

STA 130A: Mathematical Statistics: Brief Course

Lecture 16: Conditional Variance and Total Variance

Dogyoon Song

Fall 2026, UC Davis

Announcements

Homework 5 is posted

- Review the problems early and ask questions as needed.
- Submit a single PDF on time and follow the submission instructions.

Mid-course survey (Due: Sun, Nov 1)

- Please take about 10 minutes to complete the [survey on Canvas](#).
- Please share what is helping you learn, where you are struggling, and specific changes that would help.
- Concrete and detailed feedback and constructive suggestions/requests will be appreciated.

Agenda

Last time: Covariance and correlation

- Covariance describes how two variables vary together
- Correlation removes the measurement units

Today: How much variability remains after conditioning?

- Complete the affine-correlation example
- Conditional variance: variability within a group
- Total variance: within-group and between-group variability

Recap: Covariance and correlation

Throughout this lecture, assume finite second moments.

Covariance: average product of deviations from the means

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

Correlation: unit-free covariance, when both variances are positive

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}, \quad -1 \leq \rho(X, Y) \leq 1.$$

- The sign describes the direction of linear association.
- Independence implies zero covariance; the converse can fail.

Useful rules for the next example:

$$\text{Cov}(X, aX + b) = a\text{Var}(X), \quad \text{Var}(aX + b) = a^2\text{Var}(X).$$

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:

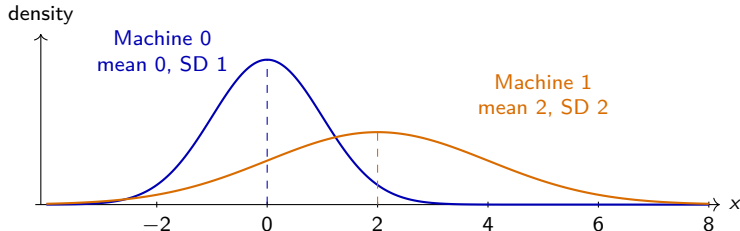
$$\rho(3X + 1, -2Y + 5) = -\rho(X, Y) = -0.6.$$

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

Motivation: Two machines, two sources of variability

Suppose that half the packages come from each of two filling machines.

Let $Y \in \{0, 1\}$ label the machine and let X be the actual weight minus the target, in grams (negative means underfilled).



- **Within a machine:** packages vary around that machine's mean.
- **Between machines:** the two mean errors differ.

Question: How much variability remains when we know the machine?

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.

Conditional variance measures the spread in that conditional distribution.

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 variance = average within-group variance + variance of group means.

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

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

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

Therefore, by the law of total variance,

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

Takeaway: Averaging conditional variances alone misses variation in the group means.

What “conditioning reduces uncertainty” means

The law of total variance implies

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

Thus conditioning cannot increase variance **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 random quantity of interest, and Y is an observed measurement. 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

$$\text{Var}(X) = \underbrace{\text{Var}(\hat{X})}_{\text{uncertainty resolved/explained by } Y} + \underbrace{\text{Var}(R)}_{\text{average remaining conditional variance}}.$$

Message: Observing Y may change 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.
(The explained term is large when conditional means vary substantially across predictor values.)
- 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

Conditional variance: spread around the conditional mean

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

- For fixed y , this is a number; $\text{Var}(X \mid Y)$ varies with Y .

Total variance: two sources of variability

$$\text{Var}(X) = \underbrace{\mathbb{E}[\text{Var}(X \mid Y)]}_{\text{average within-group variance}} + \underbrace{\text{Var}(\mathbb{E}[X \mid Y])}_{\text{variance of group means}}.$$

- Weight each group by its probability, not automatically equally.
- Conditioning cannot increase variance on average, but an individual conditional variance can be larger.

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.