The Bootstrap

Section 8 & Lecture Notes 20

Li Yuekai

City University of Hong Kong

BIOS8004 Presentation 2024.11.27

Contents

Plug-in Estimator and Monte Carlo Simulation

The Empirical Distribution Function

- **Definition 1.1:**
  - Let $X_1, \ldots, X_n \sim F$ be i.i.d. samples where F is a distribution function. The empirical distribution function $\hat{F}_n$ is defined as:

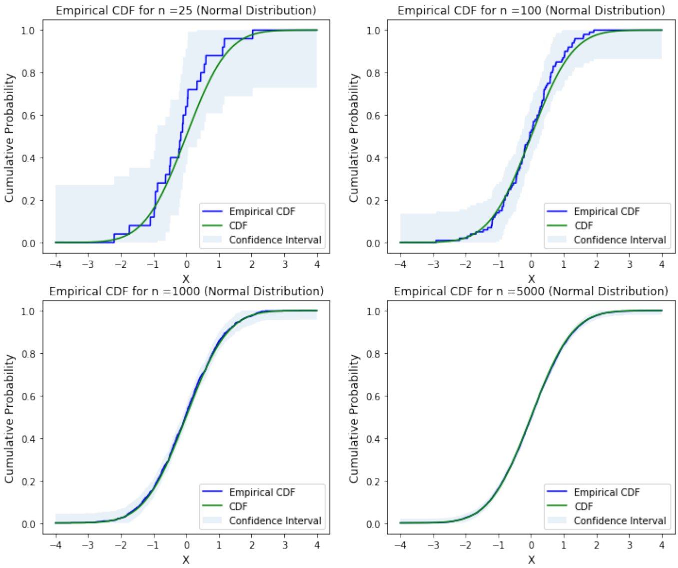
$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} I(X_i \leq x) \tag{1.1}$$

  - Here, $I(X_i \leq x)$ is the indicator function:

$$I(X_i \leq x) = \begin{cases} 1, & \text{if } X_i \leq x \\ 0, & \text{if } X_i > x \end{cases} \tag{1.2}$$

- **Interpretation:**
  - The empirical distribution function assigns a mass of $\frac{1}{n}$ to each data point X_i.
  - It is a step function that increases by $\frac{1}{n}$ at each data point.

The Empirical Distribution Function



The Glivenko-Cantelli Theorem

- **Theorem 1.1:**

  Let $X_1, \ldots, X_n \sim F$ be i.i.d. samples. Then:

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \xrightarrow{a.s.} 0 \tag{1.3}$$

  - The Glivenko-Cantelli theorem describes the asymptotic properties of the empirical distribution function of i.i.d. samples.
  - The EDF converges uniformly to the true distribution function $F(x)$ as $n \to \infty$, and this convergence is almost sure.
  - It is a stronger result than the pointwise convergence provided by the strong law of large numbers.

Statistical Functionals

- **Definition 1.2:**
  - A statistical functional ψ is a map from a distribution function F to a real number.
  - Essentially, the numerical characteristics of random variables (such as mean and variance) are statistical functionals of distribution functions.
- **Examples:**
  - Mean μ: $\psi(F) = \int x \, dF(x)$
  - Variance σ^2: $\psi(F) = \int x^2 \, dF(x) - \left(\int x \, dF(x)\right)^2$
  - Median m: $\psi(F) = F^{-1}(1/2)$

Plug-in Estimator

- **Definition 1.3:**
  - A plug-in estimator is a method of estimating a statistical functional by replacing the true distribution function with the empirical distribution function.
  - The plug-in estimator of $\theta = \psi(F)$ is defined by:

$$\hat{\theta}_n = \psi(\hat{F}_n) \tag{1.4}$$

- **Theorem 1.2:**

  The plug-in estimator for linear functional $\psi(F) = \int r(x)\,dF(x)$ is:

$$\psi(\hat{F}_n) = \int r(x)\,d\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} r(X_i) \tag{1.5}$$

Plug-in Estimator

- **Example:**
  - Let $\mu = \psi(F) = \int x \, dF(x)$, then the plug-in estimator is:

$$\hat{\mu} = \psi(\hat{F}_n) = \int x \, d\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} X_i = \bar{X} \tag{1.6}$$

