IPTW estimation

20.05.21 09:14

- Recap with estimation in regression models:

observed - $Y = X \cdot \beta + \varepsilon$ - error term, typically mean $(\mu) = 0$ and constant variance σ^2 (random noise i.e.)

→ estimating involves minimizing the sum-of-squared deviations (least squares estimator):

$$\sum_{i=1}^{n} X_i (Y_i - X_i^T \hat{\beta}) = 0 \quad (\text{for } \hat{\beta})$$

↳ comes from the least sum of squared differences formulation

Estimation in a generalized linear model: (GLM e.g. LR, log R, Poisson regr.)

$$E(Y_i \mid X_i) = \mu_i = g^{-1}(X_i^T \beta) \quad \text{involves solving: } \underbrace{\sum_{i=1}^{n} \frac{\partial \mu_i^T}{\partial \beta} V_i^{-1} \{Y_i - \mu_i(\beta)\} = 0}_{\text{"estimating equation"}}$$

$g(\cdot)$ is a link fct. (e.g. logit fct.)

- $\frac{\partial \mu_i^T}{\partial \beta}$ ... derivative of the mean μ_i w.r.t. β
- V_i ... variance term

Estimation in MSMs: (MSMs model the mean of potential outcomes)

E.g.: $E(Y_i^a) = g^{-1}(\psi_0 + \psi_1 a)$ | we set a (to whatever we want)

not observed (potential outcome)

this model is not equivalent to the regression model:

$$E(Y_i \mid A_i) = g^{-1}(\psi_0 + \psi_1 A_i)$$ | we condition on A_i

observed

because of confounding!!!

- the pseudo-population (obtained from IPTW) is free from confounding (assuming (1) ignorability (2) positivity)

⇒ So: we can estimate MSMs by solving estimating equations for the observed data of the pseudo-population

→ with IPTW, we can create a pseudo-population which is free from confounding.

- using the IPTW (weights) $W_i = \frac{1}{A_i \cdot P(A=1 \mid X_i) + (1-A_i) \cdot P(A=0 \mid X_i)}$

if $A=0$: $\frac{1}{P(A=0 \mid X_i)}$
if $A=1$: $\frac{1}{P(A=1 \mid X_i)}$ } inverse prob. of treatment received

we can estimate ψ with:

$$\sum_{i=1}^{n} \frac{\mu_i^T}{\partial \psi} V_i^{-1} W_i \{Y_i - \mu_i(\psi)\} = 0$$

Steps for estimating parameters from a MSM:

1) Estimate propensity score: $P(A=1 \mid X)$ (e.g. log Reg.)
2) Create weights
   - $\frac{1}{\text{prop. score}}$ (for treated subjects) $:= \frac{1}{P(A=1 \mid X)}$
   - $\frac{1}{1-\text{prop. score}}$ (for control subjects) $:= \frac{1}{1-P(A=1 \mid X)} = \frac{1}{P(A=0 \mid X)}$
3) Specify MSM of interest (e.g. w/ or w/o effect modification, continuous outcome (linear MSM) or count data/outcome (log-linear model))
4) Use software to fit a weighted generalized linear model
5) Use asymptotic (sandwich) variance estimator (or bootstrapping) } robust asymptotic variance estimates
   ↳ this accounts for fact that the pseudo-population might be larger than sample size

e.g.: in dataset population: 1 subject in treated group, 9 subjects in control group
⇓
BUT in IPTWeighted pseudo population: 10 subjects in each group

⇒ $n_{true} = 10 \longrightarrow n_{pseudo} = 2 \times 10 = 20$. (artificially inflated pop. size).