POLI 273: Causal Inference

Professor Umberto Mignozzetti

Lecture 02 | Some Background

Announcements

- Syllabus in on. All the readings are on now.
- PS 01 is also on. Make sure you have access to Canvas and to Gradescope.
- And let me know if you do not!
- GitHub page: https://github.com/umbertomig/POLI273

Background

Background

- This is going to be very repetitive for you now.
- But it is important to build a solid foundation. And as Leo and I were talking in the previous lecture, repetitive is not necessarily bad.
  - Ok, maybe boring...
- This background part: Probability Theory, Distributions, and Random Samples
- Mostly based on Aronow and Miller (2019)
  - A great book that we have free access here at the UCSD!

Probability Theory

- Important to establish some basic understanding of our research fundamentals.
- Helps us to formalize the quantities of interest and their uncertainty in precise terms.
- **Probability**: “Random generative process that selects an outcome among multiple possible outcomes.”

Fundamentals of Probability Theory

1. (Definition) **Sample space**: Denoted by Ω. Containing individual outcomes, denoted as $\omega \in \Omega$.

- Polity IV scores: $\Omega = [-10, 10]$
- Income: $\Omega = \mathbb{R}^+$
- GDP growth: $\Omega = \mathbb{R}$ (but not really...)
- Deaths in a civil conflict: $\Omega = \mathbb{N} \cup 0$
- And so on.
- One simple case, a coin toss: $\Omega = \{H, T\}$.

Fundamentals of Probability Theory

2. (Definition) **Event space**: Subsets of Ω (that form a σ-Algebra: check this). The set of the subsets (S) have to satisfy the following properties:

- $S \neq \emptyset$
- If $A \in S$, then $A^C \in S$
- For a countable sequence $A_1, A_2, \cdots \in S$, then $\bigcup A_i \in S$
- For a coin toss, let us consider $S = 2^{\Omega}$.

Fundamentals of Probability Theory

- A set with these two properties is called a measurable set.
- And this pretty much means that we can measure stuff in this set.
- We are going to use a special function to measure stuff. Something called a *probability measure*.

Fundamentals of Probability Theory

3. (Definition) **Probability Measure**: A function $\mathbb{P} : S \to [0, 1]$, that we are going to call *probability measure*. $\mathbb{P}$ has to satisfy three axioms (Kolmogorov):

- $\forall A \in S$, $\mathbb{P}(A) \in [0, 1]$.
- $\mathbb{P}(\Omega) = 1$.
- Let the sequence $A_1, A_2, \dots \in S$, such that $A_i \cap A_j = \emptyset$ $\forall i \neq j$. Then, $\mathbb{P}(\bigcup A_i) = \sum \mathbb{P}(A_i)$
- For a coin toss, let us consider $\mathbb{P}(A) = \frac{1}{2}|A|$.
- Lots of consequences can be reached from these simple things: e.g., $\mathbb{P}(\emptyset)$ or $\mathbb{P}(A^C) = 1 - \mathbb{P}(A)$, etc.

Fundamentals of Probability Theory

- We rarely talk about one event. In science, we talk about *how one event is related to another.*
- (Definition) **Joint Probability**: For $A, B \in S$, the joint probability of A and B is $\mathbb{P}(A \cap B)$
- (Theorem) **Addition Rule**: For $A, B \in S$

$$\mathbb{P}(A \cup B) \;=\; \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$$

Fundamentals of Probability Theory

- (Definition) **Conditional Probability**: For $A, B \in S$, the joint probability of A and B is

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

- (Theorem) **Multiplicative Law**: For $A, B \in S$, with $\mathbb{P}(B) > 0$, then $\mathbb{P}(A|B)\mathbb{P}(B) = \mathbb{P}(A \cap B)$.
- (Theorem) **Bayes Rule**: For $A, B \in S$, with $\mathbb{P}(A) > 0$ and $\mathbb{P}(B) > 0$,

