1 Counterexamples to the Bootstrap

In this section, we consider counterexamples to the bootstrap. Specifically, we will provide examples of roots R_n such that $J_n(x, \hat{P}_n)$ (the distribution of the root R_n under $\hat{P}_n$) does not converge in distribution to $J_n(x, P)$ (the distribution of the root R_n under P).

2 Extreme Order Statistic

Suppose $X_i, i = 1, \ldots, n$ are i.i.d. with distribution P. Let $\theta(P)$ denote the upper bound of the support of P. A natural estimator of $\theta(P)$ is $X_{(n)}$, where

$$X_{(1)} \leq \cdots \leq X_{(n)}$$

denote the ordered values of the data. These statistics are sometimes referred to as the order statistics of the data. Consider the root

$$R_n = n(X_{(n)} - \theta(P)) \ ,$$

where $\theta(P)$ is the upper bound of the support of P. For concreteness, suppose $P = U(0, \theta)$ where $\theta \geq 0$, so $\theta(P)$ is simply θ.

We first show that $J_n(x, P)$ converges in distribution to $J(x, P) = \Pr\{-\theta X \leq x\}$, where $X \sim \exp(1)$. Recall that $X \sim \exp(1)$ if

$$\Pr\{X \leq x\} = \begin{cases} 1 - \exp(-x) & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases} ,$$

which implies that

$$\Pr\{-X\theta \leq x\} = \Pr\{X \geq -x/\theta\} = \begin{cases} \exp(x/\theta) & \text{if } x \leq 0 \\ 1 & \text{otherwise} \end{cases} .$$

Next note that

$$\begin{aligned} \Pr\{R_n \leq x\} &= \Pr\{n(X_{(n)} - \theta) \leq x\} \\ &= \Pr\{X_{(n)} \leq \theta + \frac{x}{n}\} \\ &= \Pr\{X_i \leq \theta + \frac{x}{n}\}^n \ . \end{aligned}$$

Note that

$$\Pr\{X_i \leq \theta + \frac{x}{n}\} = \begin{cases} 0 & \text{if } x \leq -n\theta \\ \frac{1}{\theta}(\theta + \frac{x}{n}) & \text{if } -n\theta < x \leq 0 \\ 1 & \text{if } x > 0 \end{cases} .$$

Therefore,

$$\Pr\{R_n \leq x\} = \begin{cases} 0 & \text{if } x \leq -n\theta \\ (\frac{1}{\theta}(\theta + \frac{x}{n}))^n & \text{if } -n\theta < x \leq 0 \\ 1 & \text{if } x > 0 \end{cases} .$$

Note that

$$(\frac{1}{\theta}(\theta + \frac{x}{n}))^n = (1 + \frac{x}{\theta n})^n \to \exp(x/\theta)$$

because of the identity

$$\exp(x) = \lim_{n \to \infty} (1 + \frac{x}{n})^n .$$

Hence,

$$\Pr\{R_n \leq x\} \to \begin{cases} \exp(x/\theta) & \text{if } x \leq 0 \\ 1 & \text{otherwise} \end{cases} .$$

Now consider a sequence of probability distributions $P_n, n \geq 1$ such that each distribution P_n puts equal mass on n distinct points. For each n, let $X_{i,n}, i = 1, \ldots, n$ be an i.i.d. sequence of random variables with distribution P_n. The distribution $J_n(x, P_n)$ is simply the distribution of

$$n(X_{(n),n} - \theta(P_n))$$

under P_n. But, $X_{(n),n} = \theta(P_n)$ with probability

$$1 - (1 - \frac{1}{n})^n \to 1 - \exp(-1) .$$

Therefore, for any $\epsilon > 0$,

$$\begin{aligned} \Pr\{n(X_{(n),n} - \theta(P_n)) \leq -\epsilon\} &= 1 - \Pr\{n(X_{(n),n} - \theta(P_n)) > -\epsilon\} \\ &\leq 1 - \Pr\{n(X_{(n),n} - \theta(P_n)) = 0\} \to \exp(-1) . \end{aligned}$$

Choose $\epsilon > 0$ so that

$$\exp(-1) < \exp(-\epsilon/\theta) = \Pr\{-\theta X \leq -\epsilon\} \ .$$

It follows that for such ϵ,

$$\Pr\{n(X_{(n),n} - \theta(P_n)) \leq -\epsilon\} \not\to \exp(-\epsilon/\theta) \ .$$

To complete the argument, simply note that the $X_i, i = 1, \ldots, n$ are all distinct a.s., so $\hat{P}_n$ satisfies the requirements on P_n a.s. Thus, $J_n(x, \hat{P}_n)$ does not converge in distribution to $J(x, P)$ a.s.

3 Parameter on the Boundary

Suppose $X_i, i = 1, \ldots, n$ are i.i.d. with distribution P. Let $\theta(P)$ denote the mean of P. If it is known *a priori* that $\theta(P) \geq 0$, then a natural estimator of the the mean is $(\bar{X}_n)_+$, where $(x)_+ = \max\{x, 0\}$. Consider the root

$$R_n = \sqrt{n}((\bar{X}_n)_+ - (\theta(P))_+) \ .$$

(We could also consider the root $R_n = \sqrt{n}((\bar{X}_n)_+ - \theta(P))$; the results would be similar.) For concreteness, suppose $P = N(\theta, 1)$ where $\theta \geq 0$, so $\theta(P)$ is simply θ.

