Percentile bootstrap confidence intervals

Suppose that a quantity $\theta = \eta(F)$ is of interest and that

$$T_n = \eta \text{ (the empirical distribution of } Y_1, Y_2, \ldots, Y_n)$$

Based on $B$ bootstrapped values $T_{n1}^*, T_{n2}^*, \ldots, T_{nB}^*$, define ordered values

$$T_{n(1)}^* \leq T_{n(2)}^* \leq \cdots \leq T_{n(B)}^*$$

Adopt the following convention (to locate lower and upper $\frac{\alpha}{2}$ points for the histogram/empirical distribution of the $B$ bootstrapped values). For

$$k_L = \left\lceil \frac{\alpha}{2} (B + 1) \right\rceil \quad \text{and} \quad k_U = (B + 1) - k_L$$

($[x]$ is the largest integer less than or equal to $x$) the interval

$$\left[ T_{n(k_L)}, T_{n(k_U)} \right]$$

contains (roughly) the “middle $(1 - \alpha)$ fraction of the histogram of bootstrapped values.” This interval is called the (uncorrected) “$(1 - \alpha)$ level bootstrap percentile confidence interval” for $\theta$.

The standard argument for why interval (1) might function as a confidence interval for $\theta$ is as follows. Suppose that there is an increasing function $m(\cdot)$ such that with

$$\phi = m(\theta) = m(\eta(F))$$

and

$$\hat{\phi} = m(T_n) = m(\eta \text{ (the empirical distribution of } Y_1, Y_2, \ldots, Y_n))$$

for large $n$

$$\hat{\phi} \sim N(\phi, w^2)$$

Then a confidence interval for $\phi$ is

$$\left[ \hat{\phi} - zw, \hat{\phi} + zw \right]$$

and a corresponding confidence interval for $\theta$ is

$$\left[ m^{-1}(\hat{\phi} - zw), m^{-1}(\hat{\phi} + zw) \right]$$

(2)

The argument is then that the bootstrap percentile interval (1) for large $n$ and large $B$ approximates this interval (2). The plausibility of an approximate correspondence between (1) and (2) might be argued as follows. Interval (2) is

$$\left[ m^{-1}(\hat{\phi} - zw), m^{-1}(\hat{\phi} + zw) \right]$$

$$\approx \left[ m^{-1} \left( \text{lower } \frac{\alpha}{2} \text{ point of the dsn of } \hat{\phi} \right), m^{-1} \left( \text{upper } \frac{\alpha}{2} \text{ point of the dsn of } \hat{\phi} \right) \right]$$

$$= \left[ m^{-1} \left( \text{lower } \frac{\alpha}{2} \text{ point of } T_n \text{ dsn} \right), m^{-1} \left( \text{upper } \frac{\alpha}{2} \text{ point of } T_n \text{ dsn} \right) \right]$$

$$= \left[ \text{lower } \frac{\alpha}{2} \text{ point of the dsn of } T_n, \text{upper } \frac{\alpha}{2} \text{ point of the dsn of } T_n \right]$$
and one may hope that interval (1) approximates this last interval. The beauty of the bootstrap argument in this context is that one doesn’t need to know the correct transformation $m$ (or the standard deviation $w$) in order to apply it.