Introduction

The **score bootstrap** perturbs the score contributions of an M-estimator using IID, mean zero, unit variance weights. The distribution of the perturbed score statistics is then used for inference on the observed test statistic.

Motivation

Consider an M-estimator $\hat{\boldsymbol{\theta}}$ obtained by solving the estimating equation:

$$\boldsymbol{\Psi}_n(\boldsymbol{\theta}) = \sum_{i=1}^{n} \boldsymbol{\psi}_i(\boldsymbol{\theta}) \stackrel{\text{Set}}{=} \mathbf{0}$$

By first order Taylor expansion:

$$\begin{aligned}
\big(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\big) &= -\dot{\boldsymbol{\Psi}}_n(\boldsymbol{\theta}_0)^{-1}\boldsymbol{\Psi}_n(\boldsymbol{\theta}_0) + \mathcal{O}_p(n^{-1}) \\
\sqrt{n}\big(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\big) &= -\left[\frac{1}{n}\sum_{i=1}^{n}\frac{\partial \boldsymbol{\psi}_i}{\partial \boldsymbol{\theta}'}(\boldsymbol{\theta}_0)\right]^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{\psi}_i(\boldsymbol{\theta}_0) + \mathcal{O}_p(n^{-1/2})
\end{aligned}$$

Define $\boldsymbol{A}_0$ as the expected Jacobian of $\boldsymbol{\psi}_i$ w.r.t. $\boldsymbol{\theta}$:

$$\boldsymbol{A}_0 \equiv -E\left[\frac{\partial \boldsymbol{\psi}_i}{\partial \boldsymbol{\theta}'}(\boldsymbol{\theta}_0)\right]$$

By LLN:

$$-\frac{1}{n}\sum_{i=1}^{n}\frac{\partial \boldsymbol{\psi}_i}{\partial \boldsymbol{\theta}'}(\boldsymbol{\theta}_0) = \boldsymbol{A}_0 + o_p(1)$$

Thus the quantity $\sqrt{n}\big(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\big)$ is expressible as:

$$\sqrt{n}\big(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\big) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{A}_0^{-1}\boldsymbol{\psi}_i(\boldsymbol{\theta}_0) + o_p(1)$$

The limiting distribution is:

$$\sqrt{n}\big(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\big) \xrightarrow{\mathcal{L}} N\big(\mathbf{0},\ \boldsymbol{A}_0^{-1}\boldsymbol{B}_0\boldsymbol{A}_0^{-T}\big)$$

Here $\boldsymbol{B}_0$ is the expected outer product of the score equation:

$$\boldsymbol{B}_0 = E\big[\boldsymbol{\psi}_i\boldsymbol{\psi}_i'(\boldsymbol{\theta}_0)\big]$$

Let (ω_i) denote IID mean zero, unit variance weights. Form the perturbed score:

$$\tilde{\boldsymbol{\Psi}}_n(\boldsymbol{\theta}) \equiv \sum_{i=1}^{n}\omega_i\boldsymbol{\psi}_i(\boldsymbol{\theta})$$

Denote by $\tilde{\boldsymbol{\theta}}$ a solution to the equation:

$$\boldsymbol{\Psi}_n(\boldsymbol{\theta}) - \tilde{\boldsymbol{\Psi}}_n(\hat{\boldsymbol{\theta}}) \stackrel{\text{Set}}{=} \mathbf{0}$$

By first order Taylor expansion:

$$\boldsymbol{\Psi}_n(\tilde{\boldsymbol{\theta}}) = \boldsymbol{\Psi}_n(\boldsymbol{\theta}_0) + \dot{\boldsymbol{\Psi}}_n(\boldsymbol{\theta}_0)(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + \mathcal{O}_p(n^{-1})$$

Subtract $\tilde{\boldsymbol{\Psi}}_n(\hat{\boldsymbol{\theta}})$ from each side, then:

$$\mathbf{0} = \boldsymbol{\Psi}_n(\boldsymbol{\theta}_0) - \tilde{\boldsymbol{\Psi}}_n(\hat{\boldsymbol{\theta}}) + \dot{\boldsymbol{\Psi}}_n(\boldsymbol{\theta}_0)(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + \mathcal{O}_p(n^{-1})$$

