

ISyE 6404 CP.1 Part II: PH Regression Bootstrap & Large p Small n Problems

Yuan Gao, Kevin Lee, Akshay Govindaraj,
Yijun (Emma) Wan, Peter Williams, Ruixuan Zhang

Date: 2018-10-30

Contents

Workload Distribution	2
1. Bootstrap Method	3
2. Large-p-Small-n Problems in PH-Regression (30%)	3

Workload Distribution

Below is a description of tasks completed by each team member for this project:

Team Member	Task Description
Yuan Gao	...TBD
Kevin Lee	...TBD
Akshay Govindaraj	...TBD
Yijun (Emma) Wan	...TBD
Peter Williams	...TBD
Ruixuan Zhang	...TBD

1. Bootstrap Method

Use the bootstrap method to construct a 90% pointwise confidence interval (CI) of betacoefficient(s) based on the PH-regression parameter-estimate(s). With the 10000 bootstrap samples you have the entire distribution of beta-estimate(s). Thus, you can use the methods below to construct CIs.

- i) Calculate standard error (denoted as s.e.(beta-est)) of a beta-estimate from the 1000 bootstrapped beta-estimates. Use [beta-est. \pm 1.645*s.e.(beta-est.)] as the largesample CI.
- ii) If the R-program provides 90% confidence interval from either normal or chi-square approximation based on the large-sample theory, it would be worthwhile to compare the results against the ones given above to examine what R-program is doing.
- iii) If the distribution of the bootstrap samples of beta-estimate is not symmetric like normal, use $10000*0.05 = 500$ th lower and upper percentiles as the second largesample CI. This is the so-called Percentile-Bootstrap-CI.
- iv) Apply the Bias Correction (BC) Method to obtain BC-Bootstrap-CI. See Section 15.3 textbook for the details of the BC-method.
- v) A better way to construct the large-sample CI for situation in (iii) is to adjust the percentiles in two tails for having 100 bootstrap samples in total. That is, use a trialand-error to locate these two tail-percentiles (there are better methods to find them) for a non-symmetric distribution. For example, one tail might have 700 bootstrap samples, and the other tail has 300 bootstrap samples. The locations of these two tails form the CI. Students can look into the so-called Highest-Probability-Density Confidence Interval (HPD-CI) for a better version of this idea.

2. Large-p-Small-n Problems in PH-Regression (30%)

- 1) Search the Internet using the key words from the title given in Task #2.
- 2) Locate three publications with a section on real-life data analysis/example.
- 3) Briefly summarize (in an half-page report for one publication)
 - i) what is the goal of the study,
 - ii) key ideas in solving the problem, and
 - iii) what do you learn from their real-life data study

Questions?

Contact: ygao390, kylee20, ywan40, agovindaraj6, pwilliams60, rzhang438 / @gatech.edu