1. Answer:

(1) E-step:

$$p(c|x,\pi,\theta) \propto p(x|c,\theta)\cdot p(c|\pi)$$
$$\propto \prod_{i=1}^{n} p(x_i|c_i,\theta_{c_i})\, p(c_i|\pi) \; ; \qquad Z_i = \sum_{j=1}^{K} p(x_i|c_i,\theta_{c_i})\, p(c_i|\pi).$$

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

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

$$\therefore q(c) = p(c|\pi,x,\theta) \Rightarrow \prod_{i=1}^{n} q(c_i) = \prod_{i=1}^{n} p(c_i|x_i,\pi,\theta_{c_i})$$

$$L(\pi,\theta) = \sum_{i=1}^{n} E_q[\ln (x_i,c_i|\pi,\theta)] + \text{const}.$$
$$= \sum_{i=1}^{n}\sum_{j=1}^{K} q(c_i=j)\,[\ln p(x_i|c_i=j,\theta) + \ln p(c_i=j|\pi)]$$
$$= \sum_{i=1}^{n}\sum_{j=1}^{K} \phi_i(j)\,\{\ln\{\binom{20}{x_i}\theta_j^{x_i}(1-\theta_j)^{20-x_i}\} + \ln\pi_j\} + \text{const},$$
$$= \sum_{i=1}^{n}\sum_{j=1}^{K} \phi_i(j)\,[\ln C_{20}^{x_i} + x_i\ln\theta_j + (20-x_i)\ln(1-\theta_j) + \ln\pi_j] + \text{const}.$$

$$\phi_i(j) = q(c_i=j) = \frac{\pi_j\cdot\binom{20}{x_i}(\theta_j)^{x_i}(20-x_i)^{1-\theta_j}}{Z_i}$$

(2). M-step:

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

$$\nabla_\pi L = 0.$$
$$\begin{cases} \sum_{i=1}^{n}\phi_i(j)\cdot\frac{1}{\pi_j} = 0 \\ \pi_j \geq 0 \\ \sum_{j=1}^{K}\pi_j = 1 \end{cases} \Rightarrow \quad \pi_j = \frac{n_j}{n}, \quad n_j = \sum_{i=1}^{n}\phi_i(j).$$

Pseudo-code:

Input: Data $x_1, \cdots, x_n$. K clusters.

Output: parameter π, θ, and cluster distributions ϕ_{ij}.

1. Initialize π_0, θ_0 in some way.
2. for iteration $t = 1, \cdots, T$.

   (a). E-step: for $i = 1, \cdots, n$; $j = 1, \cdots, K$.

$$\phi_i^{(t)}(j) = \frac{\pi_j^{(t-1)} \binom{20}{x_i} (\theta_j^{t-1})^{x_i} (20 - \theta_j^{t-1})^{1-x_i}}{\sum_{j=1}^{K} \pi_k^{t-1} \binom{20}{x_i} (\theta_k^{t-1})^{x_i} (20 - \theta_k^{t-1})^{1-x_i}}.$$

   (2). M-step:

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

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

$$\pi_j^t = \frac{n_j^t}{n}.$$

3. Calculate $L(\pi, \theta) = \ln p(x|\pi, \theta)$. for each iteration. to access convergence of object function.

2. Answer:

We first factorize $q(\pi, \theta, c) = q(\pi)\left[\prod_{k=1}^{K} q(\theta_k)\right]\left[\prod_{i=1}^{n} q(c_i)\right]$

$p(x, c, \pi, \theta) = \left[\prod_{i=1}^{n} p(x_i \mid c_i, \theta)\, p(c_i \mid \pi)\, p(\theta_{c_i})\right] p(\pi)$

$\ln p(x, c, \pi, \theta) = \sum_{i=1}^{n}\sum_{j=1}^{K} \mathbb{1}(c_i = j)\left[\ln p(x_i \mid \theta_j) + \ln(\pi_j)\right] + \sum_{j=1}^{K}\left[\ln p(\theta_j)\right] + \ln p(\pi)$.

