Supplement – Nonparametric Clustering with Distance Dependent Hierarchies

June 16, 2014

1 Von Mises-Fisher distributions

If a d-dimensional unit vector $\mathbf{x} \in \mathbb{R}^D$ and $||\mathbf{x}||_2 = 1$ follows a von Mises-Fisher (vMF) distribution then:

$$p(\mathbf{x} \mid \mu, \kappa) = \frac{\kappa^{\frac{d}{2}-1}}{(2\pi)^{\frac{d}{2}} \mathcal{I}_{\frac{d}{2}-1}(\kappa)} e^{\kappa \mu^T \mathbf{x}} \tag{1}$$

with $||\mu||_2 = 1$, $\kappa \geq 0$ and $d \geq 2$. The parameter μ corresponds to the mean direction and κ is a concentration parameter. $\mathcal{I}_r(.)$ is a modified Bessel function of the first kind with degree r. The modified Bessel function grows exponentially fast with κ canceling out the exponential growth of the numerator in equation 1, keeping the density well behaved.

1.1 Likelihood Model - Known κ, Unknown direction μ

$$\mu \sim \text{vMF}(\mu_0, \kappa_0) \tag{2}$$

Observed vectors $\mathbf{x}_i$ are then generated according to:

$$\mathbf{x}_i \mid \mu, \kappa \sim \text{vMF}(\mu, \kappa) \tag{3}$$

1.1.1 Posterior on μ

$$\begin{aligned} & p(\mu \mid \mathbf{x}, \mu_0, \kappa_0, \kappa) \propto p(\mu \mid \mu_0, \kappa_0) p(\mathbf{x} \mid \mu, \kappa) \\ & \propto p(\mu \mid \mu_0, \kappa_0) \prod_{i=1}^{N} p(\mathbf{x}_i \mid \mu, \kappa) \\ & \propto \frac{{\kappa_0}^{\frac{d}{2}-1}}{(2\pi)^{\frac{d}{2}} \mathcal{I}_{\frac{d}{2}-1}(\kappa_0)} e^{\kappa_0 {\mu_0}^T \mu} \prod_{i=1}^{N} \frac{\kappa^{\frac{d}{2}-1}}{(2\pi)^{\frac{d}{2}} \mathcal{I}_{\frac{d}{2}-1}(\kappa)} e^{\kappa \mu^T \mathbf{x}_i} \\ & \propto \frac{\kappa_0^{\frac{d}{2}-1} \kappa^{N(\frac{d}{2}-1)}}{(2\pi)^{\frac{(N+1)d}{2}} \mathcal{I}_{\frac{d}{2}-1}(\kappa_0) (\mathcal{I}_{\frac{d}{2}-1}(\kappa))^N} e^{\kappa_0 \mu_0^T \mu + \kappa \mu^T \sum \mathbf{x}_i} \end{aligned} \tag{4}$$

$$\begin{aligned} & \propto \frac{\kappa_0^{\frac{d}{2}-1} \kappa^{N(\frac{d}{2}-1)}}{(2\pi)^{\frac{(N+1)d}{2}} \mathcal{I}_{\frac{d}{2}-1}(\kappa_0) (\mathcal{I}_{\frac{d}{2}-1}(\kappa))^N} e^{\kappa_0 \mu_0^T \mu + \kappa (\sum \mathbf{x}_i)^T \mu} \\ & \propto \frac{\kappa_0^{\frac{d}{2}-1} \kappa^{N(\frac{d}{2}-1)}}{(2\pi)^{\frac{(N+1)d}{2}} \mathcal{I}_{\frac{d}{2}-1}(\kappa_0) (\mathcal{I}_{\frac{d}{2}-1}(\kappa))^N} e^{(\kappa_0 \mu_0 + \kappa (\sum \mathbf{x}_i))^T \mu} \end{aligned} \tag{5}$$

A valid vMF distribution requires the mean vector to be a unit vector. In general, $(\kappa_0\mu_0 + \kappa \sum \mathbf{x}_i)$ will not be a unit vector and we need to explicitly normalize it.

