Joint and Conditional Distributions

Gov 2001: Quantitative Social Science Methods I

Scott Cunningham

Harvard University

Spring 2026

Today's Reading

Required

- **Aronow & Miller**, §1.3: Joint, marginal, conditional distributions (pp. 31–44)
- **Aronow & Miller**, §2.2.1–2.2.2: Covariance, correlation, independence (pp. 59–65)
- **Blackwell**, Ch. 1: Setting the stage for regression

Blackwell's Chapter 1 introduces the CEF as the target of regression—this lecture builds the foundations.

Galton's Deeper Question

From "regression to the mean" to joint distributions

Last week we met **Galton** and "regression to the mean."

But Galton faced a harder problem: *How do I mathematically describe the relationship between father's height and son's height?*

His solution required inventing new machinery:

- A way to describe **two variables together** (joint distributions)
- A way to ask **"given X, what's Y?"** (conditional distributions)
- A number summarizing **how strongly they're related** (correlation)

Today we build that machinery. By the end, you'll have the tools Galton invented to study heredity—and that political scientists use to study everything from voting to war.

Why Do Political Scientists Care?

We always observe multiple variables together

Every interesting question involves relationships:

- Education and party identification
- Income and voter turnout
- War duration and casualties
- Campaign spending and vote share

Single-variable summaries ($\mathbb{E}[X]$, $\text{Var}(X)$) miss the *relationship*.

The joint distribution captures everything: How do X and Y move together? If I know X, what does that tell me about Y?

Today's goal: Learn to extract relationships from joint distributions.

The Big Picture

So far: Single random variables

- Distribution: $f(x)$
- Summary: $\mathbb{E}[X]$, $\text{Var}(X)$

Now: Two (or more) random variables together

- How are they *jointly* distributed?
- If we know X, what does that tell us about Y?

This is where regression begins.
Regression asks: What's $\mathbb{E}[Y|X]$? To answer that, we need conditional distributions.

Joint Distribution: Discrete Case

Joint Probability Mass Function

For discrete random variables X and Y, the **joint PMF** is:

$$f(x, y) = \Pr(X = x \text{ and } Y = y)$$

Properties:

- $f(x, y) \geq 0$ for all x, y
- $\sum_x \sum_y f(x, y) = 1$

The joint PMF tells us the probability of every (x, y) combination.

Example: Education and Party ID

Survey data: Joint distribution of Education (X) and Party (Y)

	Party (Y)			
Education (X)	Dem	Ind	Rep	Row Total
No College	0.20	0.15	0.15	0.50
College	0.18	0.12	0.20	0.50
Col Total	0.38	0.27	0.35	1.00

Reading the table: $f(\text{No College}, \text{Dem}) = 0.20$
This means: 20% of the population has no college and identifies as Democrat.

Marginal Distributions

Question: What if we only care about X (ignoring Y)?

Marginal PMF

The **marginal distribution** of X is obtained by summing over Y:

$$f_X(x) = \sum_y f(x, y) = \Pr(X = x)$$

From our example:

- $f_X(\text{No College}) = 0.20 + 0.15 + 0.15 = 0.50$
- $f_X(\text{College}) = 0.18 + 0.12 + 0.20 = 0.50$

"Marginal" because these appear in the margins of the table.

Visualizing Marginalization

	Dem	Ind	Rep	$\to f_X$
No Col	0.20	0.15	0.15	**0.50**
College	0.18	0.12	0.20	**0.50**

Sum across rows $\longrightarrow$ marginal distribution of X

Conditional Distribution

Key question: Given that we *know* $X = x$, what's the distribution of Y?

Conditional PMF

The **conditional distribution** of Y given $X = x$ is:

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{\Pr(X = x, Y = y)}{\Pr(X = x)}$$

Intuition:

- Zoom in on the row where $X = x$
- Renormalize so probabilities sum to 1

This is the definitional formula for conditional distributions.

Example: Party ID Given Education

What's the distribution of Party among college graduates?

We need $f_{Y|X}(y|\text{College})$ for each party:

$$f_{Y|X}(\text{Dem}|\text{College}) = \frac{f(\text{College}, \text{Dem})}{f_X(\text{College})} = \frac{0.18}{0.50} = 0.36$$

$$f_{Y|X}(\text{Ind}|\text{College}) = \frac{0.12}{0.50} = 0.24$$

$$f_{Y|X}(\text{Rep}|\text{College}) = \frac{0.20}{0.50} = 0.40$$

Check: $0.36 + 0.24 + 0.40 = 1.00$ ✓

Among college grads: 36% Dem, 24% Ind, 40% Rep

Comparing Conditional Distributions

	Dem	Ind	Rep
$f_{Y\|X}(y\|\text{No College})$	0.40	0.30	0.30
$f_{Y\|X}(y\|\text{College})$	0.36	0.24	0.40

What do we learn?

