FACULTY OF SCIENCE
SCHOOL OF MATHEMATICS AND STATISTICS

MATH5895

NONPARAMETRIC STATISTICS

Session 1, 2018

MATH5895 – Course Outline

Basic information

Lecturer and Course Authority: Dr Gery Geenens,
Red Centre RC-2053,
email: `ggeenens@unsw.edu.au`,
tel: 02 9385 7032

Credit, Prerequisites, Exclusions:

This course counts for 6 Units of Credit (6UOC).

There are no prerequisites for this course. However, students are assumed to be acquainted with the basic principles of Probability and Statistics theory: random variables and their characteristics, estimators and their properties (bias, variance, consistency, asymptotic distribution), law of large numbers and central limit theorem, maximum likelihood methods. Moreover, they are expected to have basic knowledge of Real Analysis: functions and their properties, limits and series, differentials and integrals, Taylor expansions and function spaces.

Lectures: There will be one three-hour lecture per week (Weeks 1-13):

Tuesday	5:00-8:00	Red Centre RC-1043

Lectures run from the 27th of February (**Week 1**) to the 29th of May (**Week 13**)

No lectures in Week 6

Course details

Aim:

This course aims to introduce the "modern" nonparametric techniques in statistical analysis and the use of these techniques in a variety of disciplines. "Modern" nonparametric statistics essentially refers to smoothing procedures for curve estimation, in contrast with traditional nonparametric methods such as rank-based tests. The main idea of this course is to get the students acquainted with the fundamentals, basic properties and use of the most important recent nonparametric techniques. Another aim is to familiarise students with research questions in this domain.

Relation to other mathematics courses:

This course is an elective course for the Master and/or Honours degree of Statistics program. It gives another insight into statistical theory by addressing the problems from a *nonparametric* point-of-view, that is, by making no or few assumptions on the distribution from which the data were drawn (unlike most of the classical statistical procedures which are essentially *parametric*). As such, it can be thought of as an interesting complement to most of the other courses offered in the program.

Student Learning Outcomes:

Students taking this course will develop an appreciation of nonparametric statistics and will be expected to be able to

- understand why nonparametric techniques are needed and useful;

- explain the fundamental principles of smoothing and nonparametric curve estimation;

- estimate functions of interest (like probability density function, regression function, conditional probability function, variance function, and their derivatives) without making strong parametric assumptions about them;

- test hypotheses about these functions and construct confidence regions;

- use in practice the modern nonparametric techniques to answer concrete questions about real data sets;

- use the R software to generate output in regard to the previous point and for computing intensive methods such as the bootstrap.

Teaching strategies underpinning the course

New ideas and skills are first introduced and demonstrated in lectures, then students develop these skills by applying them to specific assessment tasks. Computing skills will be used to some extent but this is not a course in computing; the computing part is mainly used to illustrate the theory/methodology. The concepts will be illustrated with the R software, and students are strongly encouraged to use this software in their assignments. R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. You can download it from

$$\texttt{http://cran.r-project.org/}$$

Effective learning is best supported by a climate of enquiry, in which students are actively engaged in the learning process. To ensure effective learning, students should attend all classes and have prepared effectively for classes by reading through previous lecture notes. Lectures should be viewed by the student as an opportunity to learn, rather than just copy down lecture notes. The art of logically setting out mathematics is best learned by watching an expert and paying particular attention to detail. This skill is best learned by regularly attending classes.

Finally, effective learning is achieved when students have a genuine interest in the subject and make a serious effort to master the basic material.

Assessment

Assessment in this course will consist of ① **two assignments** (10% each), ② **an oral presentation** during the session (30%) and ③ **a final examination** (50%).

Knowledge and abilities assessed: All assessment tasks will assess the learning outcomes outlined above, specifically, the ability to explain the theory underlying nonparametric techniques, to apply the techniques in analysing real datasets and critically interpreting the results of analyses, and to solve theoretical problems related to modern nonparametric statistics.

Assessment criteria: The main criteria for marking assessment tasks involving explanation of theory and solution of theoretical problems will be clear and logical presentation of correct solutions. In the case of assessment tasks involving the application of techniques to the analysis of real datasets, the main criteria will be selection and justification of appropriate analysis methods; clear, logical, and well-documented computer code; well-organised output giving evidence of successful implementation; correct interpretation of results; and clear, complete, and fully justified conclusions.

① **Assignments**

Rationale: Assignments will give an opportunity for students to try their hand at more difficult problems and also introduce them to aspects of the subject which are not explicitly covered in lectures.

Assignments must be handed in by the due date and time. **Late submission will not be accepted** unless there is documentary evidence of mitigating circumstances.

All work submitted for assessment (other than formal examination scripts) will be returned with comments on the assessment where appropriate.