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}$$

Fundamentals of Probability Theory

- (Definition) **Partition**: If $\{A_1, A_2, \cdots\}$, with $A_i \in S$, are non-empty and pairwise disjoint sets such that $\Omega = \cup_i A_i$, the set $\{A_1, A_2, \cdots\}$ is called a *partition* of Ω.
- (Theorem) **Law of Total Probability**: If $\{A_1, A_2, \cdots\}$ is a partition of Ω, and $B \in S$, then

$$\mathbb{P}(B) = \sum_i \mathbb{P}(B \cap A_i)$$

And if, $\forall i$, $\mathbb{P}(A_i) > 0$, then

$$\mathbb{P}(B) = \sum_i \mathbb{P}(B|A_i)\mathbb{P}(A_i)$$

Independence

- (Definition) **Independence of Events**: Events $A, B \in S$ are *independent* if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$
- (Theorem) **Conditional Probability and Independence**: For $A, B \in S$, with $\mathbb{P}(B) > 0$, A and B are independent if, and only if, $\mathbb{P}(A|B) = \mathbb{P}(A)$.
- (Definition) **Random Variable**: A *random variable* is a function $X : \Omega \to \mathbb{R}$ such that, $\forall r \in \mathbb{R}$, $\{\omega \in \Omega : X(\omega) \leq r\} \in S$

Random Variable

- A random variable maps each state of the world ($\omega \in \Omega$) to a real number.
  - Example: the event $\{X = 1\}$ means $\{\omega \in \Omega : X(\omega) = 1\}$
- The book differentiates the prob of sets (prob spaces) and probs of random variables and random operators.
  - I will loosely keep using $\mathbb{P}$ for both, but you should understand that this is *poetic license* (see discussions on pages 16–18).

Discrete Random Variables

- (Definition) **Discrete Random Variable**: A *random variable* X is discrete if its range is a countable set.
- (Definition) **Probability Mass Function**: For a discrete random variable X, the *PMF* of X is $f(x) = \mathbb{P}[X = x]$, $\forall x \in \mathbb{R}$
  - If you look carefully, you will now understand why the book uses $\mathbb{P}$ and Pr!
- (Theorem) **Properties of Probability Mass Functions**: For a discrete variable X, with *PMF* f:
  1. $\forall x \in \mathbb{R}$, $f(x) \geq 0$
  2. $\sum_{x \in X(\Omega)} f(x) = 1$
  3. If $D \subseteq \mathbb{R}$, and $A = \{X \in D\}$, then $\mathbb{P}(A) = \mathbb{P}(X \in D) = \sum_{x \in X(A)} f(x)$

Discrete Random Variables

- (Definition) **Cumulative Distribution Function**: For a random variable X, the *CDF* of X is $F(x) = \mathbb{P}[X \leq x]$, $\forall x \in \mathbb{R}$
  - More general way to look at things than look at *PMFs*
- (Theorem) **Properties of CDFs**: For a random variable X, with *CDF* F:
  - $\lim_{x \to -\infty} F(x) = 0$
  - $\lim_{x \to \infty} F(x) = 1$
  - $\forall x \in \mathbb{R}$, $1 - F(x) = \mathbb{P}[x > x]$

Continuous Random Variables

- (Definition) **Continuous Random Variables and CDFs**: A random variable X is continuous if $\exists$ a function $F : \mathbb{R} \to [0, 1]$ such that the *CDF* of X is $F(x) = \mathbb{P}[X \leq x] = \int_{-\infty}^{x} f(u) du, \forall x \in \mathbb{R}$.
- (Definition) **Probability Density Function**: For a continuous random variable X with *CDF* $F(x)$, the *PDF* is defined as $f(x) = \left. \frac{dF(u)}{du} \right|_{u=x}, \forall x \in \mathbb{R}$.

Continuous Random Variables

- (Theorem) **Properties of Continuos Random Variables**: For a continuous random variable X with *CDF* $F(x)$ and *PDF* $f(x)$:
  1. $\forall x \in \mathbb{R}$, $f(x) \geq 0$
  2. $\int_{-\infty}^{\infty} f(x)dx = 1$
  3. $\forall x \in \mathbb{R}$, $\mathbb{P}[X = x] = 0$
  4. $\forall x \in \mathbb{R}$,
     $\mathbb{P}[X > x] = \mathbb{P}[X \geq x] = 1 - F(x) = \int_{x}^{\infty} f(u)du$
  5. $\forall a, b \in \mathbb{R}$, with $a \leq b$, $\mathbb{P}[a \leq X \leq b] = F(b) - F(a)$

Support

- (Definition) **Support of a Random Variable**: For a random variable X with *PMF/PDF* f, the *support* of X is $\text{Supp} = \{x \in \mathbb{R} : f(x) > 0\}$
    - Sounds out of nowhere, but it is actually very important for Conditional Expectations!

Bivariate Relationships

- Most of what matters for us is about more than just one Random Variable.
- The same way we care about differents sets of events, we care even more about bivariate relations
  - Or how to Random Variables relate to each other.
- Luckly, we can extend the primitives we formulated so far to account for bivariate relationships.

Bivariate Relationships

- (Definition) **Equality of Random Variables**: Let X and Y random variables. $X = Y$ if, $\forall \omega in \Omega$, $X(\omega) = Y(\omega)$.
- (Theorem) **Equality of Functions of Random Variables**: Let X a random variable and f and g functions of X. Then $f(X) = g(X) \iff \forall x \in X(\Omega)$, $g(x) = h(x)$.

Discrete Bivariate Relationships

- (Definition) **Joint Probability Mass Function**: For discrete random variables X and Y, the joint *PMF* of X and Y is $f(x, y) = \mathbb{P}[X = x, Y = y]$.
- (Definition) **Joint Cumulative Distribution Function**: For discrete random variables X and Y, the joint *CDF* of X and Y is $F(x, y) = \mathbb{P}[X \leq x, Y \leq y]$.

Discrete Bivariate Relationships

- (Theorem) **Marginal Probability Mass Function**: For discrete random variables X and Y with joint *PMF* f, the *marginal PMF* of Y is ($\forall y \in \mathbb{R}$):

$$f_Y(y) \;=\; \mathbb{P}[Y = y] \;=\; \sum_{x \in \mathrm{Supp[X]}} f(x, y)$$

.

- (Definition) **Conditional Probability Mass Function**: For discrete random variables X and Y with joint *PMF* f, the *conditional PMF* of Y given $X = x$ is ($\forall y \in \mathbb{R}$ and $\forall x \in \mathrm{Supp}[X]$):

$$f_{Y|X}(y|x) \;=\; \mathbb{P}[Y = y, X = x] \;=\; \frac{f(x, y)}{f_X(x)}$$

Discrete Bivariate Relationships

- (Theorem) **Multiplicative Law for Probability Mass Functions**: For discrete random variables X and Y with joint *PMF* f. Then, $\forall x \in \mathbb{R}$ and $\forall y \in \text{Supp}[Y]$, $f_{X|Y}(x|y)f_Y(y) \ = f(x,y)$.
- Joint distributions are essential for Causal Inference, and data analysis in general.
- Now, let's look at the continuous case.

Continuous Bivariate Relationships

- (Definition) **Jointly Continuous Random Variables**: Two random variables X and Y are jointly continuous if there exists a function $f : \mathbb{R}^2 \to [0, 1]$ such that the joint *CDF* of X and Y is

$$F(x, y) = \mathbb{P}[X \leq x, Y \leq y] = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v) dv du$$

- (Definition) **Joint Probability Distribution Function**: For two jointly continuous random variables X and Y, with joint *CDF* F, the joint *PDF* is

$$f(x, y) = \left. \frac{\partial^2 F(u, v)}{\partial u \partial v} \right|_{u=x, v=y}$$

Continuous Bivariate Relationships

- (Theorem) **Marginal Probability Distribution Function**: For two jointly continuous random variables X and Y, with joint PDF f, the *marginal PDF* of Y is

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)dx$$

- (Definition) **Conditional Probability Distribution Function**: For two jointly continuous random variables X and Y, with joint PDF f, the *conditional PDF* of Y is

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

Independence

- (Theorem) **Multiplicative Law of Probability Distribution Functions**: For two jointly continuous random variables X and Y, with joint *PDF* f. Then, $\forall x \in \mathbb{R}$ and $y \in \text{Supp[Y]}$, $f_{X|Y}(y|x)f_Y(y) \;=\; f(x,y)$
- (Definition) **Independence of Random Variables**: Let X and Y two random variables with joint *PDF/PMF* f. We say X and Y are independent if, for all $x, y \in \mathbb{R}$,

$$f(x,y) = f_X(x)f_Y(y)$$

And we write $X \perp\!\!\!\perp Y$ to denote that the random variables X and Y are independent.