Assume:

$q(\pi) \sim \text{Dirichlet}(\alpha_1', \cdots, \alpha_K')$.

$q(\theta_k) \sim \text{Beta}(a', b')$

$q(c_i) \sim \text{Discrete}(\pi)$

$q(c_i = j) \propto \exp\{E_q[\ln p(x, c, \pi, \theta)]\}$

$\propto \exp\{E_{-q}[\ln p(x_i \mid c_i, \theta) + \ln(c_i \mid \pi)]\}$

$\propto \exp\{E_q[\ln(C_{20}^{x_i}\theta_j^{x_i}(1-\theta_j)^{20-x_i}) + \ln \pi_j]\}$

$\propto \exp\{x_i E_q[\ln\theta_j] + (20-x_i)E_q[\ln(1-\theta_j)] + E_q[\ln \pi_j]\}$

$\propto \exp\{x_i(\psi(a_j) - \psi(a_j+b_j)) + (20-x_i)(\psi(b_j) - \psi(a_j+b_j)) + (\psi(\alpha_j) - \psi(\sum_k \alpha_k))$

$q(\theta_j) \propto \exp\{E_{-q}[\ln p(x, c, \pi, \theta)]\}$

$\propto \exp.\{E_{-q}[\ln p(x \mid c, \theta) + \ln p(\theta)]\}$

$\propto \exp\{\sum_{i=1}^{n} E_{-q}[\ln p(x_i \mid c_i, \theta_j) + \ln p(\theta_j)]\}$

$\propto \exp\{\sum_{i=1}^{n} E_{-q}[x_i \ln\theta_j + (20-x_i)\ln(1-\theta_j)]\}\ \cancel{\exp}\ p(\theta_j)$

$\propto \exp\{\sum_{i=1}^{n} E_{-q}[x_i \ln\theta_j + (\sum_{i=1}^{n}\phi_i(j)(20-x_i) + b_0 - 1)\ln(1-\theta_j)\}$

$\propto \exp\{(\sum_{i=1}^{n} x_i\phi_i(j) + a_0 - 1)\ln\theta_j + (\sum_{i=1}^{n}\phi_i(j)(20-x_i) + b_0 - 1)$.

$\propto \theta_j^{a_0 + \sum_{i=1}^{n} x_i\phi_i(j) - 1}(1-\theta_j)^{b_0 + \sum_{i=1}^{n}(20-x_i)\phi_i(j) - 1}$

$\therefore a_j' = a_0 + \sum_{i=1}^{n}\phi_i(j)x_i$; $b_j' = b_0 + \sum_{i=1}^{n}\phi_i(j)(20-x_i)$.

$q(\pi) \propto \exp\{E_q[\ln p(x, c, \pi, \theta)]\}$

$\propto \exp\{E_{-q}[\ln p(c \mid \pi) + \ln p(\pi)]\}$

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

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

$\alpha_j' = \alpha_0 + n_j$; $n_j = \sum_{i=1}^{n}\phi_i(j)$

Pesudo – code :

1. Inputs: Data and definitions $q(\theta)$, $q(\pi)$, $q(c|\pi)$.
2. Outputs: parameters a_j', b_j', α_j' for $j=1,\cdots,K$.
3. For iteration $t=1,\cdots,T$

(1) Update $q(c_i=j)$ by setting $\phi_i(j) = \exp\{x_i[\psi(a_{j,t-1}') - \psi(a_{j,t-1}'+b_{j,t-1}')] + (20-x_i)[\psi(b_{j,t-1}') - \psi(a_{j,t-1}'+b_{j,t-1}')] + \psi(\alpha_{j,t-1}') - \psi(\sum_K \alpha_{k,t-1}')\}$

(2) Update $q(\theta_j)$ by setting $a_{j,t}' = a_0 + \sum_{i=1}^{n} x_i\phi_i(j)$; $b_{j,t}' = b_0 + \sum_{i=1}^{n}(20-x_i)\phi_i(j)$

(3) Update $q(\pi)$ by setting $\alpha_{j,t}' = \alpha_0 + n_j$, $n_j = \sum_{i=1}^{n}\phi_i(j)$, $t+1$.

