constructed through a collection of independent priors on customer assignments to other customers, which then implies a prior on partitions. Unlike the distance dependent CRP, however, the distribution presented in Dahl (2008) requires normalization of these customer assignment probabilities. The model in Dahl (2008) may always be written as a distance dependent CRP, although the normalization requirement prevents the reverse from being true (see Section 2). We note that Dahl (2008) does not present an algorithm for sampling from the posterior, but the Gibbs sampler presented here for the distance dependent CRP can also be employed for posterior inference in that model.

There are a number of Bayesian nonparametric models that allow for dependence between (marginal) partition membership probabilities. These include the dependent Dirichlet process (MacEachern, 1999) and other similar processes (Duan et al., 2007; Griffin and Steel, 2006; Xue et al., 2007). Such models place a prior on collections of sampling distributions drawn from Dirichlet processes, with one sampling distribution drawn per possible value of covariate and sampling distributions from similar covariates more likely to be similar. Marginalizing out the sampling distributions, these models induce a prior on partitions by considering two customers to be clustered together if their sampled values are equal. (Recall, these sampled values are drawn from the sampling distributions corresponding to their respective covariates.) This prior need not be exchangeable if we do not condition on the covariate values.

Distance dependent CRPs represent an alternative strategy for modeling non-exchangeability. The difference hinges on marginal invariance, the property that a missing observation does not affect the joint distribution. In general, dependent DPs exhibit marginal invariance while distance dependent CRPs do not. For the practitioner, this property is a modeling choice, which we discuss in Section 2. Section 4 shows that distance dependent CRPs and dependent DPs represent nearly distinct classes of models, intersecting only in the original DP or CRP.

Still other prior distributions on partitions include those presented in Ahmed and Xing (2008) and Zhu et al. (2005), both of which are special cases of the distance dependent CRP. Rasmussen and Ghahramani (2002) use a gating network similar to the distance dependent CRP to partition datapoints among experts in way that is more likely to assign nearby points to the same cluster. Also included are the product partition models of Hartigan (1990), their recent extension to dependence on covariates (Muller et al., 2008), and the dependent Pitman-Yor process (Sudderth and Jordan, 2008). A review of prior probability distributions on partitions is presented in Mueller and Quintana (2008). The Indian Buffet Process, a Bayesian non-parametric prior on sparse binary matrices, has also been generalized to model non-exchangeable data by Miller et al. (2008). We further discuss these priors in relation to the distance dependent CRP in Section 2.

The rest of this paper is organized as follows. In Section 2 we develop the distance dependent CRP and discuss its properties. We show how the distance dependent CRP may be used to model discrete data, both fully-observed and as part of a mixture model. In Section 3 we show how the customer assignment representation allows for an efficient Gibbs sampling algorithm. In Section 4 we show that distance dependent CRPs and dependent DPs represent distinct classes of models. Finally, in Section 5 we describe an empirical study of three text corpora using the distance dependent CRP. We show that relaxing the assumption of exchangeability with distance dependent CRPs can provide a better fit to sequential data. We also show its alternative formulation of the traditional CRP leads to a faster-mixing Gibbs sampling algorithm than the one based on the original formulation.