EE6720 Homework 4

Name: Qi Zhao
UNI: qz2316

December 10, 2017

Problem 1

(a)

E-Step(Step 1): Use Bayes rule to calculate the conditional posterior distribution.

$$\begin{aligned} p\left(c_i = k \mid \pi, x_i, \boldsymbol{\lambda}\right) &= \frac{p\left(x_i \mid c_i = k, \pi, \boldsymbol{\lambda}\right) p\left(c_i = k \mid \pi\right)}{\sum_{j=1}^{K} p\left(x_i \mid c_i = j, \pi, \boldsymbol{\lambda}\right) p\left(c_i = j \mid \pi\right)} \\ &= \frac{\pi_k Poisson\left(x_i \mid \lambda_k\right)}{\sum_{j=1}^{K} \pi_j Poisson\left(x_i \mid \lambda_j\right)} \end{aligned}$$

So to summarize

$$q\left(\mathbf{c}\right) = p\left(\mathbf{c} \mid \pi, x, \boldsymbol{\lambda}\right) = \prod_{i=1}^{n} p\left(c_i \mid \pi, x_i, \boldsymbol{\lambda}\right) = \prod_{i=1}^{n} q\left(c_i\right)$$

.

I use notation $q\left(c_i = j\right) = \phi_i\left(j\right)$. In implementing, this means I would first set $\phi_i\left(j\right) = \pi_k Poisson\left(x_i \mid \lambda_k\right)$ for $j = 1, ..., K$ and then normalize the K-dimensional vector ϕ_i by dividing it by its sum.

E-Step(Step 2): Since in our problem, x_i only depends on c_i and all c_i are i.i.d., The expectation we need to take is

$$\mathcal{L}\left(\pi, \boldsymbol{\lambda}\right) = \sum_{i=1}^{n} E_{q(\mathbf{c})}\left[lnp\left(x_i, c_i \mid \pi, \boldsymbol{\lambda}\right)\right] + constant$$

Where the constant is w.r.t π and $\boldsymbol{\lambda}$. Further, because each (x_i, c_i) are conditionally independent of each other. We are then left with:

$$\begin{aligned}
\mathcal{L}(\pi, \boldsymbol{\lambda}) &= \sum_{i=1}^{n} E_{q(c_i)}\left[lnp\left(x_i, c_i | \pi, \boldsymbol{\lambda}\right)\right] - E_{q(\mathbf{c})}\left[q\left(\mathbf{c}\right)\right] + constant \\
&= \sum_{i=1}^{n}\sum_{j=1}^{K} \phi_i\left(j\right)\left[lnp\left(x_i|\lambda_j\right) + ln\pi_j\right] - E_{q(\mathbf{c})}\left[q\left(\mathbf{c}\right)\right] + constant \\
&= \sum_{i=1}^{n}\sum_{j=1}^{K} \phi_i\left(j\right)\left[-lnx_i! - \lambda_j + x_i ln\lambda_j + ln\pi_j\right] - \sum_{i=1}^{n}\sum_{j=1}^{K} \phi_i\left(j\right) ln\phi_i\left(j\right) + constant
\end{aligned}$$

M-Step(Step 3): Finally, I maximiza $\mathcal{L}(\pi, \boldsymbol{\lambda})$ over its parameters. I do this by taking derivatives and setting to zero.

1. $\nabla_{\lambda_j}\mathcal{L} = 0$:

$$\nabla_{\lambda_j}\mathcal{L} = \sum_{i=1}^{n} \phi_i\left(i\right)\left(-1 + \frac{x_i}{\lambda_j}\right) = 0$$

$$\Rightarrow \lambda_j = \frac{1}{n_j}\sum_{i=1}^{n} \phi_i\left(j\right) x_i$$

   Where $n_j = \sum_{i=1}^{n} \phi_i\left(j\right)$

2. $\nabla_{\pi}\mathcal{L} = 0$
   Subject to $\pi_j \geq 0$ and $\sum_{j=1}^{K} \pi_j = 1$, and after Lagrange Multipliers used here I get

$$n_j = \sum_{i=1}^{n} \phi_i\left(j\right)$$

$$n = \sum_{j=1}^{K} n_j$$

$$\pi_j = \frac{n_j}{n}$$

EM algorithm for Poisson Mixture Model:

Inputs: Data $X = (x_1, ..., x_n)$, and K.

Output: Poisson parameters $\boldsymbol{\lambda}$, π and cluster assignment distributions ϕ_i.

1. Initialize π^0 and all λ_j^0 in some way.

2. At iteration t,

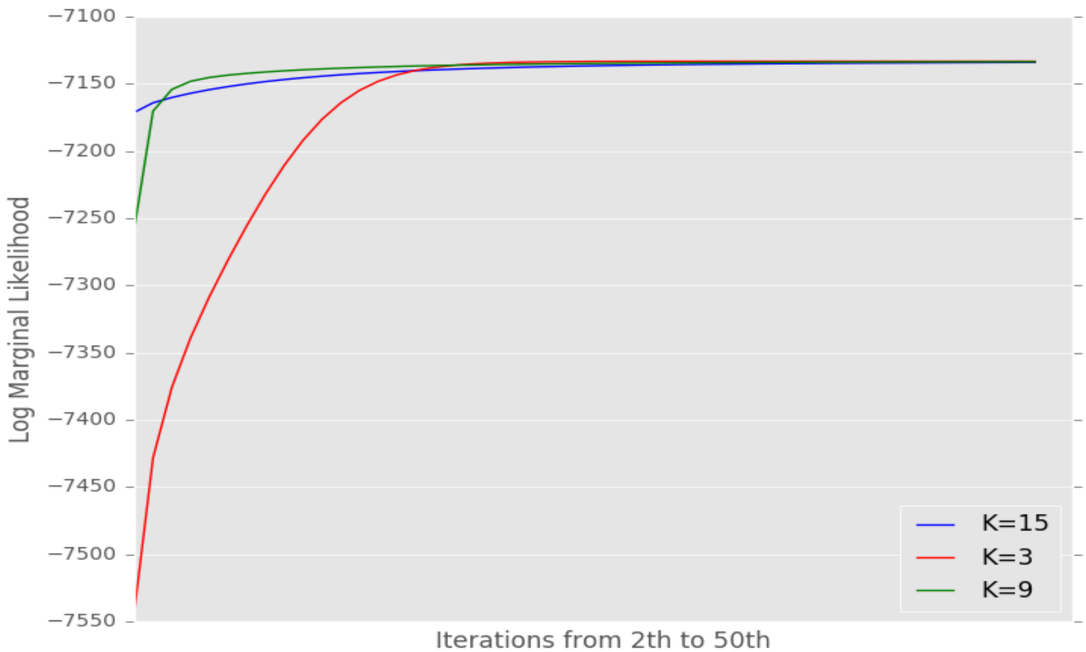

Figure 1: Log Marginal Likelihood over Iterations 2 to 50

(a) E-Step: For $i = 1, ..., n$ and $j = 1, ..., K$ set

$$\phi_i^t(j) = \frac{\pi_j^{t-1} Poisson\left(x_i | \lambda_j^{t-1}\right)}{\sum_{k=1}^{K} \pi_k^{t-1} Poisson\left(x_i | \lambda_k^{t-1}\right)}$$

(b) M-Step: Set

$$n_j^t = \sum_{i=1}^{n} \phi_i^t(j)$$

$$\lambda_j^t = \frac{1}{n_j^t} \sum_{i=1}^{n} \phi_i^t(j)\, x_i$$

$$\pi_j^t = \frac{n_j^t}{\sum_{k=1}^{K} n_k^t}$$

(c) Evaluate $\mathcal{L}\left(\pi^t, \boldsymbol{\lambda}^t\right)$ to access convergence.

(b)

The figure for $K = 3, 9, 15$ is showed below as Figure 1.

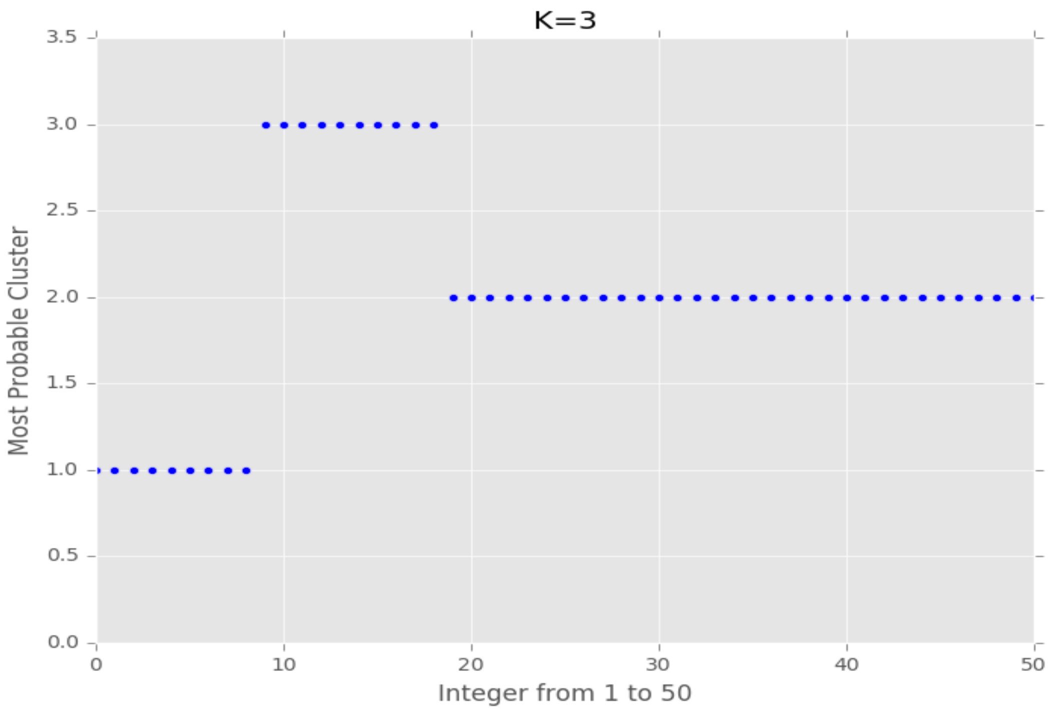

Figure 2: Most Probable Cluster for 0 to 50 When K=3

(c)

Here, what we want to maximize is $p\left(\hat{c}=j,\hat{x}|\pi,\boldsymbol{\lambda},X\right)$ for any $\hat{x}$ of interest.

$$\begin{aligned}p\left(\hat{c}=j,\hat{x}|\pi,\boldsymbol{\lambda},X\right)&=p\left(\hat{x}|\pi,c_i=j,\boldsymbol{\lambda},X\right)p\left(c_i=j|\pi,\boldsymbol{\lambda},X\right)\\&=p\left(\hat{x}|\lambda_j\right)\pi_j\\&=Poisson\left(\hat{x}|\lambda_j\right)\pi_j\end{aligned}$$

We then use the most recent updated values of $\boldsymbol{\lambda}$ and π to calculate the probabilities based on the above equation. The results are showed from Figure 2 to 4.

Problem 2

(a)

From the general approach that we can find the optimal form of each q distribution as followed.