We first derive the asymptotic behavior of $J_n(x, P)$. Since $\theta(P) \geq 0$, we have that

$$\begin{aligned} \sqrt{n}((\bar{X}_n)_+ - (\theta(P))_+) &= \max\{\sqrt{n}\bar{X}_n, 0\} - \sqrt{n}\theta(P) \\ &= \max\{\sqrt{n}(\bar{X}_n - \theta(P)), -\sqrt{n}\theta(P))\} \ . \end{aligned}$$

Under our assumptions, $\sqrt{n}(\bar{X}_n - \theta(P)) \sim Z$, where $Z \sim N(0, 1)$. It follows that $J_n(x, P)$ converges in distribution to $J(x, P)$, where

$$J(x, P) = \begin{cases} \Pr\{(Z)_+ \leq x\} & \text{if } \theta = 0 \\ \Pr\{Z \leq x\} & \text{otherwise} \end{cases} \ .$$

For each n, let $X_{i,n}, i = 1, \ldots, n$ be an i.i.d. sequence of random variables with distribution P_n. The distribution of $J_n(x, P_n)$ is simply the distribution of

$$\sqrt{n}((\bar{X}_{n,n})_+ - (\theta(P_n))_+)$$

under P_n. Even if $\theta(P_n)$ is possibly negative, we may still write

$$\sqrt{n}((\bar{X}_{n,n})_+ - (\theta(P_n))_+) = \max\{\sqrt{n}(\bar{X}_{n,n} - \theta(P_n)) + \sqrt{n}\theta(P_n), 0\} - \max\{\sqrt{n}\theta(P_n), 0\} \ .$$

Suppose $\theta(P) = 0$. Let $c > 0$ and suppose $\sqrt{n}\theta(P_n) < -c$ for all n. For such a sequence P_n,

$$\sqrt{n}((\bar{X}_{n,n})_+ - (\theta(P_n))_+) \leq \max\{\sqrt{n}(\bar{X}_{n,n} - \theta(P_n)) - c, 0\} \ .$$

If P_n converges in distribution to P, $\theta(P_n) \to \theta(P)$, and $\sigma^2(P_n) \to \sigma^2(P)$, then we know from our earlier results that

$$\max\{\sqrt{n}(\bar{X}_{n,n} - \theta(P_n)) - c, 0\} \stackrel{d}{\to} \max\{Z - c, 0\}$$

under P_n, which is dominated by the distribution of $(Z)_+$.

To complete the argument, it suffices to show that $\hat{P}_n$ satisfies a.s. the requirements on P_n in the above discussion. By the SLLN $\hat{P}_n$ converges in distribution to P a.s., $\theta(\hat{P}_n) \to \theta(P)$ a.s., and $\sigma^2(\hat{P}_n) \to \sigma^2(P)$ a.s. It remains to determine whether $\sqrt{n}\theta(\hat{P}_n) < -c$ for all n a.s. Equivalently, we need to determine whether

$$\bar{X}_n < -\frac{c}{\sqrt{n}}$$

for all n a.s. Unfortunately, the SLLN will not suffice for this purpose. Instead, we will need the following refinement of the SLLN known as the law of the iterated logarithm (LIL):

Theorem 3.1 Let $Y_i, i = 1, \ldots, n$ be an i.i.d. sequence of random variables with distribution P on $\mathbf{R}$. Suppose $\mu(P) = 0$ and $\sigma^2(P) = 1$. Then,

$$\limsup_{n \to \infty} \frac{\bar{Y}_n}{\sqrt{\frac{2 \log \log n}{n}}} = 1 \text{ a.s.}$$

Recall that for a sequence of real numbers $a_n, n \geq 1$

$$\limsup_{n \to \infty} a_n = a$$

if and only if for any $\epsilon > 0$

$$a_n > a - \epsilon \text{ i.o.}$$

and

$$a_n < a + \epsilon$$

for all n sufficiently large. An implication of the LIL therefore is that for any $\epsilon > 0$,

$$\bar{Y}_n > (1 - \epsilon)\sqrt{\frac{2 \log \log n}{n}} \text{ i.o. a.s.}$$

Since $(1 - \epsilon)\sqrt{2 \log \log n} > c$ for all n sufficiently large, it follows that

$$\bar{Y}_n > \frac{c}{\sqrt{n}} \text{ i.o. a.s.}$$

We may apply the LIL to $Y_i = -X_i$ to conclude that

$$\bar{X}_n < -\frac{c}{\sqrt{n}} \text{ i.o. a.s.}$$

In other words,

$$\sqrt{n}\theta(\hat{P}_n) < -c \text{ i.o. a.s.}$$

Thus, $\hat{P}_n$ satisfies the requirements on P_n, at least along a subsequence, a.s., which is good enough for our purposes. It follows that at least along a subsequence, $J_n(x, \hat{P}_n)$ does not converge to $J(x, P)$ a.s.