STAT 252: Bootstrap Methods
AY 2022-2023, 2nd Semester
Instructor: Michael Daniel C. Lucagbo, Ph.D.
Email: mclucagbo@up.edu.ph
Consultation hours: Monday to Friday, 2-4pm (please let me know ahead)
Class hours: Friday, 6-9pm
Course prerequisite: Stat 222/232/equiv./COI & Stat 223/233/equiv./COI
Course credit: 3 units

Course description: Empirical distribution functions; resampling and nonparametric statistical inference; optimality of the bootstrap; bootstrap in hypothesis testing; bootstrap in confidence intervals; bootstrap in regression models; bootstrap for dependent data

Course Objective

This course aims to equip students with an extensive arsenal of computational techniques needed to analyze complicated datasets. It gives graduate students a sound understanding of the bootstrap, which may prove useful in theses, dissertations, special problems, and in real-life applications.

What to expect in this course

This course is intended to give you a solid understanding of the bootstrap and the skills to use it in a wide variety of situations where data analysis is needed. Moreover, this course also trains you to implement the bootstrap in R or R Studio. I have included topics which are among the most commonly-used methods both in practical applications and even in graduate students' research. There will be lots of real-life examples sprinkled throughout the course.

The lectures will be done via Zoom. Generally, our lectures will be done synchronously, but there may be some exceptions. Below is the joining information to the classes. The meeting room is open from 6 to 9pm every Friday. Please do NOT share the joining information with other students who are not enlisted in our class. If you have friends who want to sit in on the class, please tell them to ask me first. To join, you may paste this link in your browser:

https://up-edu.zoom.us/j/91338130300
Meeting ID: 913 3813 0300 **Passcode**: statmike

This course will draw upon material from the prerequisite courses. For instance, only a brief review of the linear regression model will be given before I proceed to bootstrap in regression. You are encouraged to do a self-review of basic concepts including the CDF, PDF, the normal distribution, bias and variance, random samples, MLE and least squares estimation, standard error of an estimator, the linear regression model and its estimation.

Course Outline

I. Basics of the bootstrap
- 2.1 Standard errors and confidence intervals
- 2.2 Empirical distribution and plug-in principle
- 2.3 Bootstrap estimate of the standard error
- 2.4 Parametric vs. nonparametric bootstrap
- 2.5 Bootstrap for more complicated data structures

II. Bootstrap in regression models
- 2.1 Review of the linear regression model
- 2.2 Bootstrap estimate of the S.E. of the coefficients
- 2.3 Bootstrapping the residuals
- 2.4 Bootstrapping pairs of data
- 2.5 Bootstrap for heteroskedastic data

III. Bias estimation
- 3.1 Bias of an estimator
- 3.2 Bootstrap estimate of the bias
- 3.3 Improved bootstrap estimate of the bias
- 3.4 Jackknife estimate
- 3.5 Application to bioequivalence

IV. Confidence intervals based on the bootstrap
- 4.1 Student's *t* interval
- 4.2 Bootstrap *t* interval
- 4.3 Bootstrap percentile interval
- 4.4 Coverage probabilities
- 4.5 BC_a and ABC* methods

V. Hypothesis testing with the bootstrap and permutation tests
- 5.1 Permutation tests
- 5.2 Testing using the bootstrap
- 5.3 Relationship between the permutation test and the bootstrap
- 5.4 Applications

VI. Bootstrap for time series
- 6.1 Brief introduction to time series models
- 6.2 Model-based bootstrap
- 6.3 Sieve bootstrap
- 6.4 Subsampling
- 6.5 Block bootstrap

VII. Optimality of the bootstrap

VIII. Other topics*
- 8.1 Double bootstrap
- 8.2 Relationship between the bootstrap and the jackknife
- 8.3 Bootstrap bioequivalence

*If time permits.

Course Requirements
Exams – 50%
Homeworks – 50%

Computing

The bootstrap algorithms will be done in R. You are strongly encouraged to download and install R Studio on your computer.

Homeworks

The homeworks are an important component for this course to train you to use the bootstrap methods in real-life applications. Thus, the homework component of this course is worth 50% of your grade. The homeworks will have a written component and a programming component.

Main References

1. Efron, B. and Tibshirani, R. (1993). **Introduction to the Bootstrap**. New York: Chapman & Hall.
2. McElroy, T. and Politis, D. (2020). **Time Series: A First Course with Bootstrap Starter**. Boca Raton, FL: CRC Press.

Supplementary References

3. Davison, A. C. and Hinkley, D. V. (1997). **Bootstrap methods and their application**. UK: Cambridge University Press.
4. Efron, B. and Hastie, T. (2016). **Computer Age Statistical Inference**. UK: Cambridge University Press.
5. Givens , G.H. and J. A. Hoeting. (2005). **Computational Statistics**. Wiley, New Jersey.
6. Hesterberg, T. (2015). **What Teachers Should Know About the Bootstrap: Resampling in the Undergraduate Statistics Curriculum**. The American Statistician. 69(4): 371-386.
7. Hesterberg, T. (2021). **Bootstrap methods and permutation tests**. Notes for a talk given at the 2021 Conference on Statistical Practice (CSP) sponsored by the American Statistical Association.
8. Rizzo, M.L. (2008). **Statistical computing with R**. Boca Raton, FL: Taylor & Francis.

Note: More references may be added as the semester goes.