STAT 7650 - Computational Statistics Lecture Slides

Bootstrap

Elvan Ceyhan

Updated: April, 2025

AU

- Based on parts of: Chapter 9 in Givens & Hoeting (Computational Statistics), and Chapter 24 of Lange (Numerical Analysis for Statisticians)).

Outline

Motivation for Bootstrap Methods

Statistical Foundation:

- In hypothesis testing and confidence intervals, a key quantity is a **statistic** whose sampling distribution is required.
- For instance, to test the null hypothesis $H_0 : \mu = \mu_0$ based on a random sample from a normal distribution $N(\mu, \sigma^2)$ where σ^2 is unknown, we typically use the t-statistic:

$$T = \frac{\overline{X} - \mu_0}{S/\sqrt{n}},$$

  where $\overline{X}$ is the sample mean, S is the sample standard deviation, and n is the sample size.
- Under the stated conditions, the null distribution of T is a Student's t-distribution.

Motivation for Bootstrap Methods

Challenges with Non-Standard Conditions:

- Deviations from this basic setup, such as different distributions or assumption violations, lead to complex distributional calculations.

Bootstrap Solution:

- The goal of the bootstrap method is to provide a simple, approximate solution to these challenges using simulations.
- This method leverages resampling techniques to estimate the sampling distribution of the statistic, thereby facilitating more flexible statistical inference.

Statistical Notation and Empirical cdf

Introduction to Statistical Functionals:

- Let F be a cumulative distribution function (cdf) of a random variable.
- Consider a parameter $\theta = \varphi(F)$, written as a *functional* of F.

Examples of Functionals:

- **Mean:** $\varphi(F) = \int x \, dF(x)$
- **Median:** $\varphi(F) = \inf\{x : F(x) \geq 0.5\}$

Empirical cdf: Given data $\mathcal{X} = \{X_1, \ldots, X_n\}$ sampled from F, the empirical cdf is defined as: $\widehat{F}(x) = \dfrac{1}{n}\sum_{i=1}^{n} I(X_i \leq x), \quad x \in \mathbb{R},$ where I is the indicator function.

Estimation of θ: A natural estimate of θ is $\widehat{\theta} = \varphi(\widehat{F})$, applying the same functional φ to the empirical cdf.

Notation and Bootstrap Methodology (Cont'd)

Statistical Inference with Statistic $T(\mathcal{X}, F)$:

- For inference, a statistic $T(\mathcal{X}, F)$ is utilized, defined as:

$$T(\mathcal{X}, F) = \frac{\overline{X} - \mu_0}{S/\sqrt{n}}$$

where $\overline{X}$ is the sample mean, S is the sample standard deviation, μ_0 is a hypothesized mean, and n is the sample size.

Challenges with Sampling Distribution:

- The sampling distribution of $T(\mathcal{X}, F)$ can be complicated, unknown, or dependent on the unknown true cdf F.

Notation and Bootstrap Methodology (Cont'd)

Bootstrap Approach:

- **Replacing the True cdf:** The unknown cdf F is replaced with the empirical cdf $\widehat{F}$ for analysis.
- **Numerical Approximation:** By repeatedly iid sampling from $\widehat{F}$, produce a numerical approximation of the sampling distribution of $T(\mathcal{X}, F)$.
- This approach allows for estimating the distribution of the statistic T without knowing the true distribution F.

Bootstrap Methodology

Bootstrap Sampling:

- Let $\mathcal{X}^* = \{X_1^*, \ldots, X_n^*\}$ be an i.i.d. sample from the empirical cdf $\widehat{F}$, drawn by sampling size n (with replacement) from the original sample $\mathcal{X}$.
- Given $\mathcal{X}^*$, the statistic $T^* = T(\mathcal{X}^*, \widehat{F})$ can be evaluated.

Bootstrap Distribution:

- Repeated sampling of $\mathcal{X}^*$ produces a sample of T^*'s, $\mathcal{T}^* = \{T_1^*, \ldots, T_B^*\}$ where B is the number of bootstrap samples.
- $\mathcal{T}^*$ can be used to approximate the distribution of $T(\mathcal{X}, F)$.
- For instance, the variance of $T(\mathcal{X}, F)$ can be approximated by the variance of $\{T_1^*, \ldots, T_B^*\}$, i.e., $\mathrm{Var}(T(\mathcal{X}, F)) \approx \mathrm{Var}(\mathcal{T}^*)$

Theoretical Justification:

- **Glivenko-Cantelli Theorem:** States that $\widehat{F} \to F$ almost surely as $n \to \infty$, supporting the idea that i.i.d. sampling from $\widehat{F}$ approximates i.i.d. sampling from F when n is large.
- Numerous rigorous studies have explored the limits of this approximation, identifying conditions under which the bootstrap is effective and its potential pitfalls.

