Seminar/Talk

Designing Bayesian learning models for large genomic datasets

Matt Robinson

University of Lausanne

Host: Nick Barton

Genome-wide association studies (GWAS) have detected thousands of genomic regions associated with common complex diseases and quantitative traits, but they rely on single-marker regression approaches, which have poor estimation and prediction properties. Here, we develop a Bayesian penalised regression model that estimates genetic effects jointly from a mixture of distributions, allowing for related individuals and accounting for marker LD and population structure. We first apply this approach to 456,426 individuals from the UK Biobank dataset finding evidence for thousands of genomic regions with ?95% posterior probability of contributing ?0.001% of trait variation captured by SNP markers for body mass index (BMI, 7297 250kb genomic regions, or 63% of the genome), cardiovascular disease (CAD, 6235, 54%), type-2 diabetes (T2D, 5781, 50%) and height (HT, 4978, 43%). We then show how this model can be adapted and applied to DNA methylation data to estimate association between blood biomarkers and clinical outcomes, whilst controlling for cell-count confounding. Finally, we discuss how this regression approach can be used to formulate a Bayesian factor analysis, which when applied to genomic data may provide additional insights into population genetic differentiation either across a gradient, or between groups.

Friday, September 21, 2018 10:00am - 11:00am

Meeting room 3rd floor / Central Bldg. (I01.3OG.Meeting Room)

This invitation is valid as a ticket for the ISTA Shuttle from and to Heiligenstadt Station. Please find a schedule of the ISTA Shuttle on our webpage: https://ista.ac.at/en/campus/how-to-get-here/ The ISTA Shuttle bus is marked ISTA Shuttle (#142) and has the Institute Logo printed on the side.

www.ista.ac.at | Institute of Science and Technology Austria | Am Campus 1 | 3400 Klosterneuburg