INTERNATIONAL CENTRE OF EXCELLENCE FOR EDUCATION IN MATHEMATICS

Experimental Design for Logistic Regresssion

Jennifer Gordon, School of Mathematics and Applied Statistics, University of Wollongong

Consider a set of Bernoulli trials which have a probability of success π that is related to a single predictor variable x. Let there be n_1 and n_2 Bernoulli trials respectively at two values of x, say x_1 and x_2 $(x_1 < x_2)$. These values x_1 and x_2 are called the support points of the experiment. At x_1 and x_2, we will denote the probabilities of success by π_1 and π_2 respectively and the number of successes observed at each support points by Y_1 and Y_2 respectively.

Data collected from an experimental situation such as the one described above does not meet the assumptions of standard normal-based analyses, such as regression and Analysis of Variance (ANOVA). When this occurs, one must instead turn to the methodology of Generalized Linear Models (GLMs). Designing good experiments for any statistical analysis is of great importance but has yet to be thoroughly explored when GLMs are the analysis methods being used. This is due to a major complication that arises when attempting to find optimal experimental designs when using GLMs– the performance of the experiment depends on the parameters that the experiment is designed to estimate.

For my summer scholarship research project, under the supervision of Assoc. Prof. Ken Russell, I investigated what conditions are required to optimise the experimental design when the GLM is the simple first-order logistic model, for the case described above where there are two parameters and two support points.

For the first part of my research, I needed to complete a significant amount of reading to familiarise myself with GLMs. I next found expressions for the Maximum Likelihood (ML) estimators and Maximum Penalised Likelihood (MPL) estimators. After some initial explorations into the behaviour of these two sets of estimators, using the mathematical computer package MAPLE, I found the better set to be the MPL estimators. In changing different aspects of the experiment, I attempted to see what conditions led to the best predictions of π using the MPL estimators for any value of x one might choose.

Finally, I composed a miniature thesis, which encompasses all my findings, conclusions and MAPLE programs. I also prepared a presentation of my research project for the Big Day In (BDI) held by the CSIRO in Lindfield.

I found the ICE-EM/AMSI summer scholarship experience to be a very beneficial and worthwhile one. I'm very grateful for this opportunity I was given to extend my research skills and to also sink my teeth into an area of statistics I might never have had the chance to explore but discovered I enjoyed. I found the BDI to be a fun and eye-opening event as well. It was great to meet people, fellow uni students, statisticians, scientists and mathematicians, similar to myself and I found the research topics and presentations of all the other students to be truly amazing.

Jenn received an ICE-EM Vacation Scholarship in December 2005.
See www.ice-em.org.au/students.html#scholarships05