Problem 1

$$\ln p(X|\pi,\theta) = \sum_c q(c) \ln \frac{p(x,c|\pi,\theta)}{q(c)} + \sum_c q(c) \ln \frac{q(c)}{p(c|x,\pi,\theta)}.$$

E-step: Set $q(c) = p(c|x,\pi,\theta)$.

$$p(c|x,\pi,\theta) \propto \prod_{i=1}^{n} p(x_i|c_i,\theta)\, p(c_i|\pi).$$

$$= \prod_{i=1}^{n} \frac{p(x_i|c_i,\theta)\, p(c_i|\pi)}{Z_i} = \prod_{i=1}^{n} p(c_i|x_i,\pi,\theta) = \prod_{i=1}^{n} q(c_i).$$

where
$$p(c_i=k|x_i,\pi,\theta) = \frac{p(x_i|c_i=k,\pi,\theta)\, p(c_i=k|\pi)}{\sum_{j=1}^{K} p(x_i|c_i=j,\pi,\theta)\, p(c_i=j|\pi)}.$$

$$= \frac{\pi_k \,\text{Binomial}(x_i|\theta_k)}{\sum_{j=1}^{K} \pi_j \,\text{Binomial}(x_i|\theta_j)}.$$

$$\mathcal{L}(\pi,\theta) = \sum_{i=1}^{n} \mathbb{E}_{q(c)}[\ln p(x_i,c_i|\pi,\theta)] + \text{const}.$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{K} \phi_i(j)[\ln p(x_i|c_i=j,\pi,\theta) + \ln p(c_i=j|\pi)] + \text{const}.$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{K} \phi_i(j)[x_i \ln\theta_j + (20-x_i)\ln(1-\theta_j) + \ln\pi_j] + \text{const}.$$

M-step:

$$\nabla_{\theta_j}\mathcal{L} = 0 \qquad \theta_j = \frac{\sum_{i=1}^{n}\phi_i(j)x_i}{20\sum_{i=1}^{n}\phi_i(j)} = \frac{\sum_{i=1}^{n}\phi_i(j)x_i}{20 n_j}, \quad n_j = \sum_{i=1}^{n}\phi_i(j).$$

$$\nabla_{\pi}\mathcal{L} = 0 \qquad \pi_j = \frac{n_j}{n}, \quad n_j = \sum_{i=1}^{n}\phi_i(j).$$

$$f_t = \ln p(X|\pi,\theta) = \cancel{\sum_{i=1}^{n}\sum_{j=1}^{K} \ln\pi_j + x_i \ln}$$

$$\sum_{i=1}^{n}\sum_{j=1}^{K} \ln\pi_j + \ln \text{Binomial}(x_i|\theta_j).$$

ML-EM for BMM

Input: Data $x_1, \cdots, x_n$, $x \in R^d$. Number of clusters K.

Output: BMM parameters π, θ and cluster assignment distributions ϕ_i.

1. Initialize $\pi^{(0)}$ and $\theta^{(0)}$ in some way.

2. At iteration t,

   E-Step: For $i=1, \cdots, n$ and $j=1, \cdots, K$ set.

$$\phi_i^{(t)}(j) = \frac{\pi_j \, \text{Binomial}(x_i \mid \theta_j)}{\sum_{k=1}^{K} \pi_k \, \text{Binomial}(x_i \mid \theta_k)}$$

   M-Step: Set

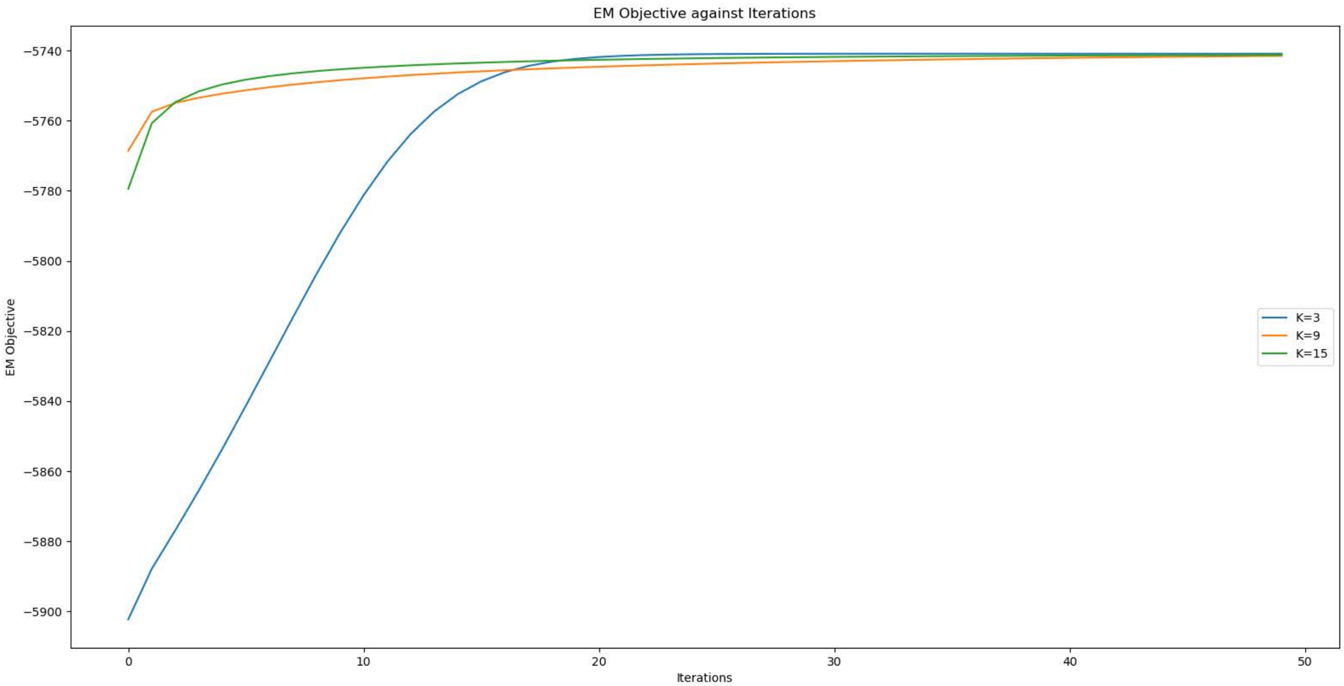
$$n_j^{(t)} = \sum_{i=1}^{n} \phi_i^{(t)}(j)$$

