

# **STA 130A: Mathematical Statistics: Brief Course**

## **Lecture 15: Covariance, Correlation, and Total Variance**

Dogyoon Song

Fall 2026, UC Davis

# Announcements

---

**Homework 4 is due tomorrow (Thu, Oct 29, 11:59 PM PT)**

**Extended office hours:** Wednesdays, 1–3 PM (MSB 1143)

- The 1–2 PM hour is shared with STA 35C, whose students may receive priority

# Agenda

---

## **Last time:** Derived distributions

- Find the distribution of  $Z = g(X)$  using its CDF
- Use the inverse-derivative formula for smooth monotone transformations

## **Today:** Describe dependence and variability without deriving a full distribution

- Covariance and variance of sums
- Correlation
- Conditional variance
- Law of total variance

## Recap: Derived distributions

---

Given a random variable  $X$  with known distribution, define  $Z = g(X)$ . Our goal is to derive the support and distribution of  $Z$ .

### General method: CDF first

1. Identify the support of  $Z = g(X)$ .
2. Rewrite the CDF in terms of  $X$ :

$$F_Z(z) = P(Z \leq z) = P(g(X) \leq z).$$

3. If  $Z$  has a PDF, differentiate where valid:

$$f_Z(z) = F'_Z(z),$$

and set  $f_Z(z) = 0$  outside its support.

### Shortcut: smooth monotone $g$

If  $g$  is strictly monotone on the support of  $X$ , with inverse

$$x = h(z) = g^{-1}(z),$$

then, on the transformed support,

$$f_Z(z) = f_X(h(z)) |h'(z)|.$$

Evaluate the original density at the corresponding input, then adjust for stretching or compression. Use CDF-first when  $g$  is not one-to-one.

## Recap: Affine transformations

For

$$Z = aX + b, \quad a \neq 0,$$

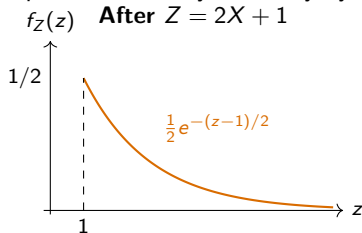
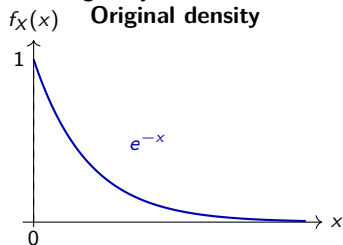
the inverse map is  $x = (z - b)/a$ , so

$$f_Z(z) = \frac{1}{|a|} f_X\left(\frac{z - b}{a}\right).$$

**Example:** Let  $X \sim \text{Exponential}(1)$  and  $Z = 2X + 1$ . Then

$$f_X(x) = e^{-x}, \quad x \geq 0, \quad f_Z(z) = \frac{1}{2} e^{-(z-1)/2}, \quad z \geq 1.$$

**Effect:** shift right by 1, stretch horizontally by 2, and compress the density vertically by  $1/2$ .



## Recall: Variance

---

The variance of a random variable  $X$  is

$$\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.$$

- $\mathbb{E}[X]$  measures the center of the distribution
- $\text{Var}(X)$  measures the spread around the center
- $\text{Var}(aX + b) = a^2\text{Var}(X)$

### Example (Why dependence matters)

Let  $X \sim \text{Bernoulli}(1/2)$ .

- If  $Y \sim \text{Bernoulli}(1/2)$ , independent of  $X$ , then  $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) = \frac{1}{2}$ .
- If we set  $Y = X$ . Then  $\text{Var}(X + Y) = \text{Var}(2X) = 4 \left(\frac{1}{4}\right) = 1$ .  
The same fluctuations are counted twice, so the variances do not simply add.

**Question:** How do we account for the way two variables vary together?

# Covariance

---

## Definition (Covariance)

For random variables  $X$  and  $Y$ , the **covariance** between  $X$  and  $Y$  is

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

Expanding the definition gives the useful computational formula

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

Interpretation:

- $\text{Cov}(X, Y) > 0$ :  $X$  and  $Y$  tend to move together
- $\text{Cov}(X, Y) < 0$ : one tends to be high when the other is low
- $\text{Cov}(X, Y) = 0$ : no linear co-movement, but not necessarily independence

## Example: Covariance from a small joint table

### Example (Illustrative morning/afternoon rain model)

Let  $X$  and  $Y$  indicate morning and afternoon rain, respectively.

| $p_{X,Y}(x,y)$ | $y = 0$ | $y = 1$ | $p_X(x)$ |
|----------------|---------|---------|----------|
| $x = 0$        | 0.4     | 0.1     | 0.5      |
| $x = 1$        | 0.1     | 0.4     | 0.5      |
| $p_Y(y)$       | 0.5     | 0.5     | 1        |

Since  $XY = 1$  exactly when both periods have rain,

$$\mathbb{E}[X] = \mathbb{E}[Y] = 0.5, \quad \mathbb{E}[XY] = 0.4,$$

$$\text{Cov}(X, Y) = 0.4 - (0.5)(0.5) = 0.15.$$

**Interpretation:** In this model,  $P(Y = 1 \mid X = 1) = 0.4/0.5 = 0.8 > 0.5 = P(Y = 1)$ .

## Self-study: A tunable covariance

### Example

Let  $p \in [0, 1]$ . Let  $X$  and  $Y$  be independent random variables such that

$$X = \begin{cases} 1 & \text{with probability } 1/2, \\ -1 & \text{with probability } 1/2, \end{cases} \quad \text{and} \quad Y = \begin{cases} 1 & \text{with probability } p, \\ -1 & \text{with probability } 1 - p, \end{cases}$$

Define  $Z = XY$ . Observe that

$$\mathbb{E}[X] = 0, \quad \mathbb{E}[Z] = \mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y] = 0, \quad XZ = X(XY) = X^2Y = Y.$$

Thus,

$$\text{Cov}(X, Z) = \mathbb{E}[XZ] - \mathbb{E}[X]\mathbb{E}[Z] = \mathbb{E}[Y] = 2p - 1.$$

and therefore

$$\text{Cov}(X, Z) \begin{cases} > 0, & p > 1/2, \\ = 0, & p = 1/2, \\ < 0, & p < 1/2. \end{cases}$$

When  $p > 1/2$ ,  $Z = XY$  tends to have the same sign as  $X$ ; when  $p < 1/2$ , the opposite.

# Properties of covariance

---

For any random variables  $X, Y, Z$  and any scalars  $a, b$ ,

- $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
- $\text{Cov}(X, X) = \text{Var}(X)$
- $\text{Cov}(X, aY + b) = a\text{Cov}(X, Y)$
- $\text{Cov}(X, Y + Z) = \text{Cov}(X, Y) + \text{Cov}(X, Z)$

The variance of a sum of random variables satisfies

$$\text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{1 \leq i < j \leq n} \text{Cov}(X_i, X_j).$$

- When  $n = 2$ ,

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

- Unlike expectations, variances add only when the covariance terms vanish

## Example: Variance of a sum and a difference

### Example

Suppose

$$\text{Var}(X) = 4, \quad \text{Var}(Y) = 9, \quad \text{Cov}(X, Y) = -3.$$

**Question:** Compute  $\text{Var}(X + Y)$  and  $\text{Var}(X - Y)$ .

Using  $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$  and  $\text{Cov}(X, -Y) = -\text{Cov}(X, Y)$ , we get

$$\text{Var}(X + Y) = 4 + 9 + 2(-3) = 7,$$

$$\begin{aligned} \text{Var}(X - Y) &= \text{Var}(X) + \text{Var}(Y) - 2\text{Cov}(X, Y) \\ &= 4 + 9 - 2(-3) = 19. \end{aligned}$$

**Interpretation:** Negative covariance reduces the variability of  $X + Y$ , but increases the variability of  $X - Y$ .

## Independence and zero covariance

---

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

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y], \quad \text{and therefore} \quad \text{Cov}(X, Y) = 0.$$

However, the converse is false:

$$\text{Cov}(X, Y) = 0 \not\Rightarrow X, Y \text{ independent.}$$

### Example

Let  $X$  be uniform on  $\{-1, 0, 1\}$ , and let

$$Y = X^2.$$

Then  $Y$  is determined by  $X$ , so  $X$  and  $Y$  are not independent.

Nevertheless,

$$\mathbb{E}[X] = 0, \quad \mathbb{E}[XY] = \mathbb{E}[X^3] = 0, \quad \text{and hence,} \quad \text{Cov}(X, Y) = 0.$$

## Pop-up quiz: Covariance

---

Suppose

$$\text{Var}(X) = 4, \quad \text{Var}(Y) = 9, \quad \text{Cov}(X, Y) = -3.$$

**Question:** What is  $\text{Var}(2X - Y)$ ?

- A) 7
- B) 19
- C) 25
- D) 37

**Answer:** D.

$$\begin{aligned}\text{Var}(2X - Y) &= 4\text{Var}(X) + \text{Var}(Y) + 2(2)(-1)\text{Cov}(X, Y) \\ &= 4(4) + 9 - 4(-3) = 37.\end{aligned}$$

**Follow-up:** What sign of  $\text{Cov}(X, Y)$  reduces  $\text{Var}(X + Y)$  and  $\text{Var}(X - Y)$  respectively?

# Correlation coefficient

## Definition

When  $\text{Var}(X) > 0$  and  $\text{Var}(Y) > 0$ , the **correlation coefficient** between  $X$  and  $Y$  is

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \text{Var}(Y)}}.$$

Covariance depends on the units of  $X$  and  $Y$ , but the correlation coefficient normalizes.

- When  $\text{Var}(X) > 0$  and  $\text{Var}(Y) > 0$ ,

$$-1 \leq \rho(X, Y) \leq 1.$$

Interpretation:

- $\rho > 0$ : positive linear association
- $\rho < 0$ : negative linear association
- $\rho = 0$ : uncorrelated, but not necessarily independent
- $|\rho| = 1$ : exact linear relationship

## Example: Correlation under an affine relationship

### Example

Let  $a, b \in \mathbb{R}$  with  $a \neq 0$ , and suppose  $\text{Var}(X) > 0$  and

$$Y = aX + b.$$

**Question:** Compute the correlation coefficient between  $X$  and  $Y$ .

By definition,

$$\text{Cov}(X, Y) = \text{Cov}(X, aX + b) = a\text{Var}(X),$$

$$\text{Var}(Y) = a^2\text{Var}(X).$$

Therefore,

$$\rho(X, Y) = \frac{a\text{Var}(X)}{\sqrt{\text{Var}(X)} a^2\text{Var}(X)} = \frac{a}{|a|}. \quad \text{and thus,} \quad \rho(X, Y) = \begin{cases} 1, & a > 0, \\ -1, & a < 0. \end{cases}$$

## Pop-up quiz: Correlation

---

Suppose  $\rho(X, Y) = 0.6$ , and define

$$U = 3X + 1, \quad V = -2Y + 5.$$

**Question:** What is  $\rho(U, V)$ ?

- A) 0.6
- B) -0.6
- C) -1.2
- D) Cannot be determined

**Answer: B.**

Adding constants does not change correlation. Multiplying one variable by a negative constant flips the sign:

$$\rho(3X + 1, -2Y + 5) = -\rho(X, Y) = -0.6.$$

**Follow-up:** What if  $V = 2Y + 5$  instead?

# Motivation: Conditional uncertainty

---

Variance measures uncertainty:

$$\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2].$$

But if we observe another random variable  $Y$ , our uncertainty about  $X$  may change.

- $\mathbb{E}[X]$ : unconditional center of  $X$
- $\mathbb{E}[X | Y]$ : center of  $X$  after observing  $Y$
- $\text{Var}(X)$ : unconditional uncertainty
- $\text{Var}(X | Y)$ : remaining uncertainty after observing  $Y$

**Goal:** decompose total uncertainty in  $X$  into what remains after conditioning and what is explained by  $Y$ .

## Conditional variance

---

Recall that  $\mathbb{E}[X \mid Y]$  is a random variable encoding the mean of  $X$  after  $Y$  is observed.

Similarly, conditional variance measures the remaining spread of  $X$  after  $Y$  is observed.

### Definition

The **conditional variance** of  $X$  given  $Y = y$  is

$$\text{Var}(X \mid Y = y) = \mathbb{E} \left[ (X - \mathbb{E}[X \mid Y = y])^2 \mid Y = y \right].$$

Equivalently,

$$\text{Var}(X \mid Y = y) = \mathbb{E}[X^2 \mid Y = y] - (\mathbb{E}[X \mid Y = y])^2.$$

The conditional variance  $\text{Var}(X \mid Y)$  is a random variable

- When  $Y = y$ , its value is  $\text{Var}(X \mid Y = y)$ .

## Example: Conditional variance in a two-group model

### Example

Let  $Y \sim \text{Bernoulli}(1/2)$ . Suppose

$$X \mid Y = 0 \sim N(0, 1), \quad X \mid Y = 1 \sim N(2, 4).$$

**Question:** Compute conditional mean and conditional variance for  $Y = 0$  and  $Y = 1$ .

If  $Y = 0$ , then

$$\mathbb{E}[X \mid Y = 0] = 0, \quad \text{Var}(X \mid Y = 0) = 1.$$

If  $Y = 1$ , then

$$\mathbb{E}[X \mid Y = 1] = 2, \quad \text{Var}(X \mid Y = 1) = 4.$$

Observe that

$$\mathbb{E}[X \mid Y] = 2Y \quad \text{and} \quad \text{Var}(X \mid Y) = 1 + 3Y.$$

**Interpretation:** observing  $Y$  reveals which group mean and within-group variance apply.

**Note:**  $\mathbb{E}[X \mid Y]$  and  $\text{Var}(X \mid Y)$  are random variables as they depend on random label  $Y$

# Law of total variance

---

## Theorem (Law of total variance)

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

- $\mathbb{E}[\text{Var}(X \mid Y)]$ : average uncertainty remaining after observing  $Y$ .
- $\text{Var}(\mathbb{E}[X \mid Y])$ : variability explained by how the conditional mean changes with  $Y$ .

Interpretation:

total variability = average within-condition variability + between-condition variability.

This is the variance analogue of the law of total expectation:

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X \mid Y]].$$

## Self-study: Proof of the law of total variance

### Theorem (Law of total variance)

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

**Proof.** Let  $m(Y) = \mathbb{E}[X \mid Y]$ . Then

$$\text{Var}(X \mid Y) = \mathbb{E}[X^2 \mid Y] - m(Y)^2.$$

Taking expectations and using the law of total expectation,

$$\mathbb{E}[\text{Var}(X \mid Y)] = \mathbb{E}[\mathbb{E}[X^2 \mid Y]] - \mathbb{E}[m(Y)^2] = \mathbb{E}[X^2] - \mathbb{E}[m(Y)^2]. \quad (1)$$

Also, since  $\mathbb{E}[m(Y)] = \mathbb{E}[X]$  by total expectation,

$$\text{Var}(\mathbb{E}[X \mid Y]) = \text{Var}(m(Y)) = \mathbb{E}[m(Y)^2] - (\mathbb{E}[m(Y)])^2 = \mathbb{E}[m(Y)^2] - (\mathbb{E}[X])^2. \quad (2)$$

Adding the two equations (1) and (2) gives

$$\mathbb{E}[\text{Var}(X \mid Y)] + \text{Var}(\mathbb{E}[X \mid Y]) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \text{Var}(X).$$

## Example: Total variance in the two-group model

### Example

Recall the two-group model:

$$Y \sim \text{Bernoulli}(1/2), \quad X \mid Y = 0 \sim N(0, 1), \quad X \mid Y = 1 \sim N(2, 4).$$

Thus,

$$\mathbb{E}[X \mid Y] = 2Y, \quad \text{Var}(X \mid Y) = 1 + 3Y.$$

Since  $Y \sim \text{Bernoulli}(1/2)$ ,

$$\mathbb{E}[\text{Var}(X \mid Y)] = \mathbb{E}[1 + 3Y] = 1 + 3 \cdot \frac{1}{2} = \frac{5}{2}.$$

Also,

$$\text{Var}(\mathbb{E}[X \mid Y]) = \text{Var}(2Y) = 4\text{Var}(Y) = 4 \cdot \frac{1}{4} = 1.$$

Therefore, by the law of total variance,

$$\text{Var}(X) = \frac{5}{2} + 1 = \frac{7}{2}.$$

## What “conditioning reduces uncertainty” means

---

The law of total variance implies

$$\mathbb{E}[\text{Var}(X \mid Y)] \leq \text{Var}(X).$$

Thus conditioning reduces uncertainty **on average**.

However, it need not reduce variance for every conditioning value.

