# ECON 626 Causal Inference

## Problem Set 1

Creator: Junjie Lei

Creation Date: Jan_28/2020

Title: Intro Causality Fundamentals

# 1 Notebook problems

## 1.1 Law of large numbers

1. Where in the code are the parameters µ and $\sigma^2$ set?

   - the `data generating sequence` is set via the `np.random.seed(94115)` → DGP

   - The $\mu$ & $\sigma^2$ is set via the line of code

   ```
   X_n = np.random.normal(loc=0, scale=1, size=(10 ** 3, 1))

   # numpy.random.normaldraws random sample from a normal(Gaussian)
   distribution;
   # loc indicates the --> mean
   # scale --> standard deviation
   ```

2. In the code, what objects correspond to

   - $X_1, X_2$ ...

   - $S_n,$

   - $\dfrac{1}{n}S_n$?

   - Why are these objects in math not exactly the same as in a computer?

   1. the code responding to $X_1, X_2$ .. is

   ```
   X_n = np.random.normal(loc=0, scale=1, size=(10 ** 3, 1))
   ```

   2. the code respoding to $S_n = \sum_{i=1}^{n} X_i$ is

```
partial_sums = X_n.cumsum(axis= 0)
```

3. the code responding to $\frac{1}{n}S_n$ is

```
partial_means = np.divide(partial_sums, n_index)
```

4. the object in math in not the *exactly* the same in a computer;
   - `python` index starts from `0` instead of `1`
   - It is a `parital mean()` and `parital sum()`. so the result is a moving average, but mathematically calculation does not require those; we just sum them and divide by the total number of observations.

3. Explain intuitively why the line is more stable at the end.

   becasue of the **Law of large numbers**; if we calculate the average of a large data set it should be very close to the theoretical mean of the underlying distribution.

## 1.2 Selection bias

```
1 df.groupby('D').mean()
```

|   | Y0 | Y1 | Y | delta | C |
|---|---|---|---|---|---|
| **D** | | | | | |
| **0** | 0.642120 | 0.642120 | 0.642120 | 0.0 | -1.066920 |
| **1** | 0.814983 | 0.814983 | 0.814983 | 0.0 | -0.827018 |

1. calculate the $E[Y^1 - Y^0]$
   - **ATE** $\rightarrow$ the average treatment effect aberaged across the entire population.
   - should be `0`
2. calculate the **NATE**

   $NATE \rightarrow E[Y^1|D=1] - E[Y^0|D=0]$

   hence it is $0.814968 - 0.642790 \rightarrow$ `0.172178`

3. calculate the **Selection bias**
   - **Selection bias** $\rightarrow E[Y^0|treatmentgrp] - E[Y^0|controlgrp]$

     hence it is $0.814968 - 0.642790 \rightarrow$ `0.172178`

4. what line(s) of code **fundamentally** craetes the selection bias? explain.

there are 2 lines of code that I think highly related

```
charitability = np.random.normal(loc=-1, scale=0.5, size=N)
prob_search_dwb = scipy.special.expit(charitability)
```

```
# 1st line of code
charitability = np.random.normal(loc=-1, scale=0.5, size=N)
```

and the following lines of code determines if the individuals search for the DWB ads and receive the `treatment`;

```
prob_search_dwb = scipy.special.expit(charitability)
```

`selction bias` is the systematic difference in $Y^0$ between the treatment and control groups. It occurs when the two groups are systematically different even in the absence of an actual treatment, in this case, it is the `charitibility` of different indiviudals creates the `prob_search_dwb`;

5. calculate the $P(D = 1)$

   use the following code can calculate the percentage of each group;

```
1 count = df.groupby('D').count().reset_index()
2 count['percentage'] =  (count['Y0']/count['Y0'].sum()) * 100
3 count
```

|   | D | Y0 | Y1 | Y | delta | C | percentage |
|---|---|-----|-----|-----|-------|-----|------------|
| 0 | 0 | 720725 | 720725 | 720725 | 720725 | 720725 | 72.0725 |
| 1 | 1 | 279275 | 279275 | 279275 | 279275 | 279275 | 27.9275 |