4. Evaluate $L(c, a_i', b_i', \alpha_i')$ to access convergence.

$$L = E_q[\ln p(x,c,\pi,\theta) - \ln q(c,\pi,\theta)]$$

$$= E_q[\sum_{i=1}^{n}\sum_{k=1}^{K}(\ln\binom{20}{x_i} + x_i\ln\theta_j + (20-x_i)\ln(1-\theta_j)]$$

$$+ E_q[\sum_{i=1}^{n}\sum_{j=1}^{K}\ln p(c_i=j|\pi)]$$

$$+ E_q[\ln p(\pi)]$$

$$+ E_q[\sum_{j=1}^{K}\ln p(\theta_j)]$$

$$- E_q[\sum_{i=1}^{n}\sum_{k=1}^{K}\ln\phi_i(j)]$$

$$- E_q[\sum_{j=1}^{K}\ln q(\theta_j)]$$

$$- E_q[\ln q(\pi)]$$

$$\Rightarrow L = \sum_{i=1}^{n}\sum_{j=1}^{K}\phi_i(j)[x_i(\psi(a_{j,t}') - \psi(a_{j,t}'+b_{j,t}')) + (20-x_i)(\psi(b_{j,t}') - \psi(b_{j,t}'+a_{j,t}'))]$$

$$+ \sum_{i=1}^{n}\sum_{j=1}^{K}\phi_i(j)[\psi(\alpha_{j,t}') - \psi(\sum_K \alpha_{k,t}')]$$

$$+ (\alpha_0-1)\sum_{j=1}^{K}[\psi(\alpha_{j,t}') - \psi(\sum_K \alpha_{i,t}')]$$

$$+ \sum_{j=1}^{K}\{(a_0-1)[\psi(a_{j,t}') - \psi(a_{j,t}'+b_{j,t}')] + (b_0-1)(\psi(b_{j,t}') - \psi(a_{j,t}'+b_{j,t}')]$$

$$- \sum_{i=1}^{n}\sum_{j=1}^{K}\phi_i(j)\ln\phi_i(j)$$

$$- \sum_{j=1}^{K}\{(a_{j,t}'-1)[\psi(a_{j,t}') - \psi(a_{j,t}'+b_{j,t}')] + (b_{j,t}'-1)[\psi(a_{j,t}'+b_{j,t}')] - \ln\frac{\Gamma(a_{j,t}')\Gamma(b_{j,t}')}{\Gamma(a_{j,t}'+b_{j,t}')}\}$$

$$- \sum_{j=1}^{K}(\alpha_{j,t}'-1)[\psi(\alpha_{j,t}') - \psi(\sum_K \alpha_{k,t}')]$$ -ln(B(a))

3. Answer:

(1). Sampling θ:

①. $p(\theta_j \mid x, c, \theta_{-j}) \propto p(x \mid \theta_j, c)\, p(\theta_j)$

$$\propto \left[\prod_{i=1}^{n} p(x_i \mid \theta_j)^{\mathbb{1}(c_i=j)}\right] p(\theta_j) \quad \text{for} \quad n_j > 0.$$

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

② if $n_j < 0$, $p(\theta_j \mid x, c, \theta_{-j}) = \text{prior } p(\theta_j)$.

(2) Sampling c:

$$p(c_i = j \mid x, \theta, c_{-i}) \propto p(x \mid c_i = j, \theta)\, p(c_i = j \mid c_{-i})$$

① For $n_j^{-i} > 0$, $p(c_i = j \mid x, \theta, c_{-i}) \propto p(x_i \mid \theta_j) \dfrac{n_j^{-(i)}}{\alpha + n - 1}$. $\quad (K \to \infty,\ \frac{\alpha}{K} \to 0)$.

② for $n_j^{-i} = 0$, $p(c_i = \text{new} \mid x, \theta, c_{-i}) = \sum_{j: n_j^{-i} = 0} p(c_i = j \mid x, \theta, c_{-i})$

$$\propto \lim_{K \to \infty} \sum_{j: n_j^{-i} = 0} p(x_i \mid \theta_j) \frac{\alpha / K}{\alpha + n - 1}$$

$$\propto \lim_{K \to \infty} \sum_{j: n_j^{-i} = 0} \frac{\alpha}{\alpha + n - 1} \cdot \frac{p(x_i \mid \theta_j)}{K} = \frac{\alpha}{\alpha + n - 1} \underbrace{\int p(x_i \mid \theta_j)\, p(\theta)\, d\theta}_{E[p(x_i \mid \theta_j)]}.$$

$$\therefore c_i = \begin{cases} j & \text{w.p.} \propto p(x_i \mid \theta_j) \dfrac{n_j^{-i}}{\alpha + n - 1} \quad \text{if } n_j^{-i} > 0 \\ \text{new} & \text{w.p.} \propto \dfrac{\alpha}{\alpha + n - 1} \displaystyle\int p(x_i \mid \theta_j)\, p(\theta)\, d\theta. \quad \text{if } n_j^{-i} = 0. \end{cases}$$

Pseudo-code:

Input: Data $x_1, \ldots, x_n$.

Output: K, parameter θ, ϕ_{ij}.

1. Initialize in some way. (We can set $c_i = 1$ for $i = 1, \ldots, n$), $\theta_1 \sim p(\theta)$.
2. For iteration $t = 1, \ldots, T$, re-index cluster go from 1 to $K^{(t-1)}$.

(a). for $i = 1, \ldots, n$:

For all j that $n_j^{-i} > 0$,

$$\phi_i(j) = p(x_i \mid \theta_j)\, n_j^{-i} / (\alpha + n + 1).$$

For ~~all~~ a new value j',

$$\phi(j') = \frac{\alpha}{\alpha + n + 1} \int p(x_i \mid \theta)\, p(\theta)\, d\theta.$$

(b). Normalize ϕ_i, and sample the index c_i from a discrete distribution with the parameter.

(c). If $c_i = j'$, generate $\theta_{j'} \sim p(\theta \mid x_i)$.

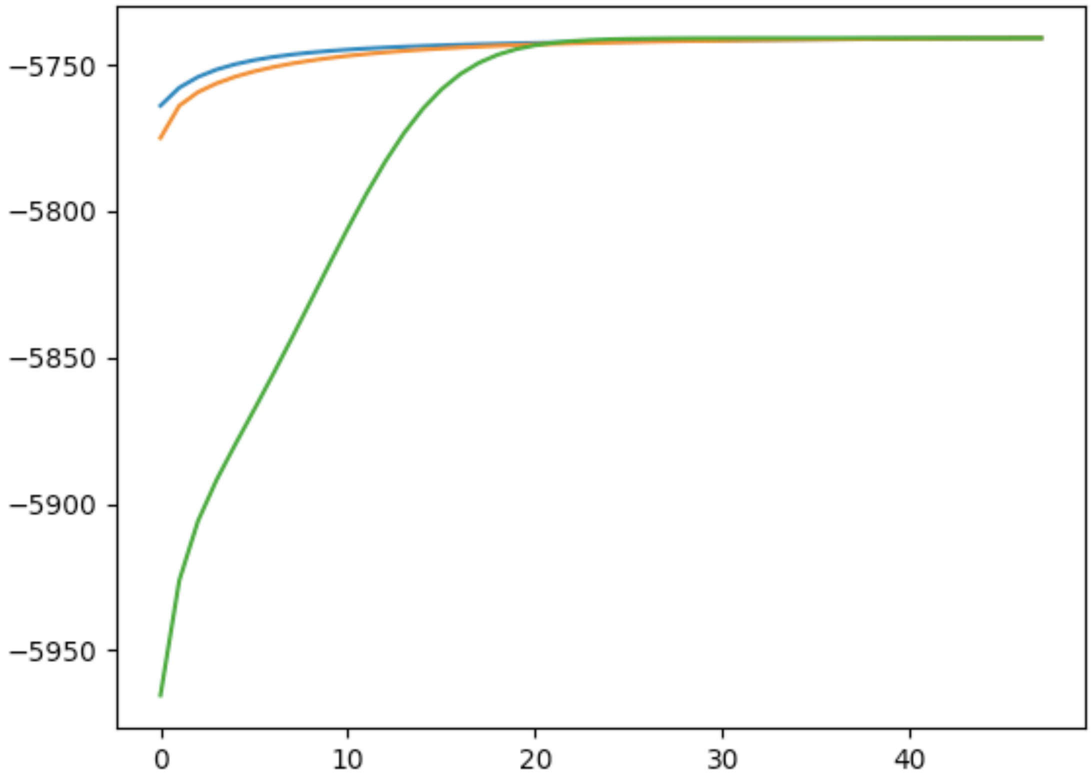
3. For $j = 1, \ldots, K^{(t)}$ generate

$$\theta_j \sim p(\theta \mid \{x_i : c_i = j\}).$$

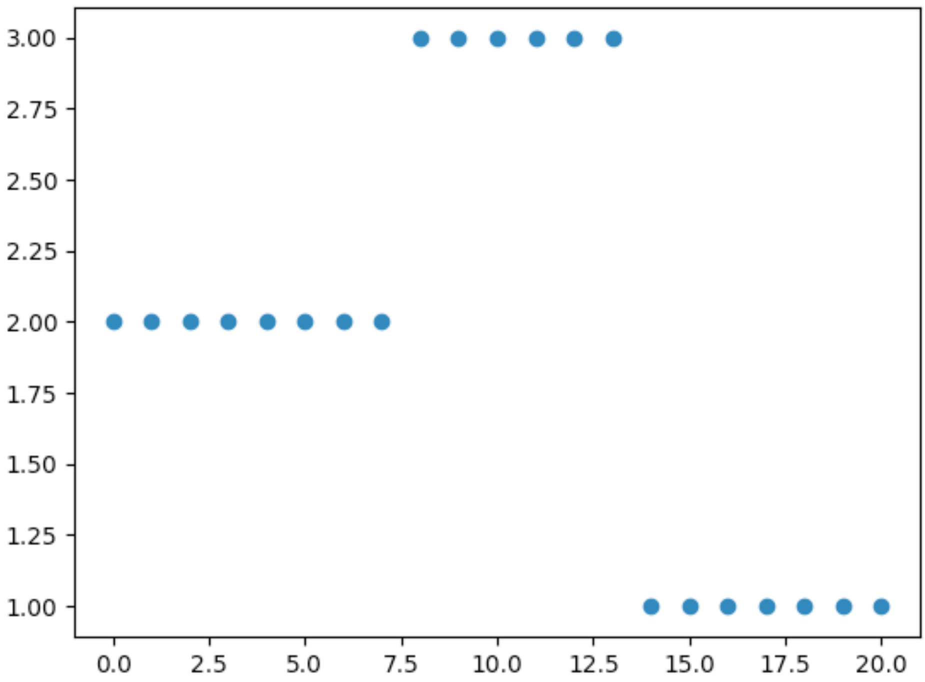
(K_t = # non-zero clusters that are re-indexed after completing 2)