$$\begin{aligned}
&\propto \frac{\kappa_0^{\frac{d}{2}-1}\kappa^{N(\frac{d}{2}-1)}}{(2\pi)^{\frac{(N+1)d}{2}}\mathcal{I}_{\frac{d}{2}-1}(\kappa_0)(\mathcal{I}_{\frac{d}{2}-1}(\kappa))^N} e^{||(\kappa_0\mu_0+\kappa(\sum \mathbf{x}_i))||_2 \frac{(\kappa_0\mu_0+\kappa(\sum \mathbf{x}_i))^T}{||(\kappa_0\mu_0+\kappa(\sum \mathbf{x}_i))||_2}\mu} \\
&\propto \frac{\kappa_0^{\frac{d}{2}-1}\kappa^{N(\frac{d}{2}-1)}}{(2\pi)^{\frac{(N+1)d}{2}}\mathcal{I}_{\frac{d}{2}-1}(\kappa_0)(\mathcal{I}_{\frac{d}{2}-1}(\kappa))^N} e^{\tilde{\kappa}\tilde{\mu}^T\mu}
\end{aligned} \tag{6}$$

$$\begin{aligned}
&\propto \frac{\kappa_0^{\frac{d}{2}-1}\kappa^{N(\frac{d}{2}-1)}}{(2\pi)^{\frac{(N)d}{2}}\mathcal{I}_{\frac{d}{2}-1}(\kappa_0)(\mathcal{I}_{\frac{d}{2}-1}(\kappa))^N} \frac{\mathcal{I}_{\frac{d}{2}-1}(\tilde{\kappa})}{\tilde{\kappa}^{\frac{d}{2}-1}} \frac{\tilde{\kappa}^{\frac{d}{2}-1}}{(2\pi)^{(\frac{d}{2})}\mathcal{I}_{\frac{d}{2}-1}(\tilde{\kappa})} e^{\tilde{\kappa}\tilde{\mu}^T\mu} \\
&\propto \text{vMF}(\tilde{\mu}, \tilde{\kappa})
\end{aligned} \tag{7}$$

where $\tilde{\kappa} = \Big|\Big|(\kappa_0\mu_0 + \kappa(\sum \mathbf{x}_i))\Big|\Big|_2$ and $\tilde{\mu} = \dfrac{(\kappa_0\mu_0 + \kappa(\sum \mathbf{x}_i))^T}{||(\kappa_0\mu_0 + \kappa(\sum \mathbf{x}_i))||_2}$

1.1.2 Marginal Likelihood

$$p(\mathbf{x} \mid \mu_0, \kappa, \kappa_0) = \int p(\mathbf{x}, \mu \mid \mu_0, \kappa, \kappa_0) d\mu$$

From equation 7 we have:

$$\begin{aligned}
\int p(\mathbf{x}, \mu \mid \mu_0, \kappa, \kappa_0) d\mu &= \int \frac{\kappa_0^{\frac{d}{2}-1}\kappa^{N(\frac{d}{2}-1)}}{(2\pi)^{\frac{(N)d}{2}}\mathcal{I}_{\frac{d}{2}-1}(\kappa_0)(\mathcal{I}_{\frac{d}{2}-1}(\kappa))^N} \frac{\mathcal{I}_{\frac{d}{2}-1}(\tilde{\kappa})}{\tilde{\kappa}^{\frac{d}{2}-1}} \frac{\tilde{\kappa}^{\frac{d}{2}-1}}{(2\pi)^{(\frac{d}{2})}\mathcal{I}_{\frac{d}{2}-1}(\tilde{\kappa})} e^{\tilde{\kappa}\tilde{\mu}^T\mu} d\mu \\
&= \frac{\kappa_0^{\frac{d}{2}-1}\kappa^{N(\frac{d}{2}-1)}}{(2\pi)^{\frac{(N)d}{2}}\mathcal{I}_{\frac{d}{2}-1}(\kappa_0)(\mathcal{I}_{\frac{d}{2}-1}(\kappa))^N} \frac{\mathcal{I}_{\frac{d}{2}-1}(\tilde{\kappa})}{\tilde{\kappa}^{\frac{d}{2}-1}} \int \frac{\tilde{\kappa}^{\frac{d}{2}-1}}{(2\pi)^{(\frac{d}{2})}\mathcal{I}_{\frac{d}{2}-1}(\tilde{\kappa})} e^{\tilde{\kappa}\tilde{\mu}^T\mu} d\mu \\
&= \frac{\kappa_0^{\frac{d}{2}-1}\kappa^{N(\frac{d}{2}-1)}}{(2\pi)^{\frac{(N)d}{2}}\mathcal{I}_{\frac{d}{2}-1}(\kappa_0)(\mathcal{I}_{\frac{d}{2}-1}(\kappa))^N} \frac{\mathcal{I}_{\frac{d}{2}-1}(\tilde{\kappa})}{\tilde{\kappa}^{\frac{d}{2}-1}} \int \text{vMF}(\mu \mid \tilde{\mu}, \tilde{\kappa}) d\mu \\
&= \frac{\kappa_0^{\frac{d}{2}-1}\kappa^{N(\frac{d}{2}-1)}}{(2\pi)^{\frac{(N)d}{2}}\mathcal{I}_{\frac{d}{2}-1}(\kappa_0)(\mathcal{I}_{\frac{d}{2}-1}(\kappa))^N} \frac{\mathcal{I}_{\frac{d}{2}-1}(\tilde{\kappa})}{\tilde{\kappa}^{\frac{d}{2}-1}} \\
&= \frac{1}{(2\pi)^{\frac{Nd}{2}}} \frac{\kappa_0^{\frac{d}{2}-1}\kappa^{N(\frac{d}{2}-1)}}{\tilde{\kappa}^{\frac{d}{2}-1}} \frac{\mathcal{I}_{\frac{d}{2}-1}(\tilde{\kappa})}{\mathcal{I}_{\frac{d}{2}-1}(\kappa_0)(\mathcal{I}_{\frac{d}{2}-1}(\kappa))^N}
\end{aligned}$$