- The two conditional distributions are *different*
- Knowing education level changes our beliefs about party ID
- The conditional distribution of Y *depends on* X

This dependence is what regression studies.

(These are stylized numbers for illustration, not real survey data.)

Conditional PDF: Interpretation

Two ways to think about joint distributions

Probability interpretation:

$$\Pr(a < Y < b \mid X = x) = \int_a^b f_{Y|X}(y|x)\, dy$$

Factorization of the joint PDF:

$$f_{X,Y}(x, y) = f_{Y|X}(y|x) \cdot f_X(x)$$

Joint = Conditional × Marginal

To sample (X, Y): First draw $X \sim f_X$, then draw $Y \sim f_{Y|X}(\cdot|X)$.

Symmetrically: $f_{X,Y}(x, y) = f_{X|Y}(x|y) \cdot f_Y(y)$

Independence of Random Variables

When does knowing X tell us nothing about Y?

Definition: Independence

X and Y are **independent**, written $X \perp\!\!\!\perp Y$, if:

$$f(x, y) = f_X(x) \cdot f_Y(y) \quad \text{for all } x, y$$

Equivalent conditions:

- $f_{Y|X}(y|x) = f_Y(y)$ for all x (conditioning doesn't change anything)
- $\Pr(X = x, Y = y) = \Pr(X = x) \cdot \Pr(Y = y)$

In our education/party example, X and Y are NOT independent—the conditional distributions differ.

Joint CDF: From Discrete to Continuous

CDF before PDF, just like the univariate case

Joint CDF

$$F(x, y) = \Pr(X \leq x,\ Y \leq y)$$

Joint PDF as partial derivative (same logic as univariate):

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x\, \partial y}$$

Why start with CDF?

- CDF always exists (even when PDF doesn't)
- PDF = rate of change of CDF in both directions
- Mirrors univariate: $F(x) \to f(x) = F'(x)$

Joint Distribution: Continuous Case

Joint Probability Density Function

For continuous X and Y, the **joint PDF** $f(x, y)$ satisfies:

$$\Pr(X \in A, Y \in B) = \iint_{A \times B} f(x, y)\, dx\, dy$$

Properties:

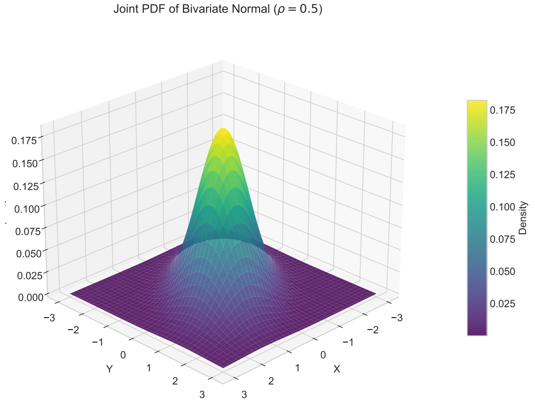
- $f(x, y) \geq 0$
- $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dx\, dy = 1$

Marginal: $f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy$

Conditional: $f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)}$

Visualizing the Joint PDF: A 3D Surface

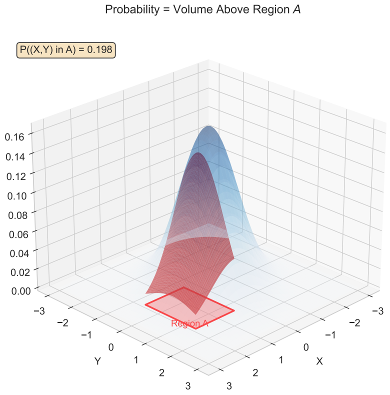
Height = density at each (x, y) point



Key insight: The joint PDF is a *surface* over the (x, y) plane. Higher = more likely.