Independence

- (Theorem) **Implications of Independence**: Let X and Y two random variables with joint *PDF*/*PMF* f. The following statements are equivalent:

1. $X \perp\!\!\!\perp Y$
2. $\forall x, y \in \mathbb{R}$, $f(x, y) = f_X(x) f_Y(y)$
3. $\forall x \mathbb{R}$ and $y \in \text{Supp}[Y]$, $f_{X|Y}(x|y) = f_X(x)$
4. $\forall D, E \subseteq \mathbb{R}$, the events $\{X \in D\}$ and $\{Y \in E\}$ are independent.
5. For all functions g of X and h of Y $g(X) \perp\!\!\!\perp h(Y)$.

Multivariate Generalizations

- (Definition) **Random Vectors**: A random vector is a function $X : \Omega \to \mathbb{R}^K$ such that, $\forall \omega \in \Omega$, $\mathbf{X}(\omega) = \Big(X_1(\omega), \cdots, X_K(\omega)\Big)$. And where X_i is a random variable.
- (Definition) **Joint Cumulative Density Function**: For a random vector $\mathbf{X}$, evaluated at $\mathbf{x}$, is denoted $F(\mathbf{x}) = \mathbb{P}[\mathbf{X} \leq \mathbf{x}]$.

Multivariate Generalizations

And from here, the extensions are evident:

- For continuous random vectors, we use multiple integrals (*CDF*s) or partial derivatives (*PDF*s)
- For discrete random vectors, we use multiple sums (*CDF*s).
- For the conditional *PDFs*/*PMFs*, we marginalize on the variables by summing up across their support.

Summarizing Distributions

Summarizing Distributions

- Statistics would not be very useful if we did not find ways to summarize the objects we work with.
- (Definition) **Expected Value**: For a random variable X with bounded variation:
    1. If X is discrete, then $\mathbb{E}[X] = \sum_x x f(x)$
    - If X is continuous, then $\mathbb{E}[X] = \int_{-\infty}^{\infty} x f(x)$

Summarizing Distributions

- (Theorem) **Expectation of a Function**: For a random variable X with *PMF/PDF* f and a function g (still assuming bounded variation):
  1. If X is discrete, then $\mathbb{E}[g(X)] = \sum_x g(x)f(x)$
  - If X is continuous, then $\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)$

Summarizing Distributions

- (Theorem) **Linearity of Expected Values**: For a random variable X, and $a, b \in \mathbb{R}$, then:

$$\mathbb{E}[aX + b] \;=\; a\mathbb{E}[X] + b$$

- (Definition) **Expectation of a Bivariate Random Vector**: For a random vector (X, Y), the expected value is $\mathbb{E}[(X, Y)] \;=\; \Big(\mathbb{E}[X], \mathbb{E}[Y]\Big)$

Summarizing Distributions

- (Definition) **Raw Moment**: For a random variable X, the j-th *raw moment* is defined as:

$$\mu_j' = \mathbb{E}[X^j]$$

- (Definition) **Central Moment**: For a random variable X, the j-th *central moment* is defined as:

$$\mu_j = \mathbb{E}\left[(X - \mathbb{E}[X])^j\right]$$

Summarizing Distributions

- (Definition) **Variance**: For a random variable X, the variance is defined as the *second central moment* of X:

$$\mathbb{V}[X] = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]$$

- (Definition) **Standard Deviation**: For a random variable X, the variance is defined as $\sigma[X] = \sqrt{\mathbb{V}[X]}$
- (Theorem) **Alternative formula for the Variance**: For a random variable X, the variance is defined as the *second central moment* of X:

$$\mathbb{V}[X] = \mathbb{E}[X^2] - [\mathbb{E}[X]]^2$$

Summarizing Distributions

- (Theorem) **Algebra of the Variance**: For a random variable X and $a \in \mathbb{R}$, $b \in \mathbb{R}$:

$$\mathbb{V}[aX + b] \;=\; a^2\mathbb{V}[X]$$

- (Corollary) **Algebra of the Standard Deviations**: Let a random variable X, $a \in \mathbb{R}$, and $b \in \mathbb{R}$. Then, $\sigma[aX + b] = |a|\sigma[X]$.