2 Video segmentation hyper-parameters.

Appearance features. The color and texture features have three hyper-paramaters each, κ, μ_0 and κ_0 controlling the within segment concentration of features, the mean direction (μ_0) and the concentration of segment directions $\mu_{z_{ji}}$ around μ_0 respectively. We set μ_0 to the mean direction of the video to be segmented, $\kappa 0$ was set to a small value 10^{-5} encouraging large variance in the segment mean directions. κ controls within segment precision. Low κ values produces segmentations with large regions while high κ values produce smaller segments. We used leave one out cross validation to determine the value of κ. We found $\kappa = 20$ to work well.

Flow features. The two dimensional flow features are modeled using normal inverse Wishart distributions. The inverse Wishart parameters n_0 and expected covariance S_0, were set to 4 and a scaled identity matrix, $s_0\mathbf{I}_{2\times 2}$. Setting n_0 to 4 allows the variance around S_0 to be as high as possible while setting $E[\Sigma_{2\times 2}] = S_0$, for $\Sigma \sim \mathcal{IW}(n_0, S_0)$. The scaling parameter $s0$ was picked to match the variance of the observed flow vectors, and was determined to be 0.25. The mean flow μ_0 was set to the vector $\mathbf{0} \in \mathbb{R}^{2\times 1}$, encoding our intuition that on average we expect super-pixels to be static. Finally, τ_0^{-1} was set to a small value 10^{-10}, to allow high variance around μ_0 thus allowing super-pixels to exhibit large motions.

3 Inference Details

This section adopts the Chinese restaurant terminology and refers to data points as customers, clusters as tables and components as dishes.

Algorithm 1: Iterative sampling of customer and table links.

```
for i ∈ 1...N do
    c*, k* ⟵ CustLinkProposal(i, x, k, c, α, D, α_0, A^0)
    Compute acceptance ratio ρ
    With probability ∝ min(1, ρ), accept c, k ⟵ c*, k*
for t ∈ T(c) do
    k_t ~ p(k_t | k_{-t}, c, x, α_0, A^0(c)) ;                /*Gibbs update k_t*/
```

