Selection Bias (Chapter 8)

Supplement

Recap

- General definition of selection bias:

  Conditioning on a common effect of

  (i) treatment A or a cause of A, and

  (ii) outcome Y or a cause of Y

- Key idea/result:

  Under certain conditions, can adjust for selection bias by inverse weighting responses of individuals that do not drop out ($C = 0$) by $\Pr[C = 0|A, L]$

  Standardization works sometimes also

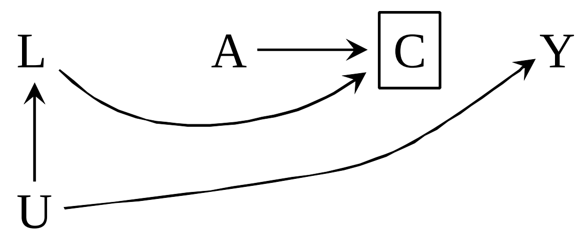
IPCW

- IP censoring weighted estimator

$$\hat{E}(Y^a) = \frac{1}{n}\sum_{i=1}^{n}\frac{I[A_i = a, C_i = 0]Y_i}{\Pr[C_i = 0|A_i, L_i]\Pr[A_i = a|L_i]}$$

unbiased estimator of $E(Y^a)$ if $Y^a \perp\!\!\!\perp A|L$ and $Y \perp\!\!\!\perp C|A, L$

- Fig 8.3



- Can also use standardization in this case

R Code Illustrating IPCW & Standardization Estimators

```
n <- 1000000
U <- rnorm(n)
L <- 1*(U>0)
Y <- U + 0.1*rnorm(n)
pA <- .5
A <- rbinom(n,1,pA)
pC.AL <- .25+.25*A+.25*L
C <- rbinom(n,1,pC.AL)

naive <- mean(Y[C==0 & A==1])
ipcw <- (1-pC.AL)*pA
ipcw.est <- 1/n*sum((Y/ipcw)[C==0 & A==1])
stand <- mean(Y[C==0 & A==1 & L==0])*mean(L==0) +
         mean(Y[C==0 & A==1 & L==1])*mean(L==1)

print(paste(naive,ipcw.est,stand))
[1] "-0.2666 -0.00119 0.00089"
```

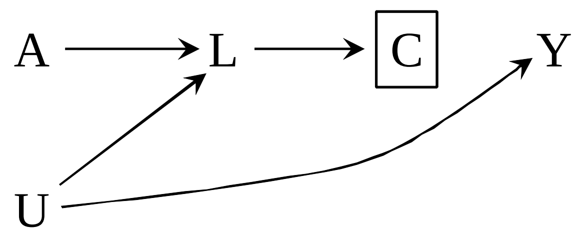
IPCW

- IP censoring weighted estimator

$$\hat{E}(Y^a) = \frac{1}{n}\sum_{i=1}^{n} \frac{I[A_i = a, C_i = 0]Y_i}{\Pr[C_i = 0|A_i, L_i]\Pr[A_i = a]}$$

unbiased estimator of $E(Y^a)$ if $Y^a \perp\!\!\!\perp A$ and $Y \perp\!\!\!\perp C|A, L$

- Fig 8.4



- Standardization? First IPCW estimator?

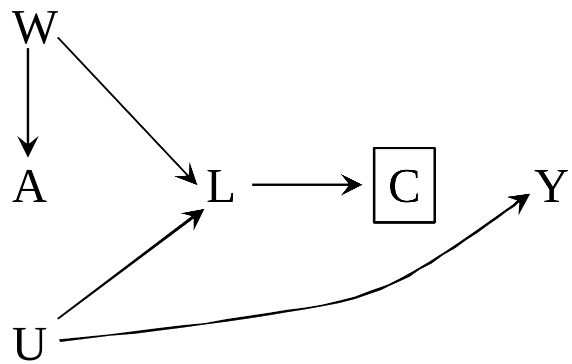
R Code Illustrating IPCW & Standardization Estimators

```
n <- 1000000
U <- rnorm(n)                       # cf Fig 8.4
pA <- .5
A <- rbinom(n,1,pA)
L <- 1*(U>0 & A >0)
Y <- U + 0.1*rnorm(n)
pC.L <- .25+.25*L
C <- rbinom(n,1,pC.L)
ipcw <- (1-pC.L)*pA

naive <- mean(Y[C==0 & A==1])
ipcw.est <- 1/n*sum((Y/ipcw)[C==0 & A==1])
stand <- mean(Y[C==0 & A==1 & L==0])*mean(L==0) +
         mean(Y[C==0 & A==1 & L==1])*mean(L==1)
print(paste(naive,ipcw.est,stand))
[1] "-0.1580 0.0016 -0.3975"
```

M-Bias Example

- What about Fig 8.6?



- $Y^a \perp\!\!\!\perp A$ and $Y \perp\!\!\!\perp C|A,L$? Then use

$$\hat{E}(Y^a) = \frac{1}{n}\sum_{i=1}^{n} \frac{I[A_i = a, C_i = 0]Y_i}{\Pr[C_i = 0|A_i, L_i]\Pr[A_i = a]}$$

- Note also true that $Y^a \perp\!\!\!\perp A|W$, which suggests

$$\hat{E}(Y^a) = \frac{1}{n}\sum_{i=1}^{n} \frac{I[A_i = a, C_i = 0]Y_i}{\Pr[C_i = 0|A_i, L_i]\Pr[A_i = a|W_i]}$$

Fig 8.6 / M-Bias / IPCW

```
n <- 1000000
U <- rnorm(n)
W <- rbinom(n,1,.5)
pA.W <- .25+.5*W
A <- rbinom(n,1,pA.W)
L <- 1*(U>0 & W>0)
pC.L <- .25+.5*L
C <- rbinom(n,1,pC.L)
Y <- U + 0.1*rnorm(n)

naive <- mean(Y[C==0 & A==1])
pA <- .5; ipcw <- (1-pC.L)*pA
ipcw.est <- 1/n*sum((Y/ipcw)[C==0 & A==1])
ipcw1 <- (1-pC.L)*pA.W
ipcw.est1 <- 1/n*sum((Y/ipcw1)[C==0 & A==1])

print(paste(naive,ipcw.est,ipcw.est1))
[1] "-0.267 5.7e-05 -0.001"
```