdot 2.15 = \mathbf{0.22}$

2 LLSE

2.1 Introduction

Theorem: Consider two random variables, X, Y with a given distribution $\text{P}[X = x, Y = y]$. Then

$$\text{L}[Y|X] = \text{E}(Y) + \frac{\text{Cov}(X,Y)}{\text{Var}(X)}(X - \text{E}(X))$$

2.2 Questions

1. Assume that

$$Y = \alpha X + Z$$

where X and Z are independent and $\text{E}(X) = \text{E}(Z) = 0$. Find $\text{L}[X|Y]$.

Solution:

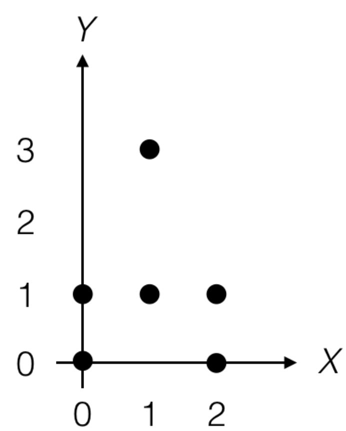
$$\begin{aligned}\text{Cov}(X,Y) &= \text{E}(XY) - \text{E}(X)\text{E}(Y)\\ &= \text{E}(X\cdot(\alpha X + Z)) = \alpha\text{E}(X^2)\end{aligned}$$

$$\begin{aligned}\text{Var}(Y) &= \alpha^2\text{Var}(X) + \text{Var}(Z)\\ &= \alpha^2\text{E}(X^2) + \text{E}(Z^2)\end{aligned}$$

Therefore,

$$\text{L}[X|Y] = \frac{\alpha\text{E}(X^2)}{\alpha^2\text{E}(X^2) + \text{E}(Z^2)}\cdot Y$$

2. The figure below shows the six equally likely values of the random pair (X, Y). Specify the functions of:
   - $L[Y \mid X]$
   - $E(X \mid Y)$
   - $L[X \mid Y]$
   - $E(Y \mid X)$



Solution: Let's calculate some useful properties of the distribution first and then see how we can use them to calculate the estimates.

$$
\begin{aligned}
|\Omega| = 6 &\implies P[\text{one point}] = \frac{1}{6} \\
\mathrm{E}(X) &= 0\left(\frac{2}{6}\right) + 1\left(\frac{2}{6}\right) + 2\left(\frac{2}{6}\right) \\
&= 1 \\
\mathrm{E}(Y) &= 0\left(\frac{2}{6}\right) + 1\left(\frac{3}{6}\right) + 30\left(\frac{1}{6}\right) \\
&= 1 \\
\mathrm{E}(XY) &= 0\left(\frac{3}{6}\right) + 1\left(\frac{1}{6}\right) + 2\left(\frac{1}{6}\right) + 3\left(\frac{1}{6}\right) \\
&= 1 \\
\mathrm{Cov}(X,Y) = 0 &\implies \mathrm{L}[Y|X] = E(Y)
\end{aligned}
$$

- $L[Y \mid X]$: Using the LLSE formula: $\mathrm{L}[Y \mid X] = \mathrm{E}[Y] + \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(Y)}(Y - \mathrm{E}[Y]) = \mathrm{E}[Y]$. Therefore $\boxed{L[Y \mid X] = 1}$
- $E[X \mid Y]$: Notice the symmetry across $X = 1$. For all values of y, $E[X|Y = y]$ is the same; therefore $\boxed{E[X|Y] = E[X] = 1}$.
- $L[X \mid Y]$: The MMSE estimator for X given Y is a linear function, therefore $\boxed{\mathrm{L}[X \mid Y] = \mathrm{E}[X \mid Y] = 1}$
- $E[Y \mid X]$ For this one we can't make use of symmetry or directly apply what we calculated above. We must go back to the definition of conditional expectation. We can calculate $\mathrm{E}[Y \mid X = x]$ for every point x, and that entirely defines the expression:

$$
\mathrm{E}(Y \mid X = x) = \begin{cases} \frac{1}{2} & \text{if } x = 0 \\ 2 & \text{if } x = 1 \\ \frac{1}{2} & \text{if } x = 2 \end{cases}
$$

The above equation is sufficient, but we can go further by realizing that these points are part of a flipped absolute value function centered around $x = 1$: $\boxed{\mathrm{E}[Y \mid X] = \frac{-3}{2}|X - 1| + 2}$. Indeed, this is not linear, which is why $\mathrm{L}[Y \mid X] \neq \mathrm{E}[Y \mid X]$.

