Section 11 Conditional Mean and Variance

TA: Yasi Zhang

May 3, 2022

1 Covariance

The covariance between X and Y is

$$Cov(X,Y) = E(X-EX)(Y-EY)$$

Covariance is a bi-linear operator:

$$\begin{aligned} Cov(X,Y) &= Cov(Y,X) \\ Cov(aX,Y) &= aCov(X,Y) \\ Cov(aX+bY,Z) &= aCov(X,Z)+bCov(Y,Z) \\ Cov(c,X) &= 0 \text{ , where c is a constant} \end{aligned}$$

$\int$ is a linear operator.

E is a linear operator.

$< \cdot, \cdot >$, the inner product of two vectors is a bi-linear operator.

Remark 1 *In most cases, we use $Cov(X,Y) = EXY - EXEY$ to find $Cov(X,Y)$ instead of using the definition.*

1.1 Correlation

$$\rho_{XY} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

Remark 2 $\rho_{XY} =: Cor(X,Y)$

Remark 3 *Pay attention to the name: Covariance v.s. Correlation*

1.2 Correlation and Linearity

Correlation can only capture linear relationship! Think about the following four cases:

1. $Y = aX + b$
2. $Y = aX + b + \epsilon$, $\epsilon \perp\!\!\!\perp X$
3. $Y = X^2$ when $X \sim N(0, 1)$, $\rho = 0$
4. $Y = X^2 - X$ when $X \sim N(0, 1)$, $\rho < 0$

Remark 4 *$\rho_{XY} = 1$ or -1 $\iff$ Deterministic linear relationship between X and Y*
Proof: Cauchy-Schwarz inequality achieve equality if and only if there exists a linear relationship.

1.3 Take-Home Practice

Let X and Y be two jointly continuous random variables with joint PDF

$$f_{XY}(x,y) = \begin{cases} 2 & y+x \leq 1, x>0, y>0 \\ 0 & \text{otherwise} \end{cases}$$

Find $\text{Cov}(X,Y)$ and $\rho(X,Y)$.

Remark 5 *Given joint distribution, derive EX, EY, EX^2, EY^2, EXY, then derive $VarX$, $VarY$, $Cov(X,Y)$*

For $0 \leq x \leq 1$, we have

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{XY}(x,y)dy \\ &= \int_0^{1-x} 2dy \\ &= 2(1-x). \end{aligned}$$

Thus,

$$f_X(x) = \begin{cases} 2(1-x) & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Similarly, we obtain

$$f_Y(y) = \begin{cases} 2(1-y) & 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Thus, we have

$$\begin{aligned} EX &= \int_0^1 2x(1-x)dx \\ &= \frac{1}{3} = EY, \\ EX^2 &= \int_0^1 2x^2(1-x)dx \\ &= \frac{1}{6} = EY^2. \end{aligned}$$

Thus,

$$Var(X) = Var(Y) = \frac{1}{18}.$$

We also have

$$\begin{aligned} EXY &= \int_0^1 \int_0^{1-x} 2xydydx \\ &= \int_0^1 x(1-x)^2dx \\ &= \frac{1}{12}. \end{aligned}$$

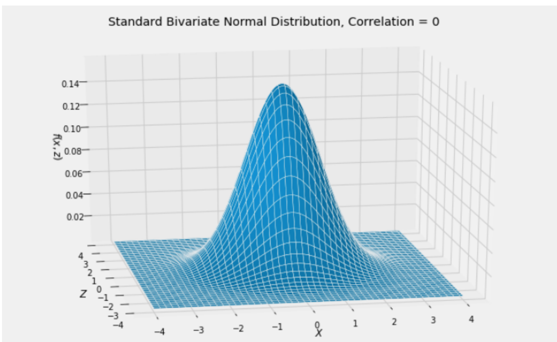
Now, we can find $\text{Cov}(X,Y)$ and $\rho(X,Y)$:

$$\begin{aligned} \text{Cov}(X,Y) &= EXY - EXEY \\ &= \frac{1}{12} - \left(\frac{1}{3}\right)^2 \\ &= -\frac{1}{36}, \\ \rho(X,Y) &= \frac{\text{Cov}(X,Y)}{\sqrt{Var(X)Var(Y)}} \\ &= -\frac{1}{2}. \end{aligned}$$

2 Bivariate Normal Distribution

If $(X, Y) \sim BN(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$, then the probability density function of the vector is

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]\right)$$



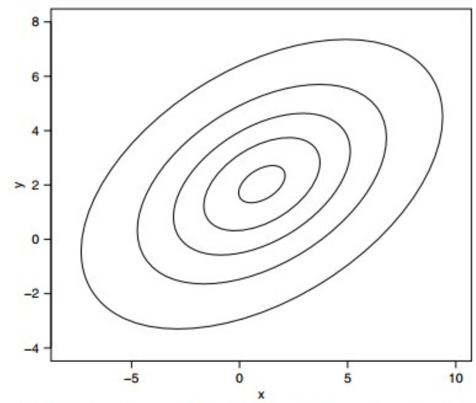


Figure 2.15 A bivariate normal distribution with parameters $\theta_X = 1$, $\theta_Y = 2$, $\sigma_X = 3$, $\sigma_Y = 2$, $\rho = 0.5$, with expanding ellipses enclosing 5%, 25%, 50%, 75% and 95% of the probability distribution.

