New LDA formulation

Let be the total number of individuals from species s in location l. Our data are given by:

Our priors are given by:

In this new formulation, we do not condition on (i.e., the number of individuals/words in each location/document). This is beneficial for several reasons:

1. can change substantially, which can potentially provide useful information. The standard LDA model, however, will not detect an effect if remains the same. This is more-or-less the situation with the dataset from the Amazonian experimental fire
2. It is simpler to interpret the meaning of slope coefficients in this formulation than in the multinomial regression formulation. In the latter, the vector indicates how community k changes relative to the reference group.

How do I select the optimal number of groups? Perhaps I could have a strong prior on the intercepts

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

Full conditional distributions:

-

We can propose new vectors using a multinomial distribution:

where

With this proposal distribution, controls how close the proposed values are to the old ones. This is similar to a random walk MH algorithm. Furthermore, this proposal distribution ensures that the natural constrain is imposed.

-

As a result:

Notice that . We can enforce few groups by choosing small and large

- for

Because we will rely on an MH algorithm anyway, I will integrate out the intercept first.

Notice that

Therefore:

Taking logs, this becomes:

- for

This implies that

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

Likelihood calculation