nodes are very likely not to be partitioned in the same group. In the ddCRP model, however, it is impossible for them to be grouped together.

We emphasize that DDP mixtures (and BDR) and distance dependent CRP mixtures are *different* classes of models. DDP mixtures are Bayesian nonparametric models, interpretable as data drawn from a random measure, while the distance dependent CRP mixtures generally are not. DDP mixtures exhibit marginal invariance, while distance dependent CRPs generally do not (see Section 4). In their ability to capture dependence, these two classes of models capture similar assumptions, but the appropriate choice of model depends on the modeling task at hand.

3. Posterior inference and prediction

The central computational problem for distance dependent CRP modeling is posterior inference, determining the conditional distribution of the hidden variables given the observations. This posterior is used for exploratory analysis of the data and how it clusters, and is needed to compute the predictive distribution of a new data point given a set of observations.

Regardless of the likelihood model, the posterior will be intractable to compute because the distance dependent CRP places a prior over a combinatorial number of possible customer configurations. In this section we provide a general strategy for approximating the posterior using Monte Carlo Markov chain (MCMC) sampling. This strategy can be used in either fully-observed or mixture settings, and can be used with arbitrary distance functions. (For example, in Section 5 we illustrate this algorithm with both sequential distance functions and graph-based distance functions and in both fully-observed and mixture settings.)

In MCMC, we aim to construct a Markov chain whose stationary distribution is the posterior of interest. For distance dependent CRP models, the state of the chain is defined by c_i, the customer assignments for each data point. We will also consider $z(\mathbf{c})$, which are the table assignments that follow from the customer assignments (see Figure 1). Let $\eta = \{D, \alpha, f, G_0\}$ denote the set of model hyperparameters. It contains the distances D, the scaling factor α, the decay function f, and the base measure G_0. Let x denote the observations.

In Gibbs sampling, we iteratively draw from the conditional distribution of each latent variable given the other latent variables and observations. (This defines an appropriate Markov chain, see Neal (1993).) In distance dependent CRP models, the Gibbs sampler iteratively draws from

$$p(c_i^{(\text{new})} \mid \mathbf{c}_{-i}, \mathbf{x}, \eta) \propto p(c_i^{(\text{new})} \mid D, \alpha) p(\mathbf{x} \mid z(\mathbf{c}_{-i} \cup c_i^{(\text{new})}), G_0). \tag{3}$$

The first term is the distance dependent CRP prior from Eq. (2).

The second term is the likelihood of the observations under the partition given by $z(\mathbf{c}_{-i} \cup c_i^{(new)})$. This can be thought of as removing the current link from the ith customer and then considering how each alternative new link affects the likelihood of the observations. Before examining this likelihood, we describe how removing and then replacing a customer link affects the underlying partition (i.e., table assignments).

To begin, consider the effect of removing a customer link. What is the difference between the partition $z(\mathbf{c})$ and $z(\mathbf{c}_{-i})$? There are two cases.