## Appendix A. A formal characterization of marginal invariance

In this section, we formally characterize the class of distance dependent CRPs that are marginally invariant. This family is a very small subset of the entire set of distance dependent CRPs, containing only the traditional CRP and variants constructed from independent copies of it. This characterization is used in Section 4 to contrast the distance dependent CRP with random-measure models.

Throughout this section, we assume that the decay function satisfies a relaxed version of the triangle inequality, which uses the notation $\bar{d}_{ij} = \min(d_{ij}, d_{ji})$. We assume: if $\bar{d}_{ij} = 0$ and $\bar{d}_{jk} = 0$ then $\bar{d}_{ik} = 0$; and if $\bar{d}_{ij} < \infty$ and $\bar{d}_{jk} < \infty$ then $\bar{d}_{ik} < \infty$.

### A.1 Sequential Distances

We first consider sequential distances. We begin with the following proposition, which shows that a very restricted class of distance dependent CRPs may also be constructed by collections of independent CRPs.

**Proposition 1** *Fix a set of sequential distances between each of $n$ customers, a real number $a > 0$, and a set $A \in \{\emptyset, \{0\}, \mathbb{R}\}$. Then there is a (non-random) partition $B_1, \ldots, B_K$ of $\{1, \ldots, n\}$ for which two distinct customers $i$ and $j$ are in the same set $B_k$ iff $\bar{d}_{ij} \in A$. For each $k = 1, \ldots, K$, let there be an independent CRP with concentration parameter $\alpha/a$, and let customers within $B_k$ be clustered among themselves according to this CRP.*

*Then, the probability distribution on clusters induced by this construction is identical to the distance dependent CRP with decay function $f(d) = a1[d \in A]$. Furthermore, this probability distribution is marginally invariant.*

**Proof** We begin by constructing a partition $B_1, \ldots, B_K$ with the stated property. Let $J(i) = \min\{j : j = i \text{ or } \bar{d}_{ij} \in A\}$, and let $\mathcal{J} = \{J(i) : i = 1, \ldots, n\}$ be the set of unique values taken by $J$. Each customer $i$ will be placed in the set containing customer $J(i)$. Assign to each value $j \in \mathcal{J}$ a unique integer $k(j)$ between 1 and $|\mathcal{J}|$. For each $j \in \mathcal{J}$, let $B_{k(j)} = \{i : J(i) = j\} = \{i : i = j \text{ or } \bar{d}_{ij} \in A\}$. Each customer $i$ is in exactly one set, $B_{k(J(i))}$, and so $B_1, \ldots, B_{|\mathcal{J}|}$ is a partition of $\{1, \ldots, n\}$.

To show that $i \neq i'$ are both in $B_k$ iff $\bar{d}_{ii'} \in A$, we consider two possibilties. If $A = \emptyset$, then $J(i) = i$ and each $B_k$ contains only a single point. If $A = \{0\}$ or $A = \mathbb{R}$, then it follows from the relaxed triangle inequality assumed at the beginning of Appendix A.

With this partition $B_1, \ldots, B_K$, the probability of linkage under the distance dependent CRP with decay function $f(d) = a1[d \in A]$ may be written

$$p(c_i = j) \propto \begin{cases} \alpha & \text{if } i = j, \\ a & \text{if } j < i \text{ and } j \in B_{k(i)}, \\ 0 & \text{if } j > i \text{ or } j \notin B_{k(i)}. \end{cases}$$

By noting that linkages between customers from different sets $B_k$ occur with probability 0, we see that this is the same probability distribution produced by taking $K$ independent distance dependent