Probability = Volume Under the Surface

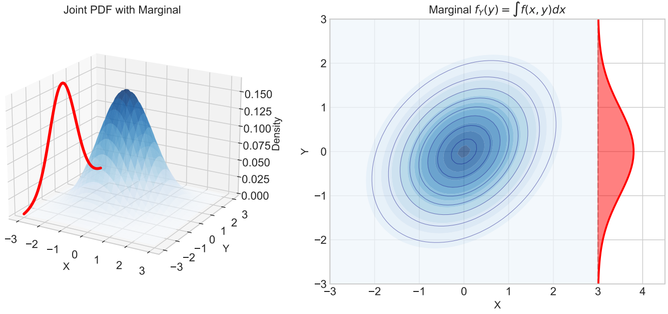
Connecting geometry to integrals



$$\Pr((X, Y) \in A) = \iint_A f(x, y)\, dx\, dy = \textbf{volume above region } A$$

Marginal Distribution = "Projection"

Collapse the surface onto one axis

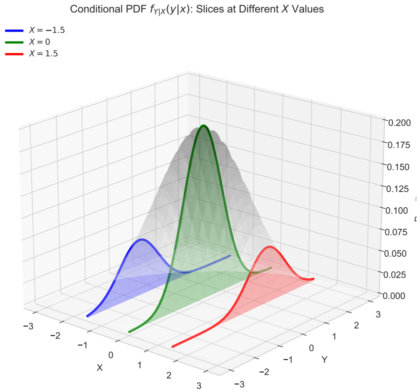


$f_Y(y) = \int f(x, y)\, dx$ — integrate out X, "project" onto the Y-axis.

Same idea as summing across rows in the discrete case.

Conditional Distribution = "Slice"

Fix $X = x$ and look at the cross-section



$f_{Y|X}(y|x)$ = the slice at $X = x$, renormalized to integrate to 1.

Key insight: The slice changes as x changes. This is what regression studies!

Multivariate Expectation and 2D LOTUS

Computing expectations of functions of two random variables

Expected value of a function of (X, Y):

$$\mathbb{E}[g(X, Y)] = \iint g(x, y) f(x, y)\, dx\, dy$$

2D LOTUS (Law of the Unconscious Statistician):

- No need to find the distribution of $g(X, Y)$
- Integrate $g(x, y)$ directly against the joint PDF

Key applications:

- $\mathbb{E}[XY]$ — needed for covariance (coming next!)
- $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$ — linearity still holds

Why Does Dependence Matter?

Independence assumptions are everywhere in statistics

We constantly assume independence:

- Poll responses assumed independent
- RCT: treatment assignment $\perp\!\!\!\perp$ background characteristics
- Regression errors assumed independent across observations

Lack of independence is a blessing or a curse:

- **Blessing**: Two variables not independent $\Rightarrow$ potentially interesting relationship
- **Curse**: In observational studies, treatment is usually not independent of background $\Rightarrow$ confounding

Question: How do we *measure* dependence?

Covariance: Measuring Dependence

How often do high values of X occur with high values of Y?

Definition: Covariance

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

Equivalent formula (often easier to compute):

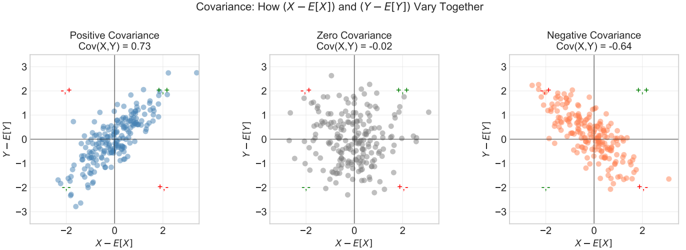
$$\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]$$

Key property: If $X \perp\!\!\!\perp Y$, then $\text{Cov}(X, Y) = 0$

But the converse is NOT true! Zero covariance does not imply independence.

Covariance: Quadrant Intuition

The sign depends on where points cluster



- **Positive Cov**: Points cluster in $(+,+)$ and $(-,-)$ quadrants
- **Negative Cov**: Points cluster in $(+,-)$ and $(-,+)$ quadrants
- **Zero Cov**: Balanced across all quadrants

Properties of Covariance

1. $\text{Cov}(X, X) = \text{Var}(X)$ (variance is covariance with itself!)
2. $\text{Cov}(X, Y) = \text{Cov}(Y, X)$ (symmetric)
3. $\text{Cov}(X, c) = 0$ for any constant c
4. $\text{Cov}(aX, Y) = a \cdot \text{Cov}(X, Y)$
5. $\text{Cov}(X + Y, Z) = \text{Cov}(X, Z) + \text{Cov}(Y, Z)$

Important Consequence

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

Unlike expected values, variances are NOT linear—covariance is the "correction term."