  - Let $\sigma^2 = \psi(F) = \int x^2 \, dF(x) - \left(\int x \, dF(x)\right)^2$, then the plug-in estimator is:

$$\begin{aligned} \hat{\sigma}^2 = \psi(\hat{F}_n) &= \int x^2 \, d\hat{F}_n(x) - \left(\int x \, d\hat{F}_n(x)\right)^2 \\ &= \frac{1}{n} \sum_{i=1}^{n} X_i^2 - (\frac{1}{n} \sum_{i=1}^{n} X_i)^2 \\ &= \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2 \end{aligned} \tag{1.7}$$

Plug-in Estimator

- **Convergence Issues:**
  - Plug-in estimators do not guarantee convergence to the true value in all cases.
  - Although the empirical distribution function $\hat{F}_n$ strongly converges to F, this does not ensure the convergence of the plug-in estimator $\psi(\hat{F}_n)$ to $\psi(F)$.
- **Infinite-Dimensional Challenges:**
  - $\hat{F}_n$ and F are infinite-dimensional because they are functions defined on $\mathbb{R}$.
  - This makes it impossible to directly apply the Continuous Mapping Theorem and then prove the convergence.
- **Practical Considerations:**
  - The effectiveness of a plug-in estimator depends on the properties of the statistical functional ψ and the distribution function F.

The application: Monte Carlo Simulation

- **The Application:**
  - A direct application of plug-in estimators is to estimate a distribution's statistical functional (such as the mean or the variance) using Monte Carlo simulation.
  - This is achievable because plug-in estimators of mean and variance converge in probability to true values, as shown by the weak law of large numbers.
- **Details:**
  - Suppose that $X \sim F$ and a large sample $X_1, \ldots, X_B$ is drawn. Then:

$$\hat{\mu} = \frac{1}{B}\sum_{i=1}^{B} X_i \xrightarrow{p} \mathbb{E}[X] \tag{1.8}$$

$$\hat{\sigma}^2 = \frac{1}{B}\sum_{i=1}^{B}(X_i - \bar{X})^2 \xrightarrow{p} \mathbb{E}[X-\mu]^2 = \mathrm{Var}[X] \tag{1.9}$$

Bootstrap Variance Estimation

Compute the Variance of a Statistic directly

- **Variance of a Statistic:**
  - Suppose that a statistic $T_n = g(X_1, \ldots, X_n)$ is a function of n i.i.d. samples from distribution F.
  - Naturally, the statistic has its own distribution, and variance $\mathbb{V}_F[T_n]$.
- **A simple example:**

  Let $T_n = \bar{X}$, then $\mathbb{V}_F[T_n] = \dfrac{\sigma^2}{n}$, where:

$$\sigma^2 = \int (x-\mu)^2 \, dF(x) \tag{2.1}$$

$$\mu = \int x \, dF(x) \tag{2.2}$$

Obviously, the variance of T_n is a functional of distribution function F.

Estimate the Variance of a Statistic with Monte Carlo

- **Monte Carlo Simulation Approach:**
  - The distribution of a statistic T_n may be complex, making direct computation of its variance $\mathbb{V}_F[T_n]$ challenging.
  - If we are able to draw samples of size n from the joint density $p(x_1, \ldots, x_n) = p(x_1) \cdots p(x_n)$ each time, we can still use Monte-Carlo simulation to estimate $\mathbb{V}_F[T_n]$.
- **Details:**
  - In other words we do the following for j in $1, \ldots, B$:

    draw $X_{1j}, \ldots, X_{nj} \sim F$, and compute $T_{nj} = g(X_{1j}, \ldots, X_{nj})$
  - Again, by the law of large numbers, as $B \to \infty$ we have:

$$\frac{1}{B} \sum_{j=1}^{B} (T_{nj} - \sum_{j=1}^{B} T_{nj})^2 \xrightarrow{p} \mathbb{V}_F[T_n] \tag{2.3}$$

Bootstrap Variance Estimation

- **Challenge:**
  - When the distribution function F is unknown and no additional samples can be obtained, estimating the $\mathbb{V}_{F}[T_n]$ becomes more challenging.
- **Bootstrap Approach:**
  - Use the available n samples to construct an empirical distribution $\hat{F}_n$.
  - Sample from $\hat{F}_n$ to simulate additional samples, compute more samples of statistic, and finally calculate $\widehat{\mathbb{V}_{\hat{F}}[T_n]}$, which is the estimator of $\mathbb{V}_{\hat{F}}[T_n]$.
- **The Approximation:**