# 1.3 Data generating process for a null experiment

1. Make a version of this data-generating process that is a true randomized experiment. The new treatment variable should be $D_{exp}$. Make the overall probability of treatment the same, i.e., $P(D = 1) = P(D_{exp} = 1)$. Show that the experiment eliminates **selection bias**. Calculate the **NATE**.

```
# size of the experiment;
N = 10 ** 6
```

```
# inherient charity;
charitability = np.random.normal(loc = -1, scale=0.5, size = N)
```

```
# AD in this case has no effect;
Y0 = np.random.lognormal(mean=(charitability + 0.5), sigma=0.1, size=N)
Y1 = Y0.copy()

# the treatment is the AD
# in this exercise, we are going to make the AD amount the same;
# prob_search_dwb = scipy.special.expit(charitability)
D = np.random.binomial(n = 1, p = 0.5, size = N)

np.unique(D,
          return_counts = True)
```



True Randomized experiment: Density of Y0 by 2 groups
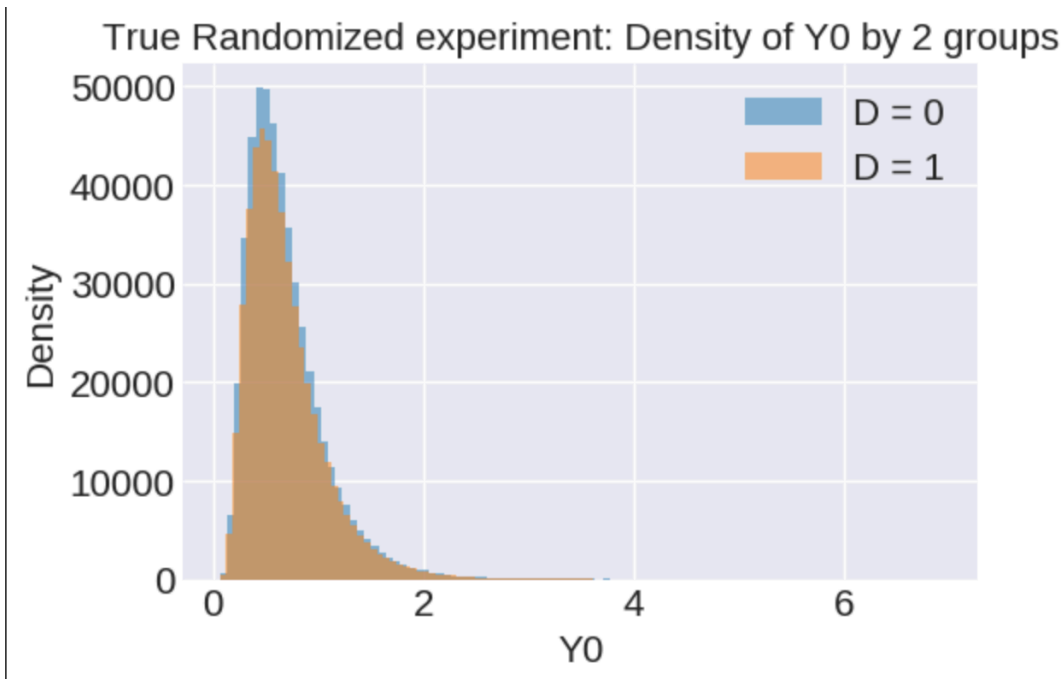
- $NATE \rightarrow E[Y^1|D=1] - E[Y^0|D=0]$

  $= -0.000705$

2. Interpret figure 1 'density of $Y_0$ by treatment group' w.r.t the independece assumption:

   $(Y^1, Y^0 \perp D)$

   even when `selection bias` is eliminated $\rightarrow$ (the potential outcomes of donation without seeing the ads are jointly indepedent of the treatment assignment.)

   the treatment & control groups bahave differently even if there is no ads

3.

