# Short Note: BCa Confidence Interval

We aim to construct a $(1-\alpha)$ confidence interval for a parameter $\theta$ using the empirical distribution of bootstrap replicates $\hat{\theta}^*$.

The BCa interval is defined as:

$$\mathrm{CI}_{1-\alpha} = \left[\hat{\theta}^*_{(\alpha_1)},\ \hat{\theta}^*_{(\alpha_2)}\right]$$

where

- $\hat{\theta}^*_{(q)}$ is the empirical *quantile* of the bootstrap distribution at level $q$,
- $\alpha_1 = \Phi\left(z_0 + \frac{z_0+z_{\alpha/2}}{1-a(z_0+z_{\alpha/2})}\right)$,
- $\alpha_2 = \Phi\left(z_0 + \frac{z_0+z_{1-\alpha/2}}{1-a(z_0+z_{1-\alpha/2})}\right)$,

and

- $\Phi$ is the standard normal CDF,
- $z_p$ is the $p$-th quantile of the standard normal,
- $z_0$ is the bias correction factor,
- $a$ is the acceleration factor.

The bias correction term $z_0$ corrects for the bias between the bootstrap distribution and the original estimate. It is given by

$$z_0 = \Phi^{-1}\left(\frac{\#(\hat{\theta}^* < \hat{\theta})}{B}\right),$$

i.e., the proportion of bootstrap replicates that are *less than* the original estimate $\hat{\theta}$, converted to a standard normal quantile.

The acceleration term $a$ adjusts for skewness in the estimator. It requires a *jackknife* resampling of the data. It is given by

$$a = \frac{\sum_{i=1}^n (\bar{\theta}_{(\cdot)} - \hat{\theta}_{(i)})^3}{6\left[\sum_{i=1}^n (\bar{\theta}_{(\cdot)} - \hat{\theta}_{(i)})^2\right]^{3/2}}$$

where

- $\hat{\theta}_{(i)}$ is the estimator computed with the $i$-th observation *left out* (jackknife sample),
- $\bar{\theta}_{(\cdot)}$ is the mean of all jackknife estimates.

**Motivation.** The plain percentile bootstrap interval simply reads off quantiles of the bootstrap distribution, ignoring any systematic shift or asymmetry in the estimator. The BCa interval corrects for both of these deficiencies: the bias-correction term $z_0$ accounts for the fact that the median of the bootstrap distribution may not coincide with the original estimate $\hat{\theta}$, while the acceleration factor $a$ adjusts for non-constant variance and skewness in the sampling distribution. These two corrections together yield intervals with markedly more accurate coverage probabilities, particularly in small samples or whenever the estimator departs substantially from normality.