Algorithm 2: CustLinkProposal

```
input  : i, x, k, c, α, D, α_0, A^0
output: k*, c*, q.(c*, k* | k, c)
K_{t_ij} = {k_s | k_s = t_ij, s ≠ t_ij} ;    /*Set of all table links pointing to t_i except
self links.*/
Set c_i = i ;
if c_i = i causes a split then
    split ⟵ TRUE;                          /*Record the occurrence of a split.*/
    k ⟵ ReassignLinks (K_{t_ij})
    k*_{t_j} ⟵ k_{t_ij} ;          /*A split table retains the current table's link.*/
Propose a new customer link: c_i ~ q(c_i)
if c_i = j* causes a merge then
    K_{t_ij*} = K_{t_i} ∪ K_{t_j*} and Update k to reflect the merge
    if split then                                                  /* Split+Merge */
        k*_{t_ij*} ⟵ k_{t_j*}
        q_sm(c*, k* | c, k, x) = (1/2)^{|K_{t_ij}|} q(c*_i)
    else                                                       /* No split + Merge */
        Move k_{t_i} to the inactive set
        k*_{t_ij*} ⟵ k_{t_j*}
        q_m(c*, k* | c, k, x) = q(c*_i)
else
    if split then                                               /* Split+No Merge */
        Sample k*_{t_i} ~ p(k*_{t_i} | α_0, A^0(c*), x, k_{-t_i}) ;
        q_s(c*, k* | c, k, x) = (1/2)^{|K_{t_ij}|} q(c*_i) p(k*_{t_i} | A^0(c*), x, k_{-t_i})
    else                                                     /* No Split+No Merge */
        /*No change to partition - Do Nothing.                                    */
        q_nc(c*, k* | c, k, x) = q(c*_i)
```

3.1 Acceptance Ratios

Notational details:

1. i is a customer, k_{t_i} denotes the link of the table containing i.
2. If two tables, one with customer i and another with j are merged, the resulting table is denoted t_{ij} and

Algorithm 3: ReassignLinks

input : $\mathcal{K}_{t_{ij}}$
output: $\mathbf{k}$

```
/*Reassign links pointing to a split table.  Links are assigned to one
  of the two split tables.                                            */
```

for $k_s \in \mathcal{K}_{t_i}$ **do**
  $b_s \sim$ Bernoulli(0.5)
  if $b_s = 1$ **then**
    $\mathcal{K}_{t_i} = \mathcal{K}_{t_i}/k_s$
    $\mathcal{K}_{t_j} = \mathcal{K}_{t_j} \cup k_s$
    $k_s = t_j$;

the corresponding table link is $k_{t_{ij}}$

3. $\mathbf{c} = \{c_1, \ldots, c_{i-1}, c_i = j, \ldots, c_N\}$
   $\mathbf{c}^* = \{c_1, \ldots, c_{i-1}, c_i = j^*, \ldots, c_N\}$ and $c_i = j^*$ is denoted as c_i^*

The proposed algorithm changes an existing partition by either

1. Merging existing tables : No new table is created when $c_i = i$ and exiting tables are merged after sampling $c_i = j^*$. The transition probability of this move is:

$$q_m(\mathbf{c}^*, \mathbf{k}^* \mid \mathbf{c}, \mathbf{k}, \mathbf{x}) = q(c_i^*) \tag{8}$$

2. Or splitting an existing table: A new table is created when $c_i = i$ and and no tables are merged after sampling $c_i = j^*$.

$$q_s(\mathbf{c}^*, \mathbf{k}^* \mid \mathbf{c}, \mathbf{k}, \mathbf{x}) = \left(\frac{1}{2}\right)^{|\mathcal{K}_{t_{ij}}|} q(c_i^*) p(k^*_{t_i} \mid A^0(\mathbf{c}^*), \mathbf{x}, \mathbf{k}^*_{-t_i}) \tag{9}$$

3. Or both merging and splitting tables : A new table is created when $c_i = i$ and tables are merged after sampling $c_i = j^*$.

$$q_{sm}(\mathbf{c}^*, \mathbf{k}^* \mid \mathbf{c}, \mathbf{k}, \mathbf{x}) = \left(\frac{1}{2}\right)^{|\mathcal{K}_{t_{ij}}|} q(c_i^*) \tag{10}$$

4. Finally, the move might not change a partition at all: No new table is created when $c_i = i$ and no tables are merged after sampling $c_i = j^*$.

$$q_{nc}(\mathbf{c}^*, \mathbf{k}^* \mid \mathbf{c}, \mathbf{k}, \mathbf{x}) = q(c_i^*) \tag{11}$$

