STATS 406 F15: Lab 08
Bootstrap

1 Estimating MSE by bootstrap

- Example setting:
  - * Suppose we want to estimate a parameter θ of the population distribution $\mathcal{F}$.
  - * We propose an estimator $\hat{\theta}$ based on a collected sample $\mathcal{X} := \{X_1, \ldots, X_n\}$.
  - * We want know how good $\hat{\theta}$ is by evaluating MSE$(\hat{\theta})$.
- If we know $\mathcal{F}$, then in principle we can calculate MSE$(\hat{\theta})$. But unless the distribution is easy to handle(e.g. normal), in general analytical formulation is intractable. Besides in practice we do not know $\mathcal{F}$.
- Now let's take an algorithmatic perspective of view, and ask:
  Suppose we can sample from $\mathcal{F}$, how to compute MSE$(\hat{\theta})$?

  1. Draw many samples, called $\mathcal{X}_1, \ldots, \mathcal{X}_K$, from $\mathcal{F}$.
  2. Compute sample statistic $\hat{\theta}_k$ for each sample $\mathcal{X}_k$.
  3. Evaluate MSE by $\sum_{k=1}^{K}(\hat{\theta}_k - \theta)^2/n$.

- **Bootstrap:** Now we don't know $\mathcal{F}$, but have only one sample $\mathcal{X}$, we

  1. Draw bootstrap samples, called $\mathcal{X}_1^*, \ldots, \mathcal{X}_K^*$, by resampling from $\mathcal{X}$.
  2. Compute bootstrap sample statistic $\hat{\theta}_k^*$ for each bootstrap sample $\mathcal{X}_k^*$.
  3. Evaluate estimated MSE by $\sum_{k=1}^{K}(\hat{\theta}_k^* - \hat{\theta})^2/n$, where $\hat{\theta}$ is the sample statistic of $\mathcal{X}$.

- **Strategy behind the bootstrap algorithm:**
  1. Use $\mathcal{X}$ to mimic $\mathcal{F}$.
  2. Use bootstrap samples $\mathcal{X}_1^*, \ldots, \mathcal{X}_K^*$ to mimic $\mathcal{X}_1, \ldots, \mathcal{X}_K$.
  3. Finally, use bootstrap sample statistic $\hat{\theta}_k^*$ to mimic $\hat{\theta}_k$ and evaluate MSE*.

2 Examples and exercises

2.1 Example 1: estimating MSE by bootstrap

Example:

- Sample $X_1, \dots, X_n$ iid from $N(\mu, \sigma^2)$. Use bootstrap to estimate $\text{MSE}(\bar{X})$.
- Set $n = 25$, $\mu = 1$, $\sigma = 1$. Draw $K = 200$ bootstrap samples $\mathcal{X}_1^*, \dots, \mathcal{X}_K^*$.
- The MSE should be close to the theoretical result: $\text{MSE}(\bar{X}) = \sigma^2/n = 1/20$.
- Also draw the histogram of $\bar{X}_k^*$'s, where $\bar{X}_k^*$ is the sample mean of $\mathcal{X}_k^*$. Check how well it approximates the theoretical distribution of $\bar{X}$, which is $N(\mu, \sigma^2/n)$.

```
# Example: see Lab_8.r
```

2.2 Example 2: bootstrap confidence interval

NOTICE that here we just introduce the simplest bootstrap confidence interval (Efron, 1981), called *bootstrap precentile* method. As reported in literature, its accuracy is not high compared to other methods*.

Example: We revisit the example in lecture notes but this time compute confidence interval.

- Sample $X_1, \dots, X_n$ iid from $\text{Exp}(\lambda)$. Use the *bootstrap percentile* method to provide a bootstrap confidence interval for estimating λ.
- Set $n = 25$, $\lambda = 1$. Draw $K = 200$ bootstrap samples.

```
# Example: see Lab_8.r
```

2.3 Exercise: bias estimation by bootstrap

Exercise:

- Consider a Poisson distribution. Recall that a Poisson distribution with parameter λ has PMF:
$$\mathbb{P}(X = k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!} \quad \text{for } k = 0, 1, 2, \dots$$
- We want to estimate $[\mathbb{P}(X = 0)]^2 = e^{-2\lambda}$ with the estimator $\widehat{p^2} := e^{-2\bar{X}}$. Use bootstrap to estimate $\text{Bias}(\widehat{p^2})$. Compute the true value of $\text{Bias}(\widehat{p^2})$ and compare these two.
- Set $n = 15, 30, 50$, $\lambda = 1$. Draw $K = 200$ bootstrap samples.

```
# Solution: see Lab_8_sol.r (will post later)
```

*The are all unfortunately well-above the level of this course. We will not introduce them. See Section 4.1 of "*The Jackknife and Bootstrap*" by Jun Shao and Dongsheng Tu for more details.

