Outline of algorithm for collapsed Gibbs sampling of a hierarchical latent Dirichlet allocation, assuming a nested Chinese restaurant prior:

1. Initialize vector according to the Chinese restaurant/Polya urn/stick breaking process (all refer to the same thing).

The CRP partitions the integers by assigning the first integer with the label ‘1’. Then, the second integer will be given the label ‘1’ with probability = where is the number of integers bearing the label ‘1’ (which is just one, so far) and gamma is a tunable parameter. Then, *n* is just the total number of integers processed thus far. The other possibility is that the next unused label, ‘2’, is assigned with probability . The process is thus repeated over all the integers in our set, with each potentially being assigned to a previously populated label or being assigned to an unused one.

In our case, the labels are the unique identifiers for a path through a tree; suppose we have several nodes which are in varying levels, each with only one parent. Then, the node’s label would identify a path from the root node to the leaf. Each node represents a topic; each path through the tree implicates several topics in a text analysis context.

1. Draw sample of per-word level allocations in each document. Intuitively, this step conditions on a path (which is just an integer value in ***c*** ) and assigns a topic label to each word in the document.
2. Draw sample from **c** conditioned on information from 2 (formula is shown in the CRP paper—it’s essentially too big to write here).
3. Go to step 2 and so on.