Bootstrap Confidence Intervals

Statistics · Module 4 — Interval Estimation

Node 4.6

Statistical Setting

Let

$$X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} P,$$

where P is unknown. Let $\theta = \theta(P)$ be a functional of P, and let $\hat{\theta}_n = \theta(\hat{P}_n)$ be the plug-in estimator, where $\hat{P}_n$ is the empirical distribution.

The bootstrap replaces the unknown sampling distribution of $\hat{\theta}_n - \theta$ with the distribution of $\hat{\theta}_n^* - \hat{\theta}_n$ under resampling from $\hat{P}_n$.

Definition — Bootstrap Distribution

A **bootstrap sample** is

$$X_1^*, \ldots, X_n^* \overset{\text{i.i.d.}}{\sim} \hat{P}_n,$$

i.e., a sample of size n drawn with replacement from $\{X_1, \ldots, X_n\}$.

Let $\hat{\theta}_n^* = \theta(\hat{P}_n^*)$ be the estimator computed on the bootstrap sample.

The **bootstrap distribution** of $\hat{\theta}_n^* - \hat{\theta}_n$ approximates the sampling distribution of $\hat{\theta}_n - \theta$.

Construction — Percentile Interval

Let $q_\alpha^*(\cdot)$ denote quantiles of the bootstrap distribution of $\hat{\theta}_n^*$.

The **percentile bootstrap interval** is

$$C_n^{\text{perc}}(X) = \left[q_{\alpha/2}^*,\ q_{1-\alpha/2}^*\right].$$

Construction — Basic Bootstrap Interval

The **basic bootstrap interval** (reverse percentile) inverts the bootstrap distribution of $\hat{\theta}_n^* - \hat{\theta}_n$:

$$C_n^{\text{basic}}(X) = \left[2\hat{\theta}_n - q_{1-\alpha/2}^*,\ 2\hat{\theta}_n - q_{\alpha/2}^*\right].$$

This corrects for asymmetry in the sampling distribution of $\hat{\theta}_n - \theta$.

Construction — Bootstrap-t Interval

Let $\hat{\sigma}_n$ be a consistent estimator of the standard deviation of $\hat{\theta}_n$. Define the studentized pivot

$$Z_n = \frac{\hat{\theta}_n - \theta}{\hat{\sigma}_n}.$$

Let $Z_n^* = (\hat{\theta}_n^* - \hat{\theta}_n)/\hat{\sigma}_n^*$ be the bootstrap analogue.

The **bootstrap-t interval** is

$$C_n^t(X) = \left[\hat{\theta}_n - q_{1-\alpha/2}^{*,t}\,\hat{\sigma}_n,\ \hat{\theta}_n - q_{\alpha/2}^{*,t}\,\hat{\sigma}_n\right],$$

where $q_p^{*,t}$ are quantiles of the bootstrap distribution of Z_n^*.

The bootstrap-t achieves second-order accuracy: coverage error is $O(n^{-1})$ rather than $O(n^{-1/2})$.

Consistency

Under regularity conditions, the bootstrap distribution of $\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n)$ converges weakly to the same limit as $\sqrt{n}(\hat{\theta}_n - \theta)$, in probability over the original sample.

Formally, if $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} L$, then

$$\sup_t \left| \mathbb{P}^*\big(\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n) \le t\big) - \mathbb{P}\big(\sqrt{n}(\hat{\theta}_n - \theta) \le t\big) \right| \xrightarrow{p} 0,$$

where $\mathbb{P}^*$ denotes probability under resampling conditional on $X_1, \ldots, X_n$.

Structural Role

Pivotal constructions (Node 4.2) require knowledge of the sampling distribution of $T_n(X, \theta)$.

The bootstrap removes this requirement by estimating the sampling distribution nonparametrically from the data.

It provides asymptotically valid confidence intervals for functionals $\theta(P)$ where no closed-form pivot exists.

Structural Compression

Invariant

The empirical distribution $\hat{P}_n$ consistently estimates P; the bootstrap distribution of $\hat{\theta}_n^* - \hat{\theta}_n$ consistently estimates the sampling distribution of $\hat{\theta}_n - \theta$.

Mechanism

Replace the unknown P with $\hat{P}_n$ in the sampling distribution calculation; simulate the resulting distribution by resampling.

Cross-System Echo

The bootstrap is a nonparametric Monte Carlo approximation to the sampling distribution; it connects to MCMC (Module 9) in that both substitute simulation for analytical distribution computation. In AI, bootstrap resampling underlies bagging and ensemble variance estimation.

Statistics · Module 4 — Interval Estimation · Node 4.6
Concepts: monte-carlo-methods, law-of-large-numbers
Prerequisites: 4.1, 4.2, 1.8, 3.1
Enables: 9.4

hbar.university — own.your.cognition