3 More on bootstrap (Optional)

3.1 Failure of bootstrap (Optional)

- **When does bootstrap work?** Bootstrap works in most applications and perhaps all examples you will see in this course.

- (Mamman, 1992) The sufficient and necesary condition for bootstrap to be consistent[†] is that the central limit property holds for $\hat{\theta}$. Formally, that is, there exists sequences of numbers $\{a_n\}$ and $\{s_n\}$, such that $(\hat{\theta} - a_n)/s_n \xrightarrow{d} N(0, 1)$.
  For more technical details, see Section 2.1 of
  `http://www.unc.edu/~saraswat/teaching/econ870/fall11/JH_01.pdf`

- Example: Suppose we bootstrap a sample drawn from $N(\mu, \sigma^2)$ to estimate μ^2 by $\widehat{\mu^2} := \left(\bar{X}\right)^2$. The bootstrap works when $\mu \neq 0$ and fails when $\mu = 0$. Because if $\mu = 0$, by delta method, $\sqrt{n}\bar{X}^2 \xrightarrow{d} 0$, which is degenerate; moreover, $n\bar{X}^2 \xrightarrow{d} \mathcal{X}_1^2$, by CLT and continuous mapping theorem. A Chi-squared distribution cannot be linearly transformed into a normal distribution, so there do not exist $\{a_n\}$ and $\{s_n\}$ that can make a CLT-type convergence to standard normal hold.

- Example: (Continuing the last example) Bootstrap also fails if we estimate $|\mu|$ with $|\bar{X}|$ when $\mu = 0$. See Example 29.7 of
  `http://www.stat.purdue.edu/~dasgupta/bootstrap.pdf`

3.2 Variant: resampling without replacement (Optional)

A natural alternative extension of the ordinary bootstrap is to sample without replacement.

- Suppose we aim at estimating the MSE of or constructing CI for an estimator $\hat{\theta}^{(m)} := T(X_1, \ldots, X_m; \mathcal{F}(\theta))$ for a population parameter θ, where $\hat{\theta}^{(m)}$ means $\hat{\theta}$ with sample size m and $\mathcal{F}(\theta)$ denotes the population distribution with parameter θ.

- We draw a large sample $\mathcal{X} = \{X_1, \ldots, X_n\}$ from $\mathcal{F}(\theta)$, where $n > m$ (sometimes $n \gg m$).

- We draw K bootstrap samples $\mathcal{X}_1, \ldots, \mathcal{X}_K$ by randomly selecting m elements of $\mathcal{X}$ **without replacement**.

- Compute statistics from $\mathcal{X}_k$'s and do whatever next steps as in the ordinary bootstrap.

Advantage: the resampling without replacement (also called "subsampling" for natural reasons) may work under some settings where ordinary bootstrap fails.

[†] A bootstrap is consistent $= EDF\left(\hat{\theta}_k^*|\text{population} = \mathcal{X}\right) \xrightarrow{d} CDF\left(\hat{\theta}_k|\text{population} = \mathcal{F}\right)$

Disadvantage: less accurate than the ordinary bootstrap, when the latter works.

For more details, see Section 2.2 of
`http://www.unc.edu/~saraswat/teaching/econ870/fall11/JH_01.pdf`

3.3 Bootstrap vs. CLT (Optional)

- Sometimes, when both bootstrap and CLT are applicable, people would prefer bootstrap for its higher accuracy. For more details, see Section 29.4 of
  `http://www.stat.purdue.edu/~dasgupta/bootstrap.pdf`

- Example: (bootstrap vs. CLT)

  * In this example, we focuses on the question "how accurate can bootstrap and CLT estimate the lower and upper bounds of CI constructed based on quantiles?" Notice that here we do not care about coverage probabilities.
  * Set the population distribution to be $\mathcal{X}_1^2$. That is, a Chi-squared distribution with 1 degree of freedom. Highly skewed distribution.
  * Set $n = 30$, df $= 1$. Draw $K = 200$ bootstrap samples.
  * In this example, compared to CLT, which always uses symmetric distributions to approximate the population distribution, bootstrap can better account for the skewness in the population distribution.

  ```
  # Example: see Lab_8.r
  ```

```
# Solution: see Lab_8.R (will post later)
```