3 Conditional Expectation

3.1 Introduction

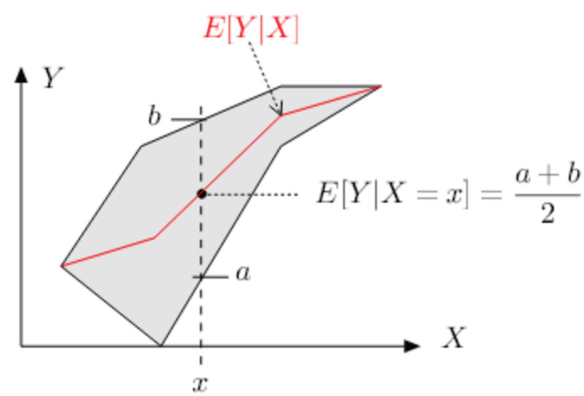
The **conditional expectation** of Y given X is defined by

$$E[Y|X=x] = \sum_y y \cdot P[Y=y|X=x] = \sum_y y \cdot \frac{P[X=x, Y=y]}{P[X=x]}$$

Properties of Conditional Expectation

$$\begin{aligned} E(a|Y)) &= a \\ E(aX+bZ|Y) &= a \cdot E(X|Y) + b \cdot E(Z|Y) \\ E(X|Y) &\geq 0 \text{ if } X \geq 0 \\ E(X|Y) &= E(X) \text{ if } X, Y \text{ independent} \\ E(E(X|Y)) &= E(X) \end{aligned}$$

Solution: Here is a picture that shows that conditioning creates a new random variable with a new distribution. Figure 9 of note 26 does so.



3.2 Questions

1. Prove $E(E(Y|X)) = E(Y)$

Solution:

$$E(E(Y|X)) = \sum_x E(Y|X=x) \cdot P[X=x]$$

$$
\begin{aligned}
&= \sum_x (\sum_y y \cdot \mathrm{P}[Y = y|X = x]) \cdot \mathrm{P}[X = x] \\
&= \sum_y y \cdot \sum_x y \cdot \mathrm{P}[X = x|Y = y]) \cdot \mathrm{P}[Y = y] \\
&= \sum_y y \cdot \mathrm{P}[Y = y] \cdot \sum_x \mathrm{P}[X = x|Y = y]) \\
&= \sum_y y \cdot \mathrm{P}[Y = y] = E[Y]
\end{aligned}
$$

2. Prove $\mathrm{E}(h(X) \cdot Y|X) = h(X) \cdot \mathrm{E}(Y|X)$

Solution:

$$
\begin{aligned}
\mathrm{E}(h(X) \cdot Y|X) &= \sum_y h(X) \cdot y \cdot \mathrm{P}[Y = y|X] \\
&= h(X) \sum_y y \cdot \mathrm{P}[Y = y|X] \\
&= h(X) \cdot \mathrm{E}[Y|X]
\end{aligned}
$$

3. Consider the random variables Y and X with the following probabilities

This table gives the probability distribution for $P[X \cap Y]$

		X		
		0	1	2
Y	0	0	.1	.2
	1	.1	.2	.1
	2	.2	.1	0

Find:

(a) $\mathrm{E}(Y|X = 0)$

Solution:

$$
\begin{aligned}
\mathrm{E}(Y|X = 0) &= \mathrm{P}[Y = 0|X = 0] \cdot 0 + \mathrm{P}[Y = 1|X = 0] \cdot 1 + \mathrm{P}[Y = 2|X = 0] \cdot 2 \\
&= \frac{0}{0 + .1 + .2} \cdot 0 + \frac{.1}{0 + .1 + .2} \cdot 1 + \frac{.2}{0 + .1 + .2} \cdot 2 \\
&= \frac{20}{12} = \frac{5}{3}
\end{aligned}
$$

(b) $\mathrm{E}(Y|X=1)$

Solution:

$$\begin{aligned}\mathrm{E}(Y|X=1) &= \mathrm{P}[Y=0|X=1]\cdot 0 + \mathrm{P}[Y=1|X=1]\cdot 1 + \mathrm{P}[Y=2|X=1]\cdot 2\\ &= \frac{0.1}{0.1+0.2+0.1}\cdot 0 + \frac{0.2}{0.1+0.2+0.1}\cdot 1 + \frac{0.1}{0.1+0.2+0.1}\cdot 2\\ &= 0 + \frac{0.2}{0.4} + \frac{0.1\cdot 2}{0.4}\\ &= 0.5 + 0.5 = 1\end{aligned}$$

(c) $\mathrm{E}(Y|X=2)$

Solution:

$$\begin{aligned}\mathrm{E}(Y|X=2) &= \mathrm{P}[Y=0|X=2]\cdot 0 + \mathrm{P}[Y=1|X=2]\cdot 1 + \mathrm{P}[Y=2|X=2]\cdot 2\\ &= \frac{0.2}{0.2+0.1+0}\cdot 0 + \frac{0.1}{0.2+0.1+0}\cdot 1 + \frac{0}{0.2+0.1+0}\cdot 2\\ &= 0 + \frac{0.1}{0.3} + 0 = \frac{1}{3}\end{aligned}$$

(d) $\mathrm{E}(Y)$

Solution: These events are disjoint, so to find $\mathrm{E}[Y]$, we can just sum up individual probabilities (note that sum of all probabilities is the sum of 1)

$$\begin{aligned}\mathrm{E}(Y) &= \mathrm{E}(Y|X=0)\cdot \mathrm{P}[X=0] + \mathrm{E}(Y|X=1)\cdot \mathrm{P}[X=1] + \mathrm{E}(Y|X=2)\cdot \mathrm{P}[X=2]\\ &= \frac{20}{12}\cdot(0+0.1+0.2) + 1\cdot(0.1+0.2+0.1) + \frac{1}{3}\cdot(0.2+0.1+0)\\ &= \frac{20}{12}\cdot\frac{3}{10} + 0.4 + \frac{0.3}{3}\\ &= \frac{60}{120} + \frac{2}{5} + \frac{1}{10}\\ &= \frac{60}{120} + \frac{48}{120} + \frac{12}{120} = \frac{120}{120} = 1\end{aligned}$$

