Uncovering Deep Hierarchies using Tree-Structured Stick Breaking

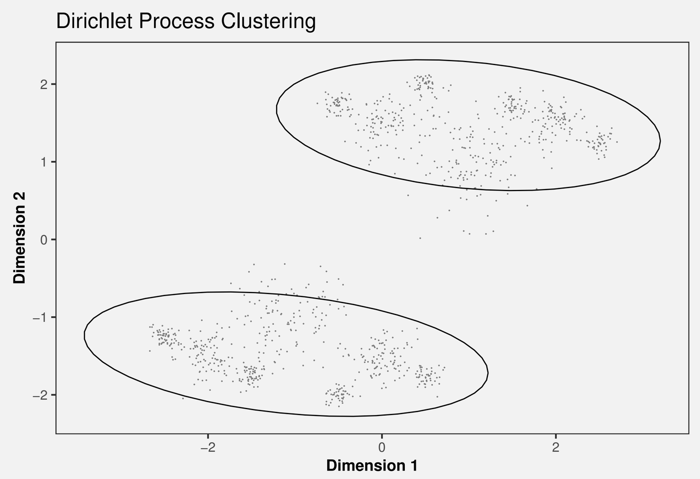
Kevin McAlister - Department of Political Science, University of Michigan

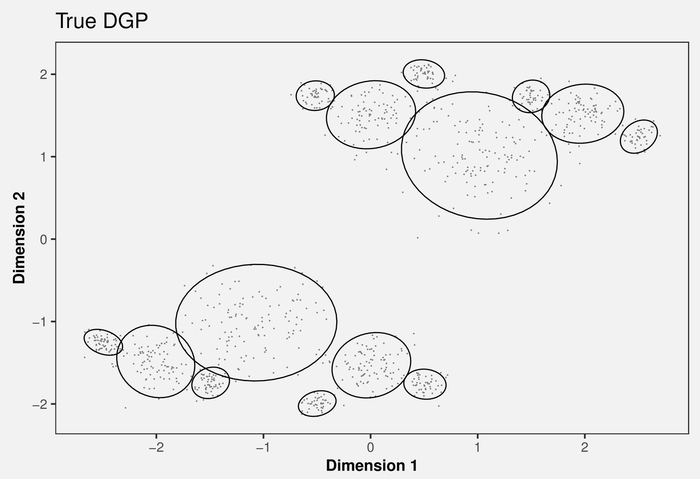
Objectives

Tree-structured priors (Adams et al., 2010) can be used as a method for uncovering hierarchical clusters or as a prior in other methods. I propose an infinite hierarchical factor analysis which allows identification of latent cluster dependencies. These dependence structures answer questions about the appropriateness of bridging and how voting coalitions evolve over time in the U.S. House.

Problem

Unsupervised clustering techniques try to find groups within data that are conditionally exchangeable. A nonparametric approach uses the Dirichlet process.

