THE BOOTSTRAP

The bootstrap is a method for estimating sampling distributions.

We usually want to estimate sampling distributions of pivot quantities, e.g.

$$\sqrt{n}(\bar{X}_n - \mu) \quad \text{or} \quad \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \quad \text{or} \quad \frac{\sqrt{n}(\bar{X}_n - \mu)}{\hat{\sigma}_n}$$

where $X_1, \dots, X_n$ are iid with $\mathbb{E}X_1 = \mu$, $\operatorname{Var} X_1 = \sigma^2$, $\bar{X}_n = n^{-1}\sum_{i=1}^n X_i$, $\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X}_n)^2$.

We will refer to the above pivot types as "unstandardized", "standardized" and "studentized."

We will focus on using these pivots to build C.I.s for μ.

Define the cdfs in the unstandardized, standardized, and studentized cases as

$$G_{n,U}(x) = P\left(\sqrt{n}(\bar{X}_n - \mu) \le x\right)$$

$$G_n(x) = P\left(\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \le x\right)$$

$$G_{n,S}(x) = P\left(\frac{\sqrt{n}(\bar{X}_n - \mu)}{\hat{\sigma}_n} \le x\right) \quad \text{for all } x \in \mathbb{R},$$

and let $G_{n,U}^{-1}, G_n^{-1}, G_{n,S}^{-1} : (0,1) \to \mathbb{R}$ be the corresponding quantile functions.

__Exercise:__ Show that a $(1-\alpha)100\%$ C.I. for μ based on $\sqrt{n}(\bar{X}_n - \mu)$ is given by

$$\left(\bar{X}_n - G_{n,U}^{-1}(1-\alpha/2)\frac{1}{\sqrt{n}},\ \bar{X}_n - G_{n,U}^{-1}(\alpha/2)\frac{1}{\sqrt{n}}\right).$$

__Solution:__ Begin by writing

$$P\left(G_{n,U}^{-1}(\alpha/2) < \sqrt{n}(\bar{X}_n - \mu) < G_{n,U}^{-1}(1-\alpha/2)\right) = 1-\alpha$$

Rearranging the probability statement gives

$$P\left(\bar{X}_n - G_{n,U}^{-1}(1-\alpha/2)\frac{1}{\sqrt{n}} < \mu < \bar{X}_n - G_{n,U}^{-1}(\alpha/2)\frac{1}{\sqrt{n}} \right) = 1-\alpha,$$

which gives the confidence interval.

The bootstrap is a method for estimating $G_{n,U}$, G_n, or $G_{n,S}$, which describe the sampling distributions of their pivot quantities, and the associated quantile functions $G_{n,U}^{-1}$, G_n^{-1}, and $G_{n,S}^{-1}$.

I.I.D. Bootstrap for the mean:

Continuing with $X_1,\dots,X_n$ iid with mean μ and variance σ^2, let

$X_1^*,\dots,X_n^* \mid X_1,\dots,X_n$ be iid with cdf $\mathbb{F}_n(x) = n^{-1}\sum_{i=1}^n \mathbb{1}(X_i \leq x)$.

Letting

$$\bar{X}_n^* = n^{-1}(X_1^* + \dots + X_n^*) \quad \text{and} \quad \hat{\sigma}_n^{*2} = \frac{1}{n}\sum_{i=1}^n (X_i^* - \bar{X}_n^*)^2$$

and noting that

$$\mathbb{E}\left[X_1^*,\dots,X_n^* \mid X_1,\dots,X_n \right] = \bar{X}_n$$

$$\mathrm{Var}\left[X_1^* \mid X_1,\dots,X_n \right] = n^{-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2 =: \hat{\sigma}_n^2,$$

we define bootstrap estimators of $G_{n,U}$, G_n, and $G_{n,S}$ as

$$\hat{G}_{n,U}(x) = P\left(\sqrt{n}(\bar{X}_n^* - \bar{X}_n) \leq x \mid X_1,\dots,X_n \right)$$

$$\hat{G}_n(x) = P\left(\sqrt{n}\frac{(\bar{X}_n^* - \bar{X}_n)}{\hat{\sigma}_n} \leq x \mid X_1,\dots,X_n \right)$$

$$\hat{G}_{n,S}(x) = P\left(\sqrt{n}\frac{(\bar{X}_n^* - \bar{X}_n)}{\hat{\sigma}_n^*} \leq x \mid X_1,\dots,X_n \right) \quad \text{for all } x \in \mathbb{R},$$

and $\hat{G}_{n,U}^{-1}$, $\hat{G}_n^{-1}$, $\hat{G}_{n,S}^{-1}$ as the inverse functions to $\hat{G}_{n,U}$, $\hat{G}_n$, and $\hat{G}_{n,S}$.