The conditional distribution of X given Y:

$$X \mid Y = y \;\sim\; \mathcal{N}\left(\mu_X + \rho\sigma_X\frac{y-\mu_Y}{\sigma_Y},\, (1-\rho^2)\sigma_X^2\right).$$

Remark 6 *Memorize the pdf of bivariate normal distribution.*

Remark 7 *How do we memorize the conditional distribution?*
$(1-\rho^2)\sigma_X^2$*: Variance Reduction*
$\frac{y-\mu_Y}{\sigma_Y}$ *represents how far y is away from its mean. Then, how far y is away from its mean has an effect on the conditional mean of X.*

2.1 Practice

Let X and Y be jointly normal random variables with parameters $\mu_X = 1, \sigma_X^2 = 1, \mu_Y = 0, \sigma_Y^2 = 4, \rho = \frac{1}{2}$.
1. Find $Cov(X+Y, 2X-Y)$.
2. Find $P(Y > 1|X = 2)$. Your answer could be an integral.
Hint: We have known that $Y|X = 2$ follows normal distribution. Find the $E(Y|X = 2)$ and $Var(Y|X = 2)$.

1. Let X and Y be jointly normal random variables with parameters $\mu_X = 1, \sigma_X^2 = 1, \mu_Y = 0, \sigma_Y^2 = 4, \rho = \frac{1}{2}$.
1. Find $Cov(X+Y, 2X-Y)$.
2. Find $P(Y > 1|X = 2)$. Your answer could be an integral.
Hint: We have known that $Y|X = 2$ follows normal distribution. Find the $E(Y|X = 2)$ and $Var(Y|X = 2)$.

b. Note that $\mathrm{Cov}(X, Y) = \sigma_X \sigma_Y \rho(X, Y) = 1$. We have

$$\begin{aligned}\mathrm{Cov}(X+Y, 2X-Y) &= 2\mathrm{Cov}(X, X) - \mathrm{Cov}(X, Y) + 2\mathrm{Cov}(Y, X) - \mathrm{Cov}(Y, Y)\\ &= 2 - 1 + 2 - 4 = -1.\end{aligned}$$

c. Using Theorem 5.4, we conclude that given $X = 2$, Y is normally distributed with

$$\begin{aligned}E[Y|X=2] &= \mu_Y + \rho\sigma_Y \frac{2-\mu_X}{\sigma_X} = 1\\ Var(Y|X=x) &= (1-\rho^2)\sigma_Y^2 = 3.\end{aligned}$$

Thus

$$P(Y > 1|X = 2) = 1 - \Phi\left(\frac{1-1}{\sqrt{3}}\right) = \frac{1}{2}.$$

3 Law of Mean and Variance

3.1 Law of Iterated Expectation

If X and Y are random variables on the same probability space, and the mean of X is finite, then

$$E(X) = E(E(X|Y))$$

- $E(X|Y)$: Within-Group Mean
- $E(E(X|Y))$: Within-Group Mean's Mean

Remark:
$E(XY) = E(E(XY|X)) = E(XE(Y|X))$
$E(XY) = E(E(XY|Y)) = E(YE(X|Y))$
$E(g(X)h(Y)) = E(E(g(X)h(Y)|X)) = E(g(X)E(h(Y)|X))$
$E(g(X)h(Y)) = E(E(g(X)h(Y)|Y)) = E(h(Y)E(g(X)|Y))$

3.2 Law of Total Variance

If X and Y are random variables on the same probability space, and the variance of X is finite, then

$$\mathrm{Var}(X) = \mathrm{E}[\mathrm{Var}(X \mid Y)] + \mathrm{Var}(\mathrm{E}[X \mid Y]).$$

Proof:

$$\begin{aligned}
\mathrm{Var}(X) &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 \\
&= \mathbb{E}(\mathbb{E}(X^2|Y)) - \mathbb{E}(\mathbb{E}(X|Y))^2 \\
&= \mathbb{E}(\mathrm{Var}(X|Y) + \mathbb{E}(X|Y)^2) - \mathbb{E}(\mathbb{E}(X|Y))^2 \\
&= \mathbb{E}(\mathrm{Var}(X|Y)) + (\mathbb{E}(\mathbb{E}(X|Y)^2) - \mathbb{E}(\mathbb{E}(X|Y))^2) \\
&= \mathbb{E}(\mathrm{Var}(X|Y)) + \mathrm{Var}(\mathbb{E}(X|Y))
\end{aligned}$$

- $Var(E(X|Y))$: Between-Group Variance
- $E(Var(X|Y))$: Within-Group Variance('s mean)
- $VarX$: Total Variance
- Total Variance = Between-Group Variance + Within-Group Variance

3.3 Practice

Suppose $X \sim U(0,1)$, and $Y|X \sim U(0,X)$.
Find EY, $VarY$.
Hint: If $X \sim U(a,b)$, its mean is $\frac{a+b}{2}$ and its variance is $\frac{1}{12}(b-a)^2$

$$EY = E(E(Y|X)) = E(\frac{X}{2}) = \frac{1}{4}$$

$$\begin{aligned}
VarY &= Var(E(Y|X)) + E(Var(Y|X)) \\
&= Var(\frac{X}{2}) + E(\frac{1}{12}X^2) \\
&= \frac{1}{4}VarX + \frac{1}{12}(VarX + (EX)^2) \\
&= \frac{1}{4} \times \frac{1}{12} + \frac{1}{12}(\frac{1}{12} + \frac{1}{2^2})
\end{aligned}$$

Remark 8 *Take-Home practice (which a very good practice): Find the distribution of Y and check whether its mean and variance are the same as above.*

4 Model

4.1 Find a Number

If we want to find a single number to represent n points, i.e. $\{x_i\}_{i=1}^n$, which number should we pick?

1. Mean
2. Median
3. Mode

Given $\{x_i\}_{i=1}^n$, find

$$\arg\min_x \sum_{i=1}^n (x_i - x)^2 \quad \rightarrow \ Mean$$

$$\arg\min_x \sum_{i=1}^n |x_i - x| \quad \rightarrow \ Median$$

$$\arg\max_x \sum_{i=1}^n I(x_i = x) \quad \rightarrow \ Mode$$

Remark 9 *You can derive the results above by letting $f'(x) = 0$. Think about what is the derivative of $|\cdot|$.*

4.2 Find a Function

If we want to find a function between X and Y, which function should we pick?

We may pick $E(Y|X)$, because it minimizes $MSE = E(Y - g(X))^2$

We may also pick $Median(Y|X)$ because it minimizes $MAE = E|Y - g(X)|$.

We may also pick $Mode(Y|X)$ because it maximizes $EI(Y = g(X))$.

However, $E(Y|X)$, $Median(Y|X)$ and $Mode(Y|X)$ seem too ugly...

Solution: Fix the type of g(X), like linear, polynomial, convex, etc, and then find the best one in this type.

Linear Model:

Assume

$$Y = a + bX + \epsilon.$$

Then, find the best a and b.

ARCH (Time Series Model):

Assume

$$R_t = \mu + \epsilon_t \sqrt{h_t}$$
$$h_t = \alpha + \beta R_{t-1}^2$$

Equivalently,

$$R_t = \mu + \epsilon_t \sqrt{\alpha + \beta R_{t-1}^2}$$

Then, find the best μ, α, and β.

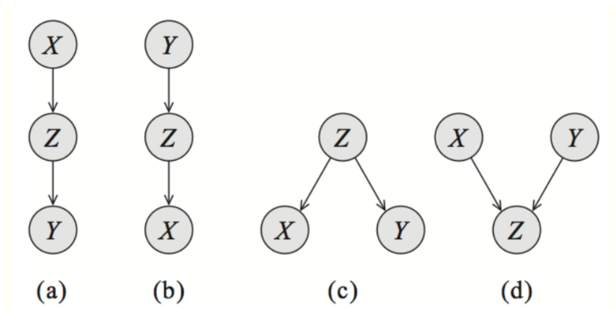
5 Causal Relationship (If you are interested)

A causal relationship exists when one variable in a data set has a direct influence on another variable. Thus, one event triggers the occurrence of another event. A causal relationship is also referred to as cause and effect.

Remark:
Causality => Correlation
Correlation $\not=>$ Causality

Remark 10 *Think about whether the models mentioned above (e.g. linear model, ARCH, etc.) are able to capture the causal relationship. Or they can only capture correlation?*

[**Bayesian Network**] Consider the relationship between X and Y in the next four cases (where '$\rightarrow$' means 'cause'):



Remark:
1. X causes Y
2. Y causes X
3. X and Y are correlated but do not have a causal relationship.
4. X and Y are independent.

An example of (c):
Z = Time spent on reviewing; X = Level of anxiety; Y = Grade.

People are likely to think about that X and Y have a causal relationship! But they do not! In economy, if you find two things correlated, you cannot conclude that they have a cause-and-effect relationship! **Economists make a great contribution because they find out the hidden variable Z!**

Detailed Explanation on the relationship in the four cases:

- *Common parent.* If G is of the form $X \leftarrow Z \rightarrow Y$, and Z is observed, then $X \perp Y \mid Z$. However, if Z is unobserved, then $X \not\perp Y$. Intuitively this stems from the fact that Z contains all the information that determines the outcomes of X and Y; once it is observed, there is nothing else that affects these variables' outcomes.
- *Cascade*: If G equals $X \rightarrow Z \rightarrow Y$, and Z is again observed, then, again $X \perp Y \mid Z$. However, if Z is unobserved, then $X \not\perp Y$. Here, the intuition is again that Z holds all the information that determines the outcome of Y; thus, it does not matter what value X takes.
- *V-structure* (also known as *explaining away*): If G is $X \rightarrow Z \leftarrow Y$, then knowing Z couples X and Y. In other words, $X \perp Y$ if Z is unobserved, but $X \not\perp Y \mid Z$ if Z is observed.