Propensity Score Weighting

Code ▾

Ruoqing Zhu

Last Updated: November 19, 2024

Example: Observational Study

In a randomized trial design, our analysis shows that, even we do not utilize the covariate information, we can still obtain an unbiased estimation of the average treatment effect. This is a very important property of RCT, and it is the reason why RCT is considered the gold standard for estimating the treatment effect. However, in many cases, the treatment cannot be assigned randomly for practical and ethical reasons. Such a scenario is the **observational study**. In that case, the difference-in-means estimator will be biased.[1] Here is a simulation study. In this case, the treatment label is generated by a logistic model with covariate X, which makes higher X values more likely to take treatment 1. Some correction will be needed to estimate the treatment effect in observational studies. We will discuss this in the next lecture.

Hide

```r
# setting parameters
n <- 200
n1 <- 100
n0 <- 100
tau <- 1 # true ATE
nsim = 500

set.seed(1)

tauhat <- rep(NA, nsim)
tauhatsd <- rep(NA, nsim)

for (i in 1:nsim) {
  # generate potential outcomes
  X <- rnorm(n)
  Y1 <- rnorm(n, mean = 0.5*X+X^2)
  Y0 <- rnorm(n, mean = 0.5*X)

  # treatment label
  A = rbinom(n, 1, exp(X) / (1 + exp(X)))

  # observed outcomes
  Y <- A * Y1 + (1 - A) * Y0

  tauhat[i] <- mean(Y[A == 1]) - mean(Y[A == 0])
  tauhatsd[i] <- sqrt(var(Y[A == 1]) / n1 + var(Y[A == 0]) / n0)
}

# make two plots on the same row
par(mfrow = c(1, 2))

# set margin of figure
par(mar = c(4, 4, 1, 1))
plot(tauhat[1:50], pch = 19, ylim = c(0, 2),
     xlab = "Simulation Runs", ylab = "Estimated ATE")
abline(h = tau, col = "red")

# adding confidence intervals
for (i in 1:50) {
  ci_lower <- tauhat[i] - 1.96 * tauhatsd[i]
  ci_upper <- tauhat[i] + 1.96 * tauhatsd[i]
  arrows(x0 = i, y0 = ci_lower, x1 = i, y1 = ci_upper, angle = 90, code = 3, leng
th = 0.05)
}

coverage = sum((tauhat - 1.96 * tauhatsd < tau) & (tauhat + 1.96 * tauhatsd > ta
u)) / nsim
legend("topright", paste("Coverage probability of", nsim, "runs:", coverage))

boxplot(tauhat, xlab = "Boxplot of Estimated ATE", ylab = "Estimated ATE")
abline(h = tau, col = "red")
```
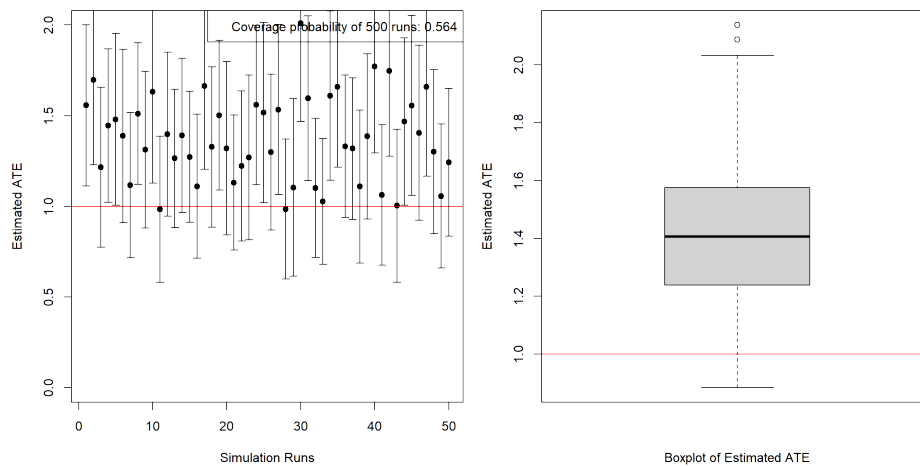
Boxplot of Estimated ATE

In this lecture, we will discuss how to deal with this bias by modeling the propensity score.

Assumptions

Recall that our ultimate goal is to estimate the **average treatment effect** (ATE), following the potential outcome notation:

$$\tau = \mathrm{E}[Y_i(1) - Y_i(0)]$$

Our goal here is to establish some properties of the new estimator $\hat{\tau}_{\mathrm{IPW}}$. To do so, we need to establish some assumptions. Previously, we had two assumptions, Independence and SUTVA. It is obvious that the independence assumption is no longer valid. But some new assumptions could help us to establish the validity of the IPW estimator, besides the SUTVA assumption.

> Unconfoundedness: $A_i \perp \{Y_i(0), Y_i(1)\} | X_i$

Essentially this is the same as the independence assumption, but now we condition on the covariate X_i. Why do we call it unconfoundedness? Because this eliminates the potential bias in our estimation if we do not conditioning on some variable. Here, conditioning on X_i is sufficient. The assumption has many different names, such as **strong ignorability** or **conditional independence**. There is an even weaker version of this called **ignorability**, readers are referred to (Rubin 1978) and (Ding 2023) for more details. The difference is rather technical and is not the focus of this lecture.

> Positivity: The treatment assignment is not deterministic for any value of X, i.e.,
> $$\eta \le P(A = 1 | X = x) \doteq e(x) \le 1 - \eta$$

This is to ensure that we have individuals from both treatment labels for each value of X so that we can compare them. This may seem to be pretty natural, but when the cardinality of X is large, this could be violated.

Inverse-propensity Weighting

Because estimating $\widehat{e}(X_i)$ comes with additional randomness, and could potentially harm the performance of the estimator, let's first consider a version with $e(X_i)$ known (e.g., by design). Then the estimator becomes

$$\widehat{\tau}_{\text{ipw}}^* = \frac{1}{n} \sum_{i=1}^n \left(\frac{A_i Y_i}{e(X_i)} - \frac{(1 - A_i)Y_i}{1 - e(X_i)} \right)$$

First, we can see that this estimator is an unbiased estimator of the ATE:

$$
\begin{aligned}
\text{E}[\widehat{\tau}_{\text{ipw}}^*] &= \text{E}\left[\frac{A_i Y_i}{e(X_i)} - \frac{(1 - A_i)Y_i}{1 - e(X_i)} \right] \quad \text{by IID} \\
&= \text{E}\left[\frac{A_i Y_i(1)}{e(X_i)} - \frac{(1 - A_i)Y_i(0)}{1 - e(X_i)} \right] \quad \text{by SUTVA} \\
&= \text{E}\left[\text{E}\left[\frac{A_i Y_i(1)}{e(X_i)} \,\Big|\, X_i \right] - \text{E}\left[\frac{(1 - A_i)Y_i(0)}{1 - e(X_i)} \,\Big|\, X_i \right] \right] \\
&= \text{E}\left[\frac{e(X_i)}{e(X_i)} \text{E}[Y_i(1)|X_i] - \frac{1 - e(X_i)}{1 - e(X_i)} \text{E}[Y_i(0)|X_i] \right] \quad \text{by Unconfoundedness} \\
&= \text{E}\left[Y_i(1) - Y_i(0) \right]
\end{aligned}
$$

Now, since $e(X_i)$ is unknown, we can estimate it by a model. A common choice is to use a logistic regression model. The estimated propensity score is $\widehat{e}(X_i) = \widehat{P}(A = 1|X_i)$. Under suitable conditions, the estimated propensity score is consistent, and hence the estimated ATE is consistent.

Example: Simulation

Hide

```r
# setting parameters
n <- 200
n1 <- 100
n0 <- 100
tau <- 1 # true ATE
nsim = 500

set.seed(1)

tauhat <- rep(NA, nsim)
tauhatsd <- rep(NA, nsim)

for (i in 1:nsim) {
  # generate potential outcomes
  X <- rnorm(n)
  Y1 <- rnorm(n, mean = 0.5*X+X^2)
  Y0 <- rnorm(n, mean = 0.5*X)

  # treatment label
  A = rbinom(n, 1, exp(X) / (1 + exp(X)))

  # observed outcomes
  Y <- A * Y1 + (1 - A) * Y0

  # fit propensity score model
  ps_model <- glm(A ~ X, family = binomial(link = "logit"))
  ehat <- predict(ps_model, type = "response")

  # ipw estimator
  tauhat[i] <- mean(Y * A / ehat - Y * (1 - A) / (1 - ehat))
}

par(mfrow=c(1,1))
boxplot(tauhat, xlab = "Boxplot of Estimated ATE", ylab = "Estimated ATE")
abline(h = tau, col = "red")
```
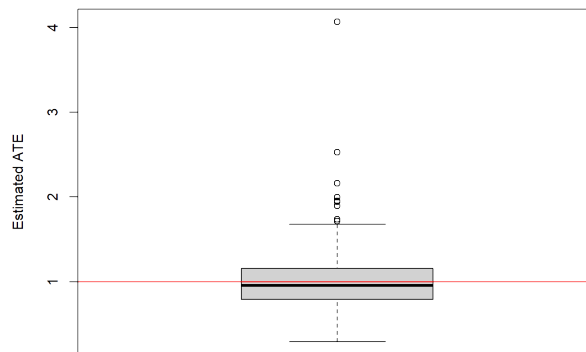

Boxplot of Estimated ATE

Example: Direct Implementation

The `lalonde` dataset in the `Maching` package is a popular dataset for causal effect. It contains the following variables:

- `treat` : a binary variable indicating whether the subject received the treatment
- `age` : the age of the subject
- `educ` : the education level of the subject
- `black` : a binary variable indicating whether the subject is black
- `hisp` : a binary variable indicating whether the subject is Hispanic
- `married` : a binary variable indicating whether the subject is married
- `nodegr` : a binary variable indicating whether the subject has a degree
- `re74` : the real earnings in 1974
- `re75` : the real earnings in 1975
- `re78` : the real earnings in 1978

LaLonde (1986) was interested in the causal effect of a job training program on earnings. This is in fact an experimental data, meaning that the participants were randomly assigned to the treatment and control groups. However, we can still use this dataset to illustrate the IPW method. If we consider a naive difference-in-means estimator, we would have:

Hide

```
library(Matching)
## Loading required package: MASS
## ##
## ##  Matching (Version 4.10-14, Build Date: 2023-09-13)
## ##  See https://www.jsekhon.com for additional documentation.
## ##  Please cite software as:
## ##   Jasjeet S. Sekhon. 2011. ``Multivariate and Propensity Score Matching
## ##   Software with Automated Balance Optimization: The Matching package for R.''
## ##   Journal of Statistical Software, 42(7): 1-52.
## ##
data(lalonde)
head(lalonde)
```

	age <int>	educ <int>	black <int>	hisp <int>	married <int>	nodegr <int>	re74 <dbl>	re75 <dbl>	re78 <dbl>
1	37	11	1	0	1	1	0	0	9930.05
2	22	9	0	1	0	1	0	0	3595.89
3	30	12	1	0	0	0	0	0	24909.50
4	27	11	1	0	0	1	0	0	7506.15
5	33	8	1	0	0	1	0	0	289.79
6	22	9	1	0	0	1	0	0	4056.49

6 rows | 1-10 of 13 columns

Hide

```
  # Calculate the average earnings for treated and control groups
  avg_treated <- mean(lalonde$re78[lalonde$treat == 1])
  avg_control <- mean(lalonde$re78[lalonde$treat == 0])

  # Calculate the difference in means
  ate_diff_means <- avg_treated - avg_control

  ate_diff_means
## [1] 1794.343
```

As we have learned in this lecture, we would like to use the IPW estimator to adjust for potential bias.

Hide

```
  ps_model <- glm(treat ~ age + educ + black + hisp + married + nodegr + re74 + re7
5,
               family = binomial(link = "logit"), data = lalonde)

  # Extract the propensity scores
  lalonde$pscore <- predict(ps_model, type = "response")
  lalonde$psweight <- ifelse(lalonde$treat == 1, 1 / lalonde$pscore, 1 / (1 - lalon
de$pscore))

  mean(lalonde$treat*lalonde$re78*lalonde$psweight - (1 - lalonde$treat)*lalonde$re
78*lalonde$psweight)
## [1] 1613.135
```

Ding, Peng. 2023. "A First Course in Causal Inference." *arXiv Preprint arXiv:2305.18793*.

Eidelman, Rachel S, Danielle Hollar, Patricia R Hebert, Gervasio A Lamas, and Charles H Hennekens. 2004. "Randomized Trials of Vitamin e in the Treatment and Prevention of Cardiovascular Disease." *Archives of Internal Medicine* 164 (14): 1552–56.

Lee, I-Min, Nancy R Cook, J Michael Gaziano, David Gordon, Paul M Ridker, JoAnn E Manson, Charles H Hennekens, and Julie E Buring. 2005. "Vitamin e in the Primary Prevention of Cardiovascular Disease and Cancer: The Women's Health Study: A Randomized Controlled Trial." *Jama* 294 (1): 56–65.

Rubin, Donald B. 1978. "Bayesian Inference for Causal Effects: The Role of Randomization." *The Annals of Statistics*, 34–58.

Stampfer, Meir J, Charles H Hennekens, JoAnn E Manson, Graham A Colditz, Bernard Rosner, and Walter C Willett. 1993. "Vitamin e Consumption and the Risk of Coronary Disease in Women." *New England Journal of Medicine* 328 (20): 1444–49.

---

1. A famous example was about vitamin E supplement and the risk of coronary heart disease. Several papers on New England Journal of Medicine in 1993 (Stampfer et al. 1993, 1993) reported that the vitamin E supplement can reduce the risk of coronary heart disease using observational study. However, later studies (Eidelman et al. 2004; Lee et al. 2005) using randomized control trial reports no significant effect.↵