Seminar 8: Bootstrap and Asymptotic Confidence Intervals

Gaetan Bakalli

November 10^{th}, 2020

Correction of HW7

Common mistakes

- $mean(FireLosses)$ instead of $mean(FireLosses[FireLosses > u])$. First compute the mean on the full dataset while the second compute the mean on the tail. Same apply for var

Introduction to Bootstap

Use of simulation to compute Statistical quantities (i.e. Confidence intervals)

Why using simulation

- Asymptotic might be a poor approximation when the sample size is small.
- Confidence intervals based on asymptotic normality may be too wide or too narrow and exhibit a wrong coverage probability.

Delta Method

$$\text{if } \sqrt{T}(\hat{\theta} - \theta) \sim \mathcal{N}(0, \Sigma), \text{ then } \sqrt{T}\Big(g(\hat{\theta}) - g(\theta)\Big) \sim \mathcal{N}\Big(0, \frac{\partial g}{\partial \theta}\Sigma\frac{\partial g'}{\partial \theta}\Big),$$

where g is a continuous function of the parameter vector θ. In practice, Σ is estimated by the estimated covariance matrix of $\hat{\theta}$ and $\frac{\partial g}{\partial \theta}$ is evaluated at $\hat{\theta}$.

You will also need the following property:

$$\begin{aligned} &\text{if } \varphi \sim \mathcal{N}(\mu, \sigma^2), \text{ then:} \\ &\sqrt{T}(\hat{\mu} - \mu) \sim \mathcal{N}(0, \sigma^2), \quad \sqrt{T}(\hat{\sigma}^2 - \sigma^2) \sim \mathcal{N}(0, \sigma^4), \quad \text{and} \\ &\qquad\qquad\qquad\qquad\qquad\qquad \hat{\mu} \text{ and } \hat{\sigma} \text{ are orthogonal.} \end{aligned}$$

HW8

Breakdown of the exercice 1

1. Use np.random.seed to fix a specific seed (so the results are reproducible) and np.random.normal to simulate data from a normal distribution.
2. Compute the empirical quantity ($\hat{\mu}, \hat{\sigma}$ and $\widehat{SR}$)
3. Delta methods:
   1. $g = \widehat{SR} = \frac{\hat{\mu}}{\sqrt{\hat{\sigma}^2}}$.
   2. $$\Sigma = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^4 \end{pmatrix}.$$
   3. Compute the sandwich formula $\frac{\partial g}{\partial \theta} \Sigma \frac{\partial g'}{\partial \theta}$. where $\hat{\boldsymbol{\theta}} = (\hat{\mu}, \hat{\sigma})$, to get the asymptotic variance.

Bootstrap

1. Start from initial data $Y_1, \ldots, Y_T$ and compute the empirical cdf.
2. Generate ST independent drawings by sampling randomly in the initial data with replacement $Y_1^s, \ldots, Y_T^s$, $s = 1, \ldots, S$ (draw ST realizations of a uniform $[0, 1]$ variable and invert the empirical cdf).
3. For each simulated sample $s = 1, \ldots, S$ of length T, compute the estimate $\hat{\theta}^s$, for example the empirical mean $\hat{\theta}^s = \hat{m}_s = \frac{1}{T}\sum_{t=1}^{T} Y_t^s$.
4. The empirical distribution of the estimates $\hat{\theta}^s$, $s = 1, \ldots, S$, is a good approximation to the true distribution of the estimator in small sample (consistent as $s \to \infty$).

HW8

Breakdown of the exercice 2

1. Use **np.random.randint** to sample $B = 2'000$ vectors of random integer with replacement from 1 to T, where T is the sample size. Those integer will be the indices representing the observation we "select" to compute our Bootstraped estimator.
2. Define a new variable X_{star}, which represent a $T \times B$ matrix of re-sampled observation.
3. Compute the statistics of interests on X_{star} (i.e. $\widehat{SR}$)
4. Construct a 95% confidence interval on $\widehat{SR}$ based on the following formula:

$$\left[2\widehat{SR} - q^{\star}(1-\alpha/2), 2\widehat{SR} - q^{\star}(\alpha/2)\right].$$