# Equations for a Dirichlet Process Mixture Model with Gamma-Poisson Observation Distribution

Aaron Schein

March 16, 2014

## 1 Generative Process

The following is a summary of the generative process for a Dirichlet Process mixture model with Gamma-Poisson observation distribution (DPMM-GP).

$$
\begin{aligned}
y_i &\sim F(\lambda_{z_i}) = \text{Poisson}(\lambda_{z_i}) && (1)\\
\lambda_k &\sim H(\gamma) = \text{Gamma}(\gamma) && (2)\\
z_i &\sim \theta && (3)\\
\theta &\sim \text{DP}(\alpha) && (4)
\end{aligned}
$$

Observations are integer-valued counts $y_1 \cdots y_n$ which are drawn from the observation distribution $F(\lambda_{z_i}) = \text{Poisson}(\lambda_{z_i})$. Observations are conditionally independent given their cluster index $z_i$ and cluster parameter $\lambda_{z_i}$. Cluster parameters are real-valued and non-negative and drawn independently from a global prior $H(\gamma) = \text{Gamma}(\gamma)$. $\theta = (\theta_1, \cdots, \theta_k)$ is a real-valued, non-negative, vector that sums to 1 and represents the mixing proportions of the clusters. The mixing proportions are drawn from a Dirichlet Process with concentration parameter $\alpha$.

# 2 Sampling Equations

## 2.1 Algorithm 3

$$\begin{aligned}
z_i^s &\sim P(z_i = c \mid \mathbf{z}_{\text{-}i}, \mathbf{y}, \gamma, \alpha) & (5)\\
&\propto P(z_i = c \mid \mathbf{z}_{\text{-}i}, \alpha)\; P(y_i \mid \mathbf{y}_{\text{-}i}, z_i = c, \gamma) & (6)
\end{aligned}$$

$$P(z_i = c \mid \mathbf{z}_{\text{-}i}, \alpha) = \begin{cases} \frac{N_{c,\,\text{-}i}}{N-1+\alpha} & \text{if } c \text{ has been seen before} \\ \frac{\alpha}{N-1+\alpha} & \text{if } c \text{ is a new cluster} \end{cases} \tag{7}$$

$$\begin{aligned}
P(y_i \mid \mathbf{y}_{\text{-}i}, z_i = c, \gamma) &= P(y_i \mid \mathbf{y}_{c,\text{-}i}, \gamma) & (8)\\
&= \frac{P(y_i, \mathbf{y}_{c,\text{-}i} \mid \gamma)}{P(\mathbf{y}_{c,\text{-}i} \mid \gamma)} & (9)
\end{aligned}$$

$$P(y_i, \mathbf{y}_{c,\text{-}i} \mid \gamma) = \int \mathbf{d}\lambda_c\; P(\lambda_c \mid \gamma)\; \left[\prod_{j \neq i : z_j = c} P(y_j \mid \lambda_c)\right]\; P(y_i \mid \lambda_c) \tag{10}$$

$$P(\mathbf{y}_{c,\text{-}i} \mid \gamma) = \int \mathbf{d}\lambda_c\; P(\lambda_c \mid \gamma)\; \left[\prod_{j \neq i : z_j = c} P(y_j \mid \lambda_c)\right] \tag{11}$$

Only the second equation is expanded further:

$$= \int \mathbf{d}\lambda_c\; \text{Gamma}(\lambda_c; \gamma)\; \left[\prod_{j \neq i : z_j = c} \text{Poisson}(y_j; \lambda_c)\right] \tag{12}$$

Here we use shape/scale parameterization: $\gamma = (\gamma_{\text{shape}}, \gamma_{\text{scale}}) = (a, b)$

$$\begin{aligned}
&= \int \mathbf{d}\lambda_c\; \frac{(\frac{1}{b})^a}{\Gamma(a)} \lambda_c^{a-1} \exp(-\frac{\lambda_c}{b})\; \left[\prod_{j \neq i : z_j = c} \frac{\lambda_c^{y_j}}{\Gamma(y_j + 1)} \exp(-\lambda_c)\right] & (13)\\
&= \frac{(\frac{1}{b})^a}{\Gamma(a)} \frac{\Gamma(a + \sum_{j \neq i} y_j)}{(\frac{1}{b} + N_{c,\text{-}i})^{(a + \sum_{j \neq i} y_j)}}\; \left[\prod_{j \neq i} \frac{1}{\Gamma(y_j + 1)}\right] & (14)
\end{aligned}$$

Plugging this expression back:

$$P(y_i \mid \mathbf{y}_{\text{-}i}, z_i = c, \gamma) = \frac{\frac{(\frac{1}{b})^a}{\Gamma(a)} \frac{\Gamma(a+y_i+\sum_{j\neq i} y_j)}{(\frac{1}{b}+1+N_{c,\text{-}i})^{(a+y_i+\sum_{j\neq i} y_j)}} \left[\prod_{j\neq i} \frac{1}{\Gamma(y_j+1)}\right] \frac{1}{\Gamma(y_i+1)}}{\frac{(\frac{1}{b})^a}{\Gamma(a)} \frac{\Gamma(a+\sum_{j\neq i} y_j)}{(\frac{1}{b}+N_{c,\text{-}i})^{(a+\sum_{j\neq i} y_j)}} \left[\prod_{j\neq i} \frac{1}{\Gamma(y_j+1)}\right]} \tag{15}$$

$$= \frac{\frac{\Gamma(a+y_i+\sum_{j\neq i} y_j)}{(\frac{1}{b}+1+N_{c,\text{-}i})^{(a+y_i+\sum_{j\neq i} y_j)}} \frac{1}{\Gamma(y_i+1)}}{\frac{\Gamma(a+\sum_{j\neq i} y_j)}{(\frac{1}{b}+N_{c,\text{-}i})^{(a+\sum_{j\neq i} y_j)}}} \tag{16}$$

$$= \frac{\Gamma(a + y_i + \sum_{j\neq i} y_j)}{\Gamma(a + \sum_{j\neq i} y_j)} \frac{(\frac{1}{b} + N_{c,\text{-}i})^{(a+\sum_{j\neq i} y_j)}}{(\frac{1}{b} + 1 + N_{c,\text{-}i})^{(a+y_i+\sum_{j\neq i} y_j)}} \frac{1}{\Gamma(y_i + 1)} \tag{17}$$

## 2.2 Algorithm 8

TODO (very similar to above)

## 2.3 Split-Merge Algorithm

$$a(\mathbf{z}^*, \mathbf{z}) = \min \left[1,\ \frac{q(\mathbf{z}|\mathbf{z}^*)}{q(\mathbf{z}^*|\mathbf{z})} \frac{P(\mathbf{z}^*|\alpha)}{P(\mathbf{z}|\alpha)} \frac{P(\mathbf{y}|\mathbf{z}^*, \gamma)}{P(\mathbf{y}|\mathbf{z}, \gamma)}\right] \tag{18}$$

### 2.3.1 Split

$$\frac{P(\mathbf{z}^{\text{split}}|\alpha)}{P(\mathbf{z}|\alpha)} = \alpha \frac{\Gamma(N_{z_i^{\text{split}}})\Gamma(N_{z_j^{\text{split}}})}{\Gamma(N_{z_i})} \tag{19}$$

$$\frac{q(\mathbf{z}|\mathbf{z}^{\text{split}})}{q(\mathbf{z}^{\text{split}}|\mathbf{z})} = \frac{1}{\left(\frac{1}{2}\right)^{N_{z_i^{\text{split}}}+N_{z_j^{\text{split}}}-2}} \tag{20}$$

$$P(\mathbf{y}|\mathbf{z}, \gamma) = \prod_{c=1}^{C} \prod_{k:z_k=c} P(y_k \mid \mathbf{y}_{c,<k}, \gamma) \tag{21}$$

Here $\mathbf{y}_{c,<k}$ represents all observations in cluster $c$ that ‘arrived’ before $k$.

$$= \prod_{c=1}^{C} \prod_{k:z_k=c} \frac{P(y_k, \mathbf{y}_{c,<k} \mid \gamma)}{P(\mathbf{y}_{c,<k} \mid \gamma)} \tag{22}$$

We've already solved this above in the Algorithm 3 section. Note $\gamma = (a, b)$. Also note that $q < k$ (in the sums) is shorthand for $q : z_q = c, q < k$:

$$= \prod_{c=1}^{C} \prod_{k:z_k=c} \frac{\Gamma(a + y_k + \sum_{q<k} y_q)}{\Gamma(a + \sum_{q<k} y_q)} \frac{(\frac{1}{b} + N_{c,\, q<k})^{(a+\sum_{q<k} y_q)}}{(\frac{1}{b} + 1 + N_{c,\, q<k})^{(a+y_k+\sum_{q<k} y_q)}} \frac{1}{\Gamma(y_k + 1)} \tag{23}$$

$$= \prod_{c=1}^{C} \frac{\Gamma(a + \sum_{k:z_k=c} y_k)}{\Gamma(a)} \frac{(\frac{1}{b})^a}{(\frac{1}{b} + N_{z_k})^{a+\sum_{k:z_k=c} y_k}} \frac{1}{\prod_{k:z_k=c} \Gamma(y_k + 1)} \tag{24}$$

$$\frac{P(\mathbf{y}|\mathbf{z}^{\text{split}}, \gamma)}{P(\mathbf{y}|\mathbf{z}, \gamma)} = \frac{\prod_{k:z_k=z_{i^{\text{split}}}} P(y_k \mid \mathbf{y}_{c,<k}, \gamma) \prod_{k:z_k=z_{j^{\text{split}}}} P(y_k \mid \mathbf{y}_{c,<k}, \gamma)}{\prod_{k:z_k=z_i} P(y_k \mid \mathbf{y}_{c,<k}, \gamma)} \tag{25}$$

For each product of terms, we plug in our result from one line above. Note that the $\frac{1}{\Gamma(y_k+1)}$ terms will all cancel, thus we ignore them here.

$$= \frac{\left( \frac{\Gamma(a+\sum_{k:z_k=z_{i^{\text{split}}}} y_k)}{\Gamma(a)} \frac{(\frac{1}{b})^a}{(\frac{1}{b}+N_{z_{i^{\text{split}}}})^{a+\sum_{k:z_k=z_{i^{\text{split}}}} y_k}} \right) \left( \frac{\Gamma(a+\sum_{k:z_k=z_{j^{\text{split}}}} y_k)}{\Gamma(a)} \frac{(\frac{1}{b})^a}{(\frac{1}{b}+N_{z_{j^{\text{split}}}})^{a+\sum_{k:z_k=z_{j^{\text{split}}}} y_k}} \right)}{\left( \frac{\Gamma(a+\sum_{k:z_k=z_i} y_k)}{\Gamma(a)} \frac{(\frac{1}{b})^a}{(\frac{1}{b}+N_{z_i})^{a+\sum_{k:z_k=z_i} y_k}} \right)} \tag{26}$$

$$= \frac{\left( \frac{\Gamma(a+\sum_{k:z_k=z_{i^{\text{split}}}} y_k)}{\Gamma(a)} \frac{(\frac{1}{b})^a}{(\frac{1}{b}+N_{z_{i^{\text{split}}}})^{a+\sum_{k:z_k=z_{i^{\text{split}}}} y_k}} \right) \left( \frac{\Gamma(a+\sum_{k:z_k=z_{j^{\text{split}}}} y_k)}{(\frac{1}{b}+N_{z_{j^{\text{split}}}})^{a+\sum_{k:z_k=z_{j^{\text{split}}}} y_k}} \right)}{\left( \frac{\Gamma(a+\sum_{k:z_k=z_i} y_k)}{(\frac{1}{b}+N_{z_i})^{a+\sum_{k:z_k=z_i} y_k}} \right)} \tag{27}$$