$$\theta_j = \frac{\sum_{i=1}^{n} \phi_i^{(t)}(j)\, x_i}{20 n_j^{(t)}}$$

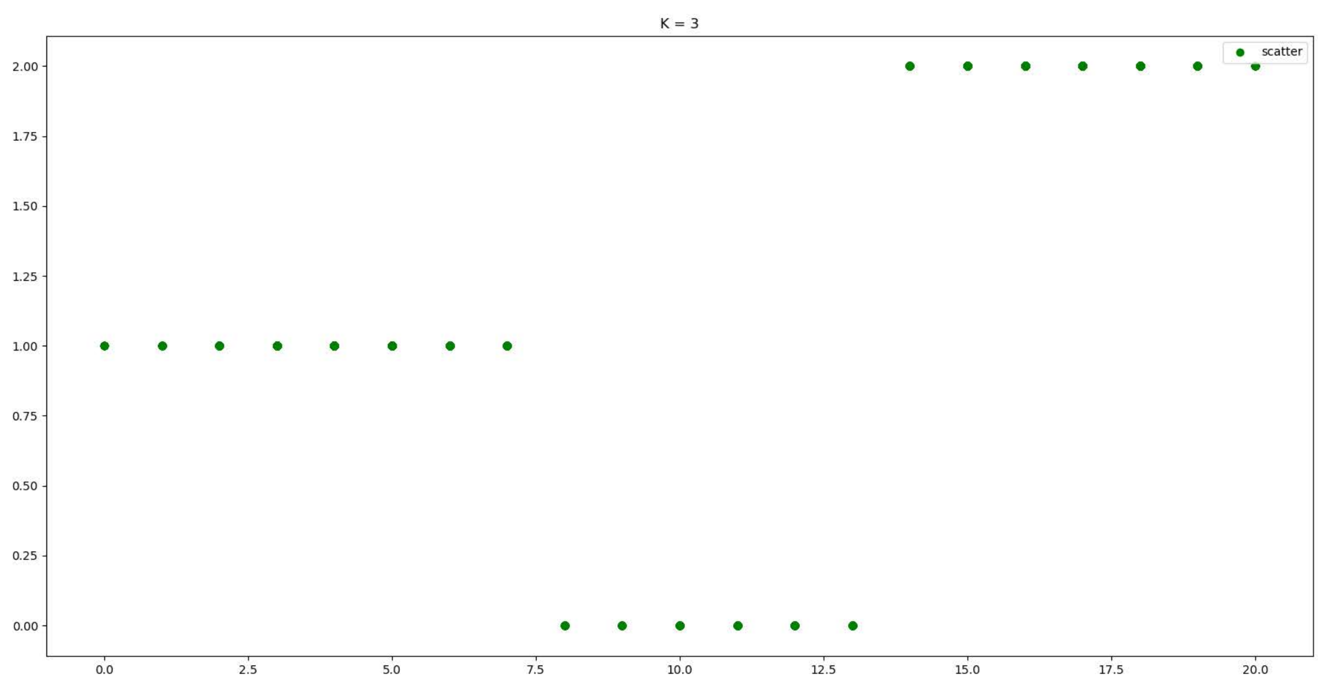
$$\pi_j = \frac{n_j^{(t)}}{n}$$

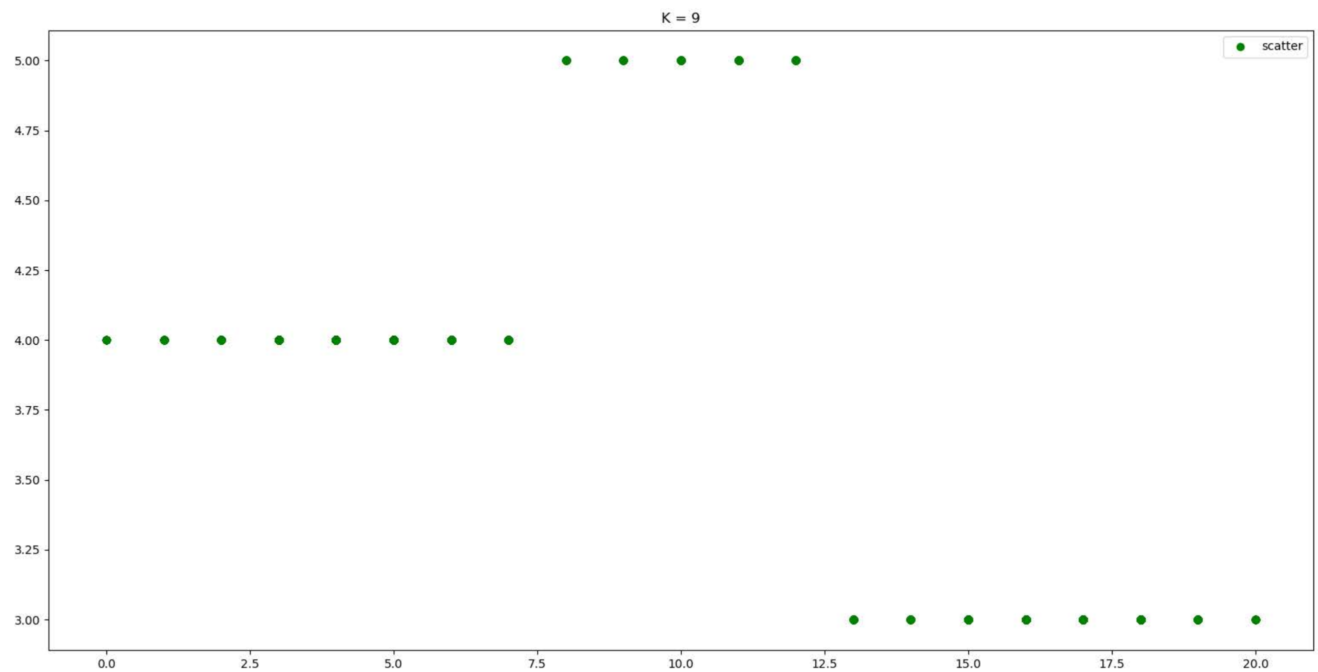
3. Calculate $f_t = \ln p(x \mid \pi, \theta)$

