

Lecture 2-4: Bootstrap Procedures & Clustering

Goal: Discuss bootstrap methods and inference with clustered data.

- Motivation
 - Uncertainty for general statistics
 - Limits of analytic sampling distributions
 - Mean vs. median inference challenges
- Bootstrap
 - Resampling principle
 - Empirical distribution approximates population
 - Nonparametric bootstrap
 - Parametric bootstrap
- Clusters
 - Clustered sampling structure
 - Within-cluster dependence issues
 - Variance decomposition: between vs. within
 - Intra-cluster correlation (ICC)
 - Estimating ICC via ANOVA mean squares
 - Adjusted standard errors and intervals
 - Cluster (block) bootstrap procedure
 - Performance comparisons across ρ

Worksheet

1. Suppose the sample consists of G independent clusters, each of size m (so $n = mG$). Consider the one-way random-effects model

$$X_{gi} = \mu + u_g + \varepsilon_{gi}, \quad u_g \stackrel{iid}{\sim} (0, \sigma_B^2), \quad \varepsilon_{gi} \stackrel{iid}{\sim} (0, \sigma_W^2),$$

with $u_g \perp \varepsilon_{gi}$. Let $\sigma^2 = \sigma_B^2 + \sigma_W^2$ and the *intra-cluster correlation* $\rho = \frac{\sigma_B^2}{\sigma_B^2 + \sigma_W^2}$.

Show that

$$\text{Var}(\bar{X}) = \frac{\sigma^2 + (m-1)\sigma_B^2}{mG} = \frac{\sigma^2}{mG} (1 + (m-1)\rho).$$

2. Under the same setup with equal cluster sizes, define the cluster means

$$\bar{X}_g = \frac{1}{m} \sum_{i=1}^m X_{gi}, \quad \bar{X} = \frac{1}{G} \sum_{g=1}^G \bar{X}_g.$$

Let

$$\text{MSW} = \frac{1}{G(m-1)} \sum_{g=1}^G \sum_{i=1}^m (X_{gi} - \bar{X}_g)^2, \quad \text{MSB} = \frac{m}{G-1} \sum_{g=1}^G (\bar{X}_g - \bar{X})^2.$$

Given that $\mathbb{E}[\text{MSW}] = \sigma_W^2$ and $\mathbb{E}[\text{MSB}] = \sigma_W^2 + m\sigma_B^2$ (proofs not required),

- (a) Propose consistent estimators for σ_W^2 , σ_B^2 , and the intraclass correlation coefficient (ICC) $\rho = \sigma_B^2 / (\sigma_B^2 + \sigma_W^2)$.
- (b) Using your results from (a), derive an estimator $\widehat{\text{Var}}(\bar{X})$ for the sampling variance of the sample mean $\bar{X}$.
- (c) Show that your estimators $\hat{\rho}$ and $\widehat{\text{Var}}(\bar{X})$ are consistent as $G \rightarrow \infty$ with fixed m (for $\widehat{\text{Var}}(\bar{X})$, which goes to zero, show that $\widehat{\text{Var}}(\bar{X}) / \text{Var}(\bar{X}) \xrightarrow{p} 1$).
- (d) Construct a $100(1 - \alpha)\%$ cluster-adjusted confidence interval for μ .