

$$\begin{aligned}
1) \quad \mathbb{E}[\hat{\mu}_c | \sum_{i=1}^n R_i > 0] &= \mathbb{E}\left[\mathbb{E}\left[\frac{\sum_i R_i Y_i}{\sum_i R_i} | R_1, \dots, R_n, \sum_i R_i > 0\right] | \sum_i R_i > 0\right] \text{ Adam's Law} \\
&= \mathbb{E}\left[\frac{1}{\sum_i R_i} \mathbb{E}\left[\sum_i R_i Y_i | R_1, \dots, R_n, \sum_i R_i > 0\right] | \sum_i R_i > 0\right] \text{ Take out what is known} \\
&= \mathbb{E}\left[\frac{1}{\sum_i R_i} \sum_i R_i \mathbb{E}[Y_i | R_1, \dots, R_n, \sum_i R_i > 0] | \sum_i R_i > 0\right] \text{ Linearity} \\
&= \mathbb{E}\left[\frac{1}{\sum_i R_i} \sum_i R_i \mathbb{E}[Y_i | R_1, \dots, R_n, \sum_i R_i > 0] | \sum_i R_i > 0\right] \text{ Take out what is known} \\
&= \mathbb{E}\left[\frac{1}{\sum_i R_i} \sum_i R_i \mathbb{E}[Y_i] | \sum_i R_i > 0\right] \quad Y \perp R, \text{ from MCAR} \\
&= \mathbb{E}\left[\mathbb{E}[Y] | \sum_i R_i > 0\right] \quad \text{cancel terms, no division by 0.} \\
&= \mathbb{E}[Y]
\end{aligned}$$

$$\begin{aligned}
2.1) \quad \mathbb{E}[\hat{\mu}_{f, full}] &= \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n f(x_i)\right] \\
&= \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n \mathbb{E}[Y | X_i]\right] \\
&= \frac{1}{n} \left[ \sum_{i=1}^n \mathbb{E}[\mathbb{E}[Y | X_i]] \right] \quad \text{linearity} \\
&= \frac{1}{n} \left[ \sum_{i=1}^n \mathbb{E}[Y] \right] \quad \text{Adam's Law} \\
&= \frac{n}{n} \mathbb{E}[Y] \\
&= \mathbb{E}[Y]
\end{aligned}$$

$$\begin{aligned}
2.2) \quad \mathbb{E}[\hat{\mu}_y] &= \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n [R_i Y_i + (1-R_i) f(x_i)]\right] \\
&= \frac{1}{n} \left[ \sum_{i=1}^n [\mathbb{E}[R_i Y_i] + \mathbb{E}[(1-R_i) f(x_i)]] \right] \quad \text{linearity} \\
&= \frac{1}{n} \left[ \sum_{i=1}^n [\mathbb{E}[\mathbb{E}[R_i Y_i | X_i]] + \mathbb{E}[\mathbb{E}[(1-R_i) f(x_i) | X_i]]] \right] \quad \text{Adam's} \\
&= \frac{1}{n} \left[ \sum_{i=1}^n [\mathbb{E}[\mathbb{E}[R_i | X_i] \mathbb{E}[Y_i | X_i]] + \mathbb{E}[f(x_i) \mathbb{E}[1-R_i | X_i]]] \right] \quad \text{MAR, Take out what is known} \\
&= \frac{1}{n} \left[ \sum_{i=1}^n [\mathbb{E}[\pi(x_i) \mathbb{E}[Y_i | X_i]] + \mathbb{E}[\mathbb{E}[Y_i | X_i] (1-\pi(x_i))]] \right] \\
&= \frac{1}{n} \left[ \sum_{i=1}^n [\mathbb{E}[\pi(x_i) \mathbb{E}[Y_i | X_i]] + (1-\pi(x_i)) \mathbb{E}[Y_i | X_i]] \right] \\
&= \frac{1}{n} \left[ \sum_{i=1}^n \mathbb{E}[\mathbb{E}[Y_i | X_i]] \right] = \frac{n}{n} \mathbb{E}[Y] = \mathbb{E}[Y]
\end{aligned}$$

$$3.1) \hat{M}_{ipw-a} = \hat{M}_{ipw}(D+a) - a \\ = \hat{M}_{ipw}(D) + \frac{a}{n} \sum_{i=1}^n \frac{R_i}{\pi(x_i)} - a$$