Introduce and remove $\hat{\boldsymbol{\theta}}$ from the equation:

$$\mathbf{0} = \boldsymbol{\Psi}_n(\boldsymbol{\theta}_0) - \tilde{\boldsymbol{\Psi}}_n(\hat{\boldsymbol{\theta}}) + \dot{\boldsymbol{\Psi}}_n(\boldsymbol{\theta}_0)(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}) + \dot{\boldsymbol{\Psi}}_n(\boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + \mathcal{O}_p(n^{-1})$$

The quantity $\dot{\boldsymbol{\Psi}}_n(\boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ is expressible as:

$$\dot{\boldsymbol{\Psi}}_n(\boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = -\boldsymbol{\Psi}_n(\boldsymbol{\theta}_0) + \mathcal{O}_p(n^{-1})$$

Therefore:

$$\dot{\boldsymbol{\Psi}}_n(\boldsymbol{\theta}_0)(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}) = \tilde{\boldsymbol{\Psi}}_n(\hat{\boldsymbol{\theta}}) + \mathcal{O}_p(n^{-1})$$

Scaling by $\sqrt{n}$:

$$\sqrt{n}\big(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}\big) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\omega_i \boldsymbol{A}_0^{-1}\boldsymbol{\psi}_i(\hat{\boldsymbol{\theta}})$$

Taking the limit conditional on $\hat{\boldsymbol{\theta}}$:

$$\sqrt{n}\big(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}\big) \xrightarrow{\mathcal{L}} N\big(\mathbf{0},\ \boldsymbol{A}_0^{-1}\boldsymbol{B}(\hat{\boldsymbol{\theta}})\boldsymbol{A}_0^{-T}\big)$$

This argument motivates approximation of the distribution of $\sqrt{n}\big(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\big)$ as:

$$\sqrt{n}\big(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\big) \overset{d}{\approx} \sqrt{n}\big(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}\big) \overset{d}{\approx} \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\omega_i \boldsymbol{A}_0^{-1}\boldsymbol{\psi}_i(\hat{\boldsymbol{\theta}})$$

Remark 2.1. Although $\tilde{\boldsymbol{\theta}}$ was introduced as a solution to the equation $\boldsymbol{\Psi}_n(\boldsymbol{\theta}) - \tilde{\boldsymbol{\Psi}}_n(\hat{\boldsymbol{\theta}}) \overset{\text{Set}}{=} \mathbf{0}$, **actually obtaining $\tilde{\boldsymbol{\theta}}$ is unnecessary**. ♦

Perturbation Distribution

3.1 Mammen Distribution

The Mammen two-point distribution takes the form:

$$\omega_i = \begin{cases} \frac{1-\sqrt{5}}{2} & , \text{ w.p. } \frac{1+\sqrt{5}}{2\sqrt{5}} \\ \frac{1+\sqrt{5}}{2} & , \text{ w.p. } 1 - \frac{1+\sqrt{5}}{2\sqrt{5}} \end{cases}$$

The Mammen distribution has expectation $E[\omega_i] = 0$, variance $E[\omega_i^2] = 1$, and third (central) moment $E[\omega_i^3] = 1$. Thus, Mammen perturbations are able to match this first three moments of a distribution, providing an asymptotic refinement over the normal approximation.

3.2 Rademacher Distribution

The Rademacher distribution is defined as:

$$\omega_i = \begin{cases} -1 & , \text{ w.p. } \frac{1}{2} \\ 1 & , \text{ w.p. } \frac{1}{2} \end{cases}$$

The Rademacher distribution has $E[\omega_i] = E[\omega_i^3] = 0$ and $E[\omega_i^2] = E[\omega_i^4] = 1$. Thus, Rademacher perturbations are able to match the first four moments of a **symmetric** distribution, providing an asymptotic refinement over the normal approximation.