The $(1-\alpha)*100\%$ bootstrap C.I.s for μ based on the unstandardized, standardized, and studentized pivots are given by

Unstandardized: $$\left(\bar{X}_n - \hat{G}_{n,U}^{-1}(1-\alpha/2)\frac{1}{\sqrt{n}},\ \bar{X}_n - \hat{G}_{n,U}^{-1}(\alpha/2)\frac{1}{\sqrt{n}}\right)$$

Standardized: $$\left(\bar{X}_n - \hat{G}_n^{-1}(1-\alpha/2)\frac{\sigma}{\sqrt{n}},\ \bar{X}_n - \hat{G}_n^{-1}(\alpha/2)\frac{\sigma}{\sqrt{n}}\right)$$

Studentized: $$\left(\bar{X}_n - \hat{G}_{n,S}^{-1}(1-\alpha/2)\frac{\hat{\sigma}_n}{\sqrt{n}},\ \bar{X}_n - \hat{G}_{n,S}^{-1}(\alpha/2)\frac{\hat{\sigma}_n}{\sqrt{n}}\right)$$

Note that only the Unstandardized and the Studentized intervals are feasible, since the Standardized uses the true value of σ.

Exercise: Verify the identities

$$\mathbb{E}\left[X_1^*,\dots,X_n^* \,\middle|\, X_1,\dots,X_n\right] = \bar{X}_n$$

$$\mathrm{Var}\left[X_1^* \,\middle|\, X_1,\dots,X_n\right] = \hat{\sigma}_n^2$$

Solution: The random variable $X_1^* | X_1,\dots,X_n$ has probability mass function

$$p_n(x) = n^{-1}\,\mathbb{1}(x \in \{X_1,\dots,X_n\}),$$

according to which

$$\mathbb{E}\left[X_1^* \,\middle|\, X_1,\dots,X_n\right] = \sum_{x\in\{X_1,\dots,X_n\}} x\cdot n^{-1} = n^{-1}\sum_{i=1}^n X_i = \bar{X}_n$$

$$\mathbb{E}\left[X_1^{*2} \,\middle|\, X_1,\dots,X_n\right] = \sum_{x\in\{X_1,\dots,X_n\}} x^2\cdot n^{-1} = n^{-1}\sum_{i=1}^n X_i^2,$$

so that

$$\mathrm{Var}\left[X_i^* \,\middle|\, X_1, \ldots, X_n\right] = n^{-1}\sum_{i=1}^n X_i^2 - \left(n^{-1}\sum_{i=1}^n X_i\right)^2$$
$$= n^{-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2.$$

Some bootstrap notation: Let P_*, $\mathbb{E}_*$, and Var_* be operators such that

$$P_*(\cdot) = P(\cdot \mid X_1, \ldots, X_n)$$
$$\mathbb{E}_*(\cdot) = \mathbb{E}(\cdot \mid X_1, \ldots, X_n)$$
$$\mathrm{Var}_*(\cdot) = \mathrm{Var}(\cdot \mid X_1, \ldots, X_n),$$

so that they represent conditional probability, expectation, and variance given $X_1, \ldots, X_n$.

Thus, P_*, $\mathbb{E}_*$, and Var_* treat the bootstrap sample $X_1^*, \ldots, X_n^*$ as random variables and the original sample $X_1, \ldots, X_n$ as fixed.