The control variate is  $\mathbb{E}\left[\frac{a}{n} \sum_{i=1}^n \frac{R_i}{\pi(x_i)}\right]$ , and its expectation is

$$\begin{aligned} \mathbb{E}\left[\frac{a}{n} \sum_{i=1}^n \frac{R_i}{\pi(x_i)}\right] &= \frac{a}{n} \sum_{i=1}^n \mathbb{E}\left[\frac{R_i}{\pi(x_i)}\right] \\ &= \frac{a}{n} \sum_{i=1}^n \mathbb{E}\left[\mathbb{E}\left[\frac{R_i}{\pi(x_i)} \mid X_i\right]\right] \\ &= \frac{a}{n} \sum_{i=1}^n \mathbb{E}\left[\frac{1}{\pi(x_i)} \mathbb{E}[R_i \mid X_i]\right] \\ &= \frac{a}{n} \sum_{i=1}^n \mathbb{E}\left[\frac{1}{\pi(x_i)} \pi(x_i)\right] \\ &= \frac{a}{n} \sum_{i=1}^n \mathbb{E}[1] \\ &= a \end{aligned}$$

3.2) According to the slides, the optimal  $a$  to choose is

$$a = \text{Corr}\left(Y, \frac{1}{n} \sum_{i=1}^n \frac{R_i}{\pi(x_i)}\right) \cdot \frac{\text{SD}(Y)}{\text{SD}\left(\frac{1}{n} \sum_{i=1}^n \frac{R_i}{\pi(x_i)}\right)}$$

4.1.1) We only expect  $\hat{f}(x) \rightarrow \mathbb{E}[Y|X=x, R=1]$  if  $\hat{f}$  is well-specified.

Reasonable: if  $\hat{f}$  is well-specified, then  $\hat{f}(x) \rightarrow \mathbb{E}[Y|X=x, R=1]$ , and  $\mathbb{E}[Y|X=x, R=1] = \mathbb{E}[Y|X=x, T=1]$ .

Not Reasonable: If  $\hat{f}$  is mispecified, then  $\hat{f}(x) \not\rightarrow \mathbb{E}[Y|X=x, R=1]$ , and  $\hat{f}(x) \not\rightarrow \mathbb{E}[Y|X=x, T=1]$ .

In order for  $\hat{f}(x)$  to perform well on cases where  $T=1$ , we should fit  $\hat{f}$  using importance weighting. This reduces the risk of  $\hat{f}$  over the  $T=1$  cases.

4.1.2) The importance weights are as follows:

$$\text{Train Distribution: } p(y, x | R=1) = \frac{p(y|x) \pi_r(x) p(x)}{P(R=1)}$$

$$\text{Target Distribution: } p(y, x | T=1) = \frac{p(y|x) \pi_t(x) p(x)}{P(T=1)}$$

$$IW = \frac{p(y, x, T=1)}{p(y, x, R=1)} = \frac{\pi_t(x) P(R=1)}{\pi_r(x) P(T=1)} \propto \frac{\pi_t(x)}{\pi_r(x)}$$

$$\text{So, } R_{IW}(\hat{f}) = \frac{\pi_t(x)}{\pi_r(x)} \ell(\hat{f}(x), Y)$$

4.2.1)

$$\text{a) } \begin{aligned} \mathbb{P}(Y=1|T=1) &= 1 \cdot \mathbb{P}(Y=1|T=1) + 0 \cdot \mathbb{P}(Y=0|T=1) \\ &= \mathbb{E}[Y|T=1] \\ &= \mathbb{E}[\mathbb{E}[Y|X=x, T=1]|T=1] \\ &= \mathbb{E}[f(x)|T=1] \end{aligned}$$

add 0  
Def of Expectation  
Adam's Law

$$\text{b) } \begin{aligned} \mathbb{E}[Tf(x)] &= \mathbb{E}[\mathbb{E}[Tf(x)|T]] \\ &= \mathbb{E}[T \mathbb{E}[f(x)|T]] \\ &= \mathbb{P}(T=1) \cdot 1 \cdot \mathbb{E}[f(x)|T=1] + \mathbb{P}(T=0) \cdot 0 \cdot \mathbb{E}[f(x)|T=0] \quad \text{Def of Expectation} \\ &= \mathbb{P}(T=1) \mathbb{E}[f(x)|T=1] \end{aligned}$$