Moves 1 and 4 are reverses of each other, while moves 2 and 3 are their own reverses.
Recall that a proposal is accepted with probability $\propto$ min$(1, \rho)$, where

$$\rho = \frac{p(\mathbf{x}, \mathbf{c}^*, \mathbf{k}^*)}{p(\mathbf{x}, \mathbf{c}, \mathbf{k})} \frac{q_{rev}(\mathbf{c}, \mathbf{k} \mid \mathbf{c}^*, \mathbf{k}^*, \mathbf{x})}{q_{fwd}(\mathbf{c}^*, \mathbf{k}^* \mid \mathbf{c}, \mathbf{k}, \mathbf{x})} \tag{12}$$

3.2 Split move

Let us first consider the split move under the prior proposal. The merge move is the reverse of a split move. Hence we have:

$$\rho_s = \frac{p(\mathbf{x}, \mathbf{c}^*, \mathbf{k}^*)}{p(\mathbf{x}, \mathbf{c}, \mathbf{k})} \frac{q_m(\mathbf{c}, \mathbf{k} \mid \mathbf{c}^*, \mathbf{k}^*, \mathbf{x})}{q_s(\mathbf{c}^*, \mathbf{k}^* \mid \mathbf{c}, \mathbf{k}, \mathbf{x})} \tag{13}$$

Substituting Equations 9 and 8 above we get:

$$\rho_s = \frac{p(\mathbf{x}, \mathbf{c}^*, \mathbf{k}^*)}{p(\mathbf{x}, \mathbf{c}, \mathbf{k})} \frac{p(c_i = j \mid \alpha, D)}{p(c_i = j^* \mid \alpha, D) p(k^*_{t_i} \mid A^0(\mathbf{c}^*), \mathbf{x}, \mathbf{k}^*_{-t_i})(0.5)^{|\mathcal{K}_{t_{ij}}|}} \tag{14}$$

Dropping dependence on α, D, A^0 for notational convenience we have:

$$\rho_s = \frac{1}{(0.5)^{|\mathcal{K}_{t_{ij}}|}} \frac{p(\mathbf{c}^*) p(\mathbf{k}^* \mid \mathbf{c}^*) p(\mathbf{x} \mid \mathbf{k}^*, \mathbf{c}^*)}{p(\mathbf{c}) p(\mathbf{k} \mid \mathbf{c}) p(\mathbf{x} \mid \mathbf{k}, \mathbf{c})} \frac{p(c_i = j)}{p(c_i = j^*) p(k^*_{t_i} \mid \mathbf{x}, \mathbf{k}^*_{-t_i}, \mathbf{c}^*)} \tag{15}$$

The customer links cancel out between the likelihood and hastings ratios:

$$\rho_s = \frac{1}{(0.5)^{|\mathcal{K}_{t_{ij}}|}} \frac{p(\mathbf{k}^* \mid \mathbf{c}^*) p(\mathbf{x} \mid \mathbf{k}^*, \mathbf{c}^*)}{p(\mathbf{k} \mid \mathbf{c}) p(\mathbf{x} \mid \mathbf{k}, \mathbf{c})} \frac{1}{p(k^*_{t_i} \mid \mathbf{x}, \mathbf{k}^*_{-t_i}, \mathbf{c}^*)} \tag{16}$$

$$\rho_s = \frac{1}{(0.5)^{|\mathcal{K}_{t_{ij}}|}} \frac{p(\mathbf{k}^* \mid \mathbf{c}^*) p(\mathbf{x} \mid \mathbf{k}^*, \mathbf{c}^*)}{p(\mathbf{k} \mid \mathbf{c}) p(\mathbf{x} \mid \mathbf{k}, \mathbf{c})} \frac{p(\mathbf{x}, \mathbf{k}^*_{-t_i}, \mathbf{c}^*)}{p(\mathbf{x}, \mathbf{k}^*, \mathbf{c}^*)} \tag{17}$$

$$\rho_s = \frac{1}{(0.5)^{|\mathcal{K}_{t_{ij}}|}} \frac{\cancel{p(\mathbf{k}^* \mid \mathbf{c}^*) p(\mathbf{x} \mid \mathbf{k}^*, \mathbf{c}^*)}}{p(\mathbf{k} \mid \mathbf{c}) p(\mathbf{x} \mid \mathbf{k}, \mathbf{c})} \frac{p(\mathbf{x} \mid \mathbf{k}^*_{-t_i}, \mathbf{c}^*) p(\mathbf{k}^*_{-t_i} \mid \mathbf{c}^*) \cancel{p(\mathbf{c}^*)}}{\cancel{p(\mathbf{x} \mid \mathbf{k}^*, \mathbf{c}^*)} \cancel{p(\mathbf{k}^* \mid \mathbf{c}^*)} \cancel{p(\mathbf{c}^*)}} \tag{18}$$