  The Bootstrap idea has two steps for approximation:

$$\widehat{\mathbb{V}_{\hat{F}}[T_n]} \xrightarrow[\text{plug-in}]{\text{Monte Carlo Simulation}} \mathbb{V}_{\hat{F}}[T_n] \xrightarrow[\text{plug-in}]{\text{Empirical Approximation}} \mathbb{V}_{F}[T_n]$$

Bootstrap Variance Estimation

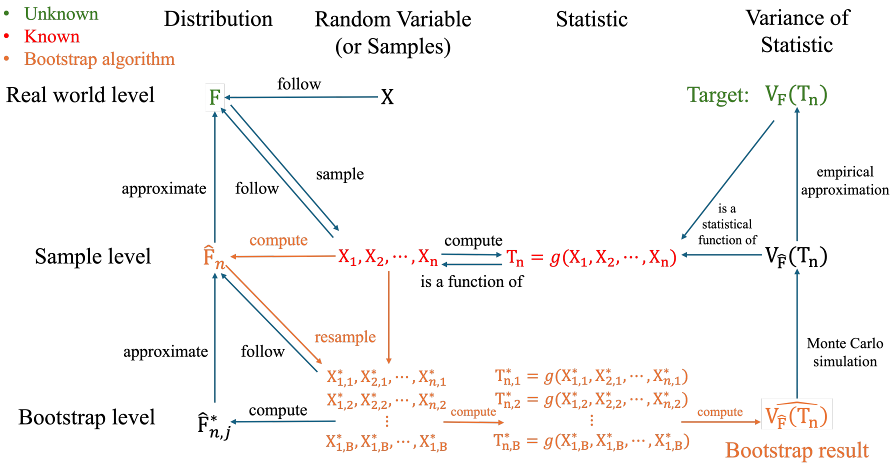
- **Resample:**
  - Notice that $\hat{F}_n$ puts a mass $\frac{1}{n}$ at each data point $X_1, \ldots, X_n$. Therefore, drawing n observations from $\hat{F}_n$ is equivalent to drawing n samples randomly from the original dataset with replacement.

Algorithm 1: Bootstrap Variance Estimation algorithm

Input: i.i.d. samples $X_1, \ldots, X_n$

for $j \leftarrow 1$ **to** B **do**
  Draw $X^*_{1j}, \ldots, X^*_{nj}$ from $X_1, \ldots, X_n$ with replacement ;
  Compute $T^*_{nj} = g(X^*_{1j}, \ldots, X^*_{nj})$;
end
Let: $\widehat{\mathbb{V}_{\hat{F}}[T_n]} = \frac{1}{n}\sum_{j=1}^{B}(T^*_{nj} - \frac{1}{B}\sum_{j=1}^{B} T^*_{nj})^2$.
return $\widehat{\mathbb{V}_{\hat{F}}[T_n]}$

Bootstrap Variance Estimation



Bootstrap Asymptotic Properties

Is Bootstrap Variance a "good" estimator?

- **Consistency of bootstrap variance estimator:**
  - Here, We focus on how the Bootstrap estimator changes with sample size n.
  - With resampling many many times, we just need to focus on whether the distribution of $T^*_{n,j}$ approximates the distribution of T_n as $n \to \infty$.
  - Unfortunately, for general functions $g()$, the bootstrap does not guarantee that $T^*_{n,j}$ converges to T_n asymptotically in all cases, which requires more conditions.
  - Next, We will discuss a simple example where Bootstrap works well and a counterexample showing the failure of Bootstrap.

Examples

- **A Simple Example: the estimate of a distribution mean**
  - Suppose that the mean of a distribution function F has the form:

$$\mu = \psi(F) = \int x \, dF(x) \tag{3.1}$$

  - Given i.i.d. samples $X_1, \ldots, X_n$ from F, the plug-in estimator is:

$$\hat{\mu}_n = \psi(\hat{F}_n) = \int x \, d\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} X_i \tag{3.2}$$

  - Given Bootstrap samples $X_1^*, \ldots, X_n^*$ from empirical distribution function $\hat{F}_n$, the Bootstrap estimator is:

$$\hat{\mu}_n^* = \psi(\hat{F}_n^*) = \int x \, d\hat{F}_n^*(x) = \frac{1}{n} \sum_{i=1}^{n} X_i^* \tag{3.3}$$

