Model formulation

Let be the total number of individuals from species s in location l. We assume that:

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.

The full model is given by:

An alternative but equivalent representation for this model can also be derived. More specifically, note that:

is equivalent to

Furthermore, notice that this multinomial expression is equivalent to:

We also know that

Thus, we can integrate out to obtain:

As a result, we can write:

Where As a result, we end up with the following alternative but equivalent representation of our model:

where . This is a useful representation because we have integrated out all the latent variables.

The posterior distribution is then given by:

where and and