Outline

Basic Setup of Nonparametric Bootstrap

Overview: The nonparametric bootstrap approximates the sampling distribution of a statistic $T(\mathcal{X}, F)$ directly using the empirical distribution derived from bootstrap samples $T_1^*, \ldots, T_B^*$.

Approximation Examples:

- **Quantiles:** Quantiles of $T(\mathcal{X}, F)$ are approximated by the sample quantiles from the bootstrap samples (i.e., from $\mathcal{T}^*$).
- **Variance:** The variance of $T(\mathcal{X}, F)$ is approximated by calculating the sample variance of the bootstrap replicates $\mathcal{T}^* = \{T_1^*, \ldots, T_B^*\}$.

Bootstrap Sample Size:

- Typically, the number of bootstrap replicates B is quite large, often around 1000, to ensure a good approximation of the distribution.
- Despite the large number, this process is computationally manageable with modern software and hardware capabilities.

Example: Variance of a Sample Median

Context and Problem:

- Example 29.4 from DasGupta (2008): Consider $X_1, \ldots, X_n$ as i.i.d. Cauchy distributed variables with median μ.
- The sample mean $\overline{X}$ is unreliable for estimating μ due to the heavy tails of the Cauchy distribution, leading to a preference for the sample median M_n.

Variance of the Sample Median:

- For an odd n, say $n = 2k + 1$, there is an exact formula for the variance:

$$\mathrm{Var}(M_n) = \frac{2\, n!}{(k!)^2\, \pi^n} \int_0^{\pi/2} x^k (\pi - x)^k (\cot x)^2 \, dx$$

- An asymptotic approximation using a Central Limit Theorem (CLT) type approach gives:

$$\widehat{\mathrm{Var}(M_n)} = \frac{\pi^2}{4n}$$

Example: Variance of Sample Median (Cont'd)

Bootstrap Approximation:

- To assess the effectiveness of the bootstrap method in approximating this variance, one can repeatedly sample from the empirical distribution of $\mathcal{X}$ and compute the variance of the resulting sample medians.
- This provides a practical approach to estimate $\text{Var}(M_n)$ when the theoretical distribution is challenging to work with.

Empirical Results with $n = 21$:

- **Theoretical Variance:** $\text{Var}(M_n) = 0.1367$
- **CLT Approximation:** $\widehat{\text{Var}(M_n)} = 0.1175$

Example: Variance of Sample Median (Cont'd)

Bootstrap Estimation:

- Using $B = 5000$ bootstrap samples, the bootstrap estimate of the variance is:

$$\widehat{\text{Var}(M_n)}_{\text{boot}} = 0.1102$$

- This result is a slight underestimate compared to the theoretical value but is quite close, demonstrating the utility of the bootstrap method.

Conclusion:

- The main advantage of the bootstrap method is obtaining a reliable estimate with minimal effort — the computer handles the computational complexity.
- **Note:** For reproducibility, the random seed was set using `set.seed(77)` in the computation.

Technical Points on Bootstrap Method

Bootstrap Consistency:

- Let $H_n(x)$ be the true distribution function for the estimator $\widehat{\theta}_n$, and $H_n^*(x)$ for $\widehat{\theta}_n^*$.
- The bootstrap is considered **consistent** if the distance between $H_n(x)$ and $H_n^*(x)$ converges to 0 in probability as $n \to \infty$, i.e., for all x where $H_n(x)$ is continuous,

$$\Pr(|H_n(x) - H_n^*(x)| > \epsilon) \to 0 \text{ as } n \to \infty.$$

Limitations of Bootstrap: The bootstrap is successful in many problems, but there are known situations when it may fail:

(i) **Support Dependency:** Support of the estimator depends on the parameter.

(ii) **Boundary Issues:** The true parameter lies on the boundary of the parameter space.

Technical Points on Bootstrap Method

Limitations of Bootstrap (Cont'd):

(iii) **Non-standard Rates:** If the estimator's convergence rate is not the standard rate of $n^{-1/2}$, bootstrap may not perform well.

- These conditions may lead to the failure of bootstrap methods.

