



Statistics and Actuarial Science Seminar

Title: Modern statistical methods for non-Mendelian disease prediction and Internet data

Speaker: Donal McMahon (Google, Dublin)

Date: Thu 31st March 2011 at 3:00PM

Location: Statistics Seminar Room- L550 Library building

Abstract: This talk shall comprise two distinct sections. Firstly, I will discuss pre-implantation disease prediction for parents contemplating IVF (in vitro fertilisation) treatment, and the models we developed to combine multiple data sources. The second part of this talk will involve a more general discussion of the role of statistics at Google. Here, I will give an overview of various applications; including model fitting for extremely large datasets (hundreds of billions of rows), an overview of the Google experimental framework, predicting economic signals early using search data and the role of statistics in developing self-driving cars.

In vitro fertilisation has become increasingly popular in the past decade. It now accounts for over 1

Between the dawn of civilisation and 2003, the human race only created five exabytes of data; now we're doing that every two days. The role of statistics changes in this "data obese" world, and methods need to be developed to both filter and understand all of this information. In this talk I will speak of some of the problems we encounter

at Google, and the initial attempts we have made at providing adequate solutions. We aim to make data-driven decisions in how we configure the basic search algorithm, in how we choose the ads to show and in all other projects. Therefore statisticians fulfill an pivotal role in many of these endeavours, and I will outline some of the current statistical research.

<mailto:nial.friel@ucd.ie>