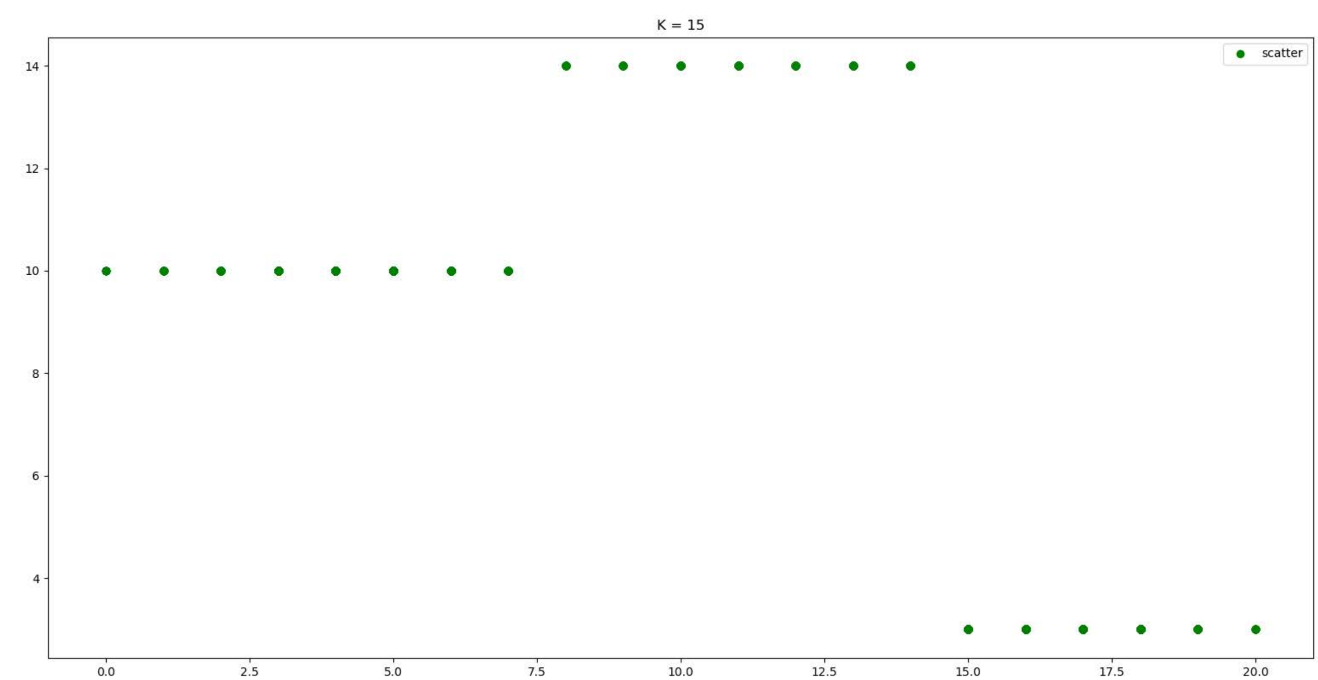
b.


EM Objective against Iterations
EM Objective
Iterations
K=3
K=9
K=15
−5740
−5760
−5780
−5800
−5820
−5840
−5860
−5880
−5900
0
10
20
30
40
50

c.


