# 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 $\Omega$. 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 $\Omega$ (that form a $\sigma$-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 $A_1, A_2, \cdots \in S$, 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 $A, B \in S$

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

▶ (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 $\mathbb{P}(B) > 0$,

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

▶ (Theorem) **Bayes Rule**: For $A, B \in S$, with $\mathbb{P}(A) > 0$ and $\mathbb{P}(B) > 0$,

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

# 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, $\mathbb{P}(A|B) = \mathbb{P}(A)$.

# Random Variable

▶ (Definition) **Random Variable**: A *random variable* is a function $X : \Omega \to \mathbb{R}$ such that, $\forall r \in \mathbb{R}$, $\{\omega \in \Omega : X(\omega) \leq r\} \in S$

▶ 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 $Pr$!

- (Facts) **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*
- (Facts) **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
$F(x) = \mathbb{P}[X \le x] = \int_{-\infty}^{x} f(u)du$, $\forall x \in \mathbb{R}$.

▶ (Definition) **Probability Density Function**: For a
continuous random variable $X$ with *CDF* $F(x)$, the *PDF*
is defined as $f(x) = \left. \dfrac{dF(u)}{du} \right|_{u=x}$, $\forall x \in \mathbb{R}$.

# Continuous Random Variables

▶ (Facts) **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

▶ (Fact) **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.

▶ (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

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

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

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

► (Fact) **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

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

► (Fact) **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

# Independence

▶ (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

▶ (Facts) **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 **X**, evaluated at **x**, is denoted $F(\mathbf{x}) = \mathbb{P}[\mathbf{X} \leq \mathbf{x}]$.

▶ 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 $\mathbb{E}[X] = \int_{-\infty}^{\infty} x f(x)$

► (Fact) **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

▶ (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

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

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

▶ (Facts) **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

▶ (Facts) **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 $c = \mathbb{E}[X]$.

Questions?

See you in the next class