POSSIBLE PROJECT TOPICS

Supervised by Jerome Sheahan, Block C (Arus De Brun), Room 102, ph. 3103; email: jerome.sheahan@nuigalway.ie

Project students are asked to study topics below that interest them and then to arrange to see Jerome

PROJECT 1: HIDDEN MARKOV MODELS

Hidden Markov Models arise in many areas, e.g., term structure of bond credits in finance, speech recognition, bioinformatics, image processing in communications (where the observed values of the underlying stochastic process are corrupted versions of the true image), etc. To initially determine if they'd be interested in this area, students should type – or copy and paste – something like
"hidden Markov models" OR "hidden Markov chains" +"x"
into Google, where "x" is a word or phrase that represents an area of application of interest to them (e.g. "financial markets" or "communication"). Sample sites are
http://www.stat.yale.edu/~jtc5/251/mrf.pdf
http://scholar.lib.vt.edu/theses/available/etd-05102000-13390004/unrestricted/AAKhan.pdf
http://www.ugcs.caltech.edu/~aldana/TransSP.pdf
This project could suit a CS student or other student with good computational skills/background who has an interest in doing computations using, e.g., the EM algorithm (or special cases thereof – e.g. Baum-Welch algorithm), or a Financial Mathematics and Economics student interested in methods of parameter estimation from a theoretical approach or applied to data they would provide.

PROJECT 2. SPATIAL PROCESSES

Spatial statistics is used to study the patterns in, and correlations between, observations in 'space' (e.g. objects observed by satellite images or images observed in brain scans). In dealing with such data, classical statistical techniques need to be modified or replaced to account for non-stationarity, spatial dependence, and the (possibly) arbitrary division of space. In this project, a student interested in statistics will find a real data set (e.g. population density and employment rates, at various geographical locations in Ireland) and analyse it with R software's spatial statistics module, or a dedicated software package. The student will learn about, and interpret a number of statistics relating to the following four categories: distance analysis (e.g. nearest neighbour statistics), spatial distribution (e.g. spatial autocorrelation indices, standard deviational ellipse), cluster identification (e.g. nearest neighbour hierarchical statistics), and kernel density interpolation (both single- and dual-variable). Note that this last technique can be used to generalise to the entire area of interest: the single-variable routine would be used to estimate the distribution of, for example, weekly amount of alcohol consumed in Ireland, while the dual-variable procedure could be used to assess the risk of heavy drinking among various socio-economic groups.
Sample web reference: http://www.esri.com/library/whitepapers/pdfs/intro-modeling.pdf

PROJECT 3: STOCHASTIC PROCESSES/TIME SERIES

This project would interest Financial Mathematics and Economics or Communication Systems Theory students who have an interest in stochastic process methods applied to the special case of time series. The project could proceed on theoretical lines (e.g. robust unit root tests, or change in structure in financial series); or it could be used by a student to exhibit an understanding of regression, ARMA and other procedures as applied to data they would provide (e.g. in modelling returns on an investment, price of some commodity etc.), or through simulations they would perform. One desirable feature would be to incorporate ARMA with ARCH/GARCH modelling. To ascertain their level of interest in this vast area, students could first type appropriate phrases into an internet search engine (e.g. copy and paste
+"time series" +"x" +pdf –course
where "x" could be "communications systems" or "economic applications" or "financial applications" or etc.; the pdf is to force files of that format because they tend to be more formal or research-oriented, and the minus sign is to avoid sites that advertises courses in the area!) Sample web sites are
http://wwwhome.math.utwente.nl/~boucherierj/twi4412/stw2000.PDF
http://cowles.econ.yale.edu/P/cp/p09a/p0925.pdf
http://www.econ.uiuc.edu/~econ472/ARCH.pdf

continued →

PROJECT 4. FORENSIC STATISTICS

Forensics is a challenging area for Science, Law, Medicine and other fields. Mathematics and Statistics play an increasingly important role, and recent court rulings suggest that without statistics, scientific validity of trace evidence is limited. In this project, a small amount of time will be spent becoming familiar with a few areas where forensic science is used (e.g. toxicology, explosives, arson). The project will primarily concentrate on learning a particular area of statistics and its associated techniques. For example, the project could concentrate on pattern recognition (to identify differences in chemicals, so that evidence found at a crime scene can be classified). In this case, some techniques that would be learned could include statistical methods for neural networks, tree-structured classifiers, unsupervised methods, and non-parametric procedures. The student would implement some of these techniques on a real data set.

Sample web site: http://www.math.ku.dk/~richard/download/courses/binf_2007/notes.pdf

PROJECT 5: PARADOXES IN PROBABILITY AND IN STATISTICS

Various paradoxes abound in both probability and statistics. Examples include the envelope problem, Bertrand's paradox, Bernstein's Paradox, Lewis Carroll's problem, the Monty Hall Problem, and Simpson's paradox [which points out lack of transitivity when, for example, statistics are pooled in a meta-analysis]. Many questions have different answers, depending on the model and assumptions adopted, but some solutions are incorrect due to fallacious probabilistic reasoning. In this project, a student interested in probability or statistics would exhibit an understanding of the scientific modelling process in obtaining his/her preferential 'resolution' to a selected few paradoxes among a number of 'acceptable' solutions that will be provided. In addition, limited simulations based on a chosen model may be performed to support the student's answers, but these will not involve programming. The bulk of the project will explore deeply one of the problems from a number of approaches. Concepts and approaches would likely include conditional expectation, likelihood functions, decision theory, and Bayesian modelling.

Sample web site: http://www.quantdec.com/envstats/homework/class_03/paradox.htm

PROJECT 6: TOPICS IN ASSESSMENT OF STUDENTS

This project involves a study of one or more techniques in the assessment of students, and in various problems that arise. One such topic is a treatise on Composite Scoring (see e.g. http://www.education.umd.edu/EDMS/MARCES/mdarch/pdf/m030833.pdf).

Other topics involve, for example, detection of cheating on multiple choice and other tests (e.g. www.natd.org/Cizek%20Symposium%20Paper.PDF)

PROJECT 7: THE KELLY CRITERION

The Kelly formula is designed to concerned with investment and betting strategies that have some desirable optimality properties. An interested student would provide a mature mathematical (mainly) description of the Kelly Criterion, apply the procedure in some area, discuss modifications, and compare with other criteria such as utility theory.

Sample references: **http://en.wikipedia.org/wiki/Kelly_criterion**
http://www.blackjacktactics.com/blackjack/articles/kelly-criterion/

PROJECT 8: Comparing athletes over time

The comparison of athletes born many years apart is a subject of interest to many. Most people have different opinions on "who is the best?" in a particular sport. Many argue that it is impossible to say that e.g. the current best Irish footballer is better or worse than the best footballer in 1960, or that Garry Kasparov is a better chess player than Robert Fischer, since they never competed against each other. Factors like the changes in the nature of the athletes and improvements in technology/equipment etc complicate the comparison. A student interested in comparing athletes from different eras would read at least the paper "Bridging Different Eras in Sports", by Berry SM, Reese CS & Larkey PD [see e.g. http://www.jstor.org/pss/2669973] and be prepared to implement some variants of their model on data the student would provide in a sport [e.g. hurling] or game [e.g. chess] of interest to him/her.

PROJECT 9:

Any problem thought up by the student that I adjudicate to be of sufficient substance in terms of depth and/or learning experience (a large variety of applied probability and statistics articles are available for perusal in my office); or any mathematics problem of mutual interest (especially involving combinatorics). Some of the many project topics in the site http://www.maths.lth.se/matstat/education/exjobb/master.pdf should interest many students.