Implementing the bootstrap:

Consider how to obtain the bootstrap estimator $\hat{G}_{n,U}$ of $G_{n,U}$. We have

$$\hat{G}_{n,U}(x) = P_*\left(\sqrt{n}\left(\bar{X}_n^* - \bar{X}_n\right) \le x\right)$$
$$= P_*\left(\bar{X}_n^* \le \bar{X}_n + x/\sqrt{n}\right)$$
$$= P_*\left(\sum_{i=1}^n X_i^* \le \sum_{i=1}^n X_i + x\sqrt{n}\right),$$

(distinct values it could take)

where $\sum_{i=1}^n X_i^*$ has a discrete distribution with as many as $\binom{2n-1}{n}$ atoms.

It is possible to evaluate cumulative probabilities for $\sum_{i=1}^{n} X_i^*$ exactly, but this can be very computationally expensive.

We therefore almost always use Monte Carlo simulation to approximate the bootstrap cdf estimators $\hat{G}_{n,U}$, $\hat{G}_n$, $\hat{G}_{n,S}$ and their quantile functions.

Here is how it goes:

<u>Monte Carlo Approximation to Bootstrap estimators $\hat{G}_{n,U}$, $\hat{G}_n$, and $\hat{G}_{n,S}$:</u>

For $b = 1, \ldots, B$, B large (≥ 500, say):

Draw $X_1^{*(b)}, \ldots, X_n^{*(b)}$ with replacement from $X_1, \ldots, X_n$

Compute

$$T_{n,U}^{*(b)} = \sqrt{n}\left(\bar{X}_n^{*(b)} - \bar{X}_n\right)$$

$$\text{or} \quad T_n^{*(b)} = \sqrt{n}\left(\bar{X}_n^{*(b)} - \bar{X}_n\right)/\hat{\sigma}_n$$

$$\text{or} \quad T_{n,S}^{*(b)} = \sqrt{n}\left(\bar{X}_n^{*(b)} - \bar{X}_n\right)/\hat{\sigma}_n^{*(b)},$$

where $\bar{X}_n^{*(b)} = n^{-1}\sum_{i=1}^{n} X_i^{*(b)}$ and $\left(\hat{\sigma}_n^{*(b)}\right)^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i^{*(b)} - \bar{X}_n^{*(b)}\right)^2$.

Then set

$$\hat{G}_{n,U}(x) = B^{-1}\sum_{b=1}^{B} \mathbb{1}\left(T_{n,U}^{*(b)} \leq x\right)$$

$$\hat{G}_n(x) = B^{-1}\sum_{b=1}^{B} \mathbb{1}\left(T_n^{*(b)} \leq x\right)$$

$$\hat{G}_{n,S}(x) = B^{-1}\sum_{b=1}^{B} \mathbb{1}\left(T_{n,S}^{*(b)} \leq x\right) \quad \text{for all } x \in \mathbb{R}.$$

After sorting the bootstrap pivot quantities such that

$$T_{n,U}^{*(1)} \leq \cdots \leq T_{n,U}^{*(B)}$$

$$T_n^{*(1)} \leq \cdots \leq T_n^{*(B)}$$

$$T_{n,S}^{*(1)} \leq \cdots \leq T_{n,S}^{*(B)},$$

set

$$\hat{G}_{n,U}^{-1}(u) = T_{n,U}^{*(\lceil Bu \rceil)}$$

$$\hat{G}_n^{-1}(u) = T_n^{*(\lceil Bu \rceil)}$$

$$\hat{G}_{n,S}^{-1}(u) = T_{n,S}^{*(\lceil Bu \rceil)} \quad \text{for all } u \in (0,1).$$

Motivation for the bootstrap:

Why not just use

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \longrightarrow N(0,1) \quad \text{in dist. as } n \to \infty$$

and build the C.I. $\bar{X}_n \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ or $\bar{X}_n \pm z_{\alpha/2} \frac{\hat{\sigma}_n}{\sqrt{n}}$?

The idea is that the bootstrap estimator of G_n may be more accurate than the approximation to G_n by the limiting cdf — the standard Gaussian cdf Φ.

If the convergence to Normality is slow, the (studentized) boostrap C.I. can achieve closer-to-nominal coverage when the sample size is small.

As a first result, we show that the bootstrap "works."

We will prove results for the "standardized" pivot here, as they are the most straightforward.

Theorem (Consistency of the bootstrap for the mean):

Let $X_1, \dots, X_n$ be iid with $\mathbb{E}X_1 = \mu$, $\operatorname{Var} X_1 = \sigma^2 < \infty$. Then

$$(😎) \qquad \sup_{x \in \mathbb{R}} \left| \hat{G}_n(x) - G_n(x) \right| \to 0 \quad \text{w.p.1.} \quad \text{as} \quad n \to \infty.$$

To prove the above result, we are going to need some tools.
But first, note that, by the central limit theorem, we have

$$\sup_{x \in \mathbb{R}} \left| G_n(x) - \Phi(x) \right| \to 0 \quad \text{w.p.1} \quad \text{as} \quad n \to \infty,$$

that is, the cdf of $\sqrt{n}(\bar{X}_n - \mu)/\sigma$ converges to that of the $N(0,1)$ distribution.
So to prove (😎), we just need to show

$$(🍦) \qquad \sup_{x \in \mathbb{R}} \left| \hat{G}_n(x) - \Phi(x) \right| \to 0 \quad \text{w.p.1} \quad \text{as} \quad n \to \infty.$$

So, we essentially prove that $\sqrt{n}(\bar{X}_n - \mu)/\sigma$ and $\sqrt{n}(\bar{X}_n^* - \bar{X}_n)/\hat{\sigma}_n$ have the same limiting distribution — the $N(0,1)$.

Our primary tool for establishing (🍦) will be the Berry-Esseen Theorem:

Berry-Esseen Theorem:

For $X_1, \dots, X_n$ iid with $\mathbb{E}X_1 = \mu$, $\operatorname{Var} X_1 = \sigma^2 < \infty$, and $\mathbb{E}|X_1|^3 < \infty$, we have

$$\sup_{x \in \mathbb{R}} \left| P\left(\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \leq x \right) - \Phi(x) \right| \leq C \cdot \frac{\mathbb{E}|X_1 - \mu|^3}{\sigma^3 \sqrt{n}}$$

for each $n \geq 1$, where $C \in \left[0, \sqrt{\frac{2}{\pi}}\left(\frac{5}{2} + \frac{12}{\pi}\right)\right]$.

[See pg 361 of Lahiri & Athreya (2006)]

We will also make use of the following strong law of large numbers:

Special instance of the Marcinkiewz - Zygmund SLLNs:

Let $Y_1, \ldots, Y_n$ be iid and let $p \in (0,1)$. Then, if $\mathbb{E}|Y_1|^p < \infty$,

$$\frac{\sum_{i=1}^{n} Y_i}{n^{1/p}} \longrightarrow 0 \quad \text{w.p. } 1.$$

We will also need a couple of inequalities; first, for a rv X, let

$$|X|_p = \left(\mathbb{E}|X|^p\right)^{1/p} \quad \text{for} \quad p \in (0,\infty).$$

This is called the L_p-norm of a random variable.

Minkowski's inequality applied to random variables:

For any rvs $X, Y \in \mathbb{R}$, $p \in (1,\infty)$

$$|X-Y|_p \leq |X|_p + |Y|_p.$$

So Minkowski's inequality gives the triangle inequality for the L_p-norm.

Jensen's inequality:

If $g: \mathbb{R} \to \mathbb{R}$ is a convex function, then for any rv X, we have

$$g(\mathbb{E}X) \leq \mathbb{E}\, g(X),$$

provided $\mathbb{E}|X| < \infty$ and $\mathbb{E}\,|g(X)| < \infty$.

Proof of ():

By the Berry-Esseen theorem we have, for each $n \geq 1$,

$$\sup_{x\in\mathbb{R}} \left| P_*\left(\frac{\sqrt{n}\,(\bar{X}_n^* - \bar{X}_n)}{\hat{\sigma}_n} \le x \right) - \Phi(x) \right| \le C \cdot \frac{\mathbb{E}_* \left| X_1^* - \bar{X}_n \right|^3}{\hat{\sigma}_n^3 \sqrt{n}} =: \Delta_n .$$

Now,

$$\left(\mathbb{E}_* \left| X_1^* - \bar{X}_n \right|^3 \right)^{1/3} \le \left(\mathbb{E}_* |X_1^*|^3 \right)^{1/3} + \left(|\bar{X}_n|^3 \right)^{1/3}, \quad \text{(Minkowski's Inequality)}$$

$$\le 2 \left(\mathbb{E}_* |X_1^*|^3 \right)^{1/3},$$

where the second inequality comes from

Jensen's with $g(x) = |x|^3$, which is a convex function

$$|\bar{X}_n|^3 = |\mathbb{E} X_1^*|^3 \le \mathbb{E} |X_1^*|^3 \implies |\bar{X}_n| \le \left(\mathbb{E} |X_1^*|^3 \right)^{1/3}$$

So we have

$$\mathbb{E}_* |X_1^* - \bar{X}_n|^3 \le 2^3\, \mathbb{E} |X_1^*|^3 = 2^3 \frac{1}{n} \sum_{i=1}^n |X_i|^3 ,$$

giving

$$\Delta_n \le C \cdot \frac{2^3}{\hat{\sigma}_n^3} \frac{1}{n^{3/2}} \sum_{i=1}^n |X_i|^3 .$$

Since $\hat{\sigma}_n^3 \to \sigma^3$ w.p.1 as $n \to \infty$, we just need to show

$$(\Diamond) \qquad \frac{1}{n^{3/2}} \sum_{i=1}^{n} |X_i|^3 \to 0 \quad \text{w.p.1} \quad \text{as} \quad n \to \infty.$$

Let $Y_i = |X_i|^3$, $i = 1, \dots, n$. Then $\mathbb{E}|Y_1|^{2/3} = \mathbb{E}(|X_1|^3)^{2/3} = \mathbb{E}|X_1|^2 < \infty$, giving

$$\frac{1}{n^{3/2}} \sum_{i=1}^{n} Y_i \to 0 \quad \text{w.p.1} \quad \text{as} \quad n \to \infty$$

by the M-Z SLLNs, which establishes $(\Diamond)$, completing the proof.

$\square$

Note that we glibly stated $\hat{\sigma}_n^3 \to \sigma^3$ w.p.1. This can be proven using Kolmogorov's SLLN:

<u>Kolmogorov's SLLN</u>:

For $X_1, \dots, X_n$

$$\bar{X}_n \to c \quad \text{w.p.1} \quad \text{for some } c \in \mathbb{R} \iff \mathbb{E}|X_1| < \infty,$$

in which case $c = \mathbb{E}X_1$.

<u>Exercise</u>: Show that for $X_1, \dots, X_n$ iid with $\mathbb{E}X_1 = \mu$ and $\mathbb{E}|X_1|^2 = \mu_2 < \infty$,

(i) $\quad n^{-1} \sum_{i=1}^{n} X_i \to \mu \quad$ w.p.1 as $n \to \infty$.

(ii) $\quad n^{-1} \sum_{i=1}^{n} X_i^2 \to \mu_2 \quad$ w.p.1 as $n \to \infty$.

<u>Solution</u>: For (i),

$$\mathbb{E}|X_1| \le \left(\mathbb{E}|X_1|^2\right)^{1/2} < \infty,$$

by Hölder's inequality, so the Kolmogorov SLLN gives (i).

For (ii) let $Y_i = |X_i|^2$, $i=1,\dots,n$. Then $\mathbb{E}|Y_1| < \infty$, so that

$$n^{-1}\sum_{i=1}^n Y_i \to \mathbb{E}Y_1 = \mathbb{E}|X_1|^2 = \mu_2 \quad \text{w.p. 1. as } n \to \infty,$$

by Kolmogorov's SLLN.

By the continuous mapping theorem, (i), (ii) $\Rightarrow$ $\hat{\sigma}_n^3 \to \sigma^3$ w.p. 1 as $n \to \infty$.

WHICH PIVOT IS BEST?

In practice we must use the unstandardized pivot or the studentized pivot to build C.I.s.

We find that the bootstrap based on the studentized pivot gives a more accurate approximation to the true sampling distribution than

- the bootstrap based on the unstandardized pivot
- the Normal limiting distribution.

But to make such comparisons, we must learn about Edgeworth expansions...