Dirichlet process mixtures are inconsistent for the number of components in a finite mixture

Jeffrey Miller, Brown University

Dirichlet process mixtures (DPMs) are often applied when the data is assumed to come from a mixture with finitely many components, but one does not know the number of components s. In many such cases, one desires to make inferences about s, and it is common practice to use the posterior distribution on the number of components occurring so far. It turns out that this posterior is not consistent for s. That is, we have proven that given unlimited i.i.d. data from a finite mixture with s_0 components, the posterior probability of s_0 does not converge to 1. Motivated by this finding, we examine an alternative approach to Bayesian nonparametric mixtures, which we refer to as a mixture of finite mixtures (MFM). In addition to being consistent for the number of components, MFMs are very natural and possess many of the attractive features of DPMs, including: efficient approximate inference (with MCMC), consistency for the density (at the optimal rate, under certain conditions), and appealing equivalent formulations ("restaurant process", distribution on partitions, stick-breaking, and random discrete measures). Our findings suggest that MFMs are preferable to DPMs when the data comes from a finite mixture.