Lecture 24
Bootstrap confidence intervals
**Confidence Intervals**

$\theta$: an unknown parameter of interest

We want to find limits $\underline{\theta}$ and $\bar{\theta}$ such that

$$P(\underline{\theta} \leq \theta) = P(\theta \leq \bar{\theta}) = 1 - \alpha$$

$G(t) = P(\sqrt{n}(\hat{\theta} - \theta) \leq t)$

If $G^{-1}(1 - \alpha)$ is known, then

$$\underline{\theta} = \hat{\theta} - G^{-1}(1 - \alpha)/\sqrt{n}$$

is an exact 100$(1 - \alpha)$% lower confidence limit for $\theta$

Traditional asymptotic approach:

$$G(t) \to \Phi(t/\sigma), \quad n \to \infty$$

$\Phi$: the standard normal cdf

$\sigma$: an unknown scale parameter

$\hat{\sigma}$: a consistent estimator of $\sigma$, $G(\hat{\sigma}t) \to \Phi(t)$

Normal approximation (NA) 100$(1 - \alpha)$% confidence limits are

$$\underline{\theta}_N = \hat{\theta} - \hat{\sigma}z_{1-\alpha}/\sqrt{n}, \quad \bar{\theta}_N = \hat{\theta} + \hat{\sigma}z_{1-\alpha}/\sqrt{n}$$

$z_a = \Phi^{-1}(a)$
Bootstrap Confidence Intervals

(1) The hybrid bootstrap (HB)
A bootstrap estimator of \( G(t) = P(\sqrt{n}(\hat{\theta} - \theta) \leq t) \) is

\[
\hat{G}(t) = P_*(\sqrt{n}(\hat{\theta}^* - \hat{\theta}) \leq t)
\]

\( G^{-1}(1 - \alpha) \) can be estimated by \( \hat{G}^{-1}(1 - \alpha) \)

HB lower and upper confidence limits:

\[
\theta_{HB} = \hat{\theta} - \hat{G}^{-1}(1 - \alpha)/\sqrt{n}
\]

\[
\hat{\theta}_{HB} = \hat{\theta} - \hat{G}^{-1}(\alpha)/\sqrt{n}
\]

If \( G(t) \) is nearly symmetric, then \(-\hat{G}^{-1}(\alpha)\) can be replaced by \(\hat{G}^{-1}(1 - \alpha)\)

Hybrid: Use bootstrap estimate in the construction of confidence interval based on the sampling distribution of \( \hat{\theta} - \theta \).
(2) Bootstrap-t (BT)

\[ G(t) = P(\sqrt{n}(\hat{\theta} - \theta) \leq t) \rightarrow \Phi(t/\sigma) \]

\( \hat{\sigma} \): a consistent estimator of \( \sigma \)

Studentized statistic:

\[ \sqrt{n}(\hat{\theta} - \theta)/\hat{\sigma} \]

\[ H(t) = P \left( \frac{\sqrt{n}(\hat{\theta} - \theta)}{\hat{\sigma}} \leq t \right) \rightarrow \Phi(t) \]

\( \hat{\sigma}^* \): a bootstrap analog of \( \hat{\sigma} \)

A bootstrap estimator of \( H(t) \):

\[ \hat{H}(t) = P_*(\sqrt{n}(\hat{\theta}^* - \hat{\theta})/\hat{\sigma}^* \leq t) \]

BT lower and upper confidence limits:

\[ \theta_{\text{BT}} = \hat{\theta} - \hat{\sigma} \hat{H}^{-1}(1 - \alpha)/\sqrt{n} \]

\[ \bar{\theta}_{\text{BT}} = \hat{\theta} + \hat{\sigma} \hat{H}^{-1}(1 - \alpha)/\sqrt{n} \]

This is also a hybrid method

It is called BT because it is based on the studentized variable

\[ \sqrt{n}(\hat{\theta} - \theta)/\hat{\sigma} \]
(3) The bootstrap percentile (BP)
Bootstrap distribution (histogram)

\[ K(t) = P_*(\hat{\theta}^* \leq t) \approx \frac{1}{B} \left( \text{# of times } \hat{\theta}^{*b} \leq t \right) \]

100(1 − \alpha)\% BP lower confidence limit

\[ \theta_{BP} = K^{-1}(\alpha) = \inf\{t : K(t) \geq \alpha\} \]

\[ \theta_{BP} \approx \alpha B \text{th ordered value of } \hat{\theta}^{*1}, ..., \hat{\theta}^{*B} \]

100(1 − \alpha)\% BP upper confidence limit

\[ \bar{\theta}_{BP} = K^{-1}(1 - \alpha) = \inf\{t : K(t) \geq 1 - \alpha\} \]

\[ \bar{\theta}_{BP} \approx (1 - \alpha) B \text{th ordered value of } \hat{\theta}^{*1}, ..., \hat{\theta}^{*B} \]

There are two improvements over the BP
Accuracy: Justification

Properties of the bootstrap percentile (BP)
Lower confidence bound (level $1 - \alpha$)

$$\theta_{BP} = K^{-1}(\alpha), \quad K(t) = P_*(\hat{\theta}^* \leq t)$$

Assumption A:
There is a monotone transformation $\phi$ such that

$$P(\hat{\phi} - \phi \leq t) = \Psi(t) \quad \text{for all } t \text{ and all } P$$

$\phi = \phi(\theta), \ \hat{\phi} = \phi(\hat{\theta})$

$\Psi$: a continuous, increasing cdf symmetric about 0

Result 1. If $\phi$ is known, then

$$\theta_E = \phi^{-1}(\hat{\phi} + z_\alpha), \quad z_\alpha = \Psi^{-1}(\alpha),$$

is an exact $100(1 - \alpha)\%$ lower confidence bound, i.e.,

$$P(\theta_E \leq \theta) = 1 - \alpha$$
Proof

\[
P(\theta_E \leq \theta) = P(\phi(\theta_E) \leq \phi(\theta))
\]
\[
= P\left(\phi(\phi^{-1}(\hat{\phi} + z_\alpha)) \leq \phi\right)
\]
\[
= P\left(\hat{\phi} + z_\alpha \leq \phi\right)
\]
\[
= P\left(\hat{\phi} - \phi \leq -z_\alpha\right)
\]
\[
= \Psi(-z_\alpha)
\]
\[
= 1 - \Psi(z_\alpha)
\]
\[
= 1 - \alpha
\]
Result 2. Under Assumption A, 

\[ \theta_{BP} = \theta_E. \]

- This means that the BP interval is exactly correct under Assumption A.
- To use the BP, we don’t need to know \( \phi \).
- The derivation of \( \phi \) is replaced by computation.
- If Assumption A holds approximately (when \( n \) is large), then the BP intervals are approximately correct.
Proof

Let \( w = \phi(\theta_{BP}) - \hat{\phi} \) and \( \hat{\phi}^* = \phi(\hat{\theta}^*) \)

Since Assumption A holds for all \( P \) (including \( P_* \)),

\[
\Psi(w) = P_*(\hat{\phi}^* - \hat{\phi} \leq w) \\
= P_*(\hat{\phi}^* - \hat{\phi} \leq \phi(\theta_{BP}) - \hat{\phi}) \\
= P_*(\phi(\hat{\theta}^*) \leq \phi(\theta_{BP})) \\
= P_*(\hat{\theta}^* \leq \theta_{BP}) \\
= K(\theta_{BP}) \\
= \alpha
\]

By the uniqueness of the quantile,

\[
w = \Psi^{-1}(\alpha) = z_\alpha
\]

Then

\[
\phi(\theta_{BP}) = \hat{\phi} + z_\alpha
\]

Hence

\[
\theta_{BP} = \phi^{-1}(\hat{\phi} + z_\alpha) = \theta_E
\]
Is Assumption A too strong?

**Assumption B:**
There is a monotone transformation \( \phi \) and a constant \( z_0 \) such that

\[
P(\hat{\phi} - \phi + z_0 \leq t) = \Psi(t) \quad \text{for all } t \text{ and all } P
\]

\( \Psi \): a continuous, increasing cdf symmetric about 0
If \( z_0 = 0 \), then Assumption B reduces to Assumption A.
\( z_0 \) is a bias correction.
Since Assumption B holds for all \( P \) (including \( P_* \)),

\[
K(\hat{\theta}) = P_*(\hat{\phi}^* \leq \hat{\theta}) = P_*(\hat{\phi}^* \leq \phi) = P_*(\hat{\phi}^* - \phi + z_0 \leq z_0) = \Psi(z_0)
\]

i.e., \( z_0 = \Psi^{-1}(K(\hat{\theta})) \)
Result 3. $\theta_E = \phi^{-1}(\hat{\phi} + z_\alpha + z_0)$ is an exact $100(1 - \alpha)\%$ lower confidence limit for $\theta$ under Assumption B

Proof: Similar to the proof of result 1.

We need to know $\phi$ and $\Psi$ in order to use $\theta_E$

The BC confidence limits are

$$\underline{\theta}_{BC} = K^{-1}(\Psi(z_\alpha + 2z_0)) = K^{-1}(\Psi(z_\alpha + 2\Psi^{-1}(K(\hat{\theta}))))$$

$$\overline{\theta}_{BC} = K^{-1}(\Psi(z_{1-\alpha} + 2\Psi^{-1}(K(\hat{\theta}))))$$

When bootstrap Monte Carlo of size $B$ is applied,

$$\underline{\theta}_{BC} \approx B\Psi(z_\alpha + 2z_0)$$th ordered value of $\hat{\theta}^*1, \ldots, \hat{\theta}^*B$

$$\overline{\theta}_{BC} \approx B\Psi(z_{1-\alpha} + 2z_0)$$th ordered value of $\hat{\theta}^*1, \ldots, \hat{\theta}^*B$

BC = BP if $K(\hat{\theta}) = 0.5$, since $\Psi^{-1}(0.5) = 0$

i.e., $\hat{\theta}$ is the median of $K(t)$ (bootstrap histogram)

Thus, BC is a bias-corrected BP

It shifts the BP bound towards left or right according to $z_0$

BC does not need the explicit form of $\phi$

But BC requires the form of $\Psi$ (typically, $\Psi = \Phi$)
Result 4. Under Assumption B,

\[ \theta_{BC} = \theta_E \]

BC is exactly correct if Assumption B holds. If Assumption B holds approximately, then BC is approximately correct

Proof: For any \( \alpha \),

\[
1 - \alpha = \Psi(z_{1-\alpha})
\]
\[
= P_*(\hat{\phi}^* - \hat{\phi} + z_0 \leq z_{1-\alpha})
\]
\[
= P_*(\hat{\phi}^* \leq \hat{\phi} + \Psi^{-1}(1 - \alpha) - z_0)
\]
\[
= P_*(\hat{\theta}^* \leq \phi^{-1}(\hat{\phi} + \Psi^{-1}(1 - \alpha) - z_0))
\]
\[
= K(\phi^{-1}(\hat{\phi} + \Psi^{-1}(1 - \alpha) - z_0))
\]

Hence, for all \( t \),

\[
K^{-1}(t) = \phi^{-1}(\hat{\phi} + \Psi^{-1}(t) - z_0)
\]

Let \( t = \Psi(z_\alpha + 2z_0) \),

\[
K^{-1}(\Psi(z_\alpha + 2z_0)) = \phi^{-1}(\hat{\phi} + z_\alpha + 2z_0 - z_0)
\]
\[
= \phi^{-1}(\hat{\phi} + z_\alpha + z_0)
\]
\[
= \theta_E
\]
The bootstrap accelerated bias-corrected percentile (BC\textsubscript{a})
Assumption B is weaker than Assumption A
Assumption B is still too strong in some cases
Weaker assumptions?

**Assumption C:**
There is a monotone transformation \( \phi \) and constants \( z_0 \) and \( a \) (acceleration constant) such that

\[
P \left( \frac{\hat{\phi} - \phi}{1 + a\phi} + z_0 \leq t \right) = \Psi(t) \quad \text{for all } t \text{ and all } P
\]

\( \Psi \): a continuous, increasing cdf symmetric about 0
If \( a = 0 \), Assumption C reduces to Assumption B.
\( a \) corrects skewness
Result 5. Under Assumption C,

\[ \theta_E = \phi^{-1} \left( \phi + \frac{(z_\alpha + z_0)(1 + a\hat{\phi})}{1 - a(z_\alpha + z_0)} \right) \]

is an exact 100(1 - \alpha)\% lower confidence limit for \( \theta \)

Proof:

\[ 1 - \alpha = \Psi(-z_\alpha) \]

\[ = P \left( \frac{\hat{\phi} - \phi}{1 + a\phi} + z_0 \leq -z_\alpha \right) \]

\[ = P \left( \hat{\phi} \leq \phi - (z_\alpha + z_0)(1 + a\phi) \right) \]

\[ = P \left( \frac{\hat{\phi} + z_\alpha + z_0}{1 - a(z_\alpha + z_0)} \leq \phi \right) \]

\[ = P \left( \hat{\phi} + \frac{(z_\alpha + z_0)(1 + a\hat{\phi})}{1 - a(z_\alpha + z_0)} \leq \phi \right) \]

\[ = P (\theta_E \leq \theta) \]
We need to know $a$, $\Psi$, and $\phi$ in order to use $\theta_E$

Under Assumption C, $z_0 = \Psi^{-1}(K(\hat{\theta}))$

First, assume $a$ is known

The $BC_a$ lower confidence bound is

$$\theta_{BC_a} = K^{-1}\left(\Psi\left(z_0 + \frac{z_{\alpha} + z_0}{1 - a(z_{\alpha} + z_0)}\right)\right)$$

The $BC_a$ upper confidence bound is

$$\bar{\theta}_{BC_a} = K^{-1}\left(\Psi\left(z_0 + \frac{z_{1-\alpha} + z_0}{1 - a(z_{1-\alpha} + z_0)}\right)\right)$$

When bootstrap Monte Carlo of size $B$ is applied,

$\theta_{BC_a} \approx B\Psi\left(z_0 + \frac{z_{\alpha} + z_0}{1 - a(z_{\alpha} + z_0)}\right)$ $\text{th}$ ordered value of $\hat{\theta}^*1, \ldots, \hat{\theta}^*B$

$\bar{\theta}_{BC_a} \approx B\Psi\left(z_0 + \frac{z_{1-\alpha} + z_0}{1 - a(z_{1-\alpha} + z_0)}\right)$ $\text{th}$ ordered value of $\hat{\theta}^*1, \ldots, \hat{\theta}^*B$

We don’t need the explicit form of $\phi$

But we need to know $a$ and $\Psi$

If $a = 0$, $BC_a$ reduces to BC
Result 6. If $a$ is known, then $\theta_{BC_a} = \theta_E$.

Hence, $BC_a$ is exactly correct if Assumption C holds. If Assumption C holds approximately, then $BC_a$ is approximately correct.

Proof.

$$1 - \alpha = \Psi(z_{1-\alpha})$$

$$= P_\ast \left( \frac{\hat{\phi} - \phi}{1 + a\hat{\phi}} + z_0 \leq z_{1-\alpha} \right)$$

$$= P_\ast \left( \hat{\phi}^* \leq \phi + (z_{1-\alpha} - z_0)(1 + a\hat{\phi}) \right)$$

$$= P_\ast \left( \hat{\theta}^* \leq \phi^{-1}(\hat{\phi} + (z_{1-\alpha} - z_0)(1 + a\hat{\phi})) \right)$$

$$= K \left( \phi^{-1}(\hat{\phi} + (z_{1-\alpha} - z_0)(1 + a\hat{\phi})) \right)$$

$$= K \left( \phi^{-1}(\hat{\phi} + (\Psi^{-1}(1 - \alpha) - z_0)(1 + a\hat{\phi})) \right)$$

This shows that

$$K^{-1}(t) = \phi^{-1}(\hat{\phi} + (\Psi^{-1}(t) - z_0)(1 + a\hat{\phi}))$$
Let \( t = \Psi \left( z_0 + \frac{z_\alpha + z_0}{1 - a(z_\alpha + z_0)} \right), \)

\[
\theta_E = \phi^{-1} \left( \hat{\phi} + \frac{(z_\alpha + z_0)(1 + a\hat{\phi})}{1 - a(z_\alpha + z_0)} \right) \\
= K^{-1} \left( \Psi \left( z_0 + \frac{z_\alpha + z_0}{1 - a(z_\alpha + z_0)} \right) \right) \\
= \theta_{BC_{a}}
\]

The constant \( a \) is unknown and has to be estimated \( \hat{a} \): an estimator of \( a \)
\( \theta_{BC_{\hat{a}}} \) and \( \overline{\theta}_{BC_{\hat{a}}} \) are BC\( \hat{a} \) confidence limits
They are also called ABC confidence limits
Asymptotic Accuracy

A confidence set $C$ is first order accurate if

$$P(\theta \in C) = 1 - \alpha + O(n^{-1/2})$$

and second order accurate if

$$P(\theta \in C) = 1 - \alpha + O(n^{-1})$$

For the case of $\hat{\theta}$ is a smooth function of sample means, we have shown the following summary:

1. The BT and bootstrap $BC_{\alpha}$ one-sided confidence intervals are second order accurate.
2. The the BP, BC, HB, and NA one-sided confidence intervals are in general first order accurate.
3. The equal-tail two-sided confidence intervals produced by all five bootstrap methods and the normal approximation are second order accurate (errors cancel each other).