drawn, and then the observations are assigned to their corresponding parameter. However, the sequential distances D guarantee that we can draw each word successively. This, in turn, means that we can easily construct a predictive distribution of future words given previous words. (See Section 3 below.)

Mixture modeling The second model we study is akin to the CRP mixture or (equivalently) the DP mixture, but differs in that the mixture component for a data point depends on the mixture component for nearby data. Again, each table is endowed with a draw from a base distribution G_0, but here that draw is a distribution over mixture component parameters. In the document setting, observations are documents (as opposed to individual words), and G_0 is typically a Dirichlet distribution over distributions of words (Teh et al., 2006). The data are drawn as follows:

1. For each document $i \in [1, N]$ draw assignment $c_i \sim$ dist-CRP(α, f, D).
2. For each table, $k \in \{1, \ldots\}$, draw a parameter $\theta_k^* \sim G_0$.
3. For each document $i \in [1, N]$, draw $w_i \sim F(\theta_{z(\mathbf{c})_i})$.

In Section 5, we will study the sequential CRP in this setting, choosing its structure so that contemporaneous documents are more likely to be clustered together. The distances d_{ij} can be the differences between indices in the ordering of the data, or lags between external measurements of distance like date or time. (Spatial distances or distances based on other covariates can be used to define more general mixtures, but we leave these settings for future work.) Again, we have not defined the generative process sequentially but, as long as D respects the assumptions of a sequential CRP, an equivalent sequential model is straightforward to define.

Relationship to dependent Dirichlet processes. More generally, the distance dependent CRP mixture provides an alternative to the dependent Dirichlet process (DDP) mixture as an infinite clustering model that models dependencies between the latent component assignments of the data (MacEachern, 1999). The DDP has been extended to sequential, spatial, and other kinds of dependence (Griffin and Steel, 2006; Duan et al., 2007; Xue et al., 2007). In all these settings, statisticians have appealed to truncations of the stick-breaking representation for approximate posterior inference, citing the dependency between data as precluding the more efficient techniques that integrate out the component parameters and proportions. In contrast, distance dependent CRP mixtures are amenable to Gibbs sampling algorithms that integrate out these variables (see Section 3).

An alternative to the DDP formalism is the Bayesian density regression (BDR) model of Dunson et al. (2007). In BDR, each data point is associated with a random measure and is drawn from a mixture of per-data random measures where the mixture proportions are related to the distance between data points. Unlike the DDP, this model affords a Gibbs sampler where the random measures can be integrated out.

However, it is still different in spirit from the distance dependent CRP. Data are drawn from distributions that are similar to distributions of nearby data, and the particular values of nearby data impose softer constraints than those in the distance dependent CRP. As an extreme case, consider a random partition of the nodes of a network, where distances are defined in terms of the number of hops between nodes. Further, suppose that there are several disconnected components in this network, that is, pairs of nodes that are not reachable from each other. In the DDP model, these