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)
- **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.

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: $f(\text{No College, Dem}) = 0.20$
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	→ f_X
No Col	0.20	0.15	0.15	**0.50**
College	0.18	0.12	0.20	**0.50**

Sum across rows ⟶ 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.)

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 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:

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

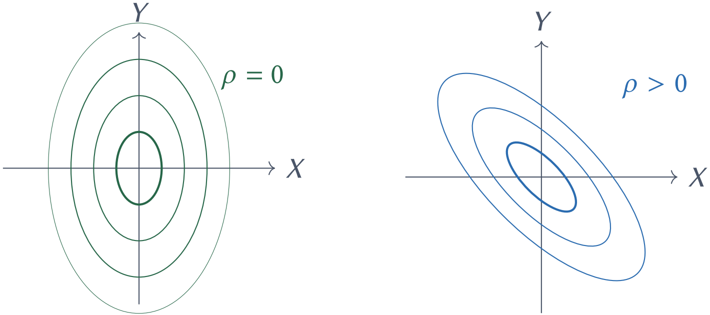
Marginal: $f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy$

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

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.

Preview: The Conditional Expectation Function

The conditional mean $\mathbb{E}[Y|X = x]$ is a function of x.

Conditional Expectation Function (CEF)

$$G_Y(x) = \mathbb{E}[Y|X = x]$$

This is the function that maps each value of x to the expected value of Y given $X = x$.

For bivariate normal: $G_Y(x)$ is linear
In general: $G_Y(x)$ can be any shape!

Regression = Finding a good approximation to $G_Y(x)$
Wednesday: Why the CEF is the best predictor, and what regression is really doing.

Blackwell's Framing (Ch. 1)

Blackwell starts here: We observe (X_i, Y_i) pairs.

The fundamental question: How does Y depend on X?

The answer: The conditional distribution $f_{Y|X}(y|x)$

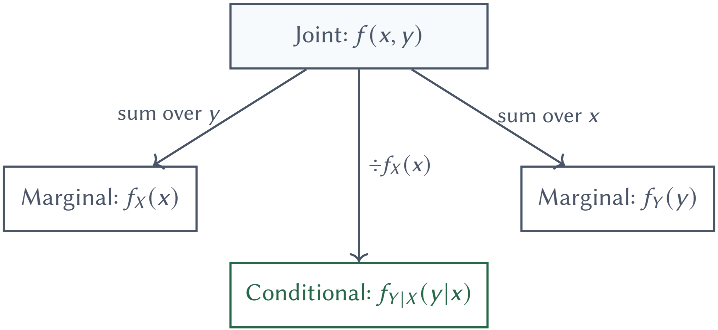
The summary: The conditional expectation $\mathbb{E}[Y|X = x]$

Blackwell's Key Insight

"Regression is about predicting Y from X. The best possible prediction—if we knew the joint distribution—would be $\mathbb{E}[Y|X]$."

We don't know the joint distribution. We have data. That's where statistics comes in.

The Hierarchy of Distributions



The conditional distribution is where the action is for regression.

The Core Formula

Conditional Distribution

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

In words: Take the joint distribution, zoom in on the slice where $X = x$, and renormalize.

This formula is the key to regression.

Everything else (marginals, independence) follows from this.

Key Takeaways

1. **Joint** → **Marginal**: Sum/integrate out what you don't care about
2. **Joint** → **Conditional**: Divide by marginal to "zoom in" on a slice
3. **The CEF** $\mathbb{E}[Y|X = x]$ summarizes the conditional distribution

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

Wednesday: Why the CEF is the best predictor.

For Wednesday

Topic: The Conditional Expectation Function

Reading:

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

The CEF is the most important concept in the probability unit.