Adam's Law  
Taking out what is known

$$\text{c) } \mathbb{P}(Y=1|T=1) = \mathbb{E}[f(x)|T=1] \quad (\text{by 4.2.1.a})$$

$$= \frac{\mathbb{E}[Tf(x)]}{\mathbb{P}(T=1)} \quad (4.2.1.b)$$

$$= \frac{\mathbb{E}[\mathbb{E}[Tf(x)|X]]}{\mathbb{P}(T=1)} \quad \text{Adam's Law}$$

$$= \frac{\mathbb{E}[f(x)\mathbb{E}[T|X]]}{\mathbb{P}(T=1)} \quad \text{Take out what is known}$$

$$= \frac{\mathbb{E}[\pi_t(x)f(x)]}{\mathbb{P}(T=1)}$$

$$\text{d) } \begin{aligned} \mathbb{P}(T=1) &= 1 \cdot \mathbb{P}(T=1) + 0 \cdot \mathbb{P}(T=0) \quad \text{add 0} \\ &= \mathbb{E}[T] \quad \text{Def of Expectation} \\ &= \mathbb{E}[\mathbb{E}[T|X]] \quad \text{Adam's Law} \\ &= \mathbb{E}[\pi_t(x)] \end{aligned}$$

Putting all these pieces together, we get  $\mathbb{P}(Y=1|T=1) = \frac{\mathbb{E}[\pi_t(x)f(x)]}{\mathbb{E}[\pi_t(x)]}$

$$4.2.2) \text{ We propose } \frac{\frac{1}{N} \sum_{i=1}^N \pi(x_i) \hat{f}(x_i)}{\frac{1}{N} \sum_{i=1}^N \pi(x_i)}$$

# Homework 2: ATE Estimation with the voter turnout data

## Introduction

In this assignment, we'll be working with some of the data described in the paper "[Social Pressure and Voter Turnout: Evidence from a Large-Scale Field Experiment](#)" by Gerber et al. (2008). We attained the data from this [Github repo](#), specifically this [file](#). It's also included in the assignment zip file. In this assignment we'll build several ATE estimators by reduction to two estimations from incomplete data, as discussed in lecture. Parts of this notebook are based on a [notebook](#) from a tutorial/course on causal inference at Stanford GSB. Although not necessary, you may find it interesting to refer to, as they give more details about the covariates, and they cover some methods that we don't get into (and vice versa).

## Data prep

In [ ]:

```
%load_ext autoreload
%autoreload 2
```

In [ ]:

```
import pandas as pd
import numpy as np
from sklearn.preprocessing import StandardScaler, PolynomialFeatures
from sklearn.linear_model import LinearRegression, LogisticRegression
from sklearn.tree import DecisionTreeRegressor
from sklearn.ensemble import RandomForestClassifier, RandomForestRegressor
from scipy.stats import norm, sem
from scipy.interpolate import UnivariateSpline
from sklearn.calibration import calibration_curve
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats.stats import pearsonr
from numpy.random import default_rng
```

In [ ]:

```
## Read the data, select some columns, and lightly process
df = pd.read_csv('sp.csv.xz')
cts_variables_names = ["yob", "hh_size", "totalpopulation_estimate",
                      "percent_male", "median_age",
                      "percent_62yearsandover",
                      "percent_white", "percent_black",
                      "percent_asian", "median_income",
                      "employ_20to64", "highschool", "bach_orhigher",
                      "percent_hispanicorlatino"]
binary_variables_names = ["sex", "g2000", "g2002", "p2000", "p2002", "p2004"]
scaled_cts_covariates = StandardScaler().fit_transform(df[cts_variables_names])
binary_covariates = df[binary_variables_names]
d = pd.DataFrame(np.concatenate((scaled_cts_covariates, binary_covariates), axis=1),
                  columns=cts_variables_names+binary_variables_names, index=df.index)
d["W"] = df["treat_neighbors"]
d["Y"] = df["outcome_voted"]
```

## Problem 1: ATE for RCT

All individuals in this experiment had an equal probability of being assigned to the treatment group, so the difference-of-means will be a reasonable estimator for the average treatment effect (ATE).

Write a function that computes the difference-of-means estimator for the treatment effect, along with an approximate 95% confidence interval. Apply it to the dataset d computed above and report the results. Save the estimate of the ATE and the radius of the confidence interval for later use.

```
In [ ]: def get_diff_of_means(d, alpha=0.05):
    # TODO

    ate, ate_CI_radius = get_diff_of_means(d, alpha=0.05)
```

## Problem 2: ATE estimation with known confounders

In this problem we're going to take a relatively small and biased subsample of our full dataset and try to use that to estimate the ATE. Our approach is to reduce ATE estimation to estimating a mean in the MAR setting.