K = 3
scatter
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0


K = 9
scatter


K = 15
scatter

Problem 2.

$$p(x, c, \pi, \theta) = \prod_{i=1}^{n} p(x_i, c_i, \pi, \theta) = \prod_{i=1}^{n} p(x_i | c_i, \pi, \theta) p(c_i|\pi) \prod_{j=1}^{K} p(\theta_j) p(\pi)$$

$\underline{q(c_i=j)}$: $q(c_i=j) \propto e^{\mathbb{E}[\ln p(x, c_i=j | \pi, \theta_j)]}$

$$\propto e^{\mathbb{E}[\ln p(x|\theta_j)] + \mathbb{E}[\ln p(c_i=j|\pi)]}$$

$$\propto e^{x_i \mathbb{E}[\ln \theta_j] + (20-x_i)\mathbb{E}[\ln(1-\theta_j)] + \mathbb{E}[\ln \pi_j]}$$

$$\propto e^{x_i[\psi(a_j') - \psi(a_j'+b_j')] + (20-x_i)[\psi(b_j') - \psi(a_j'+b_j')] + \psi(\alpha_j') - \psi(\sum_k \alpha_k')}$$

$\underline{q(\pi)}$: $q(\pi) \propto e^{\sum_{i=1}^{n} \mathbb{E}[\ln p(c_i=j|\pi)] + \ln p(\pi)}$

$$\propto e^{\sum_{i=1}^{n}\sum_{j=1}^{K} \phi_i(j) \ln \pi_j + \sum_{j=1}^{K} (\alpha-1) \ln \pi_j}$$

$$\propto \prod_{j=1}^{K} \pi_j^{\sum_{i=1}^{n} \phi_i(j) + \alpha - 1}$$

$$= \text{Dirichlet}(\alpha')$$

where $\alpha_j' = n_j + \alpha_j$, $\quad n_j = \sum_{i=1}^{n} \phi_i(j)$.

$$q(\theta_j) \propto e^{\sum_{i=1}^{n} \mathbb{E}[\ln p(x|c_i=j, \pi, \theta)] + \ln p(\theta_j)}$$

$$\propto e^{[\sum_{i=1}^{n} \phi_i(j) x_i + a - 1] \ln \theta_j + [\sum_{i=1}^{n} \phi_i(j)(20-x_i) + b - 1) \ln(1-\theta_j)}.$$

$$= \text{Beta}(a', b')$$

where $a' = \sum_{i=1}^{n} \phi_i(j) x_i + a$, $\quad b' = \sum_{i=1}^{n} \phi_i(j)(20-x_i) + b$.

$$\mathcal{L} = \mathbb{E}[\ln p(x, c, \pi, \theta)] - \mathbb{E}[\ln q]$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{K} \mathbb{E}[\mathbb{1}(c_i=j)[\ln p(x_i|\theta_j) + \ln \pi_j]] + \sum_{j=1}^{K} \mathbb{E}[\ln p(\theta_j)] + \mathbb{E}[\ln p(\pi)]$$

$$- \sum_{i=1}^{n} \mathbb{E}[\ln q(c_i)] - \sum_{k=1}^{K} \mathbb{E}[\ln q(\theta_k)] - \mathbb{E}[\ln q(\pi)]$$

$$\mathbb{E}[\mathbb{1}(c_i=j)[\ln p(x_i|\theta_j) + \ln \pi_j]] = \phi_i(j)\, \mathbb{E}[\ln p(x_i|\theta_j) + \ln \pi_j]$$

$$= \phi_i(j) \{ \mathbb{E}[\ln p(x_i|\theta_j)] + \mathbb{E}[\ln(\pi_j)]$$

$$= \phi_i(j) \{ \ln \binom{20}{x_i} + x_i[\psi(a_j') - \psi(a_j'+b_j')] + (20-x_i)[\psi(b_j') - \psi(a_j'+b_j')] + \psi(\alpha_j') - \psi(\textstyle\sum_k \alpha_k')$$

$$\mathbb{E}[\ln p(\theta_j)] = (a_j-1)[\psi(a_j') - \psi(a_j'+b_j')] + (b_j-1)[\psi(b_j') - \psi(a_j'+b_j')] - \ln B(a_j, b_j)$$

$$\mathbb{E}[\ln p(\pi)] = \textstyle\sum_i (\alpha_i - 1)[\psi(\alpha_i') - \psi(\sum_k \alpha_k')] - \ln B(\alpha_j).$$

$$\mathbb{E}[\ln q(c_i)] = \sum_{j=1}^{K} \phi_i(j) \cdot \ln \phi_i(j)$$

$$\mathbb{E}[\ln q(\theta_j)] = (a_j'-1)[\psi(a_j') - \psi(a_j'+b_j')] + (b_j'-1)[\psi(b_j') - \psi(a_j'+b_j')] - \ln B(a_j', b_j')$$

$$\mathbb{E}[\ln q(\pi_j)] = \textstyle\sum_i (\alpha_i'-1)[\psi(\alpha_i') - \psi(\sum_k \alpha_k')] - \ln B(\alpha_j').$$

where $a_j' = a_j^{(t)}$, $b_j' = b_j^{(t)}$, $\alpha_j' = \alpha_j^{(t)}$.

VI for. BMM.

Input: Data $x_1, \dots, x_n$, $x \in R^d$. Number of cluster K

Output: Parameters for $q(\pi)$, $q(\theta_j)$, $q(c_i)$.

1. Initialize $(\alpha_1^{(0)}, \dots, \alpha_K^{(0)})$, $(a_j^{(0)}, b_j^{(0)})$ in some way.

2. At iteration t,

(a) Update $q(c_i)$ for $i=1, \dots, n$:

$$\phi_i^{(t)}(j) = \frac{e^{x_i t_1(j) + (20-x_i) t_2(j) + t_3(j)}}{\sum_{k=1}^{K} e^{x_i t_1(k) + (20-x_i) t_2(k) + t_3(k)}}$$

where $t_1(j) = \psi(a_j^{(t-1)}) - \psi(a_j^{(t-1)} + b_j^{(t-1)})$ $\quad t_2 = \psi(b_j^{(t-1)}) - \psi(a_j^{(t-1)} + b_j^{(t-1)})$

$t_3 = \psi(\alpha_j^{(t-1)}) - \psi(\Sigma_k \alpha_k^{(t-1)})$.

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

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

$$\alpha_j^{(t)} = \alpha + n_j^{(t)}.$$

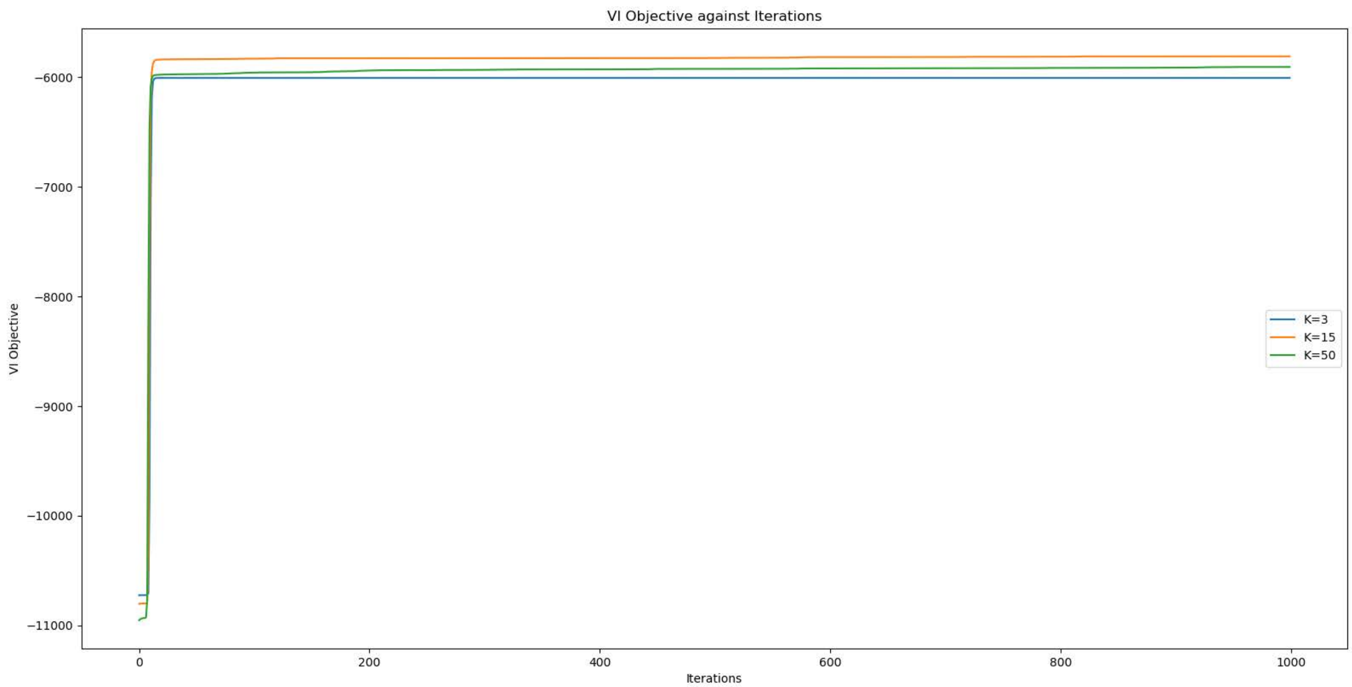
for $j=1, \dots, K$

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

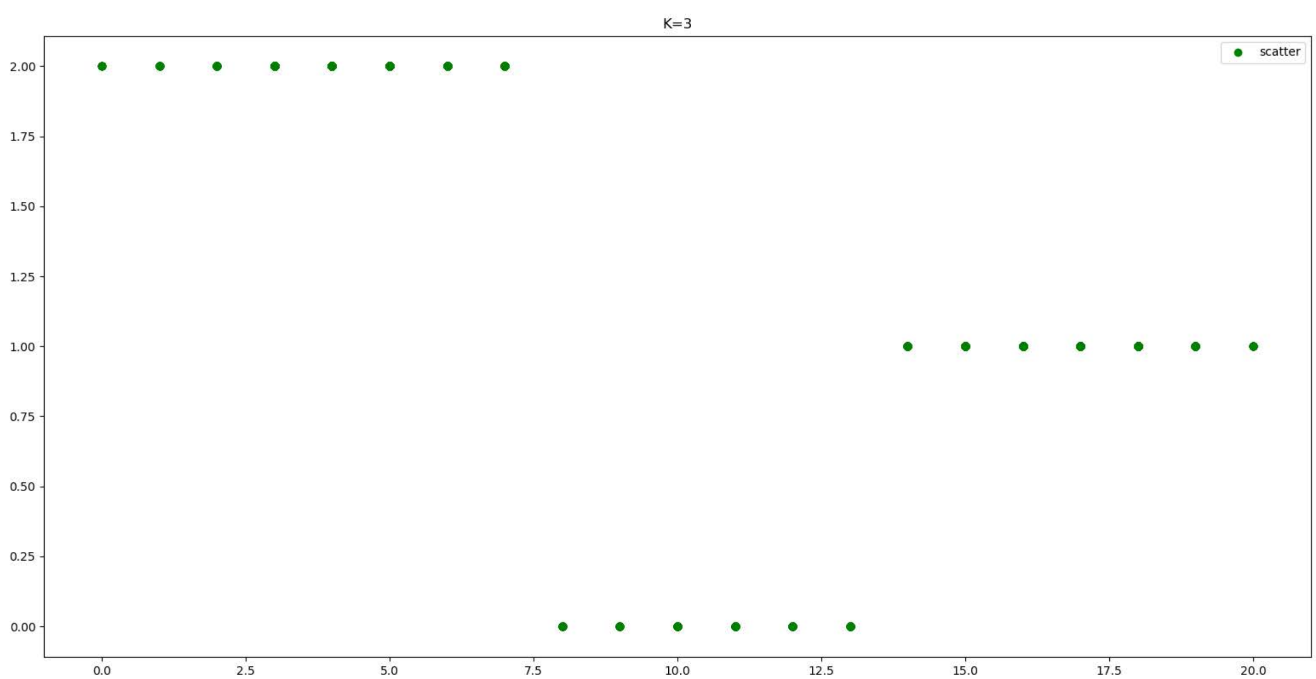
$$\cancel{a^{(t-1)} =} \; a^{(t)} = \sum_{i=1}^{n} \phi_i^{(t)}(j)\, x_i + a, \quad b^{(t)} = \sum_{i=1}^{n} \phi_i^{(t)}(j)\,(20 - x_i) + b.$$

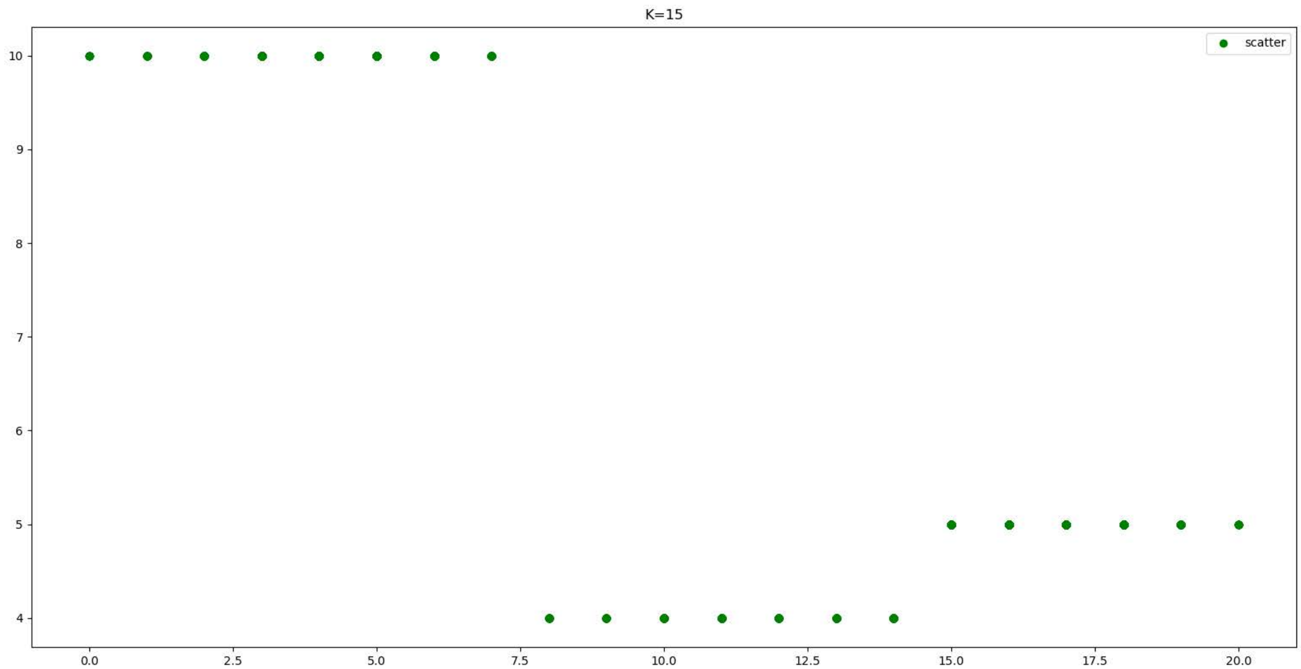
(e) Calculate $\mathcal{L} = \mathbb{E}[\ln p(x, \pi, c, \theta)] - \mathbb{E}[\ln q]$

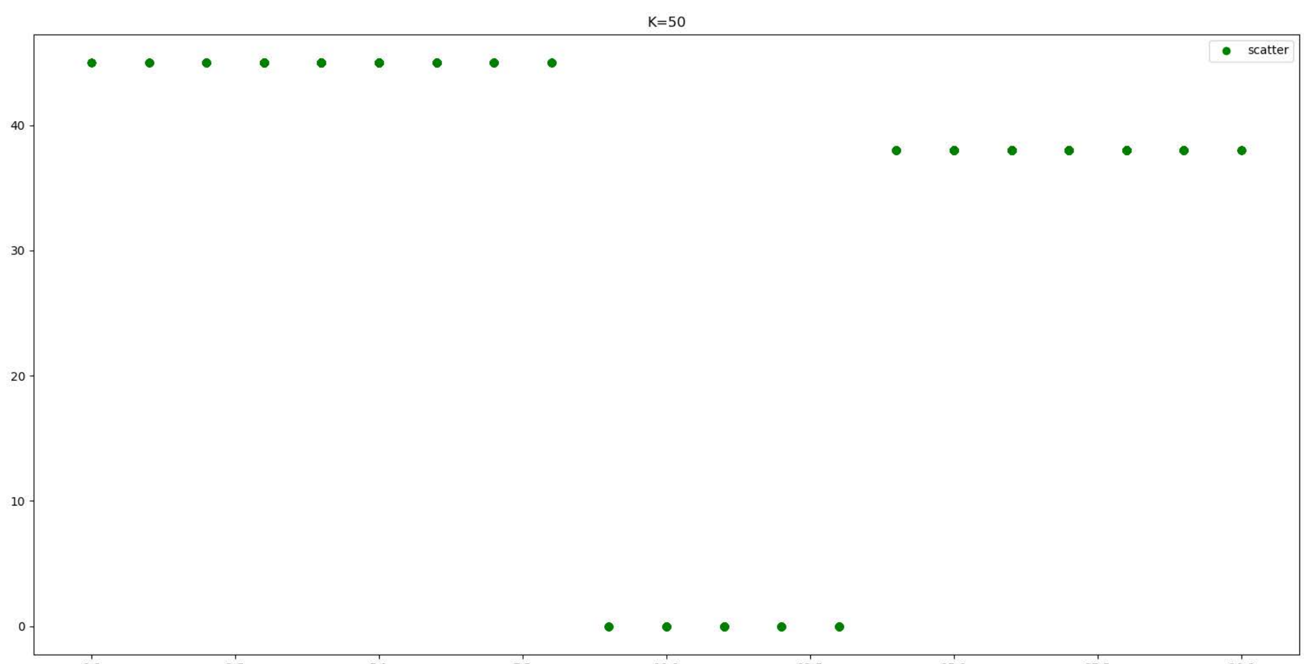
b.