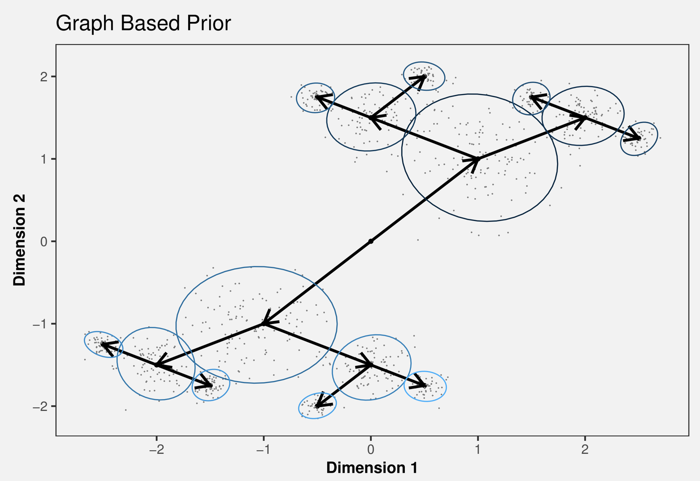


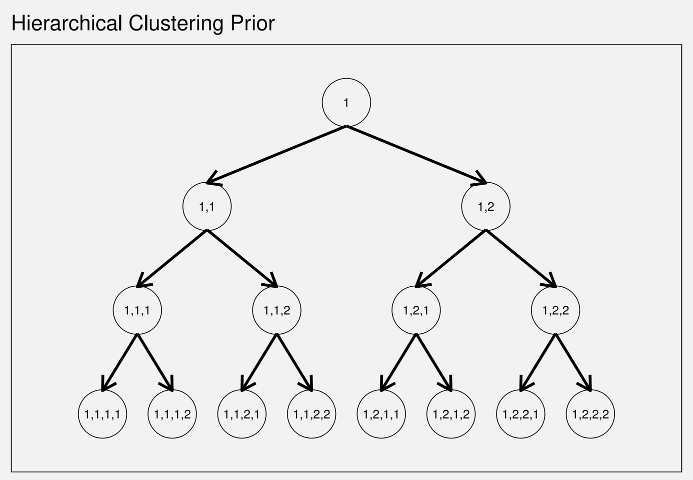


What if the true DGP is hierarchical? Understanding the conditional relationships between clusters increases the statistical power of the model and adds richness to the inference.

A Generative Solution

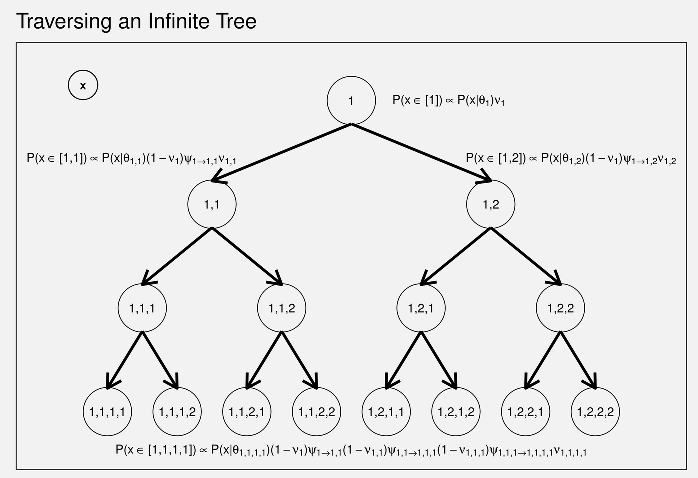
Most hierarchical clustering approaches are not generative (e.g. agglomerative hierarchical clustering). This prohibits usage as a prior in other models.

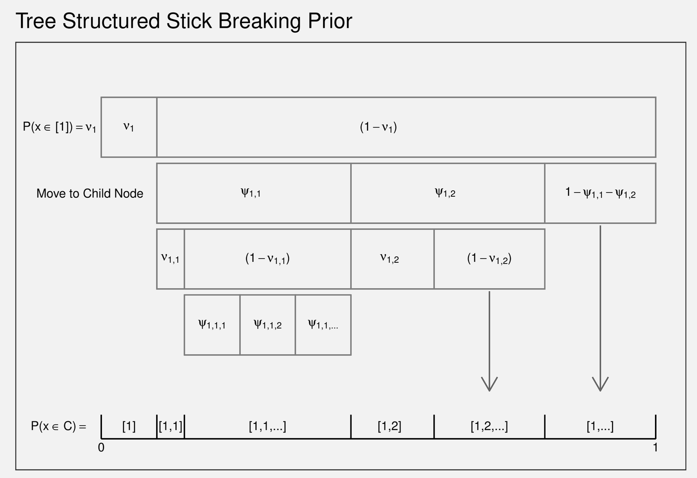




Treat the hierarchy as a graph; links between clusters are found based on a directed graph. This definition allows interpretation as a probability density over clusters. Fits within the Bayesian nonparametric framework.

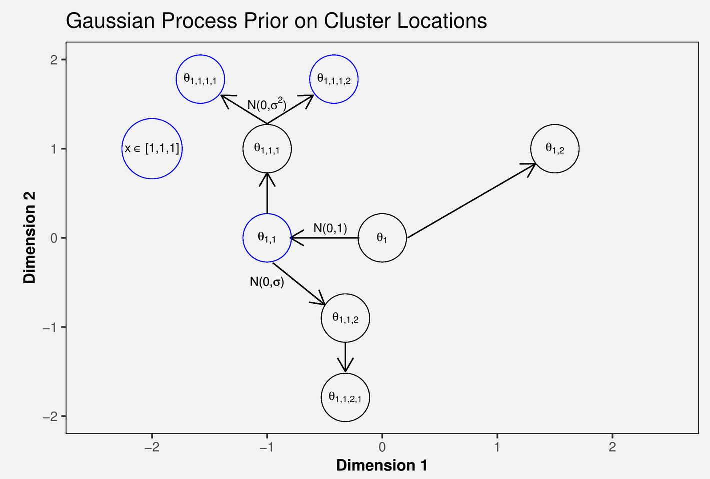
Estimation of Infinite Trees

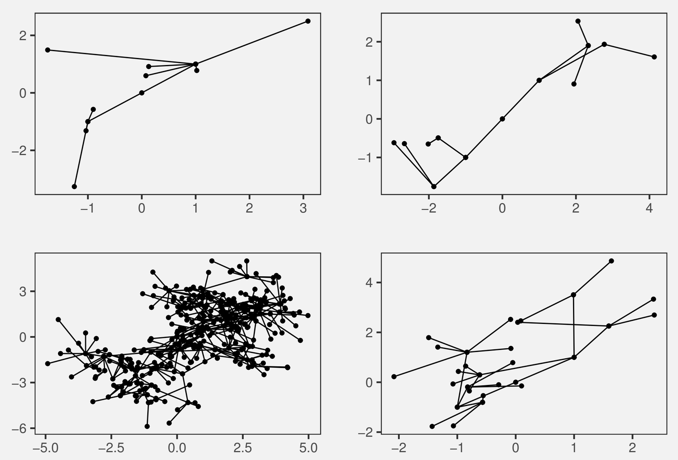




Interweave two stick breaking processes - node stopping and tree path (Adams et al., 2010).

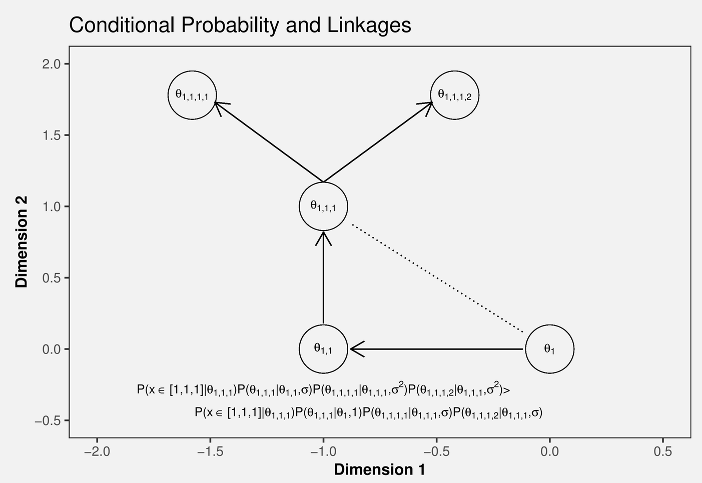
Cluster locations in the data space follow a Gaussian Process.





Cluster Links and Conditional Dependence

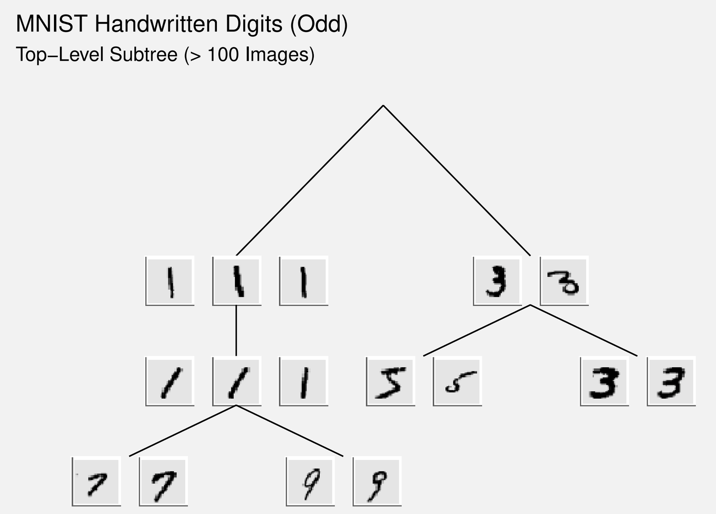
Links between clusters imply conditional dependence - knowing the location of one cluster gives information about the location of the other cluster, like in a directed acyclic graph.



Nonparametric Hierarchical Clustering for Text and Images

Change likelihood function with application - normal for continuous data, Von Mises-Fisher for text, Bernoulli for images. Find deep hierarchies within data sets using TSSBP.

2500 Odd Digits from MNIST Handwritten Numbers data set. Bernoulli likelihood over 256 binary features extracted from each image using a neural net. Tree shown represents path of $\approx$ 80% of data.



Infinite Hierarchical Latent Variable Model

ihFA:

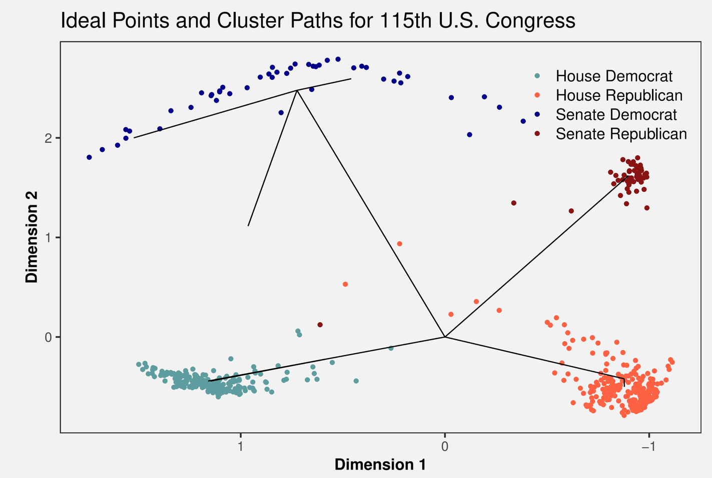
$$y_{i,j} \sim \eta(\lambda_j \omega_i - \alpha_j)$$
$$\Lambda \sim IBP(a, b)$$
$$\Omega \sim TSSBP(\alpha_0, \gamma, \lambda_0)$$

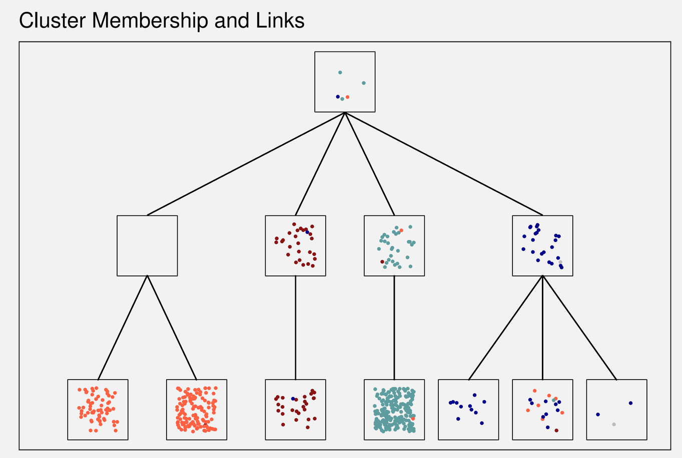
Place Indian Buffet Process prior on loadings. Place TSSBP prior on ideal point errors. Estimates dimensionality of space along with a hierarchical tree over ideal point departures from underlying quadratic loss function.

Bridging Data Sets in Latent Variable Models

Bridging combines two data sets and estimates ideal points in the same latent space. Assumes missingness at random. Should we jointly scale the U.S. House and U.S. Senate?

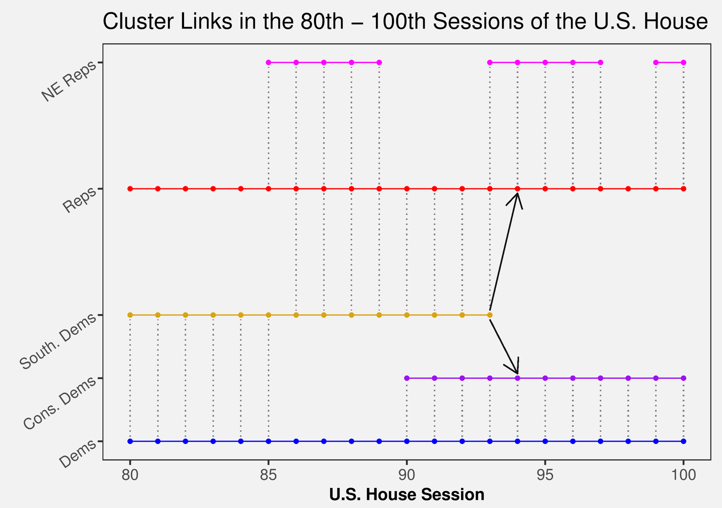
Test: Residuals of data missing at random should be *conditionally exchangeable* given the model parameters. Should link to a common non-root node in the ideal point tree.





Data breaks over party and chamber. IID assumption is violated - without further correction, problems arise similar to omitted variable bias in regression. Bridging across chambers leads to biased ideal points.

Coalitions in the U.S. House



Links show dependencies in voting across groups. Southern Democrats were a prevalent group in the 20^{th} century U.S. House. After 1961, Democrats in name only.

Still voted with Dems on some issues, but more information gained from location of Reps. Eventually split into Reps and Conservative Dem. groups. Rep. revolution essentially removed Conservative Dems. and the south transitioned to Republican representation.