1. EM_Algorithms:

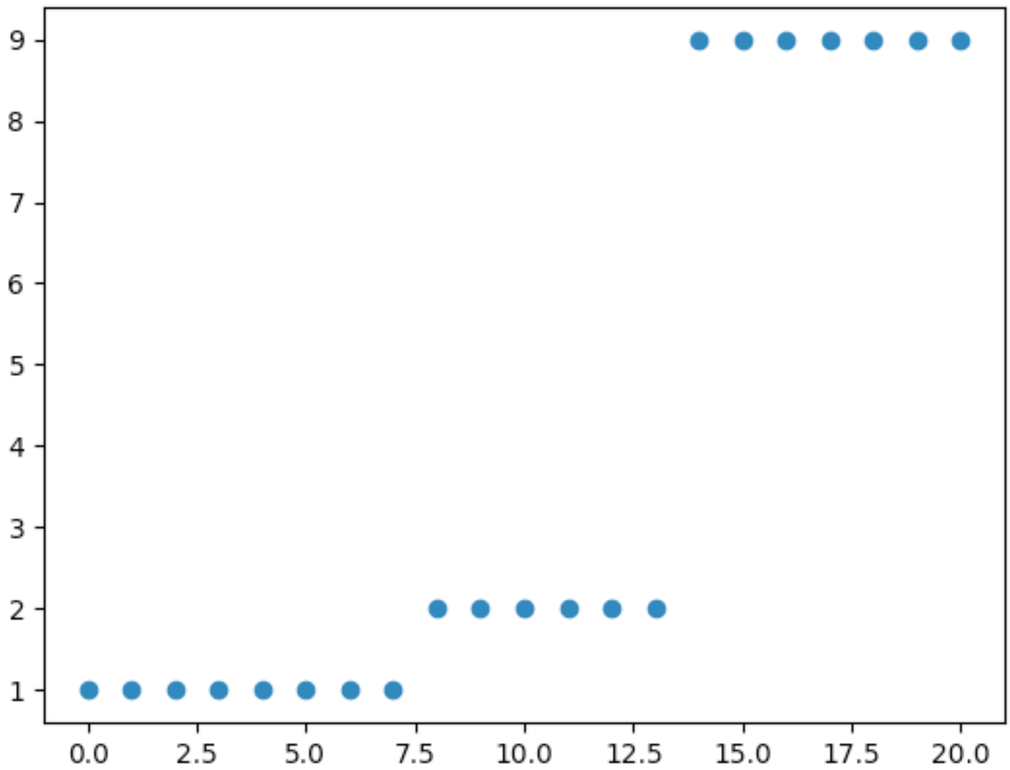
Blue: K = 15; Orange: K = 9; Green: K = 3



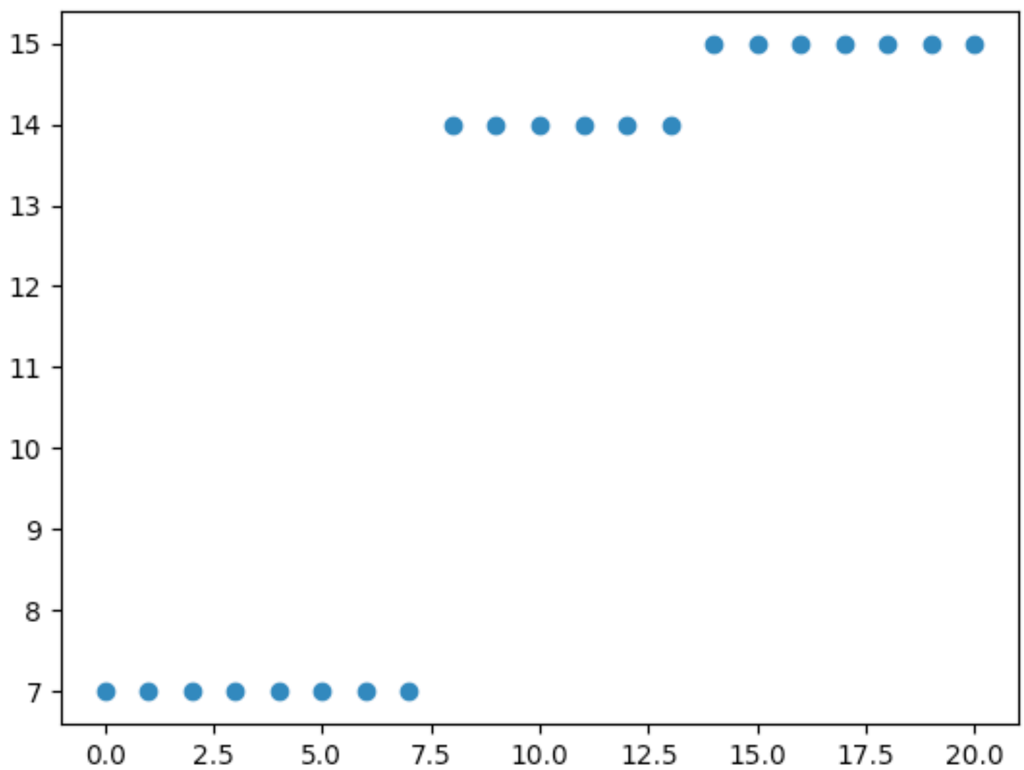
K = 3:



K = 9:

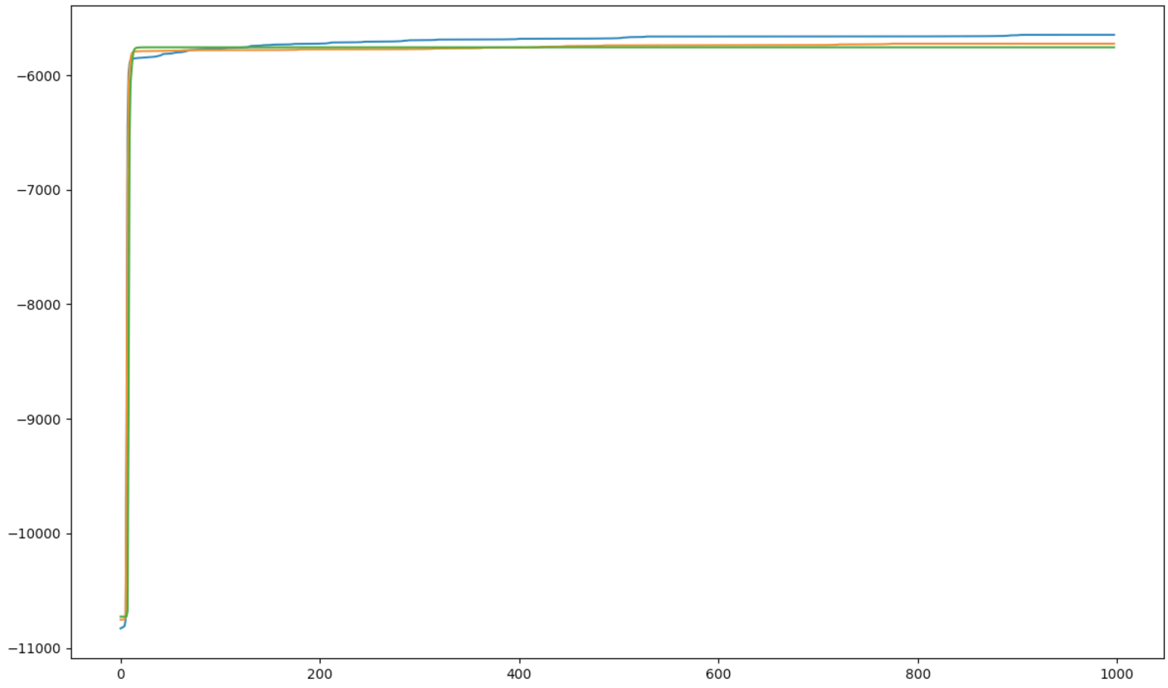

9
8
7
6
5
4
3
2
1
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0

K = 15:

