Permutation tests, the Jackknife and the Bootstrap

Gioele La Manno

École Polytechnique Fédérale de Lausanne (EPFL)

School of Life Science (SV)

May 2025

EPFL - BMI - UPLAMANNO

Contents

Computational Statistics

The Computational Revolution in Statistics

Statistics has undergone a profound transformation, driven by exponential increases in computational power.

Traditional statistical methods were constrained by:

- Need for mathematical tractability
- Simplifying assumptions about distributions
- Reliance on analytical formulas

Today, we're in "the computer age of statistics" where:

- We can tackle problems through simulation and algorithms
- We can handle violations of standard assumptions
- We can work with complex data structures

Computational Approaches to Statistical Inference

Computational approaches are especially valuable when:

- Standard parametric assumptions are violated
- Deriving analytical solutions is mathematically intractable
- Working with complex data structures (nested measurements)
- Using complex algorithms without available analytical solutions
- Dealing with small sample sizes where asymptotic approximations fail
- Analyzing censored or irregularly sampled data

Today we'll explore three fundamental approaches:

- Permutation tests
- The jackknife
- The bootstrap

Permutation Testing

The Logic of Permutation Testing

Permutation tests offer a distribution-free approach to hypothesis testing.

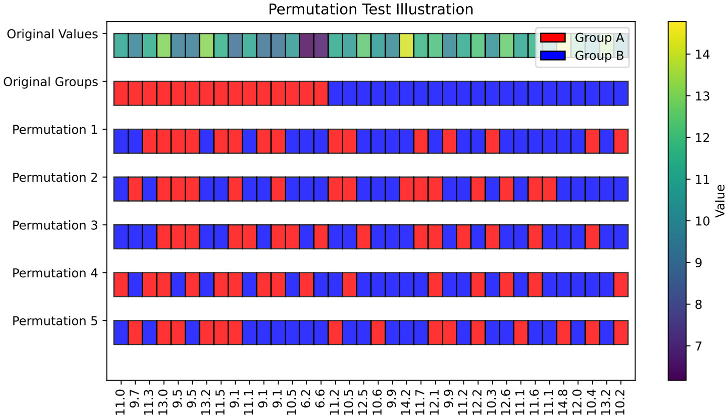
Core insight (Fisher, 1930s): If the null hypothesis is true, then group labels are essentially arbitrary.

Definition (Permutation Test)

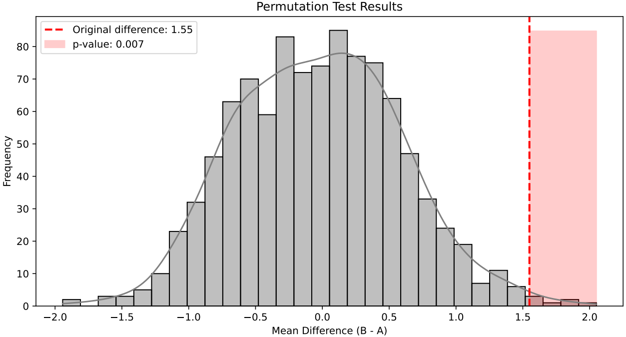
A permutation test involves:

1. Calculate test statistic T for the original data
2. Generate permutations of the data consistent with the null hypothesis
3. Calculate test statistic T_m for each permutation
4. The p-value is the proportion of permutations where T_m is as or more extreme than T

Permutation Testing Illustrated



Permutation Test Results



The p-value is the proportion of permutations with test statistics as or more extreme than the observed value.

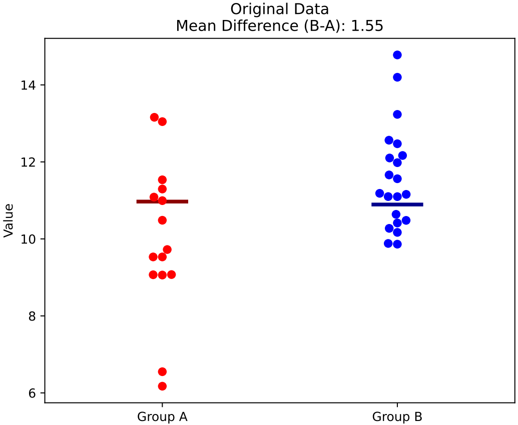
Use case of Permutation test

This approach is particularly valuable when:

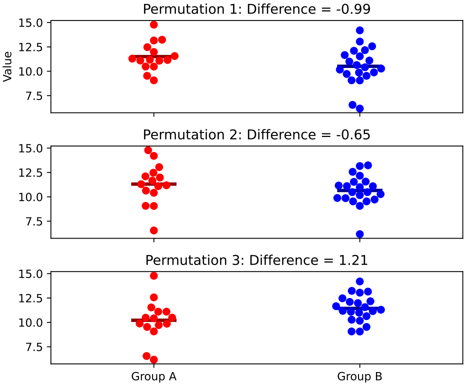
- Sample sizes are small
- Data exhibit skewness or outliers
- Variances differ between groups

Permutation Testing for Simple Comparisons

Permutation testing allows us to relax the assumptions required by traditional tests.



Permutation of Group Labels



Each permutation randomly reassigns observations to groups while maintaining the original group sizes.

Permutation Testing for Correlation

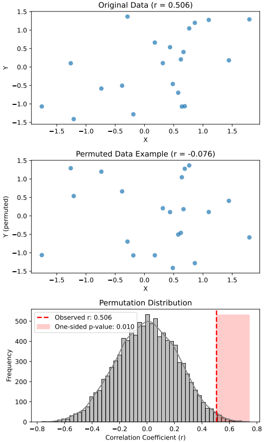
Permutation tests can also assess the significance of correlation coefficients:

1. Calculate observed correlation coefficient r_{XY} between variables
2. Randomly permute values of one variable while keeping the other fixed
3. Calculate the correlation coefficient for each permutation
4. Determine how extreme the observed correlation is relative to this distribution

This approach is particularly valuable when:

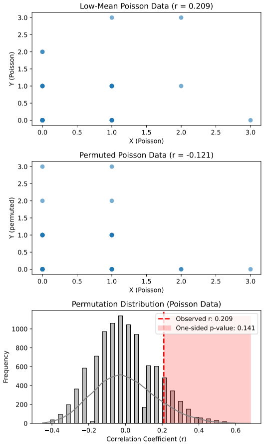
- The marginal distributions are non-normal
- The sample size is small
- We suspect a non-linear relationship

Permutation Test for Correlation - Example 1



Example with N=15 points showing a correlation of 0.5, which is statistically significant.

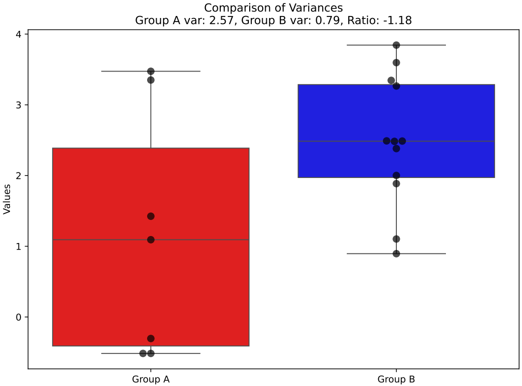
Permutation Test for Correlation - Example 2



Example with Poisson-distributed variables with low means. Note how the null distribution is not normal and not centered at zero.

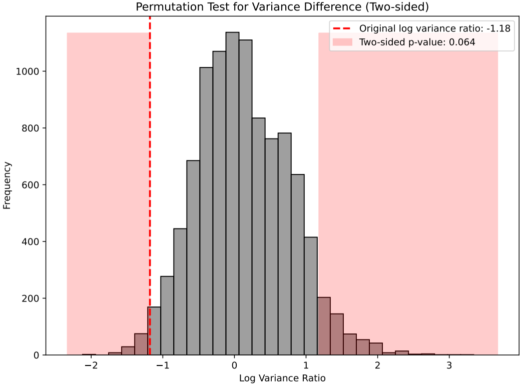
Complex Permutation Tests - Variance Comparison

Permutation tests can handle scenarios where traditional approaches struggle, like testing for differences in variance between groups.



When testing for variance differences, we need to center the data before permutation

Permutation Test for Variance - Results



The histogram shows the distribution of variance differences under the null hypothesis, with the observed value and resulting p-value.

Standard Errors: From Formulas to General Approaches

In previous lectures, we derived formulas for standard errors of specific statistics:

For the mean:

$$\hat{se}(\bar{X}) = \frac{s}{\sqrt{n}} \quad \text{where } s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \tag{1}$$

For regression coefficients:

$$\hat{se}(\hat{\beta}) = \sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}} \tag{2}$$

But how do we estimate the standard error of a statistic that is not a simple function of the data?

