Summary of "Hierarchical Dirichlet Processes"

Matt Dickenson `mcd31`
STA571/CS590.01

Teh, Jordan, Beal, and Blei (2006) develop a hierarchical variant of mixture models, in which mixture models belonging to different groups share mixture components. This method has nice properties for modeling data originating from different but related processes. Because of its relation to the Chinese restaurant process, the authors refer to this hierarchical extension as a "Chinese franchise process." Using this metaphor, each restaurant is a group, but the same "dish" can be served at multiple restaurants (and at multiple tables within a restaurant). Whenever a customer sits at a table, they are served the same dish as all other customers at that table. There are a number of sampling schemes that can be used, the efficiency of which depends on the particular application and whether reassignment of customers to a different table is allowed. As an example of a useful shortcut, in computing the posterior only the indices of the customers and tables must be sampled, because the factors ($\theta' s and \psi' s$) can be reconstructed from the global menu (ϕ).

One application of HDP is to topic modeling. Extending LDA using a Dirichlet process prior (rather than the Dirichlet distribution) is tempting, but it presents a problem. Modeling documents with LDA means that each document has its own mixing proportion between topics. Because of this the DP would have to be run for each document, but we would wish to share information across documents. This is motivates the use of HDP–to use a hierarchical strategy to make the results of each document's DP similar. Other nice applications where this model would outperform traditional LDA include modeling multiple corpora of documents, and possibly in biology for subpopulations with a shared evolutionary history.