



DSSC SEMINAR/ MATHEMATICS COLLOQUIUM

You are cordially invited to the DSSC seminar/Mathematics colloquium on
 22 May 2018, 16.00-17.00, Bernoulliborg, room 5161.0253.
 PhD and Master students are welcome!

Bayesian modeling in biological data analysis: applications to recombination and human demography

Speaker: Prof. Dr. Gerton Lunter
 Wellcome Trust Centre for Human Genetics
 University of Oxford



Abstract

Next-generation sequencing (NGS) has transformed the biological sciences, giving access to genomes, but also to myriad phenotypes including gene expression and DNA-protein interactions. The amount of data generated by NGS pose significant computational and modeling challenges. Perhaps surprisingly, Bayesian approaches remain relevant: signals can be subtle despite the volume of data, and modern Deep Learning models have many free parameters, requiring regularization to prevent overfitting.

I will show applications of Bayesian inference in two modeling problems involving genome-wide data sets. In the first we use particle filters and Variational Bayes to study human demography. Results indicate that the out-of-Africa event might be more complicated than thought, and suggest a new explanation for the relatively high ancient diversity in the African population.

In the second application we adapt a Variational Bayes techniques for Artificial Neural Networks to exploit a symmetry in DNA sequence models. Using low-resolution empirical recombination maps, we infer the binding motifs of PRDM9, a key protein regulating recombination, and obtain a predictive model that matches that of direct experimental measurement of PRDM9 activity. I will show first results of applying these techniques to broader phenotypic prediction from DNA.

Biography

Gerton Lunter is Associate Professor in Oxford, and co-founder and Head of Statistical Learning at Genomics plc. After a PhD in Mathematics from Groningen, and a foray into video data analysis at Philips Research, he moved into Bioinformatics, developing statistical models for evolution and population genetics, and later focusing on the analysis of Next-generation sequencing data. His current interest is in combining Bayesian approaches with modern Deep Learning methods to build sequence-to-phenotype models.