- In the two-group example,

$$\text{Var}(X \mid Y = 1) = 4 > \frac{7}{2} = \text{Var}(X).$$

**Takeaway:** Information can identify a particularly variable subgroup even though the average remaining uncertainty is smaller.

## Self-study: Resolved vs. remaining variability in estimation

---

**Inverse-problem setting:** Suppose  $X$  is an unknown quantity of interest, and  $Y$  is an observed measurement/data. Before observing  $Y$ , uncertainty about  $X$  is measured by  $\text{Var}(X)$ .

**Estimation of  $X$  given  $Y$ :** After observing  $Y$ , a natural prediction of  $X$  is the conditional mean

$$\hat{X} = \mathbb{E}[X \mid Y].$$

Define the residual  $R = X - \hat{X}$ . Then

$$\mathbb{E}[R \mid Y] = \mathbb{E}[X - \mathbb{E}[X \mid Y] \mid Y] = 0.$$

Thus, after using the information in  $Y$ , the residual has conditional mean zero.

Moreover, since  $\mathbb{E}[R] = 0$ ,  $\text{Var}(R) = \mathbb{E}[R^2]$ . Therefore,

$$\text{Var}(\hat{X}) = \text{Var}(\mathbb{E}[X \mid Y]) \quad \text{and} \quad \text{Var}(R) = \mathbb{E}[R^2] = \mathbb{E}[\text{Var}(X \mid Y)].$$

Therefore, the law of total variance can be read as

$$\text{Var}(X) = \underbrace{\text{Var}(\hat{X})}_{\text{uncertainty resolved/explained by } Y} + \underbrace{\text{Var}(R)}_{\text{average posterior uncertainty left after } Y}.$$

**Message:** Observing  $Y$  changes the conditional mean  $\mathbb{E}[X \mid Y]$ . The variability of these conditional means is the explained part; the residual term is the average uncertainty left after observing  $Y$ .

# Self-study: Explained vs. residual variability in regression

---

**Regression setting:** Suppose now that

$$Y = \text{response}, \quad X = \text{predictor}.$$

The conditional mean

$$m(X) = \mathbb{E}[Y \mid X]$$

is the regression function: it gives the mean response after seeing the predictor value.

The law of total variance becomes

$$\text{Var}(Y) = \underbrace{\text{Var}(m(X))}_{\text{variation explained by the predictor}} + \underbrace{\mathbb{E}[\text{Var}(Y \mid X)]}_{\text{average within-predictor noise}}.$$

**Interpretation:**

- If  $X$  is informative, then  $m(X)$  changes substantially with  $X$ , so  $\text{Var}(m(X))$  is large.
- If responses are tightly concentrated around  $m(X)$ , then  $\mathbb{E}[\text{Var}(Y \mid X)]$  is small.

This is the probability-theory version of the familiar idea behind regression:

$$\text{total variation} = \text{explained variation} + \text{residual variation}.$$

## Pop-up quiz: Total variance

---

Suppose that

$$\text{Var}(X \mid Y) = 2 \quad \text{for every value of } Y, \quad \mathbb{E}[X \mid Y] = 3Y, \quad \text{Var}(Y) = 4.$$

**Question:** What is  $\text{Var}(X)$ ?

- A) 2
- B) 6
- C) 14
- D) 38

**Answer: D.**

By the law of total variance,

$$\text{Var}(X) = \mathbb{E}[\text{Var}(X \mid Y)] + \text{Var}(\mathbb{E}[X \mid Y]) = 2 + \text{Var}(3Y) = 2 + 9 \cdot 4 = 38.$$

**Follow-up:** In

$$\text{Var}(X) = \mathbb{E}[\text{Var}(X \mid Y)] + \text{Var}(\mathbb{E}[X \mid Y]),$$

which term becomes small if observing  $Y$  makes  $X$  almost deterministic?

## Wrap-up

---

### Covariance and correlation

$$\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y], \quad \rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}.$$

- $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$
- Independence implies zero covariance; the converse can fail

### Conditional and total variance

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

- Total variability = average within-condition variability + between-condition variability

**Suggested reading:** [BT08, Sec. 4.2–4.3]

**Next lecture:** Moment generating functions and independent sums.

# References

---



Dimitri Bertsekas and John N Tsitsiklis.

*Introduction to probability*, volume 1.

Athena Scientific, 2nd edition, 2008.