1. $q\left(\pi\right)$:

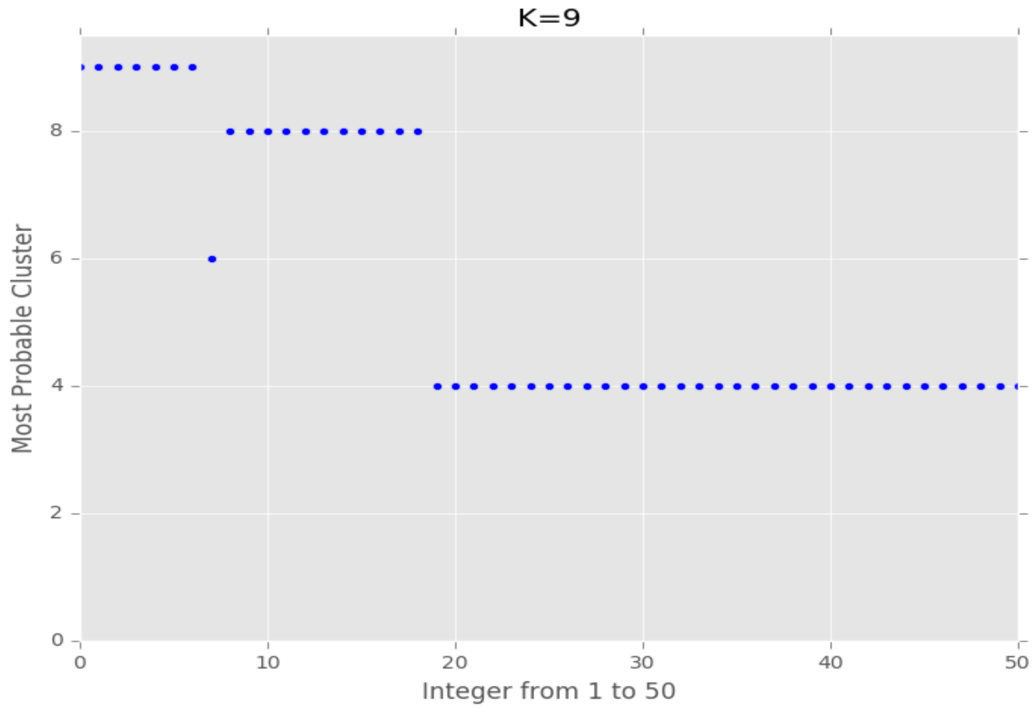

Figure 3: Most Probable Cluster for 0 to 50 When K=9

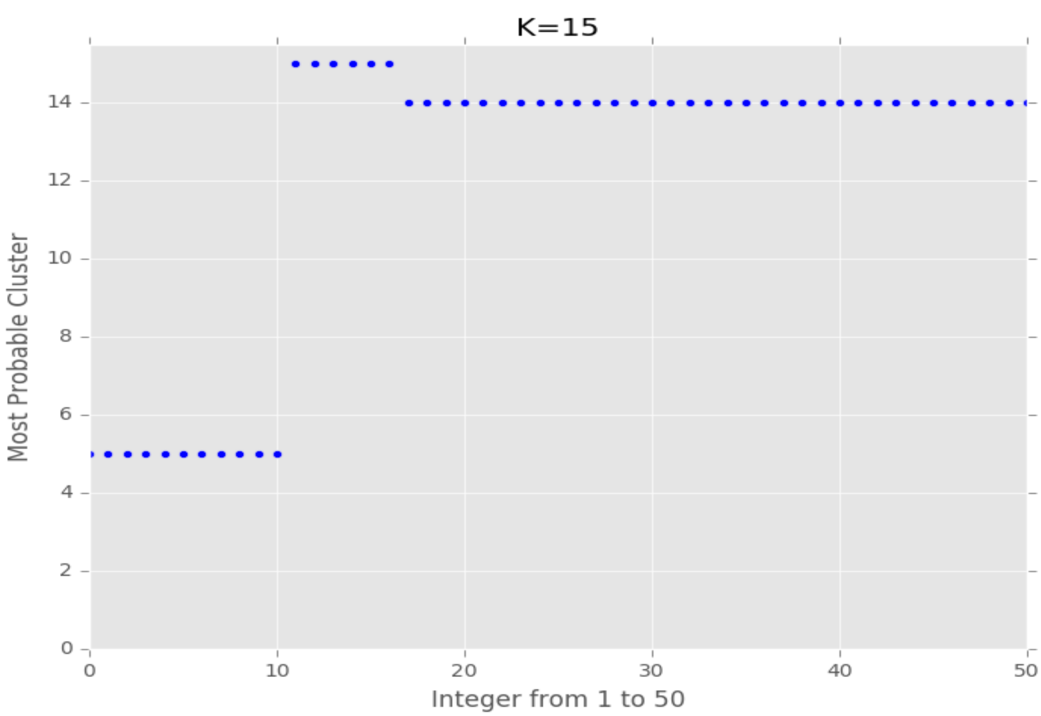

Figure 4: Most Probable Cluster for 0 to 50 When K=15

$$
\begin{aligned}
q(\pi) \propto & exp\left(E_{q(\boldsymbol{\lambda})q(\mathbf{c})}\left[lnp\left(x|\boldsymbol{\lambda},\mathbf{c}\right)+lnp\left(\mathbf{c}|\pi\right)+lnp\left(\pi\right)+lnp\left(\boldsymbol{\lambda}\right)\right]\right)\\
&\propto exp\left(\sum_{j=1}^{K}\left(\alpha-1\right)ln\pi_j+\sum_{i=1}^{n}E_{q(\mathbf{c})}\left[lnp\left(c_i|\pi\right)\right]\right)\\
&\propto exp\left(\sum_{j=1}^{K}\left(\alpha-1\right)ln\pi_j+\sum_{i=1}^{n}\sum_{j=1}^{K}\phi_i\left(j\right)ln\pi_j\right)\\
&\propto \prod_{j=1}^{K}\pi_j^{\alpha+\sum_{i=1}^{n}\phi_i(j)-1}
\end{aligned}
$$

It is clearly that we should set $q(\pi)=Dirichlet\left(\alpha_1,...,\alpha_K\right)$, where $\alpha_j=\alpha+n_j$ and $n_j=\sum_{i=1}^{n}\phi_i\left(j\right)$.

2. $q\left(\lambda_j\right)$:

$$
\begin{aligned}
q\left(\lambda_j\right) &\propto exp\left(E_q\left[lnp\left(x|\boldsymbol{\lambda},\mathbf{c}\right)+lnp\left(\boldsymbol{\lambda}\right)\right]\right)\\
&\propto exp\left(E_q\left[\sum_{i=1}^{n}I\left(c_i=j\right)lnp\left(x_i|\lambda_j\right)\right]\right)\lambda_j^{a-1}exp\left(-b\lambda_j\right)\\
&\propto \lambda_j^{a+\sum_{i=1}^{n}\phi_i(j)x_i-1}exp\left(-\left(b+\sum_{i=1}^{n}\phi_i\left(j\right)\right)\lambda_j\right)
\end{aligned}
$$

So $q\left(\lambda_j\right)=Gamma\left(a_j,b_j\right)$, where $a_j=a+\sum_{i=1}^{n}\phi_i\left(j\right)x_i$ and $b_j=b+\sum_{i=1}^{n}\phi_i\left(j\right)$.

3. $q\left(c_i=j\right)$:

$$
\begin{aligned}
q\left(c_i=j\right) &\propto exp\left(E_q\left[lnp\left(x_i|\boldsymbol{\lambda},c_i=j\right)+lnp\left(c_i=j|\pi\right)\right]\right)\\
&\propto exp\left(E_q\left[lnp\left(x_i|\lambda_j\right)+ln\pi_j\right]\right)\\
&\propto exp\left(E_q\left[-\lambda_j+x_i ln\lambda_j+ln\pi_j\right]\right)\\
&\propto exp\left(E_{q(\lambda_j)}\left(-\lambda_j\right)+x_iE_{q(\lambda_j)}\left(ln\lambda_j\right)+E_{q(\pi)}\left(ln\pi_j\right)\right)\\
&\propto exp\left(-\frac{a_j}{b_j}+x_i\left(\psi\left(a_j\right)-lnb_j\right)+\psi\left(\alpha_j\right)-\psi\left(\sum_{k=1}^{K}\alpha_k\right)\right)
\end{aligned}
$$

Variational Inference algorithm for Poisson Mixture Model:

Input: Data $X=\left(x_1,...,x_n\right)$, and K.

Output: Parameters for $q(\pi)$, $q(c_i)$ and $q(\lambda_j)$ for i from 1 to n and j from 1 to K.

1. Initialize $(\alpha_1^0, ..., \alpha_K^0)$, $(a_1^0, ..., a_K^0)$ and $(b_1^0, ..., b_K^0)$ in some way.

2. At iteration t:

   (a) Update $q(c_i)$ for i from 1 to n by setting:

$$\phi_i^t(j) = \frac{exp\left(-\frac{a_j^{t-1}}{b_j^{t-1}} + x_i\left(\psi\left(a_j^{t-1}\right) - lnb_j^{t-1}\right) + \psi\left(\alpha_j^{t-1}\right) - \psi\left(\sum_{k=1}^K \alpha_k^{t-1}\right)\right)}{\sum_{k=1}^K exp\left(-\frac{a_k^{t-1}}{b_k^{t-1}} + x_i\left(\psi\left(a_k^{t-1}\right) - lnb_k^{t-1}\right) + \psi\left(\alpha_k^{t-1}\right) - \psi\left(\sum_{k=1}^K \alpha_k^{t-1}\right)\right)}$$

   (b) Set $n_j^t = \sum_{i=1}^n \phi_i^t(j)$ for $j = 1, ...K$.

   (c) Update $q(\pi)$ by setting:

$$\alpha_j^t = \alpha + n_j^t$$

   (d) Update $q(\lambda_j)$ for $j = 1, ..., K$ by setting:

$$a_j^t = a + \sum_{i=1}^n \phi_i^t(j)\, x_i$$

$$b_j^t = b + \sum_{i=1}^n \phi_i^t(j) = b + n_j^t$$

   (e) Calculate the variational objective function to access convergence as a function of iteration:

$$
\begin{aligned}
\mathcal{L} &= E_q\left[lnp\left(x, \mathbf{c}, \boldsymbol{\lambda}, \pi\right)\right] - E\left[lnq\right] \\
&= E_q\left[\sum_{i=1}^{n}\sum_{j=1}^{K} I\left(c_i = j\right)\left[lnp\left(x_i|\lambda_j\right) + lnp\left(\pi_j\right)\right]\right] + E_q\left[lnp\left(\pi\right)\right] + \sum_{j=1}^{K} E_q\left[lnp\left(\lambda_j\right)\right] \\
&- E_q\left[lnq\left(\pi\right)\right] - \sum_{i=1}^{n} E_q\left[lnq\left(c_i\right)\right] - \sum_{j=1}^{K} E_q\left[lnq\left(\lambda_j\right)\right] \\
&= \sum_{i=1}^{n}\sum_{j=1}^{K} \phi_i\left(j\right)\left[-\frac{a_j}{b_j} - ln\left(x_i!\right) + x_i\left(\psi\left(a_j\right) - lnb_j\right) + \psi\left(\alpha_j\right) - \psi\left(\sum_{j=1}^{K}\alpha_j\right)\right] \\
&+ \left(\alpha - 1\right)\sum_{j=1}^{K}\left(\psi\left(\alpha_j\right) - \psi\left(\sum_{k=1}^{K}\alpha_k\right)\right) + ln\Gamma\left(K\alpha\right) - Kln\Gamma\left(\alpha\right) \\
&+ \sum_{j=1}^{K}\left[alnb - ln\Gamma\left(a\right) + \left(a - 1\right)\left(\psi\left(a_j\right) - lnb_j\right) - b\frac{a_j}{b_j}\right] \\
&- \sum_{j=1}^{K}\left(\alpha_j - 1\right)\left(\psi\left(\alpha_j\right) - \psi\left(\sum_{k=1}^{K}\alpha_k\right)\right) + \sum_{j=1}^{K} ln\Gamma\left(\alpha_j\right) - ln\Gamma\left(\sum_{j=1}^{K}\alpha_j\right) \\
&- \sum_{i=1}^{n}\sum_{j=1}^{K}\phi_i\left(j\right)ln\phi_i\left(j\right) \\
&+ \sum_{j=1}^{K}\left(a_j - lnb_j + ln\Gamma\left(a_j\right) + \left(1 - a_j\right)\psi\left(a_j\right)\right) \\
&+ \left(a - 1\right)\sum_{j=1}^{K}\left(\psi\left(a_j\right) - lnb_j\right) - b\sum_{j=1}^{K}\frac{a_j}{b_j}
\end{aligned}
$$

Therefore, we should evaluate $\mathcal{L}\left(a_1^t, ..., a_K^t, b_1^t, ...b_K^t, \alpha_1^t, ..., \alpha_K^t, \phi_i^t\left(j\right)\right)$ for i and j from 1 to n and 1 to K respectively.

(b)

The figure for objective function $K = 3, 15, 50$ is showed below as Figure 5.

(c)

What we need to calculate is $p\left(\hat{x}, \hat{c} = j|X\right)$ for $j = 1, ..., K$ and any value of $\hat{x}$ of interest.

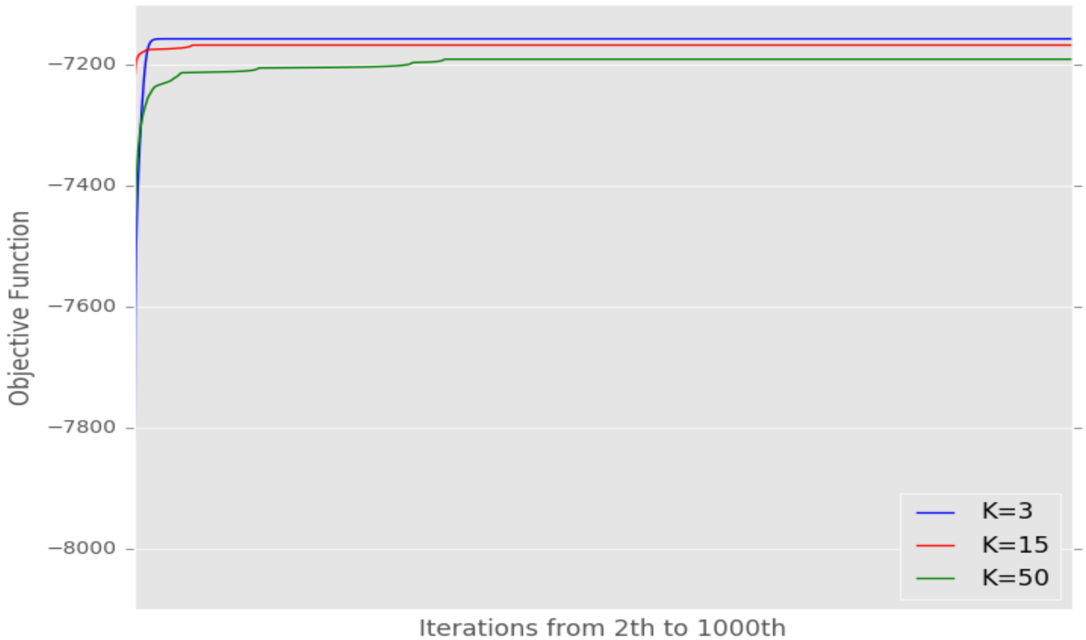

Figure 5: Objective Function over Iterations 2 to 1000

$$
\begin{aligned}
p\left(\hat{x}, \hat{c}=j | X\right) &= \int p\left(\hat{x}, \hat{c}=j | \pi, \boldsymbol{\lambda}\right) p\left(\pi | X\right) p\left(\boldsymbol{\lambda} | X\right) d\pi d\boldsymbol{\lambda} \\
&= \int p\left(\hat{x}, \hat{c}=j | \pi, \boldsymbol{\lambda}\right) q\left(\pi\right) q\left(\boldsymbol{\lambda}\right) d\pi d\boldsymbol{\lambda} \\
&= \int \pi_j q\left(\pi\right) d\pi \int q\left(\lambda_j\right) \frac{\lambda_j^{\hat{x}}}{\hat{x}!} e^{-\lambda_j} d\lambda_j \\
&= \frac{\alpha_j}{\sum_{k=1}^{K} \alpha_k} \frac{1}{\hat{x}!} \frac{b_j^{a_j}}{\Gamma\left(a_j\right)} \frac{\Gamma\left(\hat{x}+a_j\right)}{\left(b_j+1\right)^{\hat{x}+a_j}}
\end{aligned}
$$

After getting $q\left(\pi\right), q\left(\boldsymbol{\lambda}\right)$ from the last iteration of VI, we just let $\hat{x}$ equals 0 to 50 and get the indices for each of them with the maximum possibility given j from 1 to K. The results are showed from Figure 6 to 8.

Problem 3

(a)

First, we need to derive the $p\left(\lambda_k | \left\{x_i : c_i = k\right\}\right)$ and $p\left(c_i | x_i, \boldsymbol{\lambda}, c_{-i}\right)$.

