Bootstrap Implementation

Statistics · Module 9 — Computational Statistics

Node 9.4

Position in Knowledge Graph

System: Statistics

Structural Domain: Computational Statistics

Node: 9.4

The bootstrap translates the theoretical coverage guarantees of Node 4.6 into a concrete computational algorithm. It implements the plug-in principle: the empirical distribution $\hat{P}_n$ replaces the unknown population P, and resampling from $\hat{P}_n$ simulates the sampling variability of estimators. This node specifies the nonparametric and parametric bootstrap algorithms, develops bias correction, establishes consistency conditions, and identifies failure modes.

Formal Definition

Let $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} P$ and let $\hat{\theta}_n = T(\hat{P}_n)$ be a functional of the empirical distribution $\hat{P}_n = n^{-1} \sum_{i=1}^n \delta_{X_i}$.

The **nonparametric bootstrap distribution** of $\hat{\theta}_n$ is the conditional distribution of $\hat{\theta}_n^* = T(\hat{P}_n^*)$, where $\hat{P}_n^*$ is the empirical distribution of a sample drawn with replacement from $\{X_1, \ldots, X_n\}$, conditional on the observed data.

The **bootstrap principle** asserts that the distribution of $\hat{\theta}_n^* - \hat{\theta}_n$ (given the data) approximates the distribution of $\hat{\theta}_n - \theta(P)$ (over the sampling distribution).

Intuition

The bootstrap is the computational realization of a simple idea: if the empirical distribution is close to the true distribution (which Glivenko–Cantelli guarantees), then resampling from the empirical distribution should mimic sampling from the true distribution. The power of the bootstrap is its generality – it requires only the ability to compute $\hat{\theta}_n$ on resampled data, with no need for analytical variance formulas or distributional assumptions. The limitation is that the bootstrap is not magic: it fails when the empirical distribution is a poor approximation to aspects of P that matter for the distribution of $\hat{\theta}_n$.

Structural Role

Bootstrap implementation is the computational companion to the asymptotic theory of Node 4.6. It provides standard errors and confidence intervals for estimators where analytical variance formulas are unavailable or unreliable. The Monte Carlo approximation to the bootstrap distribution uses the framework of Node 9.2. In machine learning, the bootstrap underlies bagging (bootstrap aggregating), which reduces variance in ensemble methods.

Mathematical Development

Nonparametric Bootstrap Algorithm

Given observed data $x_1, \ldots, x_n$:

1. For $b = 1, \ldots, B$:
   - Draw $x_1^{(b)}, \ldots, x_n^{(b)}$ by sampling with replacement from $\{x_1, \ldots, x_n\}$.
   - Compute $\hat{\theta}_n^{(b)} = T(\hat{P}_n^{(b)})$.
2. The bootstrap distribution is $\{\hat{\theta}_n^{(1)}, \ldots, \hat{\theta}_n^{(B)}\}$.

Number of resamples. For standard errors: $B = 200$ suffices. For confidence intervals: $B = 1{,}000$ to $2{,}000$. For tail probabilities: $B \geq 5{,}000$.

Parametric Bootstrap

When the model $\{P_\theta : \theta \in \Theta\}$ is specified:

1. Compute $\hat{\theta}_n$ from the data.
2. For $b = 1, \ldots, B$: draw $X_1^{(b)}, \ldots, X_n^{(b)} \overset{\text{i.i.d.}}{\sim} P_{\hat{\theta}_n}$ and compute $\hat{\theta}_n^{(b)}$.

The parametric bootstrap samples from $P_{\hat{\theta}_n}$ rather than $\hat{P}_n$. It achieves higher accuracy when the model is correctly specified, because $P_{\hat{\theta}_n}$ is a smoother (and typically better) estimate of P than $\hat{P}_n$.

Bootstrap Consistency

Theorem (Bootstrap Consistency). Suppose $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$ and σ^2 is a smooth functional of P. Then the bootstrap is consistent:

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

where P^* denotes probability under the bootstrap measure.

