Department of Statistics and Data Science
Southern University of Science and Technology

MAT7035: Computational Statistics (计算统计)
(2021 – 2022 Autumn Semester)

Lecturer: Prof. Gary Guoliang TIAN (田国梁) 慧园三栋 Room 520
Tutors: Miss Xiuli ZHAN (詹秀丽) Email: 11930699@mail.sustech.edu.cn

Lecture Hours: Monday (每周 7– 8 节) 16:20 – 18:10
Thursday (only odd week, 单周 1 – 2 节) 08:00 – 09:50
Venue: The First Teaching Building, Room 306 (一教306)
Number of QQ: 438 486 809 (授课课件QQ群号码)

1. Course Objectives (课程目标)

This course aims to provide senior undergraduates at Department of Statistics and Data Science of SUSTech a background in modern computationally-intensive methods in statistics. It emphasizes the role of computation as a fundamental tool of discovery in data analysis, of statistical inference, and for development of statistical theory and methods.

本课程旨在为南科大统计与数据科学系高年级本科生提供常用的现代复杂计算方法。它强调计算作为一个基本工具在数据分析、统计推断、统计理论与方法的发展中的中心地位。

2. Pre-requisites (先修要求)

MA215 Probability (**MA215 概率论**)
MA204 Mathematical Statistics (**MA204 数理统计**)
MA329 Statistical Linear Models (**MA329 统计线性模型**)

3. Course Contents (教学内容)

Chapter 1: Generation of Random Variables

1.1 The inversion method
1.2 The grid method

1.3 The rejection method
1.4 The sampling/importance resampling method
1.5 The stochastic representation (SR) method
1.6 The conditional sampling method

Chapter 2: Optimization

2.1 A review of some standard concepts
2.2 Newton's method and its variants
2.3 The expectation-maximization (EM) algorithm
2.4 The ECM algorithm
2.5 Minorization-maximization (MM) algorithms

Chapter 3: Integration

3.1 Laplace approximations
3.2 Riemannian simulation
3.3 The importance sampling method
3.4 The variance reduction techniques

Chapter 4: Markov Chain Monte Carlo Methods

4.1 Bayes formula and inverse Bayes formulae (IBF)
4.2 The Bayesian methodology
4.3 The data augmentation (DA) algorithm
4.4 The Gibbs sampler
4.5 The exact IBF sampling
4.6 The IBF sampler

Chapter 5: Bootstrap Methods

5.1 Bootstrap confidence intervals
5.2 Hypothesis testing with the bootstrap

Appendix A: Some statistical distributions and stochastic processes

Appendix B: R programming

Appendix C: Introduction of latent variables

4. Learning Objectives and Outcomes (教学目标)

On successful completion of the course, students should be able to:

- understand the importance of the techniques for generating random variables in Monte Carlo integration, Bayesian statistics and bootstrapping methods;
- realize the advantages and disadvantages of the Newton-Raphson algorithm and the Fisher scoring algorithm, and apply them to fit generalized linear models;

- understand the essence and basic principle of the EM-type algorithms and MM-type algorithms, realize their ranges of application, and apply them to solve practical problems;
- apply EM-type algorithms to calculate the posterior mode and apply Markov chain Monte Carlo methods to generate posterior samples;
- apply bootstrap methods to obtain estimated standard errors of estimators and confidence intervals of parameters for both parametric and non-parametric cases.

成功完成该课的教学之后，学生可以达到：

(1) 理解产生随机变量之技术在贝氏统计、蒙特卡罗积分以及自助方法中的重要性；

(2) 了解牛顿-拉夫逊算法和费雪得分算法的优点和缺点, 且能应用这两个算法来拟合广义线性模型；

(3) 理解 EM-型算法和 MM-型算法的本质与基本原理、了解他们的应用范围, 且能应用他们来解决实际问题；

(4) 应用 EM-型算法来计算后验众数, 且能应用 MCMC 方法来产生后验样本;

(5) 对参数和非参数两种情况, 能应用自助方法得到估计量的标准误差之估计值以及参数的置信区间。

5. Teaching Hours (上课安排)

The course consists of 45 lectures (3 lectures per week; 15 weeks x 3 = 45 hours), where 11 lectures will be tutorials.

6. Assessment (成绩评定)

The assessment consists of five assignments (25%), a 2-hour written midterm/class test (25%) and a 2-hour written final examination (50%). Partially or wholly copied assignments will be penalized and/or reported as plagiarism.

7. Assignments (作业布置)

- Assignment 1 (Sep 18 – Oct 14), **4.0** weeks
- Assignment 2 (Oct 18 – Nov 15), **4.0** weeks

- Assignment 3 (Nov15 – Dec 06), **3.0** weeks
- Assignment 4 (Dec 06 – Dec 23), **2.5** weeks
- Assignment 5 (Dec 13 – Dec 31), **3.0** weeks

8. Gary's policy on absence from midterm/class test (期中考试缺考之规定)

If you are or have been unable to attend the mid-term/class test, and if you wish to have a supplementary mid-term/class test, all students should write to the course lecturer giving reasons for your absence, within **7 days** of the absence.

A supplementary midterm/class test is normally granted to those absent from the original test due to illness and with original medical certificate provided. Students absent due to other reasons, in general, are not granted a supplementary midterm/test unless with very special circumstances and with valid documental proofs provided.

Alternatively, students absent from the midterm/class test with a reason might choose the scheme that the weight of the final examination changes from 50% to 75%.

Final scores being 89, 79, 69 and 59 may be upgraded to be 90, 80, 70 and 60 depending on students' special contribution certified by Gary to the teaching/learning of this course.

9. Softwares

Mainly, the statistical package R will be used for demonstration throughout the course. For assignment problems, the package Matlab, C, Fortran, S-plus and SAS are accepted.

10. Textbooks and Main References

In this course, no single textbook can cover all the topics. Relevant references are as follows:

[1] Tan, M., Tian, G.L. and Ng, K.W. (2010). *Bayesian Missing Data Problems: EM, Data Augmentation and Non-iterative Computation*. Chapman & Hall/CRC, Boca Raton.

[2] Givens, G.H. and Hoeting, J.A. (2013). *Computational Statistics* (2nd Ed.) Wiley, New York.

[3] Gentle, J.E. (2009). *Computational Statistics* (Statistics and Computing Series) Springer, New York.

[4] Gentle, J.E. (2004). *Random Number Generation and Monte Carlo Methods* (2nd Ed.) Springer, New York.

[5] Robert, C.P. and Casella, G. (2005). *Monte Carlo Statistical Methods (2nd Ed.).* Springer, New York.

[6] Tanner, M.A. (1996). Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions (3rd Ed.). Springer, New York.

[7] McLachlan, G.J. and Krishnan, T. (1997). *The EM Algorithm and Extensions*. Wiley, New York.

[8] Gilks, W.R., Richardson, S. and Spiegelhalter, D.J. (1996). *Markov Chain Monte Carlo in Practice*. Chapman & Hall, London.

[9] Efron, B. and Tibshirani, R.J. (1993). *An Introduction to the Bootstrap*. Chapman & Hall, London.

[10] Davison, A.C. and Hinkley, D.V. (1997). Bootstrap Methods and Their Application. Cambridge University Press, New York.

[11] Lange, K. (1999). *Numerical Analysis for Statistics.* Springer, New York.

[12] Lange, K. (2004). *Optimization.* Springer, New York.

Tentative Teaching Plan for

MAT 7035: Computational Statistics

Week	Lecture/ Tutorial	Date		Time	Content	Assignment
1	L1	Sep 06	Mon	16:20 – 18:10	§1.1	
	L2				§1.2, §1.3.1	
	L3	Sep 09	Thu	08:00 – 09:50	§1.3.2 – §1.3.3	
	T1				S & E	
2	L4	Sep 13	Mon	16:20 – 18:10	§1.4.1 – §1.4.2	
	L5				§1.5.1 – §1.5.3	
	L6	Sep 18	Sat	16:20 – 18:10	§1.5.4 – §1.5.5	→A1
	T2				S & E	
	Courses in Monday of the third week will be taught in Sep 18 (Saturday) [9 月 18 (周六) 补单周(第三周)星期一的课]					
3	L7	Sep 23	Thu	08:00 – 09:50	§1.6	
	L8				Appendix B.1 – B.3	
4	L9	Sep 27	Mon	16:20 – 18:10	Appendix B.4 – B.6	
	T3				S & E	
NHW	National Holiday Week (01/10/2021 – 07/10/2021) 全国放假					
5	L10	Oct 11	Mon	16:20 – 18:10	§2.1.1 – §2.1.3	
	L11				§2.1.4 – §2.1.7	
	L12	Oct 14	Thu	08:00 – 09:50	§2.2.1 – §2.2.2	A1||4w
	T4				S & E	
6	L13	Oct 18	Mon	16:20 – 18:10	§2.2.3 – §2.2.5	→A2
	L14				§2.3.1	
7	L15	Oct 25	Mon	16:20 – 18:10	§2.3.2 – §2.3.3	
	T5				S & E	
	L16	Oct 28	Thu	08:00 – 09:50	§2.4	
	L17				§2.5.1 – §2.5.3	
8	L18	Nov 01	Mon	16:20 – 18:10	§2.5.4	
	T6				S & E	
9	L19	Nov 08	Mon	16:20 – 18:10	Appendix C.1 – C.2	
	L20				Appendix C.3 – C.4	
	L21	Nov 11	Thu	08:00 – 09:50	§3.1 – §3.2	
	T7				S & E	
10	L22	Nov 15	Mon	16:20 – 18:10	§3.3	A2||4w
	L23				§4.1	→A3
11	L24	Nov 22	Mon	16:20 – 18:10	§4.2.1 – §4.2.2	
	T8				S & E	
	L25	Nov 25	Thu	08:00 – 09:50	§4.2.3	
	L26				§4.3	
12	L27	Nov 29	Mon	16:20 – 18:10	§4.4	
	T9				S & E	

13	**L28**	Dec 06	Mon	16:20 – 18:10	§4.5	**A3||3w →A4**
	L29				§4.6	
	L30	Dec 09	Thu	08:00 – 09:50	**S & E**	
	L31				§5.1.1	
14	**L32**	Dec 13	Mon	16:20 – 18:10	§5.1.2	**→A5**
	T10				**Review**	
15	**L33**	Dec 20	Mon	16:20 – **18:20**	**Midterm/Class Test (2 hours)**	考试地点改为荔园 2 栋 101
	L34					
	L35	Dec 23	Thu	08:00 – 09:50	§5.2.1	**A4||2.5w**
	L36				§5.2.2 – §5.2.3	
16	**T11**	Dec 27	Mon	16:20 – 18:10	**S & E**	
	L37				**Final Review**	
	December 31 (Friday) is the deadline for the last assignment)					**A5||3w**
17	**Exam**	Jan 06	Thu	08:00 – 10:00	2 hours	
				Venue	荔园 1 栋 101	

NOTE:

1. §3.4 will not be taught
2. **S & E = Summary & Exercise**
3. **→A1 = 布置第一次作业**
4. **A1||4w = 交第一次作业 (你有 4 周时间来完成此次作业)**
5. 期中考试和期末考试允许带一张 A4 纸的单面笔记