Advantages and Drawbacks:

- **Detection of Skewness:** Unlike CLT approximations, the bootstrap can detect skewness in the distribution of $\widehat{\theta}_n$ and often has a **second-order accuracy** property.
- **Underestimation of Variances:** Bootstrap methods often underestimate variances, which can mislead inference, especially in cases where precise variance estimation is critical.

Bootstrap Confidence Intervals

cp

Introduction to Bootstrap CIs: Bootstrap methods are commonly used to construct confidence intervals (CIs) for statistical estimators.

Percentile Method:

- Consider $\widehat{\theta}_1^*, \ldots, \widehat{\theta}_B^*$ as a bootstrap sample of point estimators.
- A two-sided $100(1-\alpha)$

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

  where θ^*_p is the $100p$th percentile in the bootstrap sample.
- This method is straightforward and intuitive, making it a popular choice for initial bootstrap CI analyses.

Alternative Methods:

- While the percentile method is simple, there are "better" methods in terms of coverage accuracy and robustness, such

Outline

Definition: Parametric Bootstrap

Variation of Bootstrap: The parametric bootstrap is a variant of the standard (nonparametric) bootstrap, designed for use with parametric models.

Parametric Model: Consider a parametric model $F = F_\theta$, where θ represents the parameters of the model.

Methodology:

- Instead of sampling independently and identically distributed (iid) from the empirical distribution $\widehat{F}$ as in nonparametric bootstrap,
- The parametric bootstrap samples iid from $F_{\widehat{\theta}}$, where $\widehat{\theta}$ is an estimator of θ.

Complexity Considerations:

- This approach can be more complex than the nonparametric method,
- Particularly because sampling from $F_{\widehat{\theta}}$ might involve more complicated methods than sampling from $\widehat{F}$.

Example: Parametric Bootstrap for Variance of Sample Median

Setup:

- Consider a sample $X_1, \ldots, X_n$ of i.i.d. observations from a Cauchy distribution with median μ.
- Let M_n denote the sample median.

Parametric Bootstrap Method:

- Samples $X_1^*, \ldots, X_n^*$ are drawn from a Cauchy distribution with median M_n, effectively using M_n as the parameter estimate.
- This method assumes that the model parameter (median) can be accurately estimated by M_n.

Example: Parametric Bootstrap for Variance of Sample Median

Bootstrap Results:

- Using $B = 5000$ bootstrap replicates, the estimated variance of M_n using the parametric bootstrap is:

$$\widehat{\text{Var}(M_n)}_{p-\text{boot}} = 0.1356$$

- This estimate is closer to the true variance, $\text{Var}(M_n) = 0.1367$, compared to the estimate obtained via nonparametric bootstrap.

Example: Random Effects Model

Hierarchical Model Setup:

- Consider a hierarchical (random effects) model where:
  - $\nu_1, \ldots, \nu_n \stackrel{iid}{\sim} N(\mu, \sigma^2)$.
  - $Y_i|\nu_i \stackrel{ind}{\sim} N(\nu_i, \tau_i^2)$ for $i = 1, \ldots, n$, with ν_i as the random effect for each observation.

- This model is a classic example of a **hierarchical linear model** where the first level involves individual random effects:
  - Random effects $\nu_1, \ldots, \nu_n$ are assumed to be iid $N(\mu, \sigma^2)$. That is, each random effect ν_i has a mean μ and variance σ^2, representing the variation between groups or clusters.
- The second level of the model addresses the observation-specific variability:
  - Given the random effect ν_i, the observation Y_i is modeled as normally distributed as, $Y_i|\nu_i \sim N(\nu_i, \tau_i^2)$, where τ_i^2 is the known variance of each observation within its group. This captures the within-group variability.

Example: Random Effects Model

Model Parameters and Interest:

- Parameters μ and σ are unknown, but τ_i^2 known. The global parameters μ (the overall mean) and σ (the standard deviation of the random effects) are unknown and need to be estimated from the data.
- The parameter of interest is σ, particularly testing if $\sigma \approx 0$ to suggest homogeneity across groups. Estimating σ provides insights into the variability between groups:
  - A small value of σ (close to 0) would suggest **homogeneity across groups**, indicating that the random effects do not vary much from the overall mean μ, and thus, the groups are similar.
  - A larger value of σ suggests significant differences between groups, highlighting the importance of the random effects in the model.

Example: Random Effects Model (Cont'd)

Non-Hierarchical Version:

- In a **simplified non-hierarchical model**, the observations Y_i for $i = 1, \ldots, n$ are modeled as independent:
  - Each Y_i follows a normal distribution $N(\mu, \tau_i^2 + \sigma^2)$, which combines both the inherent variability (τ_i^2) within each group and the variability across groups (σ^2).
- This model structure allows for the estimation of the variance σ^2 using **maximum likelihood estimation (MLE)**:
  - MLE provides a way to estimate model parameters that maximizes the likelihood of observing the given data under the assumed model.

Example: Random Effects Model (Cont'd)

Parametric Bootstrap Application:

- To evaluate the reliability and precision of the σ estimate, we consider using the **parametric bootstrap**:
  - This technique involves repeatedly sampling from the estimated model to generate new datasets.
  - For each sampled dataset, σ is re-estimated to create a distribution of σ estimates.
  - The variability of these estimates is used to construct **confidence intervals** for σ, providing insights into the stability and confidence of our original estimate.

Random Effects Model: Coverage of Confidence Intervals

Scenario:

- Want to see what happens when $\sigma \approx 0$ (i.e., how bootstrap CIs fare for σ around 0).
- Exploring cases when the parameter $\sigma = O(n^{-1/2})$, which implies that σ decreases at the rate of $n^{-1/2}$, near the boundary of $\sigma \geq 0$.

Coverage Performance of Confidence Intervals

- We use two-sided 95% parametric bootstrap percentile CIs to estimate the coverage probability.
- Our results indicate **low coverage rates** for these intervals when σ is near zero:
  - This suggests that the CIs are often narrower than needed, failing to include the true σ value 95% of the time as expected.
  - Such poor performance persists even as the sample size n increases, highlighting a potential limitation of the parametric bootstrap method in boundary scenarios.

Random Effects Model: Coverage of Confidence Intervals

Empirical Coverage and Interval Length:

n	Coverage	Length
50	0.758	0.183
100	0.767	0.138
250	0.795	0.079
500	0.874	0.039

Table 1: Coverage and length of confidence intervals for different sample sizes.

Remedy:

- While it is possible to achieve intervals with exact coverage, common bootstrap methods may need adjustments for parameters near the boundary of the parameter space.

Outline

Paired Bootstrap in Observational Studies

Setup:

- In observational studies, pairs $\mathbf{z}_i = (x_i^\top, y_i)^\top$ are sampled from a joint predictor-response distribution.
- Let $\mathcal{Z} = \{\mathbf{z}_1, \ldots, \mathbf{z}_n\}$.

Paired Bootstrap Method:

- Following the basic bootstrap principle, repeatedly sample $\mathcal{Z}^* = \{\mathbf{z}_1^*, \ldots, \mathbf{z}_n^*\}$ with replacement from $\mathcal{Z}$.
- Use these samples to approximate the sampling distributions based on the empirical distribution from the bootstrap sample. This approach is termed the *paired bootstrap*.

Paired Bootstrap in Observational Studies

Fixed Design Complications: What about for a fixed design?

- In fixed design settings, y_i values are not i.i.d., introducing complexities.
- **Bootstrap Approach for Fixed Design:**
  - First, resample the residuals $e_i = y_i - \widehat{y}_i$ from the original least squares (LS) fit.
  - Then set $y_i^* = x_i^\top \widehat{\beta} + e_i^*$, where $\widehat{\beta}$ is the estimated coefficient from the LS model.

Example: Ratio of Slope Coefficients in Regression - Observational Data

Model and Parameter of Interest:

- Consider the simple linear regression model for **observational data**:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i = 1, \ldots, n,$$

  where ϵ_i are i.i.d. with mean zero, not necessarily normal.
- The parameter of interest is $\theta = \beta_1/\beta_0$.
- A natural estimate of θ is $\widehat{\theta} = \widehat{\beta}_1/\widehat{\beta}_0$.

Paired Bootstrap Method:

- To obtain the bootstrap distribution of $\widehat{\theta}$:
- Sample $\mathcal{Z}^* = \{\mathbf{z}_1^*, \ldots, \mathbf{z}_n^*\}$ with replacement from $\mathcal{Z}$, where $\mathcal{Z}$ consists of the pairs (x_i, y_i).
- Fit the regression model to the bootstrap sample $\mathcal{Z}^*$ to estimate $\widehat{\beta}_0^*$ and $\widehat{\beta}_1^*$.
- Evaluate $\widehat{\theta}^* = \widehat{\beta}_1^*/\widehat{\beta}_0^*$.

Bootstrap Distribution of Ratio of Slope Coefficients

Purpose of the Bootstrap:

- This bootstrap method allows us to estimate the variability and construct confidence intervals for θ under the model assumption.
- It accounts for the correlation between $\widehat{\beta}_1$ and $\widehat{\beta}_0$ by using the paired bootstrap approach.

Visualization of Bootstrap Distribution:

- The histogram below shows the bootstrap distribution of $\widehat{\theta} = \widehat{\beta}_1/\widehat{\beta}_0$, representing the ratio of slope coefficients.
- The distribution illustrates how $\widehat{\theta}$ varies across bootstrap samples.

Bootstrap Distribution of Ratio of Slope Coefficients

Bootstrap Confidence Interval:

- Based on the bootstrap samples, a 95% confidence interval for θ is calculated as: $(-0.205, -0.173)$
- This interval indicates the range in which the true ratio of slope coefficients is likely to fall with 95% confidence.

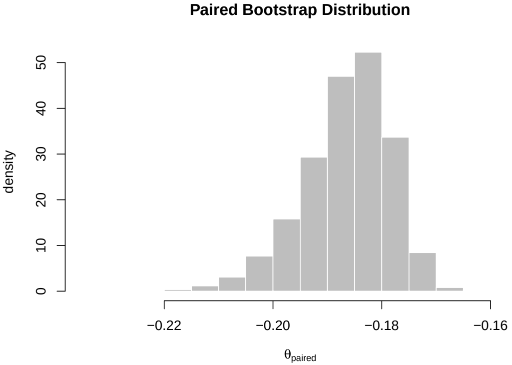

Figure 1: Bootstrap Distribution of $\widehat{\theta}$

Outline

Constructing Confidence Intervals for Parameters

Context of Confidence Intervals:

- Aim to construct a $100(1-\alpha)\%$ confidence interval (CI) for a parameter θ.
- Commonly used for the mean μ of a normal population.

Traditional t-Distribution Method:

- The formula for the t confidence interval is:

$$\overline{X} \pm t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}$$

  where $\overline{X}$ is the sample mean, S is the sample standard deviation, n is the sample size, and $t_{\alpha/2,n-1}$ are the critical values from the Student's t-distribution.

- This approach is appropriate because the standardized mean, $(\overline{X}-\mu)/(S/\sqrt{n})$, exactly follows a Student's t-distribution.

Constructing Confidence Intervals for Parameters

Bootstrap Methods for CIs:

- Previously discussed: Bootstrap Percentile Method.
- Now introducing: Bootstrap t Method.
  - Similar to the traditional t-method but uses bootstrap samples to estimate the distribution of $(\overline{X} - \mu)/(S/\sqrt{n})$.

Comparison of Methods:

- Various approaches offer different advantages depending on assumptions about the data distribution and sample size.

Bootstrap t Method for Confidence Intervals

Bootstrap t Statistics:

- Let $\widehat{\theta}_n$ be a given estimate of the parameter θ from the original sample.
- Suppose $\widehat{\sigma}_n$ is an estimate of the standard deviation of $\widehat{\theta}_n$, typically $\widehat{\sigma}_n = \frac{S}{\sqrt{n}}$ in the normal mean problem.
- Delta Method can often be used to approximate $\widehat{\sigma}_n$.

Construction of Bootstrap t-Statistic:

- Define the t-statistic as $T_n = \frac{\widehat{\theta}_n - \theta}{\widehat{\sigma}_n}$.
- Its bootstrap counterpart is $T_n^* = \frac{\widehat{\theta}_n^* - \widehat{\theta}_n}{\widehat{\sigma}_n}$, where $\widehat{\theta}_n^*$ is obtained from bootstrap sample and $\widehat{\sigma}_n$ are obtained from the original sample.
- We could also use $\widehat{\sigma}_n^*$ (obtained from bootstrap samples) instead of $\widehat{\sigma}_n$. However, using $\widehat{\sigma}_n$ stabilizes the variance estimate (which can be particularly useful when the sample size is small, or the parameter estimate is complex).

Bootstrap t Method for Confidence Intervals

Bootstrap t Confidence Interval:

- The bootstrap t confidence interval for θ is calculated as:

$$\left[\widehat{\theta}_n + c_{\alpha/2}\widehat{\sigma}_n, \widehat{\theta}_n + c_{1-\alpha/2}\widehat{\sigma}_n\right]$$

- Here, c_p is the $100p$th percentile of the bootstrap distribution of T_n^*.