VI Objective against Iterations
−6000
−7000
−8000
−9000
−10000
−11000
VI Objective
0
200
400
600
800
1000
Iterations
K=3
K=15
K=50

c.


K=3
scatter
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0


K=15
scatter
10
9
8
7
6
5
4
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0


K=50
scatter
40
30
20
10
0
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0

Problem 3.

$$p(\theta \mid \{x_i : c_i = j\}) \propto \left[\prod_{i=1}^{n} p(x_i \mid \theta_j)^{\mathbb{1}(c_i=j)}\right] p(\theta_j).$$

$$\propto \theta_j^{\sum_{i=1}^{n} \mathbb{1}(c_i=j) x_i} (1-\theta_j)^{\sum_{i=1}^{n} \mathbb{1}(c_i=j)(20-x_i)} \theta_j^{a-1} (1-\theta_j)^{b-1}$$

$$\propto \theta_j^{a+\sum_{i=1}^{n} \mathbb{1}(c_i=j) x_i - 1} (1-\theta_j)^{b+\sum_{i=1}^{n} \mathbb{1}(c_i=j)(20-x_i) - 1}.$$

$$= \mathrm{Beta}\left(a+\sum_{i=1}^{n} \mathbb{1}(c_i=j) x_i, \ b+\sum_{i=1}^{n} \mathbb{1}(c_i=j)(20-x_i)\right).$$

$$p(c_i = j \mid x_i, \theta, c_{-i}) \propto p(x_i \mid \theta_j) \frac{n_j^{(-i)}}{\alpha+n-1}.$$

$$\propto \mathrm{Binomial}(x_i \mid 20, \theta_j) \cdot \frac{n_j^{(-i)}}{\alpha+n-1}.$$

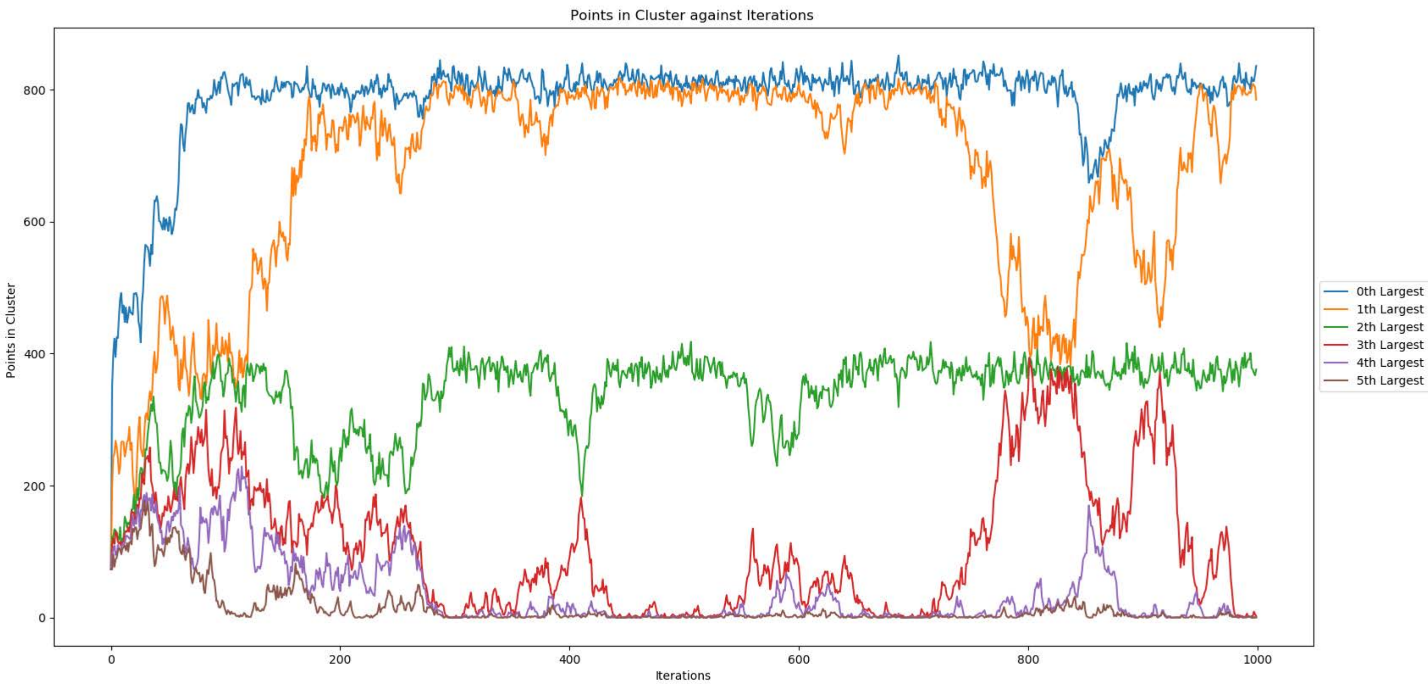
$$p(c_i = \text{new} \mid x_i, \theta, c_{-i}) \propto \frac{\alpha}{\alpha+n-1} \int \underbrace{p(x_i \mid \theta)\, p(\theta)}_{\text{Beta-Binomial distribution.}} d\theta$$

