Goal: Use truncated SBP with covariates to simultaneously perform model selection on covariates and number of groups

Model: Let be the species of individual i in location l. We assume that:

where is a latent variable indicating the group membership of individual i in location l. We assume that:

We adopt the following priors:

In relation to , we adopt a probit regression formulation of a truncated stick breaking prior:

To implement this specification, we assume that if , then

2. , and

where

Notice that, if , then . In other words, we only have to represent a maximum of K-1 latent variables for each .

#----------------------------------------

FCD’s

I will sample after integrating out all the corresponding . More specifically,I can use a categorical distribution with probabilities given by:

* For :

Where

* For :

If , then:

If , then:

If , then

Where is the set of all individuals with .

Notice that is a matrix that is formed by copying multiple times for each location.

This implies that

How can we speed some of these calculations up?

Say that we have 2, 2, and 1 individuals in locations 1-3. Therefore:

With this parametrization, it is clear that:

Notice that: . In other words, this is a diagonal matrix where the diagonal elements are given by the number of individuals in each location.

Similarly, it is clear that:

where is a vector that essentially sums all corresponding to individuals in a given location.

This implies that the summary statistics that we need are only:

1. How many individuals assigned to group in location l
2. Sum of for all individuals assigned to group in location l