Advantages of Bootstrap t Method:

- This method provides a more robust estimate of confidence intervals by incorporating variability from the bootstrap distribution, making it suitable for more complex or non-standard distributions.

Bootstrap Methods in Regression Analysis: Example Cont'd

Linear Model and Parameter Estimation:

- Linear model parameter of interest: $\theta = \beta_1/\beta_0$.
- MLE estimate: $\widehat{\theta} = \widehat{\beta}_1/\widehat{\beta}_0$.
- Use any available approximation to estimate the variance σ^2 of $\widehat{\theta}$.

Variance Estimation Using Delta Theorem:

- Delta theorem approximation gives:

$$\widehat{\sigma}^2 = \widehat{\theta}^2 \left(\frac{\widehat{\mathrm{Var}(\widehat{\beta}_1)}}{\widehat{\beta}_1^2} + \frac{\widehat{\mathrm{Var}(\widehat{\beta}_0)}}{\widehat{\beta}_0^2} - 2\frac{\widehat{\mathrm{Cov}(\widehat{\beta}_0, \widehat{\beta}_1)}}{\widehat{\beta}_0\widehat{\beta}_1} \right)$$

Bootstrap Methods in Regression Analysis: Example Cont'd

Bootstrap t-Statistics and Confidence Intervals:

- Compute bootstrap estimates $(\widehat{\theta}^*, \widehat{\sigma}^{2*})$ and t-statistics $T^* = \frac{\widehat{\theta}^* - \widehat{\theta}}{\widehat{\sigma}^*}$.
- Bootstrap Confidence Intervals for θ:
  - Percentile method: $[-0.205, -0.174]$
  - t-method: $[-0.203, -0.167]$

Implications:

- These confidence intervals provide insights into the precision and reliability of the parameter estimates under the model assumptions.

Alternative Bootstrap Confidence Intervals

Limitations of Bootstrap Percentile CIs:

- In some cases, bootstrap percentile confidence intervals may not provide accurate coverage due to inherent biases in the bootstrap distribution.

Bias-Corrected Bootstrap Percentile CI:

- To improve accuracy, one can correct for the bootstrap bias using the *bias-corrected* (BC) method.
- The two-sided $100(1-\alpha)\%$ BC bootstrap CI involves selecting different quantiles from the bootstrap distribution of $\widehat{\theta}$.
- Instead of the standard $\left[\widehat{\theta}_{\alpha/2}, \widehat{\theta}_{1-\alpha/2}\right]$, use $\left[\widehat{\theta}_{z_1}, \widehat{\theta}_{z_2}\right]$, where z_1 and z_2 are adjusted quantiles based on user-specified constants (a, b).

Alternative Bootstrap Confidence Intervals

Formulas and Parameter Choices:

- The textbook provides specific formulas for calculating z_1 and z_2 and discusses how to choose the constants (a, b) effectively.
- For detailed guidance and examples, refer to G&H, Section 9.3.2.1.

Implications:

- Using bias-corrected methods can significantly enhance the reliability and validity of confidence intervals derived from bootstrap methods, especially in skewed distributions or small sample sizes.

Outline

Example: Bootstrap Failure

Setup and Statistic Definition:

- Consider $X_1, \ldots, X_n \overset{iid}{\sim} \text{Uniform}(0, \theta)$.
- Define $T_n = n(\theta - \widehat{\theta}_n)$, where $\widehat{\theta}_n = X_{(n)} = \max\{X_i\}$.

Bootstrap Resampling Statistic:

- Let $\widehat{\theta}_n^* = X_{(n)}^*$ from the bootstrap sample $\mathcal{X}^*$.
- Define $T_n^* = n(\widehat{\theta}_n - \widehat{\theta}_n^*)$.

Example: Bootstrap Failure

Bootstrap Failure:

- The distributions of T_n^* and T_n are not close as $n \to \infty$.
- Specifically, for any $t \geq 0$,

$$P(T_n^* \leq t) \geq P(T_n^* = 0) = \left(\frac{n-1}{n}\right)^n \to 1 - e^{-1} \quad \text{as } n \to \infty.$$

- If $\theta = 1$, then as t becomes small, $P(T_n \leq t) \to 1 - e^{-t}$, approaching 0 for small t.
- Conversely, $P(T_n^* \leq t)$ remains significantly larger, approximately 0.63, indicating a failure in approximation.

Conclusion:

- This example illustrates a scenario where the bootstrap method does not effectively approximate the sampling distribution of a statistic, particularly as the sample size increases.

Remedy for Bootstrap Failures: m-out-of-n Bootstrap

Observation of Bootstrap Failures:

- In some statistical settings, the standard bootstrap method, which samples the full data set size n, fails to provide accurate results.

m-out-of-n Bootstrap Method:

- A modified approach involves taking bootstrap samples of size m where $m = o(n)$ (i.e., m grows slower than n).
- This method is termed the *m-out-of-n bootstrap*.
- It can be particularly effective in scenarios where the regular bootstrap fails, such as with certain distributions like Uniform(0, θ).

Remedy for Bootstrap Failures: m-out-of-n Bootstrap

Theoretical Justification and Further Reading:

- The m-out-of-n bootstrap has been shown to be consistent in cases where the standard bootstrap method is not.
- For a detailed theoretical foundation and empirical studies, refer to:
  - DasGupta, Sections 29.7–29.8.
  - Bickel, Gotze, and van Zwet (1997) in *Statistica Sinica*.

Implications:

- Adopting the m-out-of-n bootstrap method can improve the reliability of bootstrap confidence intervals and other statistical estimates in challenging scenarios.

Bootstrap Failure and m-out-of-n Bootstrap Remedy

Revisiting the Uniform Distribution Problem:

- Reconsidering the bootstrap failure in the uniform distribution example.
- The previous example used full sample size n for bootstrapping.

Modified Bootstrap Approach:

- Introduce m-out-of-n bootstrap where $m = o(n)$ (grows slower than n).
- Define the statistic:

$$T^*_{n,m} = m(\widehat{\theta}_n - \widehat{\theta}^*_{n,m}),$$

where $\widehat{\theta}^*_{n,m} = \max\{X_1^*, \ldots, X_m^*\}$.

Bootstrap Failure and m-out-of-n Bootstrap Remedy

Consistency of m-out-of-n Bootstrap:

- Theory guarantees that the distributions $P(T^*_{n,m} \leq t)$ and $P(T_n \leq t)$ become similar as $n \to \infty$, i.e., m-out-of-n bootstrap is consistent.
- This approach addresses the consistency issues observed in the regular bootstrap for this specific problem.

Practical Implementation:

- A rule of thumb for choosing m is approximately $m \approx 2\sqrt{n}$.
- Simulation details will be presented on the next slide to illustrate the effectiveness of this approach.

Bootstrap Failure Cont'd: Uniform Distribution Example

Simulation Setup:

- Revisiting the uniform distribution problem with sample size $n = 100$ and bootstrap sample size $m = 20$.

Histogram Comparison:

- The left histogram shows the distribution of T_n^* (ordinary bootstrap), and the right histogram shows $T_{n,m}^*$ (modified m-out-of-n bootstrap).
- Both histograms include the true limiting distribution of T_n overlaid for comparison.

Bootstrap Failure Cont'd: Uniform Distribution Example

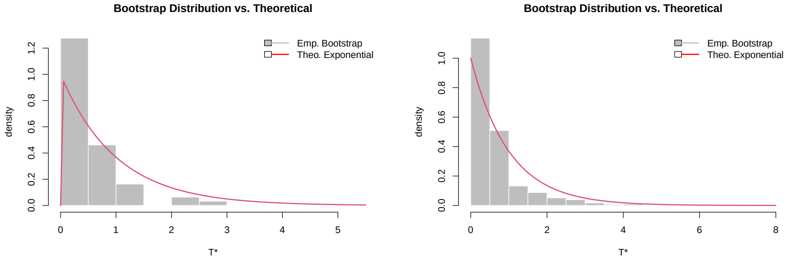

Figure 2: Left: Histogram for regular bootstrap. Right: Histogram for *m*-out-of-*n* bootstrap.

Observations:

- Note how the histograms compare to the theoretical distribution.
- Compare the effectiveness of the *m*-out-of-*n* method in addressing bootstrap failure observed with the ordinary method.

Outline

Bootstrap Methods in Hypothesis Testing

Role of Bootstrap in Hypothesis Testing (HT):

- Hypothesis testing often requires knowledge of the sampling distribution under the null hypothesis to set rejection thresholds or compute p-values.

Limitations of Nonparametric Bootstrap:

- Since the nonparametric bootstrap samples from the observed data, it may not accurately reflect the null hypothesis distribution if the null does not correspond to the observed data distribution.
- This makes nonparametric bootstrap inappropriate for most HT where the null distribution is distinct from the observed.

Bootstrap Methods in Hypothesis Testing

Parametric Bootstrap Approach:

- A more suitable approach in hypothesis testing is the parametric bootstrap, which involves sampling from P_{θ_0}, the assumed true distribution under the null hypothesis.
- Aligns with Monte Carlo approximation techniques, providing a rigorous way to simulate the null distribution.

Adjustments for Nonparametric Bootstrap:

- Section 9.3.3 in G&H discusses adjustments to nonparametric bootstrap methods that can make them suitable for HT.
- These adjustments often involve modifications to ensure that the bootstrap distribution mirrors the theoretical null distribution more closely.

Implications: Using bootstrap methods in HT can enhance the accuracy of p-values and decision rules but requires careful consideration of the bootstrap type and adjustments for the specific testing scenario.

Related Idea: Permutation Tests for Testing for Identical Distributions

Hypothesis of Interest: The goal is to test whether two groups are "identical" in terms of their statistical distributions.

Conceptual Framework:

- Consider two datasets, $\mathcal{X} = \{X_1, \ldots, X_n\}$ and $\mathcal{Y} = \{Y_1, \ldots, Y_m\}$.
- Under the null hypothesis that both groups are from the same distribution, any re-grouping of these combined data into two new groups of sizes n and m should be statistically equivalent.

Implementation of Permutation Tests:

- By permuting the group labels of the combined dataset $\{\mathcal{X}, \mathcal{Y}\}$, one can generate new samples that respect the null hypothesis.
- Test statistics (like means, variances, etc.) are computed for each permutation, forming a distribution of these statistics under the null hypothesis.

Permutation Tests: Testing for Identical Distributions

Approximation of Sampling Distributions:

- The permutation test approximates the sampling distribution of the test statistic by considering all (or a significant sample of) possible partitions of the combined data.
- This method is particularly effective for assessing the significance of observed differences between the two groups without making assumptions about the underlying distributions.

Practical Considerations:

- This test is non-parametric and does not require the assumption of normal distributions or equal variances, making it robust and widely applicable.

Bootstrapping Dependent Data

Review of Bootstrap Methods:

- Previously, we discussed bootstrap methods for:
  - Independent and identically distributed (iid) data.
  - Independent but not identically distributed (not-iid) data, such as in regression problems.
- In these cases, the order of data presentation does not affect the analysis.

Challenges with Dependent Data:

- For dependent data, such as time series, the sequence or order of data points is crucial.
- Standard bootstrap methods that reshuffle data points lose the dependency structure, making them inappropriate.

Bootstrapping Dependent Data

Adjustments for Dependent Data:

- Adjustments involve bootstrapping blocks of data to preserve the internal dependency.
- Known as "block bootstrapping," this method works with sequences of data rather than individual points.

Further Reading:

- For more detailed theoretical and practical guidance on bootstrap methods for dependent data:
  - See Section 9.5 in G&H.

Note:

- The specifics of bootstrapping for dependent data can be complex, and still an open area of research.

Conclusion: Reflections on Bootstrap Methods

Bootstrap as a Tool:

- Bootstrap is a powerful statistical tool that provides a simple and effective way to approximate sampling distributions numerically.
- This method circumvents the need for complex analytical calculations, making it accessible and practical for a wide range of applications.

Cautions in Using Bootstrap:

- Despite its apparent simplicity and automation, bootstrap methods require careful application:
  - **Sample Size Considerations:** The effectiveness of bootstrap approximations heavily depends on the sample size (n). Small sample sizes may not provide reliable results.
  - **Potential for Failure:** There are scenarios where bootstrap methods can fail, particularly when the underlying assumptions required for the bootstrap are not met.

Conclusion: Reflections on Bootstrap Methods

The Need for Theoretical Understanding:

- It is crucial not to use bootstrap methods blindly. A solid understanding of the underlying statistical theory and the specific data context is essential to ensure valid results.
- Researchers and practitioners must remain vigilant and assess the suitability of bootstrap for their specific problem, considering both theoretical and practical aspects.

Final Thoughts:

- While bootstrap is an invaluable tool in statistics, its application must be guided by both statistical theory and empirical evidence to avoid misuse and to harness its full potential effectively.