The Delta Method

The delta method uses Taylor series expansion to approximate the standard error of a function of a statistic.

If our statistic $t(\mathbf{X})$ can be written as $g(\hat{\theta})$, where $\hat{\theta}$ has known standard error:

$$\hat{se}(g(\hat{\theta})) \approx \sqrt{\nabla g(\hat{\theta})^T \cdot \Sigma \cdot \nabla g(\hat{\theta})} \tag{3}$$

For the univariate case, this simplifies to:

$$\hat{se}(g(\hat{\theta})) \approx |g'(\hat{\theta})| \cdot \hat{se}(\hat{\theta}) \tag{4}$$

Delta Method Example: Log of the Mean

For the logarithm of the sample mean:

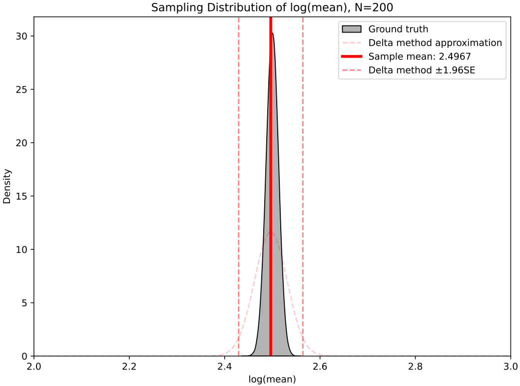
$$t(\mathbf{x}) = \log(\bar{x}) \tag{5}$$

The delta method gives us:

$$\hat{se}(\log(\bar{x})) \approx \frac{1}{\bar{x}} \cdot \hat{se}(\bar{x}) = \frac{s}{\bar{x}\sqrt{n}} \tag{6}$$

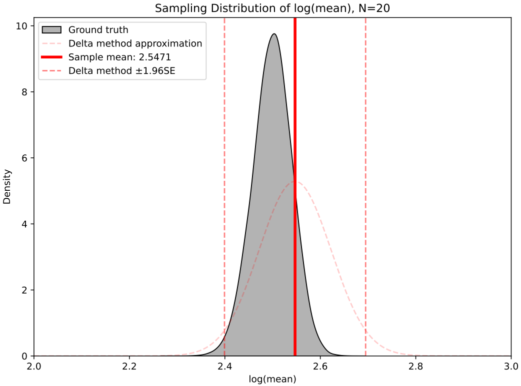
This approximation works well for large samples but may be less accurate for small samples or highly skewed distributions.

Delta Method Performance (N=200)



With a large sample size (N=200), the delta method approximation closely matches the true sampling distribution of $\log(\bar{x})$.

Delta Method Performance (N=20)



With a smaller sample size (N=20), the approximation is less accurate, especially in capturing the asymmetry of the sampling distribution.

The Jackknife

The Jackknife: Origins and Basic Concept

The jackknife, introduced by Quenouille (1949) and developed by Tukey (1958), represents an early computational approach to statistics.

Core idea: To understand the stability of a statistic, systematically leave out each observation one at a time and recalculate.

Definition (Jackknife Estimate)

Given a dataset $\mathbf{X} = \{X_1, X_2, \ldots, X_n\}$ and a statistic $\hat{\theta} = t(\mathbf{X})$, the jackknife estimates are:

$$\hat{\theta}_{(i)} = t(\mathbf{X}_{(i)})$$

where $\mathbf{X}_{(i)}$ represents the dataset with the i-th observation removed.

The Jackknife Method - Algorithm

Algorithm Jackknife Method

```
Compute full-sample estimate $\hat{\theta} = t(\mathbf{X})$
**for** $i = 1$ to $n$ **do**
    Create leave-one-out sample $\mathbf{X}_{(i)} = \{X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n\}$
    Compute jackknife replicate $\hat{\theta}_{(i)} = t(\mathbf{X}_{(i)})$
**end for**
Compute average of jackknife estimates: $\hat{\theta}_{(\cdot)} = \frac{1}{n} \sum_{i=1}^{n} \hat{\theta}_{(i)}$
Compute jackknife standard error: $\widehat{se}_{jack} = \sqrt{\frac{n-1}{n} \sum_{i=1}^{n} (\hat{\theta}_{(i)} - \hat{\theta}_{(\cdot)})^2}$
Compute jackknife bias estimate: $\widehat{bias}_{jack} = (n-1)(\hat{\theta}_{(\cdot)} - \hat{\theta})$
```

Estimating Standard Errors with the Jackknife

The jackknife estimate of standard error is:

$$\hat{se}_{jack}(\hat{\theta}) = \sqrt{\frac{n-1}{n}\sum_{i=1}^{n}(\hat{\theta}_{(i)} - \hat{\theta}_{(\cdot)})^2}$$

Where $\hat{\theta}_{(\cdot)} = \frac{1}{n}\sum_{i=1}^{n}\hat{\theta}_{(i)}$ is the average of the jackknife estimates.