True Randomized experiment: Density of Y0 by 2 groups


Figure 1: Density of Y0 by treatment group
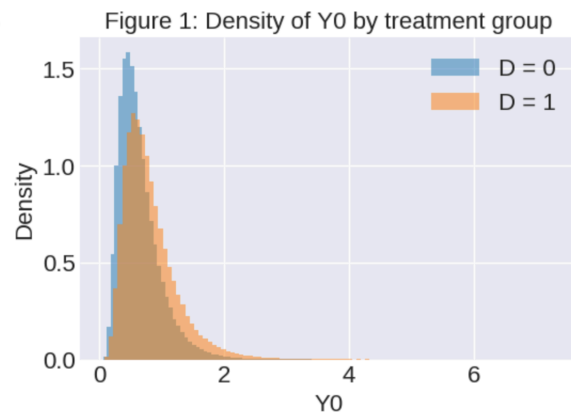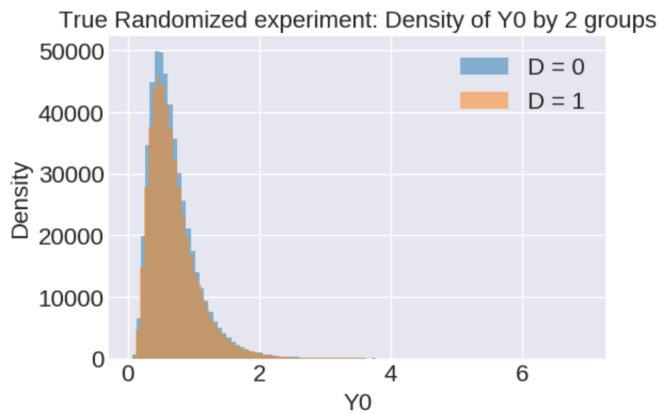
the overlap becomes samller compared with that in `figure 1`;

## 1.4 Data-generating process for an experiment with a treatment effect

1. make a version of the biased data-generating process with an $ATE = 0.3$

   the line of the code that I changed is

   ```
   Y1 = Y0 + 0.3
   NATE -->
   1.11414 - 0.64309 = 0.47104999999999986
   ```

2. Repeat the Dexp experiment to show that NATE=ATE=0.3.


True Randomized experiment Y1=Y0+0.3: Density of Y0 by 2 groups
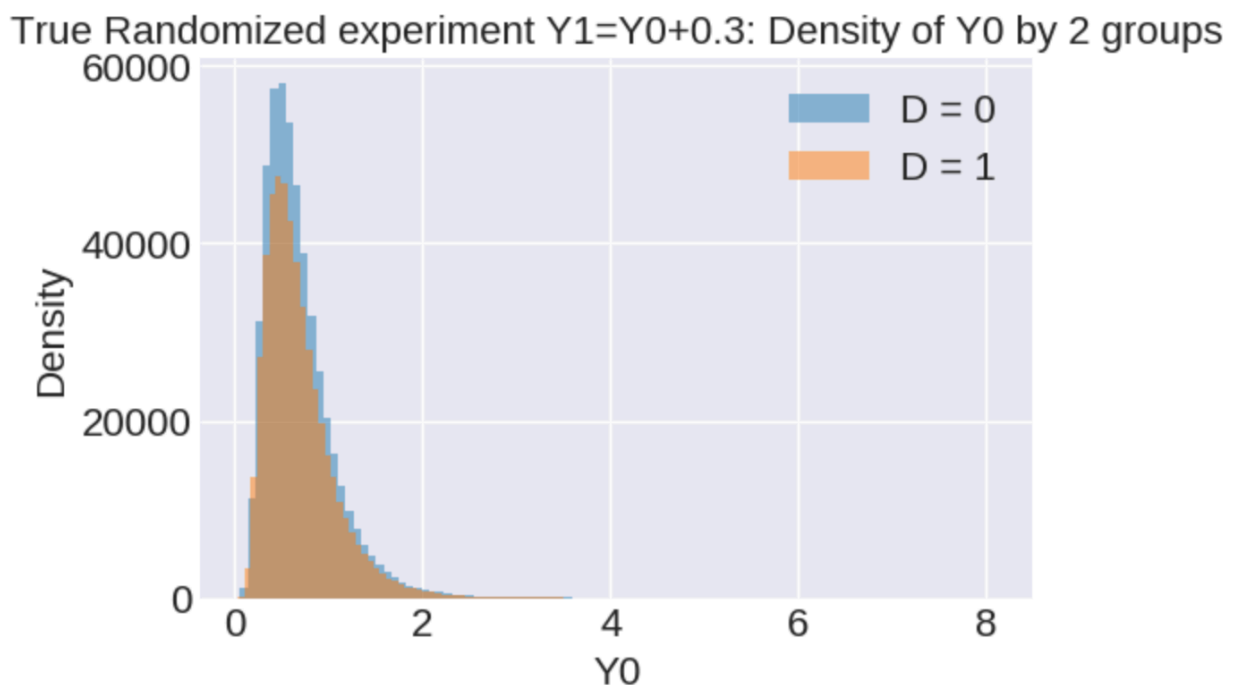
# 2 Causality warm-up

## 1.

1. Selection bias;

   - Control for all other factors, `selection bias` exisits when there are systematic differences in 2 groups. i.e. certain charateristics crucial to the output of the experiment exists in one group but not in another; even without any treament, 2 groups of individuals make very different decisions;

   - mathmatically

     NATE = ATE + selection bias + differential bias;

     $$E[Y^0|treatment] - E[Y^0|control]$$

   - it is a problem, because of the bias exists in the system, hence the true treatment effect cannot be observed or misguided;

2. 
   - It could be `cultural differences`; people with different background, religious, origins hold very different opinions on various situations or when facing the experiment;
   - it could also exists in the `sample selection process`; also selection bias can occur if the experimenters have some biases when selecting the data;

3. random variable to be independent;

   - one variable's change ($\Delta$ )does not rely/depend on all other variables' movements;
   - mathematically; it uses the symbol ($\perp$)

4. why do experiments remove `selection bias?`

   becasue **experiments** can use various techniques, simulations... to create an indepedent random process to make the random variables truely independent. Ultimately make the treatment effective when realeased to the population;

# 3 In-class study discussion

The ariticle that I selected was the `Study 5: preschooler's use of phones and tablets`

- this study is facing the sample of $402$ mothers of $3 - 5$ year-olds, the observed data are all fron online survey; So each $i$ indicates/represents the $3 - 5$ year old child;

- What are the $Y^0$ & $Y^1$ in this study?

  $Y^0 \rightarrow$ the average of **bedtime, wake time, sleep time, napping behaviors** of the children who do not have the habit of using electonic device at bed time

$Y^1 \rightarrow$ the average of **bedtime, wake time, sleep time, napping behaviors** of the children who have the habit of using electonic device at bed time.

- What is the treatment $(D)$

  The treatment in this case is the `electronic media use

- A mechanism that can casue bias in this study; which direction;

  In this study, the sample observation is relatively small;

  hence there are many ways can casue `selection bias`;

  - the parents are heavy phone, computer, social media user; which can casue the bias to go up. even without the `electonic media use`, the child can also be affect by their parents' noise and sleep late, wake up late...

  - the observed kid is a `nocturnal kid` who has a sleeping pattern that very different from peers; also causes the bias to go up;

- Describe a mechanics that could cause differntial effect bias; which direction does the bias go?

  The mothers in hoping for healthy lifestyle search online for related articles and recommanded with this survey by the reasearchers or Google;

  which can casue the differential bias to go down; because the differential effect bias is calculated with `P(treatment)` $\times$ `(ATT - ATC)` in this case, the mothers hoping for a healthy life style maybe eduacte their kids to spend less time on social media. So the preceived percetage of treatment decreases.

- describe what would it mean if the independence assumption is true in this study.

  The `independece assumption` means the potential outcomes are independet of treatment assginment $X \rightarrow (Y^0, Y^1) \perp D$; In this study, it means the chidren's sleeping pattern are indepedent from the electronic media use; Without any interfence of electronic media, the sleeping pattern of 2 groups should be exactly the same. With the using of electronic meida use, the sleeping pattern of 2 groups should also the same after the treatment.

- Describe how you would do a true randomized experiments to study this causal relationship.

  I might do the survey in the Orphanges. The reason is becasue within the Orphanges, the children might be accessble to eclectonics so often. Also becasue it is under regulation of adults, the sleeping time, wake up time, napping behavior can be more precisely observed. Also, all the children are trainned under the same way after they are received in the Orphanges.

# 4 Critical reading

1. The article that I am interested in is [Understanding Why Crime Fell in the 1990s: Four Factors that Explain the Decline and Six that Do Not](#) by professor Steven D. Levitt

2. What are the units in the study?

   the units in this study is the `FBI violent and property crime index` & `homicide rates` and other national trends in specifc categories of crime.