Below we give a function that takes a biased sample using an approach similar to [this one](#). The details aren't important for what follows.

```
In [ ]: likely_voters = (((d['g2002'] == 1) & (d['sex'] == 0)) | (d['p2004'] == 1))
unlikely_voters_control = (~likely_voters) & (d["W"] == 0)
likely_voters_treatment = likely_voters & (d["W"] == 1)

def get_biased_sample(d, overall_subsample_rate=.03, bias_rate=.4, rng=default_rng(0)):
    keep_prob = overall_subsample_rate * np.ones(len(d))
    keep_prob[unlikely_voters_control] *= bias_rate
    keep_prob[likely_voters_treatment] *= bias_rate
    keep = rng.random(len(d)) <= keep_prob
    d_bias = d[keep]
    return d_bias

d_bias = get_biased_sample(d)
print(f"We've sampled {len(d_bias)} instances out of {len(d)}.")
```

### Part A

Below we provide the function `get_basic_MAR_estimators`, which computes the complete-case mean and the IPW mean in the MAR setting with known propensity scores. You are to complete the `get_ATE_estimators` function below so that it *uses the `get_basic_MAR_estimators` function* and computes one ATE estimate for each type of estimator produced by `get_basic_MAR_estimators` (i.e. the complete-case mean and the IPW mean). Use logistic regression to estimate the propensity scores from the data provided to the function. Run the existing code block below to display the results.

```
In [ ]: def get_basic_MAR_estimators(Y, X, pi, R):
```

```
"""

```

Args:

Y (pd.Series): the measurement of the outcome  $Y_i$   
 X (pd.DataFrame): covariates, rows correspond to entries of Y  
 R (pd.Series): boolean series indicating whether Y was observed  
 pi (pd.Series): propensity scores corresponding to observations

Returns:

dict of estimator names and estimates for EY

```
"""

```

```
est = {}
n = len(Y)
## All the estimators below assume we know the pi (i.e. "missing by design")
est["mean"] = np.mean(Y[R])
est["ipw_mean"] = np.sum(Y[R] / pi[R]) / n
return est

```

In [ ]:

```
def get_ATE_estimators(Y, X, W, get_MAR_estimators=get_basic_MAR_estimators):
    """
    Args:
        Y (pd.Series): the measurement of the outcome  $Y_i$ 
        X (pd.DataFrame): covariates, rows correspond to entries or Y
        W (pd.Series): 0/1 series indicating control (0) or treatment (1) assignment
        get_MAR_estimators: function behaving like get_basic_MAR_estimators above

    Returns:
        dict of ATE estimator names and estimates, same format as for get_MAR_estimators
    """
    # TODO

```

In [ ]:

```
## Run the following and report
d_bias = get_biased_sample(d, overall_subsample_rate=.03, bias_rate=.4, rng=default_rng)
X = d_bias.drop(columns=['W', 'Y'])
ate_est = get_ATE_estimators(Y=d_bias["Y"], X=X, W=d_bias["W"], get_MAR_estimators=get_
print(ate_est)
```

## Part B

In this part we significantly expand our ATE estimators and see how they perform over repeated trials. Complete get\_MAR\_estimators below to include

- Complete-case mean (done for you)
- IPW mean (done for you)
- Self-normalized IPW mean
- Linear regression imputation
- IPW linear regression imputation (as defined in lecture)
- IW linear regression imputation (as defined in lecture)
- Augmented IPW using linear regression
- [Optional (not for credit): nonlinear regression, or any other variations you'd like to try]

