**For socio-economic information:**

Notice that this is a different model than what we had in the EcoCluster package because of the parameters and because we are combining Bernoulli and multinomial random variables.

For individual i at location l, we assume that:

For variable j associated with individual i at location l, we assume that:

Notice that we can get binary variables if is a vector of length 2.

Finally, we assume the following priors:

To automatically determine the optimal number of groups, we assume that

This implies that

**Alternative models**

What is the benefit of having the parameters ? Why not just do hard-clustering on the individuals and have a single parameter? I believe that having multiple helps to create relatively parsimonious descriptions at each location (i.e., few groups per location) because the prior is strong (i.e., it shows up repeatedly, instead of just once).

**FCD’s**

* For
* For

This implies that:

* For

For an existing group,

For a new group, we have to integrate out the parameters. Based on the derivations below, we end up with the following results for a new group:

For a multinomial random variable, we have that:

Say that . Then: