Recall ... simple bootstrap ...

Some common applications are

1) Estimation of \( \sqrt{\text{Var}_F T_n} \)

\[
\sqrt{\frac{1}{B-1} \sum_{i=1}^{B} (T_{ni} - \bar{T}_n)^2}
\]

This gets us a standard error for \( T_n \)

2) Bias of \( T_n \) as an estimator so \( \theta = \eta(F) \)

\[
\text{Bias}_F(T_n) = \mathbb{E}_F T_n - \eta(F)
\]

A standard methodology is to estimate \( F \) (with \( \hat{F} \)) and get \( B \) bootstrap versions of \( T_n \) and use

\[
T_{n^*} - \eta(\hat{F})
\]

as an estimated bias - and a "bias-corrected" version of \( T_n \) is then

\[
T_n - (T_{n^*} - \eta(\hat{F}))
\]
Note that in the case $\hat{F}$ is the histogram/empirical CDF and $\hat{T}_n = \eta(\hat{F})$. This bias correction suggests a version of $\hat{T}_n$

$$\hat{T}_n - \left( \bar{T}_n^* - \eta(\hat{F}) \right) = 2\bar{T}_n - \bar{T}_n^*$$

"Examples" (of this bias-correction stuff)

a) $T_n = \text{sample median}$

$\Theta = \eta(F) = E_F \bar{Y}$ = "population" mean

an estimated bias of $\bar{T}_n$ for $\Theta$ is

$$\bar{T}_n^* - \eta(\hat{F}) = \text{mean of } \hat{F}$$

So a bias-corrected version of $\bar{T}_n$ is thus

$$\hat{T}_n = (\bar{T}_n^* - \eta(\hat{F}))$$

b) $T_n = \text{sample median}$

$\Theta = \eta(F) = F^{-1}(0.5) = \text{population median}$

Here $T_n = \eta(\text{empirical CDF}) = \eta(F)$

Estimated bias of $\bar{T}_n$ for $\Theta$ is

$$\text{average bootstrap } \bar{T}_n^* - \eta(\hat{F}) = \text{sample median for data set sample bootstrap and a bias-corrected estimate}$$
2 \bar{T}_n = \frac{\bar{T}_n^*}{\hat{F}_n} \quad \text{average bootstrap sample median}

3) Confidence Limits: For some \( \Theta = \eta(F) \) – suppose that

\[ T_n = \eta \left( \text{empirical dsn of} \ Y_1, Y_2, \ldots, Y_n \right) \]

Based on \( B \) bootstrapped values \( \bar{T}_{n1}^*, \bar{T}_{n2}^*, \ldots, \bar{T}_{nB}^* \) in order, these

\[ \bar{T}_{n(1)}^* \leq \bar{T}_{n(2)}^* \leq \ldots \leq \bar{T}_{n(B)}^* \]

and find lower and upper \( \frac{\alpha}{2} \) points of this dsn of \( \bar{T}_{n}^* \)'s to use as end-points of an interval for \( \Theta \)

\[ \text{For } K_L = \left\lfloor \frac{K}{2} (B+1) \right\rfloor \]

\[ = \text{largest integer} \leq \frac{K}{2} (B+1) \]

and \( K_U = (B+1) - K_L \)

(these give indices of roughly \( \frac{K}{2} \) points of the generated dsn of \( \bar{T}^* \)'s) – my interval for \( \Theta \) is then

\[ \left[ \bar{T}_{n(K_L)}^*, \bar{T}_{n(K_U)}^* \right] \]

and this is often an approximate \((1-\alpha)\) level CI for \( \Theta = \eta(F) \)

?? ?? Why should this work – see panels 112-1124 of Kochlev
or Ch B of Efroim + Tibshirani or "handout" posted on 511 Web site — here's an outline

Suppose that there is a function \( m() \) s.t.

for \( \phi = m(\Theta) = m(\phi(F)) \)

and \( \hat{\phi} = m(T_n) = m(\phi(\gamma_1, \ldots, \gamma_n)) \)

for large \( n \)

\( \hat{\phi} \sim N(b, a^2) \)

then a \( CI \) for \( \phi \) is

\( (\hat{\phi} - zc, \hat{\phi} + zc) \)

and a corresponding \( CI \) for \( \Theta \) is

\( (\hat{\Theta} - zc, \hat{\Theta} + zc) \)

The argument you'll find in Kochler or handout is that

\[ \left[ T_n^*(K_c), T_n^*(K_u) \right] \]

approximates this (without having to know or use \( m() \))

This is called the "empirical percentile" bootstrap interval for \( \Theta \)