Note: $k \in z_c$ is shorthand for $k : z_k = z_c$:

$$= \frac{(\frac{1}{b})^a}{\Gamma(a)} \frac{\Gamma(a + \sum_{k\in z_{i^{\text{split}}}} y_k)}{(\frac{1}{b} + N_{z_{i^{\text{split}}}})^{a+\sum_{k\in z_{i^{\text{split}}}} y_k}} \frac{\Gamma(a + \sum_{k\in z_{j^{\text{split}}}} y_k)}{(\frac{1}{b} + N_{z_{j^{\text{split}}}})^{a+\sum_{k\in z_{j^{\text{split}}}} y_k}} \frac{(\frac{1}{b} + N_{z_i})^{a+\sum_{k\in z_i} y_k}}{\Gamma(a + \sum_{k\in z_i} y_k)} \tag{28}$$

### 2.3.2 Merge

$$\frac{P(\mathbf{z}^{\text{merge}}|\alpha)}{P(\mathbf{z}|\alpha)} = \frac{1}{\alpha} \frac{\Gamma(N_{z_i^{\text{merge}}})}{\Gamma(N_{z_i})\Gamma(N_{z_j})} \tag{29}$$

$$\frac{q(\mathbf{z}|\mathbf{z}^{\text{merge}})}{q(\mathbf{z}^{\text{merge}}|\mathbf{z})} = (\frac{1}{2})^{N_{z_i}+N_{z_j}-2} \tag{30}$$

$$\frac{P(\mathbf{y}|\mathbf{z}^{\text{merge}},\gamma)}{P(\mathbf{y}|\mathbf{z},\gamma)} = \frac{\prod_{k:z_k=z_i\text{merge}} P(y_k \mid \mathbf{y}_{c,<k},\gamma)}{\prod_{k:z_k=z_i} P(y_k \mid \mathbf{y}_{c,<k},\gamma) \prod_{k:z_k=z_j} P(y_k \mid \mathbf{y}_{c,<k},\gamma)} \tag{31}$$

$$= \frac{\Gamma(a)}{(\frac{1}{b})^a} \frac{\Gamma(a+\sum_{k\in z_i\text{merge}} y_k)}{(\frac{1}{b}+N_{z_i\text{merge}})^{a+\sum_{k\in z_i\text{merge}} y_k}} \frac{(\frac{1}{b}+N_{z_i})^{a+\sum_{k\in z_i} y_k}}{\Gamma(a+\sum_{k\in z_i} y_k)} \frac{(\frac{1}{b}+N_{z_j})^{a+\sum_{k\in z_j} y_k}}{\Gamma(a+\sum_{k\in z_j} y_k)} \tag{32}$$

## 2.4 Split-Merge Algorithm with Restricted Gibbs Scans

### 2.4.1 Gibbs Scan: Split

$$P(z_{k^{\text{split}}} \mid \mathbf{z}_{\text{-}k}, y_k, \mathbf{y}_{\text{-}k}) = \frac{N_{\text{-}k,\, z_{k^{\text{split}}}} P(y_k \mid z_{k^{\text{split}}}, \mathbf{y}_{\text{-}k})}{N_{\text{-}k,\, z_{i^{\text{split}}}} P(y_k \mid z_{k^{\text{split}}} = z_{i^{\text{split}}}, \mathbf{y}_{\text{-}k}) + N_{\text{-}k,\, z_{j^{\text{split}}}} P(y_k \mid z_{k^{\text{split}}} = z_{j^{\text{split}}}, \mathbf{y}_{\text{-}k})} \tag{33}$$

$$P(y_k \mid z_{k^{\text{split}}}, \mathbf{y}_{\text{-}k}) = \frac{\Gamma(a+y_k+\sum_{q\neq k\in z_{k^{\text{split}}}} y_q)}{\Gamma(a+\sum_{q\neq k\in z_{k^{\text{split}}}} y_q)} \frac{(\frac{1}{b}+N_{\text{-}k,\, z_{k^{\text{split}}}})^{(a+\sum_{q\neq k\in z_{k^{\text{split}}}} y_q)}}{(\frac{1}{b}+1+N_{\text{-}k,\, z_{k^{\text{split}}}})^{(a+y_k+\sum_{q\neq k\in z_{k^{\text{split}}}} y_q)}} \frac{1}{\Gamma(y_k+1)} \tag{34}$$