Assignments must be **your own work**, or severe penalties will be incurred. You should consult the University web page on plagiarism:

Schedule:

Task	Date Avail.	Date Due	Weighting
Assignment 1	Tue 13/3, 8:00PM (Week 3)	Thu 29/3, 5:00PM (Week 5)	10%
Assignment 2	Tue 24/4, 8:00PM (Week 8)	Fri 11/5, 5:00PM (Week 10)	10%

② | Oral presentation |

Rationale: It is required that the student get acquainted with the literature in this area. Each student is asked to give a classroom presentation on a recent research paper in nonparametric estimation.

The students are asked to choose a recent research paper in nonparametric estimation (paper published in an international journal, from a selected list of journals provided by the lecturer). The topic can be suggested by the lecturer and has to be approved by him. Then, they are asked to orally present this topic to the classroom. The oral presentation should be $\pm$ 15 minutes long, plus 5 minutes for a possible discussion.

These presentations will be held during the last weeks of the semester. Further details about the oral presentation will be given in class closer to the time.

Weighting: 30% of the final mark.

③ | Final examination |

Duration: Two hours.

Rationale: The final examination will assess student mastery of the material covered in the lectures.

Weighting: 50% of the final mark.

Further details about the final examination will be given in class closer to the time.

Resources and support

Slides

A copy of the slides shown during the lectures are available from Moodle.

Textbooks

There is no set textbook for this course, which will be based on lectures. However, some chapters in the books listed below are related to the course and will be rec-

ommended as an additional reading.

Tsybakov, A.B. (2009), *Introduction to Nonparametric Estimation*, Springer.

Fan, J. and Gijbels, I. (1996), *Local Polynomial Modelling and Its Applications*, Chapman and Hall.

Härdle, W., Müller, M., Sperlich, S. and Werwatz, A. (2004), *Nonparametric and Semiparametric Models - An Introduction*, Springer-Verlag.

Takezawa, K. (2006), *Introduction to Nonparametric Regression*, Wiley.

Wand, M.P. and Jones, M.C. (1995), *Kernel smoothing*, Chapman and Hall.

Wasserman, L. (2006), *All of Nonparametric Statistics*, Springer.

Computer laboratories

Computer laboratories (RC-M020 and RC-G012) are open 9am-5pm Monday-Friday on teaching days. RC-M020 has extended teaching hours (usually 8.30am-9pm Monday-Friday and 9am-5pm Monday-Friday on non-teaching weeks).

Webpage

The course web site will be made available through the Moodle web portal:

$$\texttt{http://moodle.telt.unsw.edu.au/}$$

Most course materials, including lectures slides and assessment material, will be available there. Students should check regularly for updates and new materials.

Announcements

Announcements may be made in lectures or through the course webpage.

Administrative matters

Students must read and understand the School of Mathematics and Statistics Policies regarding additional assessment policies, attendance requirements, advice concerning special consideration in the event of illness or misadventure, etc., which are available via the School of Mathematics and Statistics Web page at

$$\texttt{http://www.maths.unsw.edu.au/currentstudents/assessment-policies}$$

The set of general policies, procedures and guidelines currently in place relating to UNSW students is to be found at

$$\texttt{https://student.unsw.edu.au/policy}$$

Plagiarism is the presentation of the thoughts or work of another as one's own. Plagiarism is a serious academic misconduct. Students must be aware of UNSW's policies on academic integrity and plagiarism, available at

`https://student.unsw.edu.au/plagiarism`

Students are also reminded that careful time management is an important part of study, and one of the identified causes of plagiarism is poor time management. Students should allow sufficient time for all assessment tasks.

Course schedule

Below is the intended course schedule. However, lectures may fall slightly behind or get slightly ahead of this timetable. Any variation from this will be indicated by the lecturer.

1. Presentation and Introduction to nonparametric methods, **Week 1**

2. Nonparametric estimation of the Cumulative Distribution Function (cdf) and bootstrap, **Weeks 2-3**

3. Nonparametric estimation of the probability density, including histograms, kernel smoothing, other smoothers (orthogonal series, splines, wavelets), derivatives estimation and multivariate problems, **Weeks 3-7**
   Note: there is no class in Week 6.

4. Nonparametric regression, including kernel smoothing, other smoothers (orthogonal series, splines, wavelets), smoothing parameter selection, goodness-of-fit tests and multivariate problems – curse of dimensionality, **Weeks 7-10**

5. Semiparametric regression, including dimension reduction, single-index models, generalised additive models and projection pursuit regression, **Weeks 10-11**

6. Other topics (e.g. nonparametrics in time series context, nonparametrics in survival analysis context, classification problems, nonparametric frontiers estimation, etc.), **Week 11** (if time permits)

7. Students presentations, **Weeks 12 and 13** (depending on the number of students doing the course)