COURSE SYLLABUS

CSC15007 – Computational Statistics and Applications

1. GENERAL INFORMATION

Course name:	Computational Statistics and Applications
Course name (in Vietnamese):	Thống kê máy tính và ứng dụng
Course ID:	CSC15007
Knowledge block:	General Education Knowledge
Number of credits:	4
Credit hours for theory:	45
Credit hours for practice:	30
Credit hours for self-study:	90
Prerequisite:	Probability and Statistics
Prior-course:	none
Instructors:	

2. COURSE DESCRIPTION

Computational Statistics (or Statistical Computing) is a field that combines Statistics with Computer Science. It provides methods that leverage the computing power of computers for statistical inference and data analysis. This course covers such computationally intensive statistical methods, for example: resampling methods, Monte Carlo methods, and Markov chain Monte Carlo methods. The course also introduces tools to help students practice these methods within the Python computing ecosystem.

3. COURSE GOALS

At the end of the course, students are able to

ID	Description	Program LOs
G1	Be able to work on an individual and team level.	2.2, 2.3.1
G2	Be able to explain terminologies in statistics/data science.	2.4.3, 2.4.5
G3	Be able to explain terminologies in computer science.	1.4, 3.3
G4	Be able define applied mathematical problems.	1.3.6, 1.4
G5	Be able to apply math knowledge to solve problems.	5.1.1, 5.1.3, 5.2.1, 5.2.2, 5.3.1, 6.1.1
G6	Python programming.	1.3.6

4. COURSE OUTCOMES

CO	Description	I/T/U
G1.1	Establish, organize, operate, and manage the team.	I
G1.2	Participate in group discussions.	I, T
G1.3	Writing a technical report.	U
G2.1	Be able to explain terminologies.	I
G2.2	Reading English lectures and textbooks.	I
G3.1	Be able to explain basic concepts.	I
G3.2	Ethics.	I
G3.3	Be able to self study.	I
G4.1	Be able to use the learned models.	I, T
G5.1	Be able to describe the learned mathematical and statistical methods.	I, T
G5.2	Be able to design an algorithm.	I, T, U

5. TEACHING PLAN

ID	Topic	Course outcomes	Teaching/Learning Activities (samples)
Unit 1 (1 week)	Review of Basic Probability and Introduction to Computational Statistics - The birthday problem - Review of basic probability - Probabilistic approximation by simulation - The Monty Hall problem - Review of conditional probability		Lecturing Demonstration Discussion Coding Q&A
Unit 2 (1 week)	Review of Random Variables and Introduction to Computational Statistics - The coupon collector's problem - Review of random variables - Zipf's law and Truyen Kieu - Nguyen Du - Limit theorems - Distribution approximation by simulation		Lecturing Demonstration Discussion Coding Q&A A1#1
Unit 3 (2 weeks)	Random number generation - Introduction - Pseudo random number generators - Discrete distributions - The inverse transform method - Rejection sampling - Transformation of random variables - Special-purpose methods		Lecturing Demonstration Discussion Coding Q&A A1#2
Unit 4 (2 weeks)	Monte Carlo methods - Introduction - Monte Carlo method - Monte Carlo estimate		Lecturing Demonstration Discussion Coding

ID	Topic	Course outcomes	Teaching/Learning Activities (samples)
	- Variance reduction methods - Applications to statistical inference		Q&A A1#3
Unit 5 (3 weeks)	Computational Bayesian Statistics - Bayesian inference - Binomial model - Conjugate priors - Markov chain Monte Carlo - Probabilistic programming - Generalized linear model		Lecturing Demonstration Discussion Coding Q&A A1#4
Unit 6 (2 weeks)	Resampling - Introduction - Permutation resampling - Bootstrapping - Cross-validation		Lecturing Demonstration Discussion Coding Q&A A1#5

Accompanying the theoretical lectures are programming tutorials in Python language. Students do and submit assignments as Jupyter notebook files.

6. ASSESSMENTS

ID	Topic	Description	Course outcomes	Ratio (%)
A1	**Assignments**			**60%**
	5 assignments by topic	A1#1: Review of Probability A1#2: Random number generation A1#3: Monte Carlo methods A1#4: Computational Bayesian		

ID	Topic	Description	Course outcomes	Ratio (%)
		Statistics A1#5: Resampling		
A2	**Exams**			**40%**
	Final exam	Describe the understanding of different topics, analyze & design algorithm to solve problems		

7. RESOURCES

Textbooks

- Jochen Voss. *An Introduction to Statistical Computing - A Simulation-based Approach*. John Wiley & Sons, 2014.

Reference books

- Morris H. DeGroot, Mark J. Schervish. *Probability and Statistics*. Addison-Wesley, 2012.
- H. Pishro-Nik. *"Introduction to probability, statistics, and random processes"*, available at https://www.probabilitycourse.com. Kappa Research LLC, 2014.
- J. S. Dagpunar. *Simulation and Monte Carlo - With applications in finance and MCMC*. John Wiley & Sons, 2007.
- John K. Kruschke. *Doing Bayesian Data Analysis – A Tutorial with R, JAGS, and Stan*. Elsevier, 2015.
- Thomas M. Carsey, Jeffrey J. Harden. *Monte Carlo Simulation and Resampling Methods for Social Science*. SAGE Publications, 2014.

- Dirk P. Kroese, Joshua C.C. Chan. *Statistical Modeling and Computation*. Springer, 2014.

8. GENERAL REGULATIONS & POLICIES

- All students are responsible for reading and following strictly the regulations and policies of the school and university.
- Students who are absent for more than 3 theory sessions are not allowed to take the exams.
- For any kind of cheating and plagiarism, students will be graded 0 for the course. The incident is then submitted to the school and university for further review.
- Students are encouraged to form study groups to discuss on the topics. However, individual work must be done and submitted on your own.