This measures how much the statistic fluctuates when individual observations are removed, scaled to reflect the appropriate sampling variation.

The factor $\frac{n-1}{n}$ ensures that $\hat{se}_{jack}$ exactly equals the analytical formula for the standard error of the mean when applied to that statistic.

The Jackknife for Bias Estimation

Beyond standard errors, the jackknife can also estimate the bias of a statistic:

$$\hat{bias}_{jack}(\hat{\theta}) = (n-1)(\hat{\theta}_{(\cdot)} - \hat{\theta})$$

This allows us to construct bias-corrected estimates:

$$\hat{\theta}_{corrected} = \hat{\theta} - \hat{bias}_{jack}(\hat{\theta}) = n\hat{\theta} - (n-1)\hat{\theta}_{(\cdot)}$$

Bias correction is particularly valuable for ratio estimators, variance components, and other statistics that exhibit systematic bias in finite samples.

Theoretical Foundations of the Jackknife

The jackknife can be viewed as an automatic numerical way to compute the Taylor approximation in the delta method.

Rewriting the jackknife standard error:

$$
\begin{aligned}
\hat{se}_{jack}(\hat{\theta}) &= \sqrt{\frac{n-1}{n}\sum_{i=1}^{n}(\hat{\theta}_{(i)} - \hat{\theta}_{(\cdot)})^2} \\
&= \frac{1}{n}\sqrt{\sum_{i=1}^{n} D_i^2}
\end{aligned}
$$

Where D_i are approximate *directional derivatives*, measuring how fast the statistic changes as we decrease the weight on data point X_i.

This connects directly to the delta method formula:

Limitations of the Jackknife

The jackknife has important limitations:

- **Non-smooth statistics:** For statistics like the median, removing a single observation might not change the estimate at all if it's not near the median, leading to artificially small variance estimates.
- **Insufficient exploration:** The jackknife creates only n resamples, all with exactly $n-1$ observations. This limited exploration might not adequately capture the full sampling distribution.
- **Edge effects:** For statistics sensitive to extreme values, the jackknife can either overestimate variance (if removing an extreme value drastically changes the estimate) or underestimate it (if the statistic is robust to such removal).

These limitations motivated the development of more comprehensive resampling approaches.

The Bootstrap

The Bootstrap: Resampling with Replacement

The bootstrap, introduced by Bradley Efron in 1979, marked a paradigm shift in computational statistics.

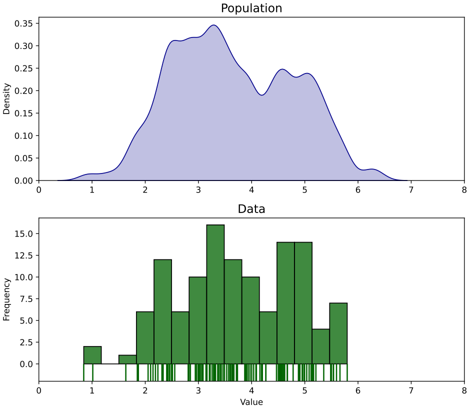
Core insight: We can simulate the sampling process by resampling with replacement from the original data, treating our observed sample as a stand-in for the unknown population.

Definition (Bootstrap Sample)

Given a dataset $\mathbf{X} = \{X_1, X_2, \ldots, X_n\}$, a bootstrap sample $\mathbf{X}^* = \{X_1^*, X_2^*, \ldots, X_n^*\}$ is created by sampling n observations with replacement from $\mathbf{X}$.

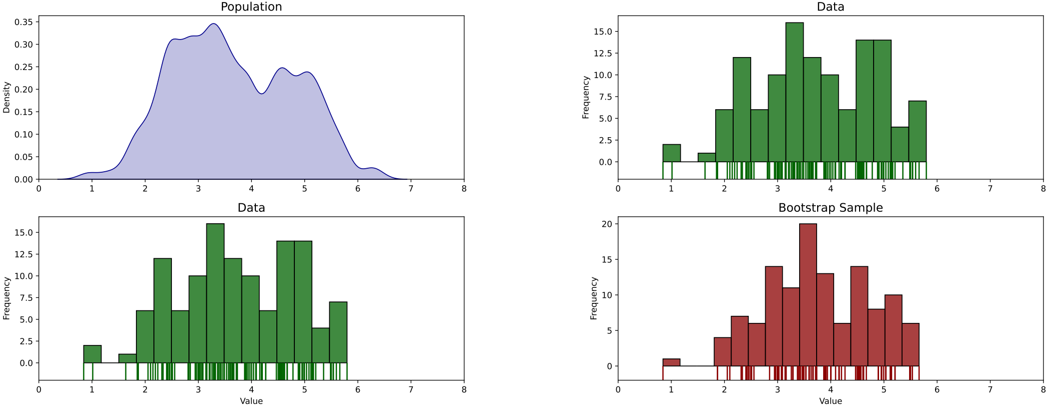
By resampling with replacement, some original observations appear multiple times in a bootstrap sample while others are omitted, mimicking natural sampling variation.

Bootstrap Sampling Illustrated



The measurement process samples from a population

Bootstrap Sampling Illustrated



The measurement process samples from a population

In bootstrap, we resample the data with replacement to simulate a population.

The Bootstrap Algorithm

The general bootstrap procedure for estimating the sampling distribution of a statistic $\hat{\theta} = t(\mathbf{X})$:

1. Generate B bootstrap samples $\mathbf{X}^{*1}, \mathbf{X}^{*2}, \ldots, \mathbf{X}^{*B}$ by sampling n observations with replacement from the original dataset $\mathbf{X}$.
2. Calculate the statistic of interest for each bootstrap sample: $\hat{\theta}^{*b} = t(\mathbf{X}^{*b})$ for $b = 1, 2, \ldots, B$.
3. Use the empirical distribution of $\hat{\theta}^{*1}, \hat{\theta}^{*2}, \ldots, \hat{\theta}^{*B}$ to approximate the sampling distribution of $\hat{\theta}$.

The Bootstrap Algorithm

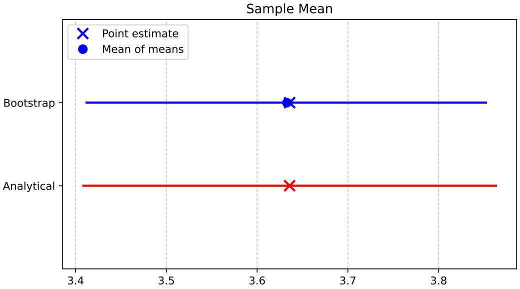
The general bootstrap procedure for estimating the sampling distribution of a statistic $\hat{\theta} = t(\mathbf{X})$:

1. Generate B bootstrap samples $\mathbf{X}^{*1}, \mathbf{X}^{*2}, \ldots, \mathbf{X}^{*B}$ by sampling n observations with replacement from the original dataset $\mathbf{X}$.
2. Calculate the statistic of interest for each bootstrap sample: $\hat{\theta}^{*b} = t(\mathbf{X}^{*b})$ for $b = 1, 2, \ldots, B$.
3. Use the empirical distribution of $\hat{\theta}^{*1}, \hat{\theta}^{*2}, \ldots, \hat{\theta}^{*B}$ to approximate the sampling distribution of $\hat{\theta}$.

From this distribution, we can estimate:

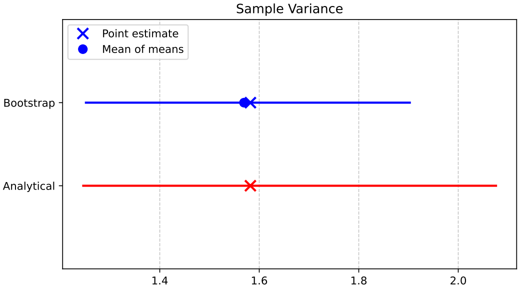
- Standard errors
- Confidence intervals
- Bias

Bootstrap vs. Analytical - Sample Mean



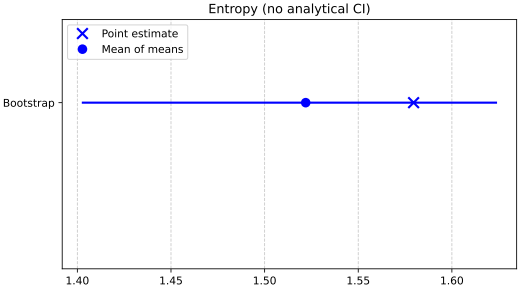
For the sample mean, bootstrap confidence intervals closely match analytical results, validating the bootstrap approach.