Run the code below to assess performance of these estimators.

```
In [ ]: def get_MAR_estimators(Y, X, pi, R):
    """
    Args:
        Y (pd.Series): the measurement of the outcome Y_i
        X (pd.DataFrame): covariates, rows correspond to entries or Y
        R (pd.Series): boolean series indicating whether Y was observed
        pi (pd.Series): propensity scores corresponding to observations

    Returns:
        dict of estimator names and estimates for EY
    """
    est = {}
    n = len(Y)
    ## All the estimators below assume we know the pi (i.e. "missing by design")
    est["mean"] = np.mean(Y[R])
    est["ipw_mean"] = np.sum(Y[R] / pi[R]) / n

    ## TODO
```

```
In [ ]: def get_estimator_stats(estimates, true_parameter_value=None):
    """
    Args:
        estimates (pd.DataFrame): each row corresponds to collection of estimates for a
            each column corresponds to an estimator
        true_parameter_value (float): the true parameter value that we will be comparing

    Returns:
        pd.DataFrame where each row represents data about a single estimator
    """

    est_stat = []
    for est in estimates.columns:
        pred_means = estimates[est]
        stat = {}
        stat['stat'] = est
        stat['mean'] = np.mean(pred_means)
        stat['SD'] = np.std(pred_means)
        stat['SE'] = np.std(pred_means) / np.sqrt(len(pred_means))
        if true_parameter_value:
            stat['bias'] = stat['mean'] - true_parameter_value
            stat['RMSE'] = np.sqrt(np.mean((pred_means - true_parameter_value) ** 2))
        est_stat.append(stat)

    return pd.DataFrame(est_stat)
```

```
In [ ]: def run_experiments(sampler, num_repeats=10, rng=default_rng(0)):
    data_list = []
    num_obs_list = []
    for i in range(num_repeats):
        d = sampler(rng)
        X = d.drop(columns=['W', 'Y'])
        ate_est = get_ATE_estimators(Y=d["Y"], X=X, W=d["W"], get_MAR_estimators=get_MAR_estimators)
        data_list.append(ate_est)
    results = pd.DataFrame(data_list)
    return results
```

```
In [ ]:
def sampler(rng):
    return get_biased_sample(d, overall_subsample_rate=.06, bias_rate=.4, rng=rng)
rng = default_rng(0)
results = run_experiments(sampler, num_repeats=500, rng=rng)
```

```
In [ ]:
results_eval = get_estimator_stats(results, true_parameter_value=ate)
results_eval
```

## Part C

You should see that the ATE estimate based on the IPW mean is significantly biased, and the bias seems to be driving most of the RMSE. In class we showed that the IPW mean is an unbiased estimator for EY in the MAR setting, and the corresponding ATE estimator is also unbiased. Why are we seeing bias in this experiment?

## Problem 3: Bootstrap confidence intervals for ATE

Let's now consider a more realistic scenario, in which we only have a single sample to work with. In this case it's very important to be able to include some form of uncertainty measure with our estimates. In this problem we'll do this using the normal approximated bootstrap confidence intervals described in our module on CATEs. However, in this setting, rather than estimating the CATE, we're just estimating the ATE.

## Part A

Complete the function `get_stratified_bootstrap_CI` to generate 95% normal approximated bootstrap confidence intervals for each of the ATE estimators we've developed above. Execute the code below to test the function and produce the results as a table and in a plot.

```
In [ ]:
def get_bootstrap_stats(boot_estimates, full_estimates, alpha=0.05):
    est_stat = []
    signif_level = -norm.ppf(alpha/2)
    for est in full_estimates:
        est_boot = np.array(boot_estimates[est])
        stat = {}
        stat['estimator'] = est
        stat['estimate'] = full_estimates[est]
        #stat['mean_boot'] = np.mean(est_boot)
        stat['SD'] = np.std(est_boot)
        stat['CI_radius'] = signif_level * stat['SD']
        stat['lower_ci'] = stat['estimate'] - stat['CI_radius']
        stat['upper_ci'] = stat['estimate'] + stat['CI_radius']
        est_stat.append(stat)

    return pd.DataFrame(est_stat)

def get_stratified_bootstrap_CI(Y, X, W, estimators, num_bootstrap=10, alpha=0.05):
    """
    Returns:
        pd.DataFrame with
            column "estimator" with the name of the estimator,
```

```
        column "estimate" which is the estimate based on (Y,X,W)
        which serves as the center of our bootstrap confidence intervals
        column "CI_radius" which is the radius of the confidence interval
        (additional columns may be included as you wish)
"""
#TODO
```

In [ ]:

```
rng = default_rng(8)
d_samp = sampler(rng) # our single biased sample
X = d_samp.drop(columns=['W','Y'])
estimators = lambda Y, X, W: get_ATE_estimators(Y, X, W, get_MAR_estimators=get_MAR_est)
ci = get_stratified_bootstrap_CI(Y=d_samp["Y"], X=X, W=d_samp["W"], estimators=estimators)
print(f"A 95% confidence interval for the ATE based on the full dataset is {ate}+/-{ate_ci}")
print(ci)
to_plot = ci[['estimator','estimate','CI_radius']]
to_plot = to_plot.append({'estimator':'Full data estimate','estimate':ate,'CI_radius':0})
fig, ax = plt.subplots(1,1, figsize=(5,7))
ax.errorbar(to_plot['estimate'], np.arange(len(to_plot)), \
            xerr=to_plot['CI_radius'], \
            fmt='o', elinewidth=3, capsize=5)
ax.grid('on')
ax.set_yticks(np.arange(len(to_plot)))
ax.set_yticklabels(to_plot["estimator"])
```