$$\propto \frac{\alpha}{\alpha+n-1} \binom{20}{x_i} \frac{B(a+x_i, b+20-x_i)}{B(a,b)}.$$

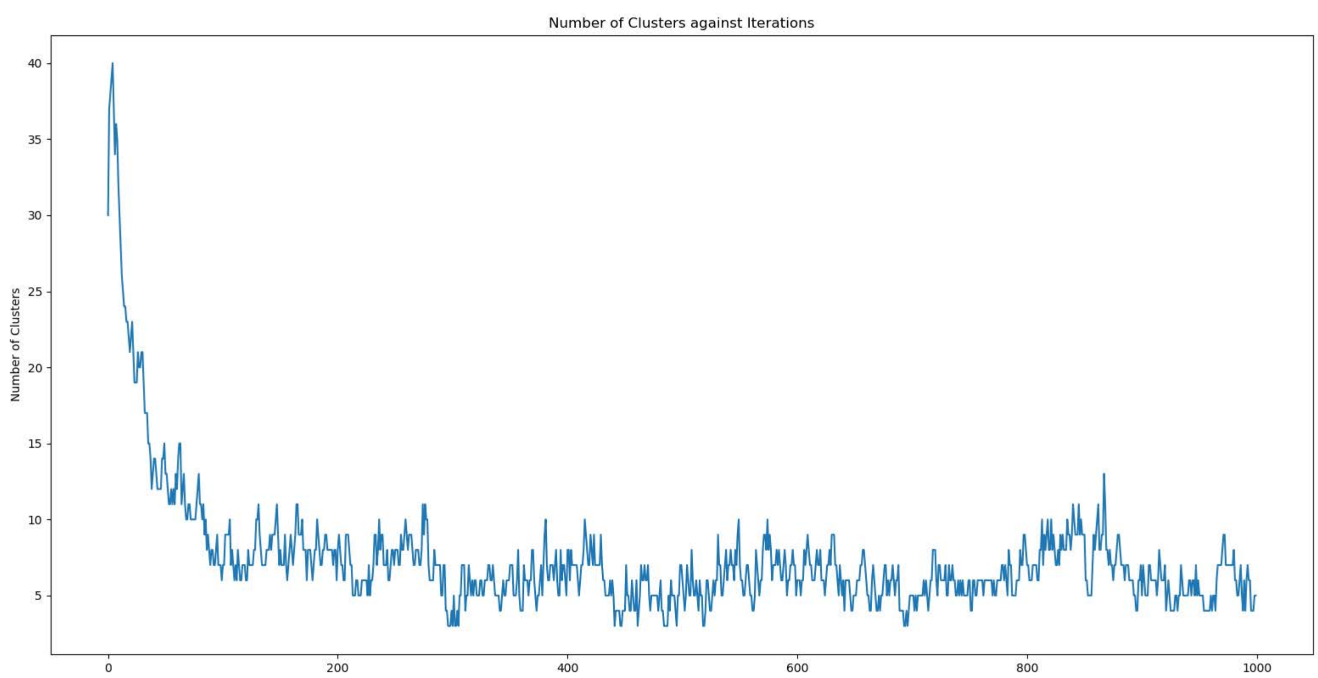
Dirichlet process ~~model~~ mixture model:

1. ~~Initer~~ Initialize in some way.
2. At iteration t, re-index clusters. Sample all variables below using the most recent values of the other values.
   1. For $i = 1, \cdots, n$.
      (a) For all j such that $n_j^{(-i)} > 0$, set:
      $$\vec{\phi}_i(j) = p(x_i \mid \theta_j)\, n_j^{(-i)} / (\alpha+n-1) = \frac{n_j^{(-i)}}{\alpha+n-1} \cdot \mathrm{Binomial}(x_i \mid 20, \theta_j).$$
      (b) For a new value j', set.
      $$\vec{\phi}_i(j') = \frac{\alpha}{\alpha+n-1} \binom{20}{x_i} \frac{B(a+x_i, b+20-x_i)}{B(a,b)}.$$
      (c). Normalize $\vec{\phi}_i$ and sample the index c_i from a discrete distribution.
      (d) If $c_i = j'$, generate $\theta_{j'} \sim \mathrm{Beta}(a+x_i, b+20-x_i)$.
   2. For $j = 1, \cdots, K^{(t)}$ generate
   $$\theta_j \sim \mathrm{Beta}\left(a+\sum_{i=1}^{n} \mathbb{1}(c_i=j) x_i, \ b+\sum_{i=1}^{n} \mathbb{1}(c_i=j)(20-x_i)\right).$$

b.


Points in Cluster against Iterations
Points in Cluster
Iterations
0th Largest
1th Largest
2th Largest
3th Largest
4th Largest
5th Largest
0
200
400
600
800
1000

c.


Number of Clusters against Iterations
Number of Clusters
Iterations
5
10
15
20
25
30
35
40
0
200
400
600
800
1000