3. the potential outcome of $Y^1$ & $Y^0$ and the treament $D$

   $Y^1 \rightarrow$ the potential crime rate if the policing strategy were put into practise;

   $Y^0 \rightarrow$ the potential crime rate if the policing strategy were rejected and not put into practise;

   $D \rightarrow$ new and strict policy strategies such as gun control laws, laws allowing the carrying of concealed weapons ...

4. What is the ATE in the study? If there is no number given just identify how they characterize the treatment effect.

   The Average Treatment Effect was given in the section `The Mafnitude of the Decline`; data was presented in the `Table 3`

   The ATE was $-4.3\%$; The treatment effect was characterized using the homicide rate data from 1991 to 2000, they observed that the homicide rates per capita fell from $9.8\% \rightarrow 5.5\%$ per $100,000$ population.

5. How is the $D$ treatment assigned;

   There are many treatments in this study.

   - One good example of how the treatment is assigned is from the `Gun Control Laws`. The Violence Prevention Act of 1993 instituted stricter requirements for background checks before the gun is sold.

   - another example might be in the section `Four Factors That Explain the Decline in Crime`

     the treament here is the increases in the number of police

6. what kind of selection bias do you think might be in the study.

   how would it affects the results;

   - Some cities in the middle of US such as **Iowa**. It has a very low population density, the climate is harsh. The crime rate is low when compared with San Francisco which has a much higher population density and nicer climate. So there is a systematic difference

between the observed cities. Also, in some cities, the funding might not be that much to increase the policing power.
  - o in this case, the selection bias will casue the observed crime rate to be lower.

7. How would you design an experiment to get an unbiased estimate of the **ATE**?

   it is a city level experiment, hence it is very hard/impossible to accomplish. It requires the similar cities within US. similar crime rate, mass, population level and practise the same policy strategies.

8. Why did you choose this study?
   - o **novelty**

     this study studies the human social behavior when respoding to new policies;
   - o This study demostrated that some **policies** might not woking but they are highly correlated with the observed data & pattern/time series trend. Also, there are 2 subtle factors such as increased incarceration, teh decline of crack and legalized aboration that are actually driving the decline of crime rates;

# 5 Selection bias vs. big data

1. write down the $NATE$ decomposition equation and write the value of each of the 4 components

   $$\mathbf{NATE} = \mathbf{ATE}(E[Y^1 - Y^0]) + \mathbf{SelectionBias} + \mathbf{Differnential\ Bias}$$
   - o NATE --> $E[Y1|D = 1] - E[Y0|D = 0] = 0.17217832824951973$
   - o ATE --> $E[Y^1 - Y^0] = 0$
   - o S.B --> $0.17217832824951973$
   - o D.B --> $0$

2. make a function `selection_bias(n)` that calculates the selection bias in just the first n units...

