The empirical bootstrap (continued)

Recall from last time: we consider a class $\mathcal{F}$ of functions, assumed to be Donsker. As usual, $\mathbb{P}_n$ is the empirical measure, and we introduced last time the following quantities related to the empirical bootstrap process:

$$
\mathbb{P}_n^* \equiv \frac{1}{n} \sum_i \delta_{X_i^*},
$$

$$
\mathbb{G}_n^* \equiv \sqrt{n}(\mathbb{P}_n^* - \mathbb{P}_n)
$$

$$
= \frac{1}{\sqrt{n}} \sum_i (M_{n_i} - 1)\delta_{X_i}, M_{n_i} \sim \text{Mult} \left( n; \frac{1}{n}, \ldots, \frac{1}{n} \right)
$$

By our earlier results on generalized CLT, we have convergence of the original empirical process to a brownian bridge $\mathbb{G}_p$:

$$
\mathbb{G}_n \equiv \sqrt{n}(\mathbb{P}_n - P) \xrightarrow{L} \mathbb{G}_p
$$

we end our discussion on the bootstrap with a functional delta result using Hadamard differentiability.

**Definition 1.** We define the concept of weak conditional convergence as:

$$
\sup_{h \in BL} |E_M h(\hat{\theta}_n^* - \theta_n) - Eh(t)| \xrightarrow{P} 0,
$$

where $BL$ is the class of bounded Lipschitz functions, and $E_M$ the expectation with respect to the $M_{n_i}$ vectors.
Theorem 1. If \( \mathcal{F} \) is assumed to be a Donskar class that satisfies an envelope condition, then

\[
\sup_{h \in BL_1(l^\infty(\mathcal{F}))} |E_M h(G_n^*) - E h(G_p)| \xrightarrow{P} 0
\]

Theorem 2. Let \( \phi \) be Hadamard differentiable at \( \theta \) tangentially \( \mathbb{D}_0 \). Assume that \( \sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} T \) and

\[
\sup_{h \in BL(\mathbb{D})} |E_M h(\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n)) - E h(T)| \xrightarrow{P} 0.
\]

Then, we have conditionally in probability on \( X_1, \ldots, X_n \) that

\[
\sqrt{n}(\phi(\hat{\theta}_n^*) - \phi(\hat{\theta}_n)) \xrightarrow{d} \phi'_\theta(T).
\]

Example 1. Empirical cumulative distribution function. It is known that the class of functions \( \{(-\infty, t]\} \) is Donskar, so the conditions for theorem 2 are satisfied, using theorem 1. Therefore, the bootstrap works for Hadamard differentiable \( \phi \) for

\[
\sqrt{n}(\phi(\hat{F}_n^*) - \phi(\hat{F}_n)).
\]

This includes quantiles, trimmed means\(^1\), etc.

Testing and Asymptotic Relative Efficiency (ARE)

We now consider a testing setup constituted of

- a null hypothesis \( H_0 : \theta \in \Theta_0 \),
- an alternative hypothesis \( H_1 : \theta \in \Theta_1 \),
- a critical region \( K_n \),
- a statistic \( T_n \).

\(^1\)Rejecting the top and bottom 25% of the data is an example of a trimmed mean statistics. The goal is to obtain a more robust statistics.
We will need the following concepts

**Definition 2.** The power function is defined as:

\[ \pi_n(\theta) = P_{\theta}(T_n \in K_n); \]

a test is said to be asymptotic of level \( \alpha \) if

\[ \limsup_{n \to \infty} \sup_{\theta \in \Theta_0} \pi_n(\theta) \leq \alpha \]

We now consider the problem of comparing tests. Clearly, if

\[ \pi_n(\theta) \leq \tilde{\pi}_n(\theta) \text{ for } \theta \in \Theta_0 \]
\[ \pi_n(\theta) \geq \tilde{\pi}_n(\theta) \text{ for } \theta \in \Theta_1 \]

then \( \pi \) is better than \( \tilde{\pi} \).

These quantities are hard to compute in practice, so we will want to work with asymptotics instead. The problem is that the asymptotic distribution will always be a step function for reasonable statistics, giving no hint on the relative performances of the two testing procedures.

