PART 1: averaged casual effect $E(Y^{a=1}=1)-E(Y^{a=0}=1)$
following are the estimators

- 1.1 IPTW:

$$\frac{1}{n}\sum_{i=1}^{n}\frac{A_iY_i}{P(A_i=1\mid L_i)}-\frac{1}{n}\sum_{i=1}^{n}\frac{(1-A_i)Y_i}{P(A_i=0\mid L_i)}$$

  note: after we assign weight to each individual, the number of individuals in the treatment groups is approxmately the number of study samples.

- 1.2 normalized IPTW:

$$\left(\sum_{i=1}^{n}\frac{A_i}{P(A_i=1\mid L_i)}\right)^{-1}\sum_{i=1}^{n}\frac{A_iY_i}{P(A_i=1\mid L_i)}-\left(\sum_{i=1}^{n}\frac{1-A_i}{P(A_i=0\mid L_i)}\right)^{-1}\sum_{i=1}^{n}\frac{(1-A_i)Y_i}{P(A_i=0\mid L_i)}$$

  three versions of weight:

  - $\frac{1}{P(A_i=1|L_i)}$ and $\frac{1}{P(A_i=0|L_i)}$: note that average weight is 2, so pseudo-population is twice that of the original study samples (page 161)
  - stablized: $\frac{P(A_i=1)}{P(A_i=1|L_i)}$ and $\frac{P(A_i=0)}{P(A_i=0|L_i)}$ narrower 95%CI than nonstablized weights (Hernan & Robin page 162)
  - if censoring is present $\frac{P(A_i=1)}{P(A_i=1|L_i)}\times\frac{P(C_i=0|A=1)}{P(C_i=0|L_i,A_i=1)}$ and $\frac{P(A_i=0)}{P(A_i=0|L_i)}\times\frac{P(C_i=0|A=0)}{P(C_i=0|L_i,A_i=0)}$

- 2.1 direct standardization (g-formula /g-computation/g-standardization)

$$\begin{aligned}E[Y_a]&=E_L[E(Y_a\mid L)]\\&=E_L[E(Y_a\mid A=a,L)]\quad\text{[conditional exchangeability and positivity]}\\&=E_L[E(Y\mid A=a,L)]\quad\text{[consistency]}\end{aligned}$$

$$\sum_{\ell}E[Y\mid A=1,\ell]P(L=\ell)-\sum_{\ell}E[Y\mid A=0,\ell]P(L=\ell)$$

- 2.2 model-based standardization

$$\sum_{\ell}E[Y\mid A=1,\ell]P(L=\ell)=E_L(E[Y\mid A=1,L])$$

  - $E[Y\mid A=1,L]$ estimated from regression model $E[Y\mid A,L]$ (could specify different function forms in the regression model)
  - had everyone(n) treated (i.e., A=1), we can compute the estimated $\hat{E}[Y\mid A=1,L=\ell]$ for every level of $L=\ell$. page 172 chapter 13.3)

– compute

$$E(E[Y \mid A=1,L]) - E(E[Y \mid A=0,L])$$
$$= \frac{1}{n}\sum_{i=1}^{n}\left\{E[Y_i \mid A_i=1, L_i;\hat{\phi}] - E[Y_i \mid A_i=0, L_i;\hat{\phi}]\right\}$$

$$\boxed{\text{IPTW and standardization are numerically equivalent with saturated treatment model } P(A|L) \text{ and outcome model } E(Y|A,L)}$$

- 3 doubly robust estimator

$$\frac{1}{n}\sum_{i=1}^{n}\left[E[Y_i \mid A_i=a, L_i;\hat{\phi}] + \frac{\mathbf{1}_{\{A_i=a\}}}{P(A_i=a \mid L_i;\hat{\gamma})}(Y_i - E[Y_i \mid A_i=a, L_i;\hat{\phi}])\right]$$

- 4 marginal structural model

$$\log\left(\frac{P(Y_a=1 \mid \theta)}{1-P(Y_a=1 \mid \theta)}\right) = \theta_0 + \theta_1 a$$

  - the procedure is exactly the same as propensity score weighting. However, Rubin prposed propensity score weighting, and Robin proposed marginal structural model.

PART 2: conditional casual effect/Heterogeneity of the effect

$$E(Y^{a=1}=1|V=v) - E(Y^{a=0}=1|V=v)$$

- 1. standardization

$$E[Y_a \mid V=v] = \sum_{\ell} E[Y \mid A=a, L=\ell, V=v]P(L=\ell \mid V=v)$$

  - add interaction term in the outcome model.

- 2. potential outcome model

$$E(Y_a|V,\boldsymbol{\theta}) = \theta_0 + \theta_1 a + \theta_2 V a + \theta_3 V$$