- (Theorem) **Chebyshev Inequality**: Let a random variable X and $\sigma[X] > 0$. Then, $\forall \epsilon > 0$:

$$\mathbb{P}\Big[\big|X - \mathbb{E}[X]\big| \geq \epsilon\sigma[X]\Big] \leq \frac{1}{\epsilon^2}$$

Summarizing Distributions

- (Definition) **Normal Distribution**: A continuous random variable X follows a normal distribution (denoted by $X \sim N(\mu, \sigma^2)$) if:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- (Theorem) **Mean and Standard Deviation of a Normal Distribution**: Let $X \sim N(\mu, \sigma^2)$. Then, $\mathbb{E}[X] = \mu$ and $\sigma[X] = \sigma$

Summarizing Distributions

- (Theorems) **Algebra of Normal Distributions**: Let $X \sim N(\mu_X, \sigma_X^2)$, $Y \sim N(\mu_Y, \sigma_Y^2)$, and $a, b \in \mathbb{R}$, $a \neq 0$. Then:
  1. If $W = aX + b$, then $W \sim N(a\mu_X + b, a^2\sigma_X^2)$
  2. If $X \perp\!\!\!\perp Y$, and $Z = X + Y$, then $Z \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$
- This proof should be straightforward.

Summarizing Distributions

- And one of the most important objects: suppose we want to say how well a random variable X approximates a given value $c \in \mathbb{R}$.
- The most common metric we use for this purpose is the *Mean Squared Error*.
- (Definition) **Mean Squared Error**: The *MSE* of a random variable X about c is equal to: $\mathbb{E}[(X-c)^2]$.

Summarizing Distributions

- (Theorem) **Alternative formulation for the Mean Squared Error**: The *MSE* of a random variable about c is equal to:

$$\mathbb{E}[(X-c)^2] \;=\; \mathbb{V}[X] + (\mathbb{E}[X]-c)^2$$

- (Theorem) **Mean Squared Error Minimization**: The value c that minimizes the *MSE* of a random variable X about c is $c = \mathbb{E}[X]$.

Summarizing Distributions (Joint Distributions)

- (Definition) **Covariance**: The covariance of two random variables is defined as

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

- (Theorem) **Variance Rule**: Let X and Y two r.v. Then,

1. $$\mathbb{V}[X + Y] = \mathbb{V}[X] + 2\text{Cov}[X, Y] + \mathbb{V}[Y]$$
2. And if $a, b, c \in \mathbb{R}$, then

   $$\mathbb{V}[aX + bY + c] = a^2\mathbb{V}[X] + 2ab\text{Cov}[X, Y] + b^2\mathbb{V}[Y]$$

Summarizing Distributions (Joint Distributions)

- (Theorem) **Algebra of Covariance**: Let X, Y, Z, W four r.v.s and $a, b, c, d \in \mathbb{R}$. Then:

1. $\text{Cov}[X, c] = \text{Cov}[c, X] = \text{Cov}[d, c]$
2. $\text{Cov}[X, Y] = \text{Cov}[Y, X]$
3. $\text{Cov}[X, X] = \mathbb{V}[X]$
4. $\text{Cov}[X + W, Y + Z] =$
   $\text{Cov}[X, Z] + \text{Cov}[X, Y] + \text{Cov}[W, Y] + \text{Cov}[W, Z]$

Summarizing Distributions (Joint Distributions)

- (Definition) **Correlation**: The correlation of two random variables X and Y, with $\sigma[X] > 0$ and $\sigma[Y] > 0$ is defined as

$$\rho[X, Y] = \frac{\text{Cov}[X, Y]}{\sigma[X]\sigma[Y]}$$

- (Theorem) **Correlation and Linear Dependence**: Let two random variables X and Y, with $\sigma[X] > 0$ and $\sigma[Y] > 0$. Then:

1. $\rho[X, Y] \in [-1, 1]$
2. For $a, b \in \mathbb{R}$, and $Y = aX + b$:

- $\rho[X, Y] = 1$ if $a > 0$
- $\rho[X, Y] = -1$ if $a < 0$

Summarizing Distributions (Joint Distributions)

- (Theorem) **Properties of Correlation**: Let random variables X, Y, and Z, all with variance higher than zero. Let also $a, b, c, d \in \mathbb{R}$.