$$\rho_s = \frac{1}{(0.5)^{|\mathcal{K}_{t_{ij}}|}} \frac{p(\mathbf{k}^*_{-t_i} \mid \mathbf{c}^*)}{p(\mathbf{k} \mid \mathbf{c})} \frac{p(\mathbf{x} \mid \mathbf{k}^*_{-t_i}, \mathbf{c}^*)}{p(\mathbf{x} \mid \mathbf{k}, \mathbf{c})} \tag{19}$$

Note that the number of table links in $\mathbf{k}$ = number of table links in $\mathbf{k}^*_{-t_i}$, and depending on the distance between tables the ratio of table links may be further simplified. For instance, for the intersection over union distance used in the video segmentation algorithm all but the links of the affected tables cancel out with the above ratio simplifying to:

$$\frac{p(\mathbf{k}^*_{-t_i} \mid \mathbf{c}^*)}{p(\mathbf{k} \mid \mathbf{c})} = \frac{\prod_{k_s \in \mathcal{K}_{t_{ij}}} p(k^*_s \mid \mathbf{c}^*)}{\prod_{k_s \in \mathcal{K}_{t_{ij}}} p(k_s \mid \mathbf{c})} \tag{20}$$

Similarly, likelihood terms of dishes not affected by the split cancel out. If the two tables serve different dishes we have

$$\frac{p(\mathbf{x} \mid \mathbf{k}^*_{-t_i}, \mathbf{c}^*)}{p(\mathbf{x} \mid \mathbf{k}, \mathbf{c})} = \frac{p(\mathbf{x}_{\mathbf{z}=z_j} \mid \mathbf{k}^*_{-t_i}, \mathbf{c}^*, \lambda) p(\mathbf{x}_{\mathbf{z}=z_{j^*}} \mid \mathbf{k}^*_{-t_i}, \mathbf{c}^*, \lambda)}{p(\mathbf{x}_{\mathbf{z}=z_i} \mid \mathbf{k}, \mathbf{c}, \lambda)} \tag{21}$$

where $\mathbf{x}_{\mathbf{z}=z_j}$ refers to all customers sharing a dish with j. Note that since in the pre spilt state the i and j are sitting at the same table, $\{\mathbf{x}_{\mathbf{z}=z_j} \mid \mathbf{c}, \mathbf{k}\} = \{\mathbf{x}_{\mathbf{z}=z_i} \mid \mathbf{c}, \mathbf{k}\}$, but $\{\mathbf{x}_{\mathbf{z}=z_j} \mid \mathbf{c}^*, \mathbf{k}^*\}$ may not equal $\{\mathbf{x}_{\mathbf{z}=z_i} \mid \mathbf{c}^*, \mathbf{k}^*\}$ in the new split state. If the two tables do serve the same dish we have:

$$\frac{p(\mathbf{x} \mid \mathbf{k}^*_{-t_i}, \mathbf{c}^*)}{p(\mathbf{x} \mid \mathbf{k}, \mathbf{c})} = \frac{p(\mathbf{x}_{\mathbf{z}=z_i} \mid \mathbf{k}^*_{-t_i}, \mathbf{c}^*, \lambda)}{p(\mathbf{x}_{\mathbf{z}=z_i} \mid \mathbf{k}, \mathbf{c}, \lambda)} \tag{22}$$

Note that due to the missing k_{t_i} link in the numerator the two sets of customers in general will not be the same.

Pseudo Gibbs Proposals. Recall that we sample the customer link from the following proposal distribution:

$$q(c^*_i) \propto p(c^*_i \mid \alpha, D) \Gamma(\mathbf{x}, \mathbf{z}, \lambda), \tag{23}$$

where

$$\Gamma(\mathbf{x}, \mathbf{z}, \lambda) = \begin{cases} \dfrac{p(\mathbf{x}_{\mathbf{z}(\Delta)=m_a} \cup \mathbf{x}_{\mathbf{z}(\Delta)=m_b} \mid \lambda)}{p(\mathbf{x}_{\mathbf{z}(\Delta)=m_a} \mid \lambda) p(\mathbf{x}_{\mathbf{z}(\Delta)=m_b} \mid \lambda)} & \text{if } c^*_i \text{ merges dishes } m_a \text{ and } m_b \\ p(c^*_i \mid \alpha, D) & \text{otherwise,} \end{cases} \tag{24}$$

where $\Delta = \{\mathbf{c}_{-i}, k_{t_i} = t_i, \mathbf{k}_{-t_i}\}$ and $\mathbf{k}_{-t_i}$ represents the set of all table links excluding k_{t_i}.

