

*El Departamento de Estadística del ITAM*

anuncia la siguiente sesión de

***EL SEMINARIO ALEATORIO***

que con el título

***Sesión 1***

A new statistical method for haplotype reconstruction

*impartirá*

**Matthew Stephens**

**University of Washington, Seattle**

**(Joint work with Nicholas Smith and Peter Donnelly, University of Oxford)**

**RESUMEN**

Haplotype information is an essential ingredient in most analyses of fine-scale molecular genetics data, but routine genotyping methods typically do not provide this haplotype information. We describe the problem and present a new statistical method for inferring haplotypes. The method compares favourably with competing methods, with error rates reduced by up to 50%.

***Fecha: Martes 5 de junio***

***Hora: 12:30***

***Salón: 201***

***Sesión 2***

Inferring population structure from genetic data via Bayesian model-based clustering

*impartirá*

**Matthew Stephens**

**University of Washington, Seattle**

**(Joint work with Jonathan K Pritchard and Peter Donnelly, University of Oxford)**

**RESUMEN**

One method of searching for disease genes from genetic data is to run a "Case-control" study, which involves looking for different patterns of DNA variation amongst individuals with the disease ("cases"), and individuals without the disease ("controls"). The underlying idea is that such differences will typically only occur in regions of DNA which are connected with the disease. However, population "structure" (or "stratification") can lead to such differences occurring in many different regions, making the problem of identifying genes associated with the disease very much more difficult. This problem motivated us to look at methods of inferring something about the underlying population structure from genetic data. We assume a model in which the population actually consists of  $K$  "subpopulations" (where  $K$  may be unknown). We use MCMC methods to cluster together individuals, who are assigned (probabilistically) to subpopulations, or jointly to two or more populations if they appear to have inherited genes from more than one of the subpopulations. We describe some of the inferential problems, particularly the problem of estimating the number of subpopulations  $K$ , and illustrate the performance of the method on some simulated data, and real data where the "answers" appear to be known.

***Fecha: Miércoles 6 de junio***

***Hora: 12:30***

***Salón: 201***

*El Seminario Aleatorio está destinado tanto a profesores como a estudiantes, por lo que el Departamento de Estadística agradece a los profesores que colaboren invitando a sus alumnos a estas sesiones.*