1. $p\left(\lambda_k | \left\{x_i : c_i = k\right\}\right)$:

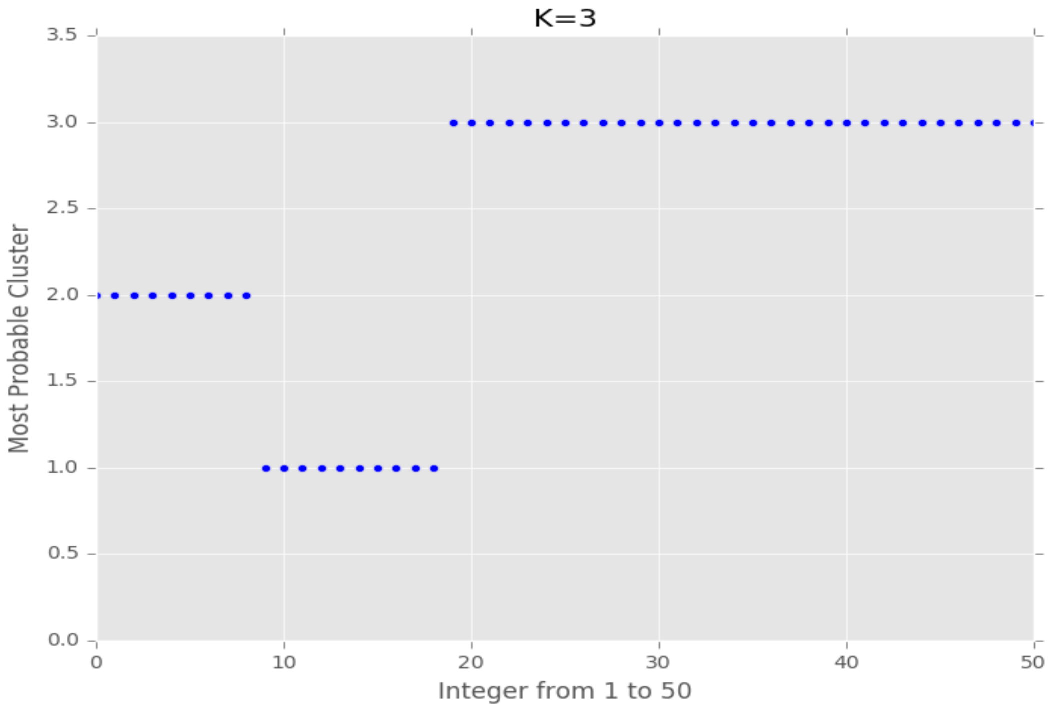

Figure 6: Most Probable Cluster for 0 to 50 When K=3

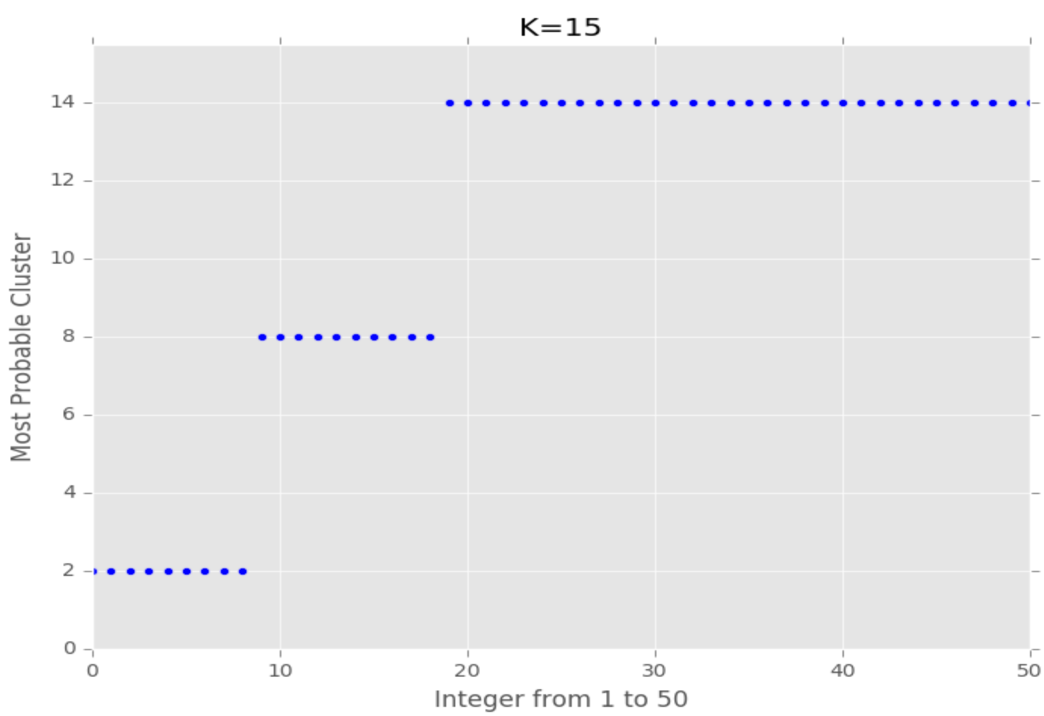

Figure 7: Most Probable Cluster for 0 to 50 When K=15

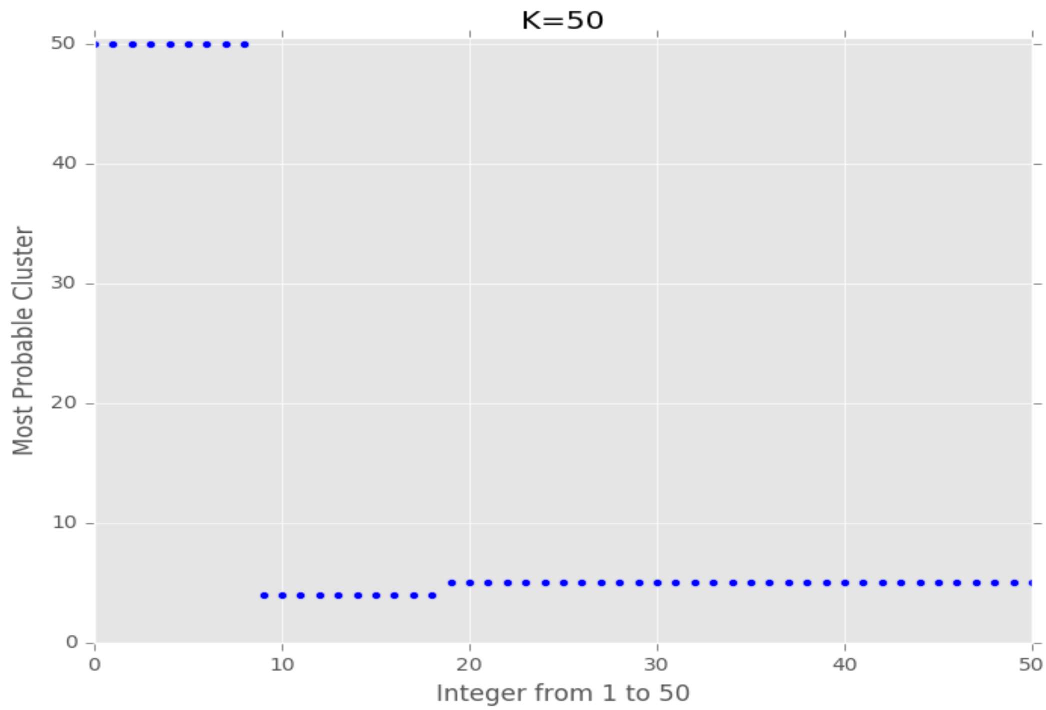

Figure 8: Most Probable Cluster for 0 to 50 When K=50

$$
\begin{aligned}
p\left(\lambda_k \middle| \left\{x_i : c_i = k\right\}\right) &\propto p\left(\lambda_k\right) \prod_{i=1}^{n} p\left(x_i|\lambda_k\right)^{I(c_i=k)} \\
&\propto \lambda_k^{a+\sum_{i=1}^{n} x_i I(c_i=k)-1} e^{-(n_k+b)\lambda_k}
\end{aligned}
$$

Where $n_k = \sum_{i=1}^{n} I\left(c_i = k\right)$. If $n_k > 0$, we sample λ_k as above, which is just $Gamma\left(a + \sum_{i=1}^{n} x_i I\left(c_i = k\right), b + n_k\right)$. If $n_k = 0$, we sample λ_k from $p\left(\lambda|x_i\right)$:

$$
p\left(\lambda|x_i\right) \sim Gamma\left(a + x_i, b + 1\right)
$$

2. $p\left(c_i|x_i, \boldsymbol{\lambda}, c_{-i}\right)$:

Case 1 $n_j^{(-i)} > 0$

$$
p\left(c_i = j|x_i, \boldsymbol{\lambda}, c_{-i}\right) \propto p\left(x_i|\lambda_j\right) \frac{n_j^{(-i)}}{\alpha + n - 1}
$$

Case 2 $n_j^{(-i)} = 0$

$$p\left(c_i = j'|x_i, \boldsymbol{\lambda}, c_{-i}\right) \propto \frac{\alpha}{\alpha+n-1}\int p\left(x_i|\lambda\right)p\left(\lambda\right)d\lambda$$
$$\propto \frac{\alpha}{\alpha+n-1}\frac{b^a}{\Gamma\left(a\right)x_i!}\frac{\Gamma\left(a+x_i\right)}{\left(b+1\right)^{a+x_i}}$$

Marginalized Sampling Method for the Dirichlet Process Poisson Mixture Model:

• Initialize in some way, e.g., set all $c_i = 1$ for $i = 1, ..., n$, and sample $\lambda_1 \sim p(\lambda) = Gamma\,(a, b)$.

• At iteration t: Re-index clusters to go from 1 to $K^{(t-1)}$, where $K^{(t-1)}$ is the number of occupied clusters after the previous iteration. Sample all the variables below using the most recent values of the other variables.

1. For $i = 1, ..., n$

   (a) For all j such that $n_j^{(-i)} > 0$, set

   $$\hat{\phi}_i\left(j\right) = Poisson\left(x_i|\lambda_j\right)\frac{n_j^{(-i)}}{\alpha+n-1}$$

   (b) For a new value j', set

   $$\hat{\phi}_i\left(j'\right) = \frac{\alpha}{\alpha+n-1}\frac{b^a}{\Gamma\left(a\right)x_i!}\frac{\Gamma\left(a+x_i\right)}{\left(b+1\right)^{a+x_i}}$$

   (c) Normalize $\hat{\phi}_i\left(j\right)$ and sample the index c_i from a discrete distribution with this parameter.

   (d) If $c_i = j'$, generate $\lambda_{j'} \sim Gamma\left(a + x_i, b+1\right)$

2. For $k = 1, ..., K^t$ (K^t is the number of non-zero clusters that are re-indexed after completing Step 1), generate $\lambda_k \sim Gamma\left(a + \sum_{i=1}^n x_i I\left(c_i = k\right), b + n_k\right)$,where $n_k = \sum_{i=1}^n I\left(c_i = k\right)$

(b)

6 different color lines in Figure 9 show the result of top 6 probably clusters. Figure 10 shows the begining part of Figure 9.

(c)

After around 600 the algorithm converged and only 3 clusters left, among the three main clusters, one cluster only has few datapoint in it. Figure 11 shows the number of total clusters.

