Chapter from Handbook of Econometrics to be posted soon. (Horowitz)

Bootstrap

$\{X_i, i \leq n\}$ data

$F_0 \in \mathcal{F}$ is dgp

$\mathcal{F} = \{F(x;\theta), \theta \in \Theta\}$ if parametric

$\mathcal{F} = \{F$ is a cdf $\}$. if nonparametric

$T_n = T_n(X_1, \ldots, X_n)$ statistic of interest

e.g. $T_n = t_j$ t-statistic based on $\hat{\beta}_j$ in linear regression

$G_n(t, F_0) = P_{F_0}(T_n \leq t)$ under true distr. fcn. F_0

- would like to work with this

$G_n(t, F) = P_F(T_n \leq t)$ under distr. fcn F

We want distributions so we can estimate the α-quantiles of $G_n(t, F_0)$: $G_n^{-1}(\alpha, F_0)$ for hypothesis testing.

p-values: $1 - G_n(t, F_0)\big|_{t=T_n}$

Defn: $G_\infty(t, F_0) = \lim_{n\to\infty} G_n(t, F_0)$ asymptotic df

$G_\infty(t, F) = \lim_{n\to\infty} G_n(t, F)$

We will estimate $G_n(\cdot, F_0)$ by $G_\infty(\cdot, F_0)$
- asymptotic principle.

Bootstrap principle: approximate $G_n(\cdot, F_0)$ by $G_n(\cdot, \hat{F})$

- Let $\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} 1\{X_i \le x\}$ nonparametric bootstrap
- Let $\hat{F}_0(x) = F(x, \hat{\theta})$ parametric bootstrap

Monte Carlo algorithm to tabulate $G_n(t, \hat{F})$

1. $\forall j = 1, \dots, B$, generate $\{X_{ij}^*, i = 1, \dots, n\}$ by sampling from $\hat{F}$ randomly. (with replacement)
   - Practically we have $B \approx 200$
2. Compute $T_{nj}^* = T_n(X_{1j}^*, \dots, X_{nj}^*)$ $j \le B$
3. Use $\{T_{nj}^*, j \le B\}$ to compute empirical probability of $\{T_n^* \le t\}$

e.g. $\{X_1, X_2, X_3\}$ observations
Draw randomly from this set. Get samples

$\{X_{21}^*, X_{31}^*, X_{21}^*\}$ $j=1$

$\{X_{12}^*, X_{12}^*, X_{12}^*\}$ $j=2$

$\vdots$

- Idealistically, if we knew F_0, we would just tabulate the distribution function $G_n(\cdot, F_0)$.
- sometimes, $G_n(t, F_0) = G_n(t, F)$ $\forall F \in \mathcal{F}$ for our statistic. Such a statistic is referred to as <u>pivotal</u>. We can use <u>any</u> $F \in \mathcal{F}$

e.g. GM:

$\boxed{1}$ $Y = X\beta + \varepsilon$, $\beta \in \mathbb{R}^k$

$\boxed{2}$ $\varepsilon | X \sim N(0, \sigma^2 I)$

$\boxed{3}$ $\text{rank}(X) = k$

Then $t_j = \dfrac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \sim t(n-k)$

irrespective of β and σ^2.

- t_j is <u>pivotal</u>.

- If we don't know F_0, and T is not pivotal, we can put bounds on $G_n(t, F_0)$ using

$$B_n = \left[\inf_{F \in \hat{\mathcal{F}}} G_n(t, F), \sup_{F \in \hat{\mathcal{F}}} G_n(t, F) \right]$$

For example, we can set (in the parametric case)

$$\hat{\mathcal{F}} = \{ F(\cdot, \theta) : \theta \in CI_{1-\beta_n}(\theta_0) \}, \quad \beta_n \to 0$$

e.g. $CI = \{ \hat{\theta} \pm se(\hat{\theta}) \log n \}$

asymptotically, this approach becomes equivalent to bootstrap

1] Start with $\theta_1 = \theta' \in CI_{1-\beta_n}(\theta_0)$. Tabulate critical value $G_n^{-1}(\alpha, F(\cdot, \theta_1))$

2] Draw nearby value $\theta_2 = \theta_1 + \eta \in CI_{1-\beta_n}(\theta_0)$, η random. Tabulate $G_n^{-1}(\alpha, F(\cdot, \theta_2))$

3] Repeat until convergence.