Proof sketch. By the continuous mapping theorem applied to the functional T and the convergence $\hat{P}_n \to P$ in the Levy–Prokhorov metric, the bootstrap distribution converges to the same Gaussian limit as the sampling distribution. The key condition is that T is Hadamard differentiable at P, which ensures that the linearization $T(\hat{P}_n^*) - T(\hat{P}_n) \approx T'_P(\hat{P}_n^* - \hat{P}_n)$ is valid. $\square$

Bootstrap Standard Error and Confidence Intervals

The bootstrap standard error is

$$\widehat{\mathrm{SE}}_B = \sqrt{\frac{1}{B-1} \sum_{b=1}^{B} \left(\hat{\theta}_n^{(b)} - \bar{\theta}^*\right)^2}, \quad \bar{\theta}^* = \frac{1}{B} \sum_{b=1}^{B} \hat{\theta}_n^{(b)}.$$

Percentile interval:

$$C^{\text{perc}} = \left[q_{\alpha/2}^*, q_{1-\alpha/2}^*\right],$$

where q_p^* is the p-quantile of $\{\hat{\theta}_n^{(b)}\}$.

Basic (reverse) interval:

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

Bootstrap-t interval:

$$C^t = \left[\hat{\theta}_n - q^{*,t}_{1-\alpha/2}\,\hat{\sigma}_n,\ \hat{\theta}_n - q^{*,t}_{\alpha/2}\,\hat{\sigma}_n\right],$$

where $q^{*,t}_p$ are quantiles of the studentized bootstrap statistics $(\hat{\theta}^{(b)}_n - \hat{\theta}_n)/\hat{\sigma}^{(b)}_n$. The bootstrap-$t$ achieves second-order accuracy (coverage error $O(n^{-1})$) compared to $O(n^{-1/2})$ for the percentile interval. The bootstrap-t requires a nested computation: for each bootstrap resample, an internal standard error $\hat{\sigma}^{(b)}_n$ must be computed, either analytically or by a second-level bootstrap.

Coverage error comparison. Let C_n denote a nominal $(1-\alpha)$ confidence interval. The coverage error is $|P(\theta \in C_n) - (1-\alpha)|$:

- Normal approximation: $O(n^{-1/2})$.
- Percentile bootstrap: $O(n^{-1/2})$ (same order, different constant).
- Bootstrap-t and BCa: $O(n^{-1})$ (second-order accurate).
- Double bootstrap: $O(n^{-2})$ (third-order accurate).

BCa (Bias-Corrected and Accelerated) Method

The **BCa interval** corrects for both bias and skewness:

$$C^{\mathrm{BCa}} = \left[q^*_{\alpha_1}, q^*_{\alpha_2}\right],$$

where the adjusted quantile levels are

$$\alpha_1 = \Phi\left(z_0 + \frac{z_0 + z_{\alpha/2}}{1 - a(z_0 + z_{\alpha/2})}\right), \quad \alpha_2 = \Phi\left(z_0 + \frac{z_0 + z_{1-\alpha/2}}{1 - a(z_0 + z_{1-\alpha/2})}\right).$$

Here $z_0 = \Phi^{-1}(\#\{\hat{\theta}^{(b)}_n < \hat{\theta}_n\}/B)$ is the bias correction and a is the acceleration constant, estimated from the jackknife:

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

where $\hat{\theta}_{(-i)}$ is the estimate with observation i deleted.

When the Bootstrap Fails

The bootstrap fails when $\hat{P}_n$ does not capture the features of P relevant to the distribution of $\hat{\theta}_n$:

1. **Heavy tails without finite variance.** If $\mathrm{Var}_P(X) = \infty$, the bootstrap variance is finite (since $\hat{P}_n$ has finite support) and underestimates the true variability.
2. **Extremes and boundary estimators.** For $\hat{\theta}_n = \max(X_1, \ldots, X_n)$, the bootstrap distribution concentrates on the observed maximum, failing to capture the correct $n(X_{(n)} - \theta)$ scaling.
3. **Non-smooth functionals.** If T is not Hadamard differentiable (e.g., the mode), the bootstrap may be inconsistent.
4. **Dependent data.** The i.i.d. bootstrap is invalid for time series; block bootstrap or sieve bootstrap are required.

