In the mixture model, we compute the marginal probability that the set of observations from each table are drawn independently from the same parameter, which itself is drawn from G_0. Each term is

$$p(\mathbf{x}_{z^k(\mathbf{c})} \mid G_0) = \int \left(\prod_{i \in z^k(\mathbf{c})} p(x_i \mid \theta) \right) p(\theta \mid G_0) d\theta. \tag{7}$$

Because this term marginalizes out the mixture component θ, the result is a collapsed sampler for the mixture model. When G_0 and $p(x \mid \theta)$ form a conjugate pair, the integral is straightforward to compute. In nonconjugate settings, an additional layer of sampling is needed.

Prediction. In prediction, our goal is to compute the conditional probability distribution of a new data point x_{new} given the data set $\mathbf{x}$. This computation relies on the posterior. Recall that D is the set of distances between all the data points. The predictive distribution is

$$p(x_{\text{new}}|\mathbf{x}, D, G_0, \alpha) = \sum_{c_{\text{new}}} p(c_{\text{new}} \mid D, \alpha) \sum_{\mathbf{c}} p(x_{\text{new}}|c_{\text{new}}, \mathbf{c}, \mathbf{x}, G_0) p(\mathbf{c}|\mathbf{x}, D, \alpha, G_0). \tag{8}$$

The outer summation is over the customer assignment of the new data point; its prior probability only depends on the distance matrix D. The inner summation is over the posterior customer assignments of the data set; it determines the probability of the new data point conditioned on the previous data and its partition. In this calculation, the difference between sequential distances and arbitrary distances is important.

Consider sequential distances and suppose that x_{new} is a future data point. In this case, the distribution of the data set customer assignments $\mathbf{c}$ does not depend on the new data point's location in time. The reason is that data points can only connect to data points in the past. Thus, the posterior $p(\mathbf{c} \mid \mathbf{x}, D, \alpha, G_0)$ is unchanged by the addition of the new data, and we can use previously computed Gibbs samples to approximate it.

In other situations—nonsequential distances or sequential distances where the new data occurs somewhere in the middle of the sequence—the discovery of the new data point changes the posterior $p(\mathbf{c} \mid \mathbf{x}, D, \alpha, G_0)$. The reason is that the knowledge of where the new data is relative to the others (i.e., the information in D) changes the prior over customer assignments and thus changes the posterior as well. This new information requires rerunning the Gibbs sampler to account for the new data point. Finally, note that the special case where we know the new data's location in advance (without knowing its value) does not require rerunning the Gibbs sampler.

4. Marginal invariance

In Section 2 we discussed the property of *marginal invariance*, where removing a customer leaves the partition distribution over the remaining customers unchanged. When a model has this property, unobserved data may simply be ignored. We mentioned that the traditional CRP is marginally invariant, while the distance dependent CRP does not necessarily have this property.

In fact, the traditional CRP is the *only* distance dependent CRP that is marginally invariant.[7] The details of this characterization are given in the appendix. This characterization of marginally invariant

7. One can also create a marginally invariant distance dependent CRP by combining several independent copies of the traditional CRP. Details are discussed in the appendix.