



## Working Group on Statistical Learning Seminar

**Title:** The effect on inferences of population size of the sampling scheme for intraspecific DNA sequences

**Speaker:** Suzy Whoriskey (University College Dublin)

**Date:** Mon 24th February 2020 at 12:00PM

**Location:** Seminar Room SCN 1.25

**Abstract:** Variation in samples of DNA sequences from within one species can be extremely informative about the demographic processes that have affected that species, revealing signals of migration patterns and population size changes in the past. The demographic models that are fitted to the data might vary, as might the way the data are used, but one almost ubiquitous assumption is that the samples sequenced in the study are randomly chosen. Yet this is rarely plausible either because random sampling is practically impossible to perform or indeed because the samples for analysis are very consciously selected in some non-random way.

This talk will cover work that explored the robustness of a particular flexible class of models used for inference of variable population size, the so-called skyline plot methods, to non-randomness of sampling by taking a simulation approach. The particular sampling scheme investigated takes sequences belonging to one subtree (or haplogroup) of the genealogy of a non-recombining locus. Pitfalls of analyses ignoring the sampling scheme are reported and a recommendation for the interpretation of such analyses is made. The recommendation suggested is a correction to population sizes that have been inferred from the non-random sample that allows the

researcher to acknowledge violations of assumptions caused by the non-random sample and to obtain a more convincing estimate of the population size in applications to mitochondrial DNA.