



**The Hong Kong University of Science and Technology**

**Department of Mathematics**

**Seminar on Statistics**

**Robust genetic prediction of complex traits with latent  
Dirichlet process regression models**

**By**

***Prof. Xiang ZHOU***

***University of Michigan***

**Abstract**

There has been a growing interest in using genotype data to perform genetic prediction of complex traits. Accurate genetic prediction can facilitate genomic selection in animal and plant breeding programs, and can aid in the development of personalized medicine in humans. Because most complex traits have a polygenic architecture and are each influenced by many genetic variants with small effects, accurate genetic prediction requires the development of polygenic methods that can model all genetic variants jointly. Many recently developed polygenic methods make parametric modeling assumptions on the effect size distribution and different polygenic methods differ in such effect size assumption. However, depending on how well the effect size distribution assumption matches the unknown truth, existing polygenic methods can perform well for some traits but poorly for others. To enable robust phenotype prediction performance across a range of phenotypes, we develop a novel polygenic model with a flexible assumption on the effect size distribution. We refer to our model as the latentDirichlet Process Regression (DPR). DPR relies on the Dirichlet process to assign a prior on the effect size distribution itself, is non-parametric in nature, and is capable of inferring the effect size distribution from the data at hand. Because of the flexible modeling assumption, DPR is able to adapt to a broad spectrum of genetic architectures and achieves robust predictive performance for a variety of complex traits. We compare the predictive performance of DPR with several commonly used polygenic methods in simulations. We further illustrate the benefits of DPR by applying it to predict gene expressions using cis-SNPs, to conduct PrediXcan based gene set test, to perform genomic selection of four traits in two species, and to predict five complex traits in a human cohort. Our method is implemented in the DPR software, freely available at [www.xzlab.org/software.html](http://www.xzlab.org/software.html).

***Date: Monday, 14 May, 2018***

***Time: 10:00a.m.-11:00a.m.***

***Venue: Room 5508, Academic Building,  
(near Lifts 25&26), HKUST***

***All are welcome!***