**Example 2.** Sign test: let \( X_i \) be iid from a distribution with unique median \( \theta \). Consider the sign statistic \( T_n \equiv \frac{1}{n} \sum_i \mathbb{1}[X_i > 0] \), and let \( F(x - \theta) \) be the cumulative distribution function. We have:

\[ ET_n = 1 - F(-\theta) \equiv \mu(\theta) \]
\[ \text{Var } T_n = \frac{(1 - F(-\theta))(F(-\theta))}{n} \equiv \frac{\sigma^2(\theta)}{n} \]

By CLT,

\[ \sqrt{n}(T_n - \mu(\theta)) \xrightarrow{L} N(0, \sigma^2(\theta)) \]
under $H_0$, $\mu(0) = \frac{1}{2}$, $\sigma^2(0) = \frac{1}{4}$. Hence, under $H_0$,

$$\sqrt{n}(T_n - \frac{1}{2}) \xrightarrow{d} N(0, \frac{1}{4})$$

The test that rejects the null hypothesis if

$$\sqrt{n}(T_n - \frac{1}{2}) > \frac{1}{2}z_\alpha$$

has power function:

$$\pi_n(\theta) = P_\theta(\sqrt{n}(T_n - \mu(\theta)) - \frac{1}{2}z_\alpha > \sqrt{n}(\mu(\theta) - \mu(0)))$$

$$= 1 - \Phi\left(\frac{\frac{1}{2}z_\alpha - \sqrt{n}(F(0) - F(-\theta))}{\sigma(\theta)}\right) + o(1)$$

Since $F(0) - F(-\theta) > 0$, for $\alpha = \alpha_n \to 0$ sufficiently slowly, we have

$$\lim_{n \to \infty} \pi_n(\theta) \to \begin{cases} 0 & \text{for } \theta = 0 \\ 1 & \text{for } \theta > 0 \end{cases}$$

We now move toward Pitman’s ARE. The idea is to make the hypothesis harder to distinguish as $n \to \infty$. If the rate at which this occurs is chosen properly, the asymptotic distributions become nontrivial, and give a way to rank testing procedures.

The new setup is the following: We define the local limiting power function as

$$\pi(h) \equiv \lim_{n \to \infty} \pi_n\left(\frac{h}{\sqrt{n}}\right)$$

and assume that the following holds:

$$\frac{\sqrt{n}(T_n - \mu(\theta_n))}{\sigma(\theta_n)} \frac{\theta_n}{\theta_0} \xrightarrow{d} N(0, 1).$$

Where the notation means convergence in law, under the distribution $\theta_n$. In particular, under the null hypothesis,
\[
\sqrt{n}(T_n - \mu(0)) \xrightarrow{\mathcal{L}} N(0, \sigma^2(0))
\]

implies that the test that reject if

\[
\sqrt{n}(T_n - \mu(0)) > \sigma(0)z_{\alpha}
\]

is asymptotic of level \(\alpha\). The power function is:

\[
\pi_n(\theta_n) = P_{\theta_n}(\sqrt{n}(T_n - \mu(\theta_n)) > \sigma(0)z_{\alpha} - \sqrt{n}(\mu(\theta_n) - \mu(0)))
\]

For \(\theta_n = \frac{h}{\sqrt{n}}\), \(\sqrt{n}(\mu(\theta_n) - \mu(0))\) converges to \(h\mu'(0)\) for differentiable \(\mu(\cdot)\).

Now, if \(\sigma(\theta_n) \to \sigma(0)\), then we have

\[
\pi_n\left(\frac{h}{\sqrt{n}}\right) \to 1 - \Phi\left(z_{\alpha} - h\frac{\mu'(0)}{\sigma(0)}\right)
\]

**Definition 3.** The Pitman ARE is defined to be the ratio of the squares of the slopes.

**Example 3.** Going back to the sign test example, we obtained that:

\[
\mu(\theta) = 1 - F(-\theta), \sigma^2(\theta) = (1 - F(-\theta))F(-\theta)
\]

and hence (omitting the details of the derivation) the slope is:

\[
\frac{f(0)}{(1/2)}
\]

Comparing this with the t-test for different distributions, we obtain:

<table>
<thead>
<tr>
<th>Distribution</th>
<th>(\frac{\sigma^2}{\sigma^2})</th>
</tr>
</thead>
<tbody>
<tr>
<td>logistic</td>
<td>(\frac{\pi^2}{12})</td>
</tr>
<tr>
<td>normal</td>
<td>(\frac{\pi}{12})</td>
</tr>
<tr>
<td>laplace</td>
<td>2</td>
</tr>
<tr>
<td>uniform</td>
<td>(\frac{1}{2})</td>
</tr>
</tbody>
</table>

where a ratio less than one indicates the superiority of the t-test, and vice-versa.