Bootstrap vs. Analytical - Sample Variance



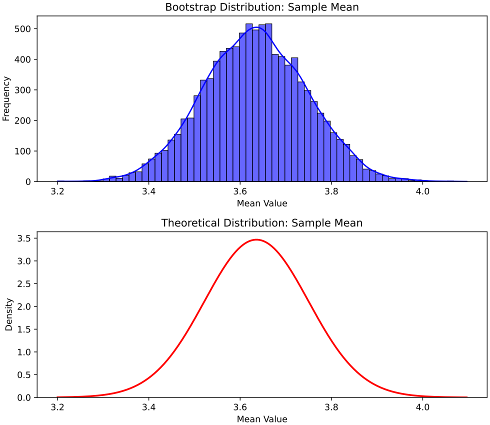
For the sample variance, bootstrap confidence intervals again provide similar results to analytical methods.

Bootstrap Confidence Interval - Entropy



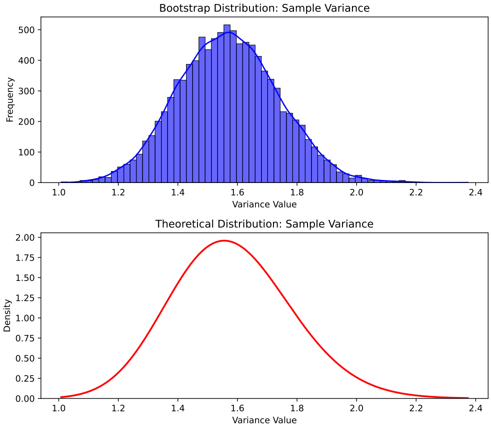
For complex statistics like entropy where analytical approaches are challenging, the bootstrap provides a straightforward method for confidence interval construction.

Bootstrap Sampling Distribution - Sample Mean



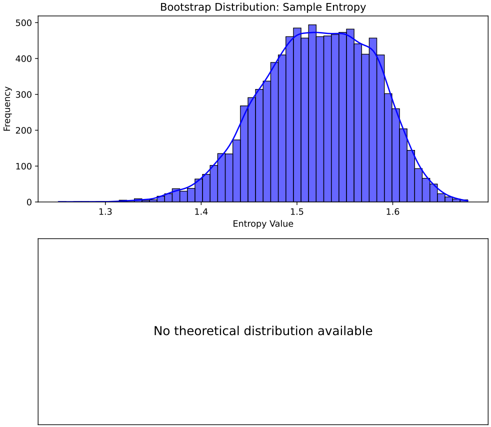
The bootstrap distribution (top) approximates the theoretical sampling distribution (bottom) for the sample mean.

Bootstrap Sampling Distribution - Sample Variance



For the sample variance, the bootstrap (top) captures the asymmetry of the theoretical sampling distribution (bottom).

Bootstrap Sampling Distribution - Entropy



For entropy, the bootstrap provides an empirical sampling distribution when theoretical distributions are difficult to derive.

Bootstrap Confidence Intervals

Beyond standard error estimation, the bootstrap excels at constructing confidence intervals.

Percentile Method:
The simplest approach uses percentiles of the bootstrap distribution directly:

$$\mathrm{CI}_{1-\alpha} = [\hat{\theta}^*_{\alpha/2}, \hat{\theta}^*_{1-\alpha/2}]$$

where $\hat{\theta}^*_{\alpha/2}$ and $\hat{\theta}^*_{1-\alpha/2}$ are the $\alpha/2$ and $1 - \alpha/2$ percentiles of the bootstrap distribution.

Advantages:

- Intuitive and simple to implement
- Preserves range restrictions (e.g., correlations between -1 and 1)
- Accounts for asymmetry in the sampling distribution

Parametric Bootstrap

When sample sizes are small, the standard (nonparametric) bootstrap may not adequately capture population variability.

Definition (Parametric Bootstrap)

The parametric bootstrap involves:

1. Fit a parametric model to the observed data and estimate its parameters $\hat{\theta}$
2. Generate synthetic datasets by sampling from the fitted model with parameters $\hat{\theta}$
3. Calculate the statistic of interest for each synthetic dataset
4. Use the distribution of these statistics to approximate the sampling distribution