4 Markov Chains

P is a **transition probability matrix** if:

1. All of the entries are non-negative.
2. The sum of entries in each row is 1.

A **Markov chain** is defined by four things: ($\mathscr{X}$, π_0, P, $\{X_n\}_{n=0}^{\infty}$)

$\mathscr{X}$	Set of states
π_0	Initial probability distribution
P	Transition probability matrix
$\{X_n\}_{n=0}^{\infty}$	Sequence of random variables where: $\mathrm{P}[X_0 = i] = \pi_0(i), i \in \mathscr{X}$ $\mathrm{P}[X_{n+1} = j \mid X_n = i, X_{n-1}, \ldots, X_0] = \mathrm{P}(i,j), \forall n \geq 0, \forall i,j \in \mathscr{X}$

A Markov chain is **irreducible** if we can go from any state to any other state, possibly in multiple steps.

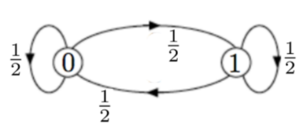
Define value $d(i)$ for each state i as:

$$d(i) := g.c.d\{n > 0 \mid P^n(i,i) = \mathrm{P}[X_n = i \mid X_0 = i] > 0\}, i \in \mathscr{X}$$

If $d(i) = 1$, then the Markov chain is **aperiodic**. If $d(i) \neq 1$, then the Markov chain is periodic and its **period** is $d(i)$.

A distribution π is **invariant** if $\pi \cdot P = \pi$.

Theorem 24.3: A finite irreducible Markov chain has a unique invariant distribution.
Theorem 24.4: All irreducible and aperiodic Markov chains converge to the unique invariant distribution. If a Markov chain is finite and reducible, the amount of time spent in each state approaches the invariant distribution as n grows large
Equations that model what will happen at the next step are called **first step equations**

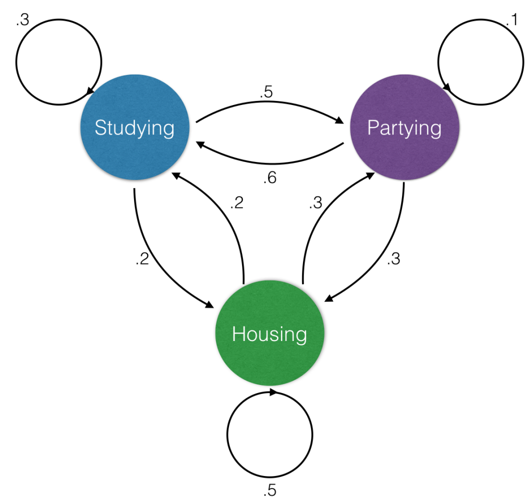


Denote $\beta(i,j)$ as the expected amount of time it would take to move from i to j. $\beta(0,1) = 1 + \frac{1}{2} \cdot \beta(0,1)$ $\beta(1,1) = 0$

4.1 Questions

1. **Life of Alex**
   Alex is enjoying college life. She spends a day either studying, partying, or looking for housing for the next year. If she is studying, the chances of her studying the next day are 30%, the chances of her partying the next day are 50%, and the chances of her looking for housing the next day are 20%. If she is partying, the chances of her partying the next day are 10%, the chances of her studying the next day are 60%, and the chances of her looking for housing the next day are 30%. If she is looking for housing, the chances of her looking for housing the next day are 50%, the chances of her partying the next day are 30% and the chances of her studying the next day are 20%.

   (a) Draw a Markov chain to visualize Alexs life.

   Solution:

   (b) Write out a matrix to represent this Markov chain.

   Solution:

$$\begin{bmatrix} .3 & .5 & .2 \\ .6 & .1 & .3 \\ .2 & .3 & .5 \end{bmatrix}$$

   (c) If Alex studies on Monday, what is the chance that she is partying on Friday? (Don't do the math, just write out the expression that you would use to find it.)

Solution: If P is the matrix above, then it is $[1, 0, 0] \cdot P^4$

(d) What percentage of her time should Alex expect to use looking for housing?

Solution: Solve the following system of equations: (first step equations)

$$\begin{aligned} S &= .3S + .6P + .2H \\ P &= .5S + .1P + .3H \\ H &= .2S + .3P + .5H \\ S + P + H &= 1 \end{aligned}$$

(e) If Alex parties on Monday, what is the chance of Alex partying again before studying?

Solution: Set up the following equations:

$$\begin{aligned} H1 &= 0 \\ H2 &= .6(H1) + .1(1) + .3(H3) \\ H3 &= .2(H1) + .3(1) + .5(H3) \end{aligned}$$

Solving for $H2$, we get 0.28