

Bayesian Inference for Latent Biologic Structure with Determinantal Point Processes (DPP)

PETER MUELLER¹, Yanxun Xu², Donatello Telesca³

¹*University of Texas at Austin, USA*

²*Johns Hopkins University, USA*

³*UCLA, USA*

Abstract

We discuss the use of the determinantal point process (DPP) as a prior for latent structure in biomedical applications, where inference often centers on the interpretation of latent features as biologically or clinically meaningful structure. Typical examples include mixture models, when the terms of the mixture are meant to represent clinically meaningful subpopulations (of patients, genes, etc.). Another class of examples are feature allocation models. We propose the DPP prior as a repulsive prior on latent mixture components in the first example, and as prior on feature-specific parameters in the second case. We argue that the DPP is in general an attractive prior model for latent structure when biologically relevant interpretation of such structure is desired. We illustrate the advantages of DPP prior in three case studies, including inference in mixture models for magnetic resonance images (MRI) and for protein expression, and a feature allocation model for gene expression using data from The Cancer Genome Atlas. An important part of our argument are efficient and straightforward posterior simulation methods. We implement a variation of reversible jump Markov chain Monte Carlo simulation for inference under the DPP prior, using a density with respect to the unit rate Poisson process.