Correlation: Scale-Free Dependence

Covariance depends on units; correlation doesn't

Problem: Covariance depends on the scale of X and Y.

Definition: Correlation

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

Equivalent form:

$$\rho(X, Y) = \text{Cov}\left(\frac{X - \mathbb{E}[X]}{\text{SD}(X)}, \frac{Y - \mathbb{E}[Y]}{\text{SD}(Y)}\right)$$

Correlation = covariance of standardized variables

Correlation: Properties

Key properties:

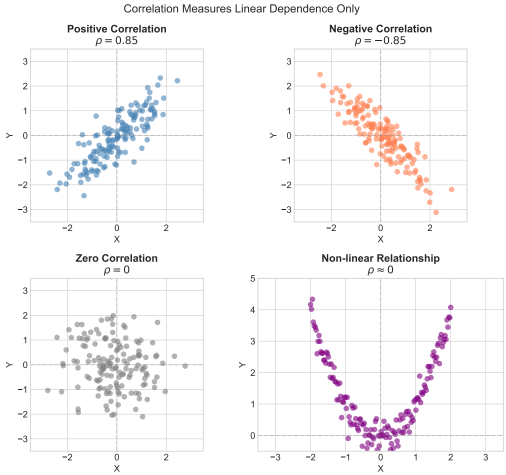
- $-1 \leq \rho(X, Y) \leq 1$
- $|\rho(X, Y)| = 1$ if and only if $Y = a + bX$ for some constants
  - $\rho = 1$: perfect positive linear relationship
  - $\rho = -1$: perfect negative linear relationship
- $\rho = 0$: no *linear* relationship

Critical caveat: Correlation measures **linear** dependence only.

Two variables can have $\rho = 0$ but still be strongly dependent! (e.g., $Y = X^2$ where X is symmetric around 0)

Correlation: Examples

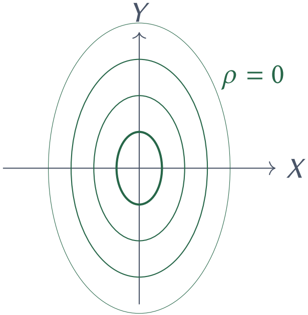
What different ρ values look like

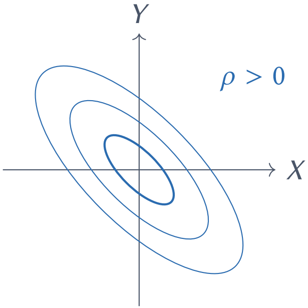


Example: Bivariate Normal

The most important continuous joint distribution

$(X, Y) \sim N(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$





Key fact: For bivariate normal, the conditional distribution $Y|X = x$ is also normal.

Conditional Distribution: Bivariate Normal

If (X, Y) is bivariate normal, then:

Conditional Distribution

$$Y|X = x \sim N\left(\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X),\ \sigma_Y^2(1 - \rho^2)\right)$$

Key observations:

- Conditional mean is **linear in** x: $\mathbb{E}[Y|X = x] = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X)$
- Conditional variance is **constant** (doesn't depend on x)
- If $\rho = 0$, conditional mean = μ_Y (knowing X tells us nothing)

This is regression! The conditional mean is a line through (x, y) space.

Key Takeaways

1. **Joint $\rightarrow$ Marginal**: Sum/integrate out what you don't care about
2. **Joint $\rightarrow$ Conditional**: Divide by marginal to "zoom in" on a slice
3. **The CEF** $\mathbb{E}[Y|X = x]$ summarizes the conditional distribution
4. **Covariance**: Measures how X and Y move together
5. **Correlation**: Scale-free measure of *linear* dependence

The big idea: Regression is about finding $\mathbb{E}[Y|X]$ from data.

Next Time: Conditional Expectation and LIE

Key topics for next lecture:

- The Conditional Expectation Function (CEF): $\mathbb{E}[Y|X = x]$
- Law of Iterated Expectations (LIE): $\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y|X]]$
- Why the CEF is the "best predictor" of Y given X
- How regression connects to the CEF

Reading:

- A&M §2.2.3–2.2.4 (CEF, LIE, best predictor property)
- Blackwell Ch. 1 (continue)

PS3 was due yesterday (Feb 17). Hope it went well!