   ●

   ```
   # The same df as we did in class;
   N = 10 ** 6
   charitability = np.random.normal(loc=-1, scale=0.5, size=N)
   Y0 = np.random.lognormal(mean=(charitability + 0.5), sigma=0.1, size=N)
   Y1 = Y0.copy()
   prob_search_dwb = scipy.special.expit(charitability)
   D = np.random.binomial(n=1, p=prob_search_dwb)
   ```

```python
Y = Y1 * D + Y0 * (1 - D)
df = pd.DataFrame({
    'D': D,
    'Y0': Y0,
    'Y1': Y1,
    'Y': Y,
    'delta': Y1 - Y0,
    'C': charitability,
})

# and then we make a selection_bias(n)
import random
n = random.randint(0,10000)  # here we radonmly select the value for n;

def selection_bias (n):
  df_pt = df[:n]  # slice the original dataframe and insert with biased data
point;
  bias = df_pt.query("D == 1")['Y0'].mean() - df_pt.query("D == 0")
['Y0'].mean()
  return bias
x = selection_bias(n)
print("Selection bias for (n = ", n, ") = ", x)


# and then we create the line plt;
my_lst = []
for i in range(0,10000):
  x = selection_bias(i)
  my_lst.append(x)


line_plt = pd.Series(my_lst)
line_plt.plot.line()
```
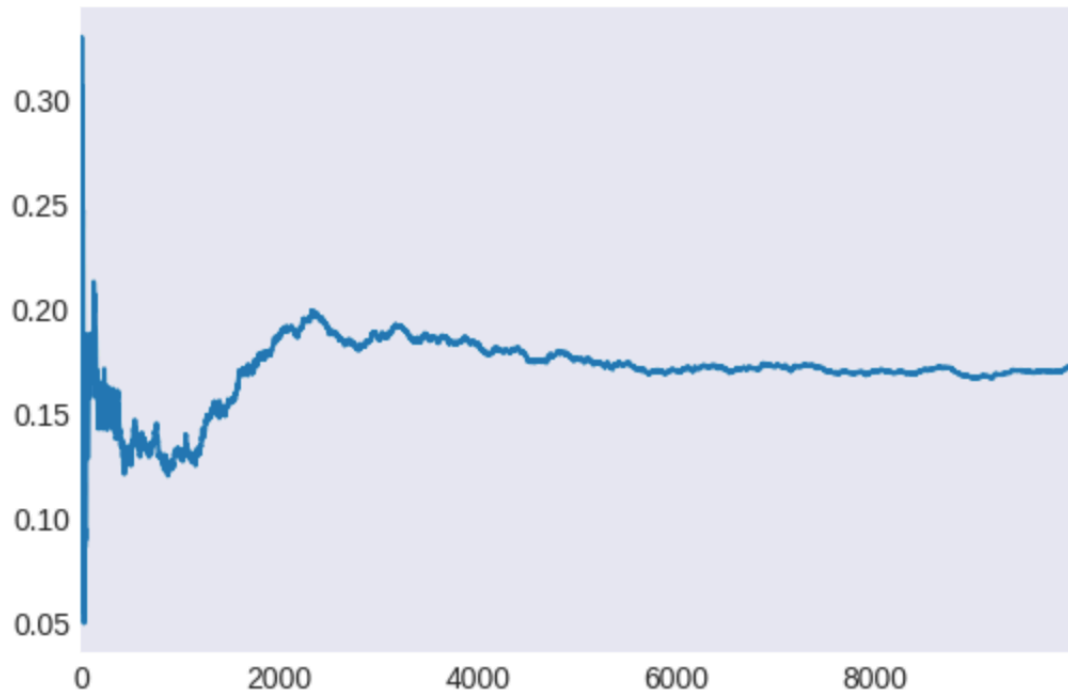
- if the n is 100 for instance. then we are going to get the plt like this;

- as we ca see form the graph that the the line is coverging to the value of selection bias → `1.7`
  hence, the conclusion is that, the ***the law of large numbers*** applied to the slection bias.

# Bonus 1

1. mathmatically, if we naively contrast the average outcomes in the two treatment group;
   it can be written in the form of $\dfrac{1}{N_1} \sum_{i:d_i=1} Y_i - \dfrac{1}{N_0} \sum_{i:d_i=0} Y_i$
   which all can be observed form the dataset; the expectation of this quntity is the Naive Average Treatment Effect (NATE)

2. without observing the potential outcome which is the $Y^1$ & $Y^0$ , we cannot directly estimate the ATE, selection bias and differential effects; Only one euqation but with several unknows;

3. If the independence assumption is not true, in this equation it will have 3 unknows → `ATE` & `Selection Bias` & `Differentianl treatment effects`;
   the solutions can be

~~NATE = ATE~~

$NATE = ATE + Selection Bias$

$NATE = ATE + Differential Bias$

$NATE = ATE + selectionbias + differentialbias$

4. If the independence assumption holds, then we will get the `NATE = ATE` becasue independence assumption eliminates selection bias and differential bias;

   the assumption is $(Y^0, Y^1) \perp D$

   and the additional equations we obtained are

   $E[Y|D = d] = E[Y^0]; d = 0, 1$

   $E[Y|D = 1] = E[Y^0|D = 0] = 0$

5. in this case, we have 4 possible equations if we including the independet assumption;

# Bonus 2

**NATE** = $E[Y^1|D = 1] - E[Y^0|D = 0]$

**ATT** = $E[Y^1 - Y^0|D = 1] \rightarrow E[Y^1|D = 1] - E[Y^0|D = 1]$

**Selection Bias** = $E[Y^0|D = 1] - E[Y^0|D = 0]$

after we add them altogether we can prove that $\rightarrow NATE = ATT + Sel.B$

my Colab for reference;