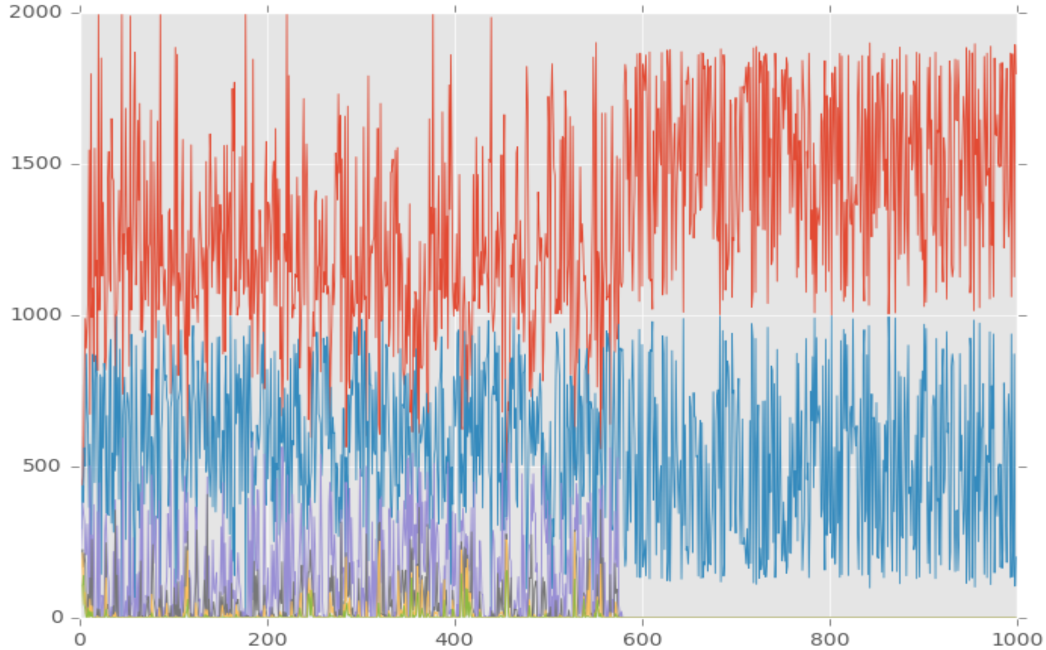

Figure 9: Iterations 1000

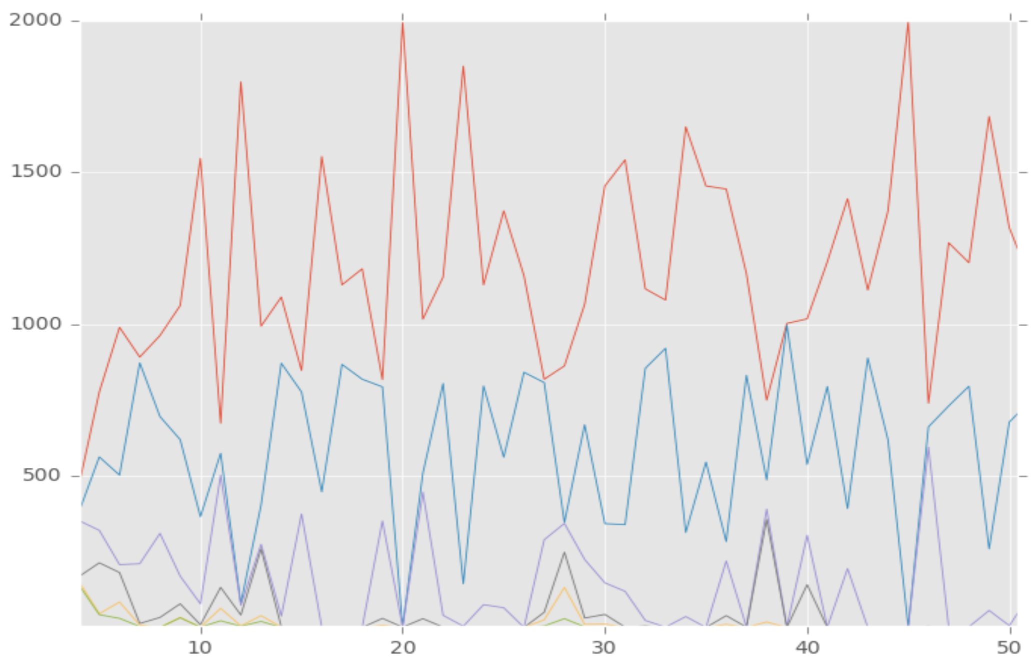

Figure 10: Iterations 1000

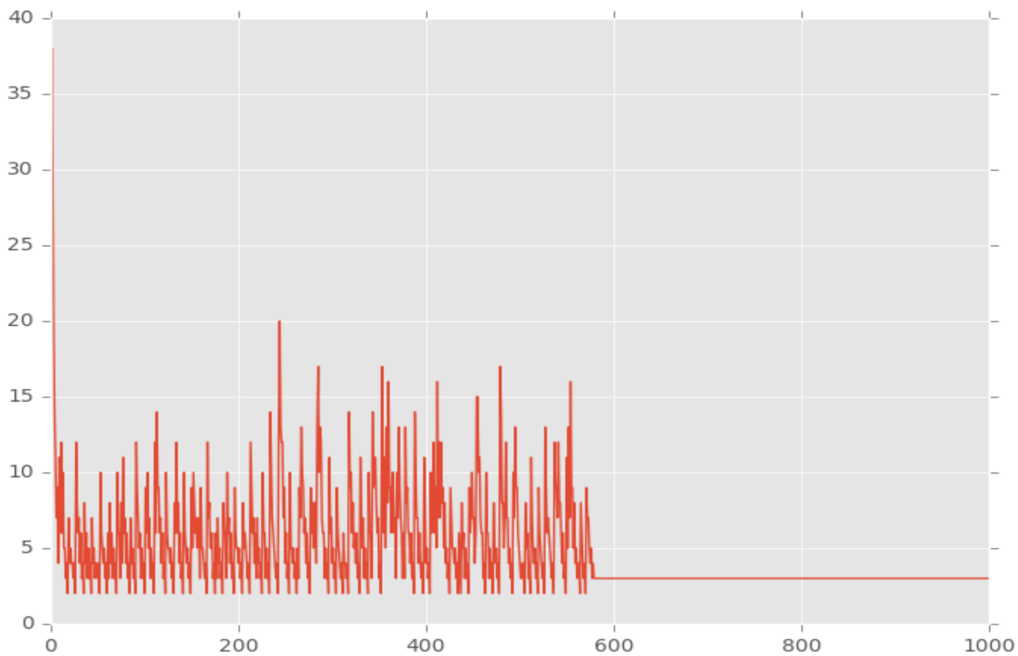

Figure 11: Iterations 1000