

Abstract: Dirichlet process mixture models (DPMMs) are a cornerstone for Bayesian non-parametric inference and are predominantly used for unsupervised tasks like clustering. Their non-parametric nature enables one to side-step the problem of knowing how many components are present in the data by assuming infinitely many components. DPMMs can also be utilised as a way to perform flexible density estimation; further by specifying hierarchical priors over all the component parameters one can encode rich structure and capture complex properties that arise in real densities like heavy tails, skewness and multi-modality. In this short talk I want to give a flavour for how probabilistic inference can be performed in such models yielding a distribution over densities rather than a single density estimate. The distribution over densities enables the construction of robust uncertainty intervals which can be used to answer fundamental questions that arise in density estimation tasks. For instance, one could in principle distinguish whether an excess(bump) in the observed data is interesting or not based on whether it repeatedly falls inside the predicted uncertainty intervals implied by the posterior distribution of densities. Such techniques can be used to complement the significance based tests currently popular in high-energy physics.