Split move acceptance ratio for pseudo Gibbs proposals First, observe that when a particular link proposal doesn't cause two dishes to merge (either because it doesn't merge tables or because it merges two tables serving the same dish) the prior proposal and the pseudo Gibbs proposals are identical. This implies that when a split is proposed that splits a table but not a dish the pseudo Gibbs acceptance ratio is equal to the prior proposal acceptance ratio given in Equation (19).

Now let us consider the case when the proposed customer link causes dishes to be split. The reverse move must then cause two distinct dishes (m_a and m_b) to be merged. Thus the reverse transition probability is:

$$q_m(\mathbf{c}, \mathbf{k} \mid \mathbf{c}^*, \mathbf{k}^*, \mathbf{x}) = \frac{1}{\mathcal{C}_i} p(c_i = j \mid \alpha, D) \frac{p(\mathbf{x}_{m_a} \cup \mathbf{x}_{m_b} \mid \lambda)}{p(\mathbf{x}_{m_a} \mid \lambda) p(\mathbf{x}_{m_b} \mid \lambda)}, \tag{25}$$

where $\mathcal{C}_i$ is the appropriate normalization constant for the discrete pseudo Gibbs proposal. Under, the pseudo Gibbs proposal, the probability of a link that doesn't cause a merge of dishes is given by $\frac{1}{\mathcal{C}_i} p(c_i = j^* \mid \alpha, D)$ which is then combined with the probability of sampling a new table link to give the forward transition probability for the split move:

$$q_s(\mathbf{c}^*, \mathbf{k}^* \mid \mathbf{c}, \mathbf{k}, \mathbf{x}) = (0.5)^{|\mathcal{K}_{t_{ij}}|} \frac{1}{\mathcal{C}_i} p(c_i = j^* \mid \alpha, D) p(k^*_{t_i} \mid A^0(\mathbf{c}^*), \mathbf{x}, \mathbf{k}_{-t_i}) \tag{26}$$

Plugging these values in Equation (12) leads to the following ratio:

$$\eta_s = \frac{1}{(0.5)^{|\mathcal{K}_{t_{ij}}|}} \frac{p(\mathbf{k}^*_{-t_i} \mid \mathbf{c}^*)}{p(\mathbf{k} \mid \mathbf{c})} \tag{27}$$

Finally the acceptance ratio for the split move under the pseudo Gibbs proposal is:

$$\rho_s^{pg} = \begin{cases} \rho_s & \text{if the split tables share the same dish} \\ \eta_s & \text{otherwise} \end{cases} \tag{28}$$

where η_s is Merge move ratios are analogously computed.

3.3 Split+Merge moves

These moves allow for customers to shift between tables. The acceptance ratio for the prior proposal works out to

$$\rho_{sm} = \left(\frac{1}{2}\right)^{|\mathcal{K}_{t_{ij^*}}| - |\mathcal{K}_{t_{ij}}|} \frac{p(\mathbf{k}^* \mid \mathbf{c}^*) p(\mathbf{x} \mid \mathbf{k}^*, \mathbf{c}^*)}{p(\mathbf{k} \mid \mathbf{c}) p(\mathbf{x} \mid \mathbf{k}, \mathbf{c})} \tag{29}$$

again with table links and likelihood terms not affected by the move canceling out. The pseudo Gibbs acceptance ratio works out to a simple ratio of table links:

$$\rho_{sm}^{pg} = \left(\frac{1}{2}\right)^{|\mathcal{K}_{t_{ij^*}}| - |\mathcal{K}_{t_{ij}}|} \frac{p(\mathbf{k}^* \mid \mathbf{c}^*)}{p(\mathbf{k} \mid \mathbf{c})} \tag{30}$$

Finally, moves which change customer links but do not cause a change to the partition structure have acceptance ratios of 1 and are always accepted under either proposals.