Examples

- **A Simple Example: the estimate of a distribution mean**
  - It is clear from the Central Limit Theorem that as $n \to \infty$:

$$\hat{\mu}_n \xrightarrow{d} N(\mu, \frac{\sigma^2}{n}) \tag{3.4}$$

$$\hat{\mu}_n^* \xrightarrow{d} N(\int x\, d\hat{F}_n(x), \frac{1}{n}[\int x^2\, d\hat{F}_n(x) - (\int x\, d\hat{F}_n(x))^2]) \tag{3.5}$$

  - According to the law of large numbers:

$$\int x\, d\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p} \mu \tag{3.6}$$

$$\int x^2\, d\hat{F}_n(x) - (\int x\, d\hat{F}_n(x))^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 \xrightarrow{p} \sigma^2 \tag{3.7}$$

  - This shows that $\hat{\mu}_n$ and $\hat{\mu}_n^*$ are identically distributed as $n \to \infty$.

Examples

- **Theorem 3.1(Bootstrap Theorem):**

  Suppose that $X_1, \ldots, X_n \sim F$ where X_i has mean μ and variance σ^2.

  Let $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i$, $\hat{\mu}^* = \frac{1}{n}\sum_{i=1}^{n} X_i^*$ and define:

$$G_n(t) = \mathrm{P}(\sqrt{n}(\hat{\mu}_n - \mu) \leq t) \tag{3.8}$$

$$\hat{G}_n(t) = \mathrm{P}(\sqrt{n}(\hat{\mu}_n^* - \hat{\mu}) \leq t) \tag{3.9}$$

  Suppose that $\mathbb{E}|X_i|^3 < \infty$. Then:

$$\sup_t |\hat{G}_n(t) - G_n(t)| = O_p(\frac{1}{\sqrt{n}}). \tag{3.10}$$

Examples

- **A Counterexample: the estimate of a Uniform distribution parameter**
  - Suppose that $X_1, \ldots, X_n \sim U[0, \theta]$, and we want to bootstrap the MLE estimator $X_{(n)}$ of θ. We have the distribution of $X_{(n)}$ and as $n \to \infty$:

$$\mathrm{P}(X_{(n)} < x) = (\frac{x}{\theta})^n \to \mathrm{I}(x \geq \theta) \tag{3.11}$$

  - However, for the Bootstrap estimator $X^*_{(n)}$:

$$\mathrm{P}(X^*_{(n)} = X_{(n)}) = 1 - (1 - 1/n)^n \to 1 - 1/e \tag{3.12}$$

  - Notice that the bootstrap distribution puts mass $1/e$ at $X_{(i)}$ $(i < n)$. Despite $n \to \infty$, we still have $X_{(i)} < X_{(n)} \leq \theta$, which shows that the bootstrap distribution does not resemble the true distribution in this case.

Bootstrap Confidence Intervals

The Normal Bootstrap Interval

- **Overview:**
  - Bootstrap can be used to get a confidence interval for parameter estimation.
  - Let $\theta = \psi(F)$ and plug-in estimator $\hat{\theta} = \psi(\hat{F}_n)$. We can get $100(1-\alpha)\%$ confidence intervals of θ with the following methods.
- **The Normal Bootstrap Interval:**
  - Suppose $\hat{\text{se}}_{\text{boot}} = \sqrt{\hat{\mathbb{V}}_{\text{boot}}}$ and $\hat{\mathbb{V}}_{\text{boot}}$ is the Bootstrap estimate of $\text{Var}[\hat{\theta}]$.
  - The Normal Bootstrap Confidence Interval is given by:

$$C_n = [\hat{\theta} - z_{\alpha/2} \cdot \hat{\text{se}}_{\text{boot}},\ \hat{\theta} + z_{\alpha/2} \cdot \hat{\text{se}}_{\text{boot}}] \tag{4.1}$$

where $z_{\alpha/2} = \Phi^{-1}(1-\alpha/2)$.

The Normal Bootstrap Interval

- **Derivation:**

$$
\begin{aligned}
&\mathrm{P}(\hat{\theta} - z_{\alpha/2} \cdot \hat{\mathrm{se}}_{\mathrm{boot}} < \theta < \hat{\theta} + z_{\alpha/2} \cdot \hat{\mathrm{se}}_{\mathrm{boot}}) \\
=&\mathrm{P}(-z_{\alpha/2} < \frac{\hat{\theta} - \theta}{\hat{\mathrm{se}}_{\mathrm{boot}}} < z_{\alpha/2}) \\
\approx&\mathrm{P}(-z_{\alpha/2} < \frac{\hat{\theta} - \theta}{\mathrm{se}(\hat{\theta})} < z_{\alpha/2}) \\
\approx&\mathrm{P}(-z_{\alpha/2} < \frac{\hat{\theta} - \mathbb{E}[\hat{\theta}]}{\mathrm{se}(\hat{\theta})} < z_{\alpha/2}) \\
\approx&1 - \alpha
\end{aligned}
\tag{4.2}
$$

- **Limitations:**
  - The interval is not accurate unless the distribution of $\hat{\theta}$ is an unbiased estimator for θ and close to Normal.

The Percentile Bootstrap Interval

- **The Percentile Bootstrap Interval:**
  - Suppose $\hat{\theta}_1^*, \ldots, \hat{\theta}_B^*$ are Bootstrap statistics, the percentile Bootstrap interval is given by:

$$C_n = [\hat{\theta}^*_{(\alpha/2)},\ \hat{\theta}^*_{(1-\alpha/2)}] \tag{4.3}$$

  where $\hat{\theta}^*_{(\alpha/2)}$ is the $\alpha/2$ quantile of $\hat{\theta}_1^*, \ldots, \hat{\theta}_B^*$ and $\hat{\theta}^*_{(1-\alpha/2)}$ is the $1-\alpha/2$ quantile of $\hat{\theta}_1^*, \ldots, \hat{\theta}_B^*$

- **Derivation:**

$$\mathrm{P}(\hat{\theta}^*_{(\alpha/2)} < \theta < \hat{\theta}^*_{(1-\alpha/2)}) \approx \mathrm{P}(\hat{\theta}^*_{(\alpha/2)} < \hat{\theta} < \hat{\theta}^*_{(1-\alpha/2)}) \approx 1-\alpha \tag{4.4}$$

The Basic Bootstrap Interval

- **The Basic Bootstrap Interval:**
  - Suppose $\hat{\theta}_1^*, \ldots, \hat{\theta}_B^*$ are Bootstrap statistics, the basic Bootstrap interval is given by:

$$C_n = [2\hat{\theta} - \hat{\theta}^*_{(1-\alpha/2)},\ 2\hat{\theta} - \hat{\theta}^*_{(\alpha/2)}] \tag{4.5}$$

where $\hat{\theta}^*_{(\alpha/2)}$ is the $\alpha/2$ quantile of $\hat{\theta}_1^*, \ldots, \hat{\theta}_B^*$ and $\hat{\theta}^*_{(1-\alpha/2)}$ is the $1-\alpha/2$ quantile of $\hat{\theta}_1^*, \ldots, \hat{\theta}_B^*$

- **Derivation:**

$$\begin{aligned}
&\mathrm{P}(2\hat{\theta} - \hat{\theta}^*_{(1-\alpha/2)} < \theta < 2\hat{\theta} - \hat{\theta}^*_{(\alpha/2)}) \\
=&\mathrm{P}(\hat{\theta}^*_{(\alpha/2)} - \hat{\theta} < \hat{\theta} - \theta < \hat{\theta}^*_{(1-\alpha/2)} - \hat{\theta}) \\
\approx&\mathrm{P}([\hat{\theta} - \theta]_{(\alpha/2)} < \hat{\theta} - \theta < [\hat{\theta} - \theta]_{(1-\alpha/2)}) \\
\approx&1 - \alpha
\end{aligned} \tag{4.6}$$

An explanation of The Basic Bootstrap Interval

- **Definition 4.1:**
  Suppose $\hat{\theta}_1^*, \ldots, \hat{\theta}_B^*$ are Bootstrap statistics, the bias is defined as:

$$\mathrm{Bias}(\hat{\theta}) = \mathbb{E}[\hat{\theta} - \theta] = \mathbb{E}[\hat{\theta}] - \theta \tag{4.7}$$

- **Definition 4.2:**
  Define the Bootstrap estimator of bias:

$$\widehat{\mathrm{Bias}}(\hat{\theta}) = \frac{1}{B}\sum_{j=1}^{B} \hat{\theta}_j^* - \hat{\theta} \tag{4.8}$$

- **Definition 4.3:**
  Define the the estimator with bias correction:

$$\tilde{\theta} = \hat{\theta} - \widehat{\mathrm{Bias}}(\hat{\theta}) = 2\hat{\theta} - \frac{1}{B}\sum_{j=1}^{B} \hat{\theta}_j^* \approx 2\hat{\theta} - \mathbb{E}[\hat{\theta}] \tag{4.9}$$

Thank you!