Procedure

4.1 General Case

1. Obtain $\hat{\boldsymbol{\theta}}$ satisfying $\boldsymbol{\Psi}_n(\boldsymbol{\theta}) = \mathbf{0}$.

   i. Calculate the individual score contributions evaluated at $\hat{\boldsymbol{\theta}}$:

   $$\hat{\boldsymbol{\psi}}_i = \boldsymbol{\psi}_i(\hat{\boldsymbol{\theta}})$$

   ii. Estimate the Jacobian at $\hat{\boldsymbol{\theta}}$:

   $$\hat{\boldsymbol{A}}_n = -\frac{1}{n}\sum_{i=1}^{n}\frac{\partial \boldsymbol{\psi}_i}{\partial \boldsymbol{\theta}'}(\hat{\boldsymbol{\theta}})$$

   iii. Calculate the normalized score:

   $$\hat{\boldsymbol{U}}_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{\boldsymbol{A}}_n^{-1}\hat{\boldsymbol{\psi}}_i$$

   iv. Estimate the outer product:

   $$\hat{\boldsymbol{B}}_n = \frac{1}{n}\sum_{i=1}^{n}\hat{\boldsymbol{\psi}}_i\hat{\boldsymbol{\psi}}_i'$$

   v. Construct the observed test statistic:

   $$T_{\text{obs}} = \hat{\boldsymbol{U}}_n'\big(\hat{\boldsymbol{A}}_n^{-1}\hat{\boldsymbol{B}}_n\hat{\boldsymbol{A}}_n^{-T}\big)^{-1}\hat{\boldsymbol{U}}_n$$

2. For $b = 1, \cdots, B$:

   i. Draw the perturbation weights $\big(\omega_i^{(b)}\big)$.

   ii. Generate the perturbed score:

   $$\boldsymbol{U}_n^{(b)} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\omega_i^{(b)}\hat{\boldsymbol{A}}_n^{-1}\hat{\boldsymbol{\psi}}_i$$

   iii. Estimate the perturbed outer product:

   $$\boldsymbol{B}_n^{(b)} = \frac{1}{n}\sum_{i=1}^{n}\big(\omega_i^{(b)}\big)^2\hat{\boldsymbol{\psi}}_i\hat{\boldsymbol{\psi}}_i'$$

   iv. Calculate the perturbed test statistic:

   $$T^{(b)} = \big(\boldsymbol{U}_n^{(b)}\big)'\big(\hat{\boldsymbol{A}}_n^{-1}\boldsymbol{B}_n^{(b)}\hat{\boldsymbol{A}}_n^{-T}\big)^{-1}\boldsymbol{U}_n^{(b)}$$

3. Use the empirical distribution of $\{T^{(b)}\}$ to approximate that of T.

   - To estimate of p-value for T_{obs}:

   $$\hat{p} = \frac{1}{B}\sum_{b=1}^{B} I\big[T^{(b)} \geq T_{\text{obs}}\big]$$

   - To construct a confidence interval, use the empirical $\alpha/2$ and $1-\alpha/2$ percentiles of $\{T^{(b)}\}$.

4.2 Maximum Likelihood

1. Obtain the MLE $\hat{\boldsymbol{\theta}}$ satisfying $\boldsymbol{S}_n(\boldsymbol{\theta}) = \mathbf{0}$.

   i. Calculate the individual score contributions evaluated at $\hat{\boldsymbol{\theta}}$, and the overall score:

   $$\hat{\boldsymbol{s}}_i = \frac{\partial \ell_i}{\partial \boldsymbol{\theta}}(\hat{\boldsymbol{\theta}})$$

   ii. Aggregate the individual score contributions:

   $$\hat{\boldsymbol{S}}_n = \sum_{i=1}^{n} \hat{\boldsymbol{s}}_i$$

   iii. Estimate the observed information at $\hat{\boldsymbol{\theta}}_n$:

   $$\hat{\boldsymbol{J}}_n = -\sum_{i=1}^{n} \frac{\partial^2 \ell}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}(\hat{\boldsymbol{\theta}})$$

   iv. Calculate the test statistic:

   $$T_{\text{obs}} = \hat{\boldsymbol{S}}_n' \hat{\boldsymbol{J}}_n^{-1} \hat{\boldsymbol{S}}_n$$

2. For $b = 1, \cdots, B$:

   i. Draw the perturbation weights $\big(\omega_i^{(b)}\big)$.

   ii. Generate the perturbed score:

   $$\boldsymbol{S}_n^{(b)} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \omega_i^{(b)} \hat{\boldsymbol{s}}_i$$

   iii. Calculate the perturbed test statistic:

   $$T^{(b)} = \big(\hat{\boldsymbol{S}}_n^{(b)}\big)' \hat{\boldsymbol{J}}_n^{-1} \hat{\boldsymbol{S}}_n^{(b)}$$

3. Use the empirical distribution of $\{T^{(b)}\}$ to approximate that of T.