



**UCD School of
Mathematics and Statistics**

University College Dublin
Belfield, Dublin 4, Ireland

Tel +353 1 716 2580
Fax +353 1 716 1196

**Scoil na
Matamaitice agus na Staitisticí UCD**

An Coláiste Ollscoile, Baile Átha Cliath
Belfield, Baile Átha Cliath 4, Éire

Email seminars@maths.ucd.ie
Web maths.ucd.ie/seminars

Working Group on Statistical Learning Seminar

Marie Galligan

will speak on

Application of Compositional Models for Glycan HPLC Data

Fri 11th March 2011 at 1:00PM

Location: Statistics Seminar Room- L550 Library building

Compositional data consist of positive real variables which sum to a constant. The sample space for a p dimensional vector of such data is a $p-1$ simplex. This constraint requires careful consideration in the statistical analysis of compositional data. We are currently modelling the data in a simplex sample space, using a generalization of the Dirichlet distribution, known as the Nested Dirichlet distribution. However, finding the model structure from this data proves to be computationally intense, which has encouraged us to seek a method for reducing the dimension of this data before model fitting. As we wish differentiate between disease groups in our data, this approach could also be useful for directly identifying disease biomarkers. We are using a simple variable selection algorithm, similar to that used in Murphy et al. (2010). We are specifying the beta distribution as the marginal density for each compositional variable. We base our decision to add or remove a variable to/from the model by looking at the difference in BICs for that variable when modelled with and without group information.

This talk is part of the **Working Group on Statistical Learning** series. For more, see <https://maths.ucd.ie/seminars>