1. $\rho[X, Y] = \rho[Y, X]$
2. $\rho[X, X] = 1$
3. If $ab > 0$, then $\rho[aX + c, bY + d] = \rho[X, Y]$
4. If $ab < 0$, then $\rho[aX + c, bY + d] = -\rho[X, Y]$

Independence

- (Theorem) **Independence**: Let X and Y two independents r.v.s. Then:

1. $\rho[X, Y] = 0$
2. $\text{Cov}[X, Y] = 0$
3. $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$
4. $\mathbb{V}[X + Y] = \mathbb{V}[X] + \mathbb{V}[Y]$

- However, these statements are not equivalent, since you may have $\rho[X, Y] = 0$ without having independence!
  - See pages 65 and 66 of the Aronow and Miller book (especially the footnote on pg 66 for a notable exception).

Conditional Expectation Functions

- (Definition) **Conditional Expectation**: For two random variables X and Y, the conditional expectation of Y given $X = x$ is:

1. If X and Y are discrete, and $x \in \text{Supp}[X]$, then

$$\mathbb{E}[Y|X = x] = \sum_{y} y f_{Y|X}(y|x)$$

2. If X and Y are continuous, and $x \in \text{Supp}[X]$, then

$$\mathbb{E}[Y|X = x] = \int_{-\infty}^{\infty} y f_{Y|X}(y|x) dy$$

Conditional Expectation Functions

- (Definition) **Conditional Expectation of a function**: For two random variables X and Y and a function of X and Y, the conditional expectation of $h(X, Y)$ given $X = x$ is:

1. If X and Y are discrete, and $x \in \text{Supp}[X]$, then

$$\mathbb{E}[h(X, Y)|X = x] = \sum_y h(x, y) f_{Y|X}(y|x)$$

2. If X and Y are continuous, and $x \in \text{Supp}[X]$, then

$$\mathbb{E}[h(X, Y)|X = x] = \int_{-\infty}^{\infty} h(x, y) f_{Y|X}(y|x) dy$$

Conditional Expectation Functions

- (Theorem) **Linearity of Conditional Expectation**: Let X and Y rvs. If g and h are functions (with $x \in \text{Supp}[X]$ is:

$$\mathbb{E}[g(X)Y + h(X)|X = x] = g(x)\mathbb{E}[Y|X = x] + h(x)$$

- (Definition) **Conditional Expectation Function**: Let X and Y rvs with joint distribution f ($x \in \text{Supp}[X]$) is:

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

Conditional Expectation Functions

- (Theorem) **Law of Iterated Expectations**: Let X and Y rvs. $\mathbb{E}[Y] = \mathbb{E}\big[\mathbb{E}[Y|X]\big]$
- (Theorem) **Law of Total Variance**: Let X and Y rvs. $\mathbb{V}[Y] = \mathbb{E}\big[\mathbb{V}[Y|X]\big] + \mathbb{V}\big[\mathbb{E}[Y|X]\big]$

Conditional Expectation Functions

- (Theorem) **Properties of Deviations from the CEF**: Let X and Y rvs and let $\epsilon = Y - \mathbb{E}[Y|X]$.

1. $\mathbb{E}[\epsilon|X] = 0$
2. $\mathbb{E}[\epsilon] = 0$
3. If g is a function of X, $\text{Cov}\big[g(X), \epsilon\big] = 0$
4. $\mathbb{V}[\epsilon|X] = \mathbb{V}[Y|X]$
5. $\mathbb{V}[\epsilon] = \mathbb{E}\big[\mathbb{V}[Y|X]\big]$

Conditional Expectation Functions

- (Theorem) **Properties of Deviations from the CEF**: Let X and Y rvs and let $\epsilon = Y - \mathbb{E}[Y|X]$.

1. $\mathbb{E}[\epsilon|X] = 0$
2. $\mathbb{E}[\epsilon] = 0$
3. If g is a function of X, $\text{Cov}\big[g(X), \epsilon\big] = 0$
4. $\mathbb{V}[\epsilon|X] = \mathbb{V}[Y|X]$
5. $\mathbb{V}[\epsilon] = \mathbb{E}\big[\mathbb{V}[Y|X]\big]$

Questions?

See you in the next class