Double Bootstrap

For calibrated confidence intervals with coverage error $O(n^{-2})$, the **double bootstrap** uses nested resampling: for each bootstrap sample $b = 1, \ldots, B_1$, generate B_2 second-level bootstrap resamples to estimate the coverage of the first-level interval, then adjust the quantile accordingly.

The double bootstrap is computationally intensive ($B_1 \times B_2$ total evaluations of T) but achieves second-order accuracy without requiring analytical calculation of the acceleration constant a in the BCa method.

Exact Bootstrap Distribution

The exact nonparametric bootstrap distribution averages over all $\binom{2n-1}{n}$ possible resamples (each corresponding to a multinomial draw of size n from n categories). The Monte Carlo bootstrap with B resamples is an approximation to this exact distribution. The Monte Carlo error (from finite B) is distinct from the bootstrap error (from $\hat{P}_n \neq P$); the former is $O(B^{-1/2})$ and can be made arbitrarily small by increasing B, while the latter is an intrinsic statistical limitation.

Residual Bootstrap for Regression

For the linear model $Y = X\beta + \varepsilon$, the **residual bootstrap** resamples the residuals $\hat{e}_i = Y_i - X_i^\top \hat{\beta}$ and constructs bootstrap responses $Y_i^{(b)} = X_i^\top \hat{\beta} + \hat{e}_{\pi(i)}$, where π is a random permutation with replacement. This preserves the fixed design while resampling the error distribution, and is appropriate when the regression structure is assumed correct but the error distribution is unknown.

Worked Examples

Example 1: Bootstrap Standard Error of the Median

Let $x_1, \ldots, x_{20}$ be a sample from an unknown distribution with observed median $\hat{m} = 3.7$. There is no simple formula for $\mathrm{SE}(\hat{m})$ without distributional assumptions. Generate $B = 2{,}000$ bootstrap resamples, compute the median of each, and take the standard deviation: $\widehat{\mathrm{SE}}_B = 0.42$. The 95% percentile interval is $[2.9, 4.5]$. Contrast this with the normal-based interval $\hat{m} \pm 1.96 \cdot \widehat{\mathrm{SE}}_B = [2.88, 4.52]$, which is similar here but would differ for skewed bootstrap distributions.

Example 2: Parametric vs. Nonparametric Bootstrap for Normal Variance

Let $x_1, \ldots, x_{15} \sim \mathcal{N}(5, 4)$ with observed sample variance $s^2 = 3.8$. The exact sampling distribution gives $(n-1)s^2/\sigma^2 \sim \chi^2_{14}$, yielding a 95% CI for σ^2 of $[2.09, 8.44]$.

Nonparametric bootstrap: resample with replacement, compute $s^{2(b)}$ for $B = 2{,}000$ resamples. The percentile interval is approximately $[1.8, 7.1]$ – shorter than the exact interval because the empirical distribution has lighter tails than the normal.

Parametric bootstrap: draw $X^{(b)} \sim \mathcal{N}(\bar{x}, s^2)$, compute $s^{2(b)}$. The percentile interval is approximately $[2.1, 8.3]$, closely matching the exact interval. The parametric bootstrap is superior here because the normality assumption is correct and the parametric model estimates P more accurately than $\hat{P}_n$ for tail behavior.

Example 3: Bootstrap for Correlation Coefficient

Let $(X_i, Y_i)_{i=1}^{30}$ be bivariate data with sample correlation $r = 0.72$. Fisher's z-transformation gives an approximate confidence interval, but the bootstrap provides a nonparametric alternative. Resample pairs $(X_i^{(b)}, Y_i^{(b)})$ with replacement and compute $r^{(b)}$ for $B = 2{,}000$. The bootstrap distribution of r is left-skewed (bounded above by 1). The BCa interval $[0.48, 0.87]$ accounts for this skewness, while the percentile interval $[0.45, 0.86]$ does not. The acceleration constant $a = -0.03$ reflects the negative skewness induced by the upper bound.

Cross-Links

- **Linear Algebra:** The bootstrap covariance matrix of a vector estimator $\hat{\theta}_n \in \mathbb{R}^k$ is the sample covariance of the bootstrap replicates, providing a nonparametric estimate of the asymptotic covariance matrix that would otherwise require the delta method and Fisher information.
- **AI:** Bagging (bootstrap aggregating) generates an ensemble of predictors trained on bootstrap resamples; the variance reduction from averaging parallels the bootstrap's variance estimation. Random forests extend bagging with feature subsampling.
- **Economics:** The bootstrap is the standard tool for inference in structural econometric models where the asymptotic distribution of estimators (GMM, two-stage least squares) depends on nuisance parameters in complex ways.
- **Quantum:** Quantum state tomography uses bootstrap resampling to quantify uncertainty in reconstructed density matrices when analytical error bars are unavailable.

Structural Compression

Invariant: The empirical distribution $\hat{P}_n$ is a consistent estimate of P; resampling from $\hat{P}_n$ consistently estimates the sampling distribution of $T(\hat{P}_n)$ whenever T is sufficiently smooth (Hadamard differentiable).

Mechanism: Drawing with replacement from the observed sample generates a Monte Carlo approximation to the bootstrap distribution; quantiles of this approximation yield confidence interval endpoints. The BCa correction adjusts for bias and skewness, achieving second-order accuracy.

Cross-System Echo: The bootstrap implements the same structural principle as the plug-in estimator: replace the unknown P with its empirical estimate and propagate uncertainty through the functional T. This principle – that consistent estimation of the input yields consistent estimation of smooth functionals of the input – is the computational form of the continuous mapping theorem and appears in every system where simulation replaces analysis.

Exercises

1. Prove that the bootstrap estimate of the mean, $\bar{X}^* = n^{-1}\sum_{i=1}^n X_i^*$, has conditional variance $\hat{\sigma}^2/n$ where $\hat{\sigma}^2 = n^{-1}\sum(X_i - \bar{X})^2$, matching the plug-in variance estimator.

2. For the sample variance S^2, derive the bootstrap distribution analytically when $n = 3$ by enumerating all $3^3 = 27$ bootstrap resamples.

3. Show that the percentile bootstrap interval is transformation-respecting: if C^{perc} is an interval for θ, then $g(C^{\text{perc}})$ is a valid interval for $g(\theta)$ for any monotone g.

4. Explain why the bootstrap fails for $\hat{\theta}_n = X_{(n)} = \max(X_1, \ldots, X_n)$ when P is uniform on $[0, \theta]$. What is the correct rate of convergence, and why does the bootstrap miss it?

5. Derive the bias-correction constant z_0 in the BCa method and show that when $z_0 = 0$ and $a = 0$, the BCa interval reduces to the percentile interval.

6. Compare the parametric bootstrap (assuming normality) and the nonparametric bootstrap for the sample median from a skewed distribution. Generate data from Exp(1), $n = 30$, and compare coverage of 95% intervals over 500 simulation replications.

7. Argue, structurally, why the bootstrap's reliance on the plug-in principle means it inherits both the strengths and limitations of empirical distribution estimation: it succeeds whenever the empirical distribution captures the relevant features of P (bulk behavior, smooth functionals) and fails whenever

the relevant features are in the tails or boundaries that the empirical distribution cannot resolve with n observations.

Statistics · Module 9 — Computational Statistics · Node 9.4
Concepts: monte-carlo-methods, simulation
Prerequisites: 4.6, 9.2

hbar.university — own.your.cognition