**The Probability Calculations**

The end goal is to calculate for each point in the trellis at each given time t. But since the probability for all these points must add up to 1, we can simply find the joint probability rather than the conditional:. We should also keep in mind that since this is a recursive algorithm, we can assume that we have from the previous step. In the following equations we try to make this term appear in the calculations.

Independence assumption: rt doesn’t depend on the data

iid assumption: xt doesn’t depend on rt-1 and those x’s that aren’t in this run

Have from the previous step

The first one is the prior over rt and the second one is the likelihood model.

iid assumption states that within each run (a segment that is created by two changepoint), the data is iid and therefore the likelihood model only depends on the data from this segment. So instead of using all the datapoint (, we use the datapoints in this segment (). In the paper, is called .

**The Likelihood**

In the twitter example, x is all the tweets in the timestep that are put together in one document and the likelihood model is , where ||D|| is the size of the dictionary (all possible words) and . Note that we are assuming a unigram language model and defining a multinomial distribution for θ. If document x is broken down into wi words, then the unigram, bag-of-words likelihood model will become . However we want to define a distribution on θ as well. Since they have a multinomial distribution, by defining a Dirichlet distribution on θ’s, we can indulge in some mathematical conveniences in the future. The concentration parameters (the parameters of the Dirichlet distribution,) will have a uniform distribution themselves.

*where n(i,D) is the number of times θi appeared in the document(s) D*

D: Dictionary of all the words

Dt: The documents at time t plus all the documents in the current run

So for the first time (t = 0), Dt is only the current document, for t = 1 is the current document plus the previous one, and so on.

Note:

1. For now, D will consist of all the words in the corpus and not an English dictionary. Because many words in the tweets won’t be in a typical dictionary, or they are proper nouns, especially current events words that we are interested in (such as Charlie Hebdo, or Leonard Nimoy).
2. The likelihood model is the same for all the points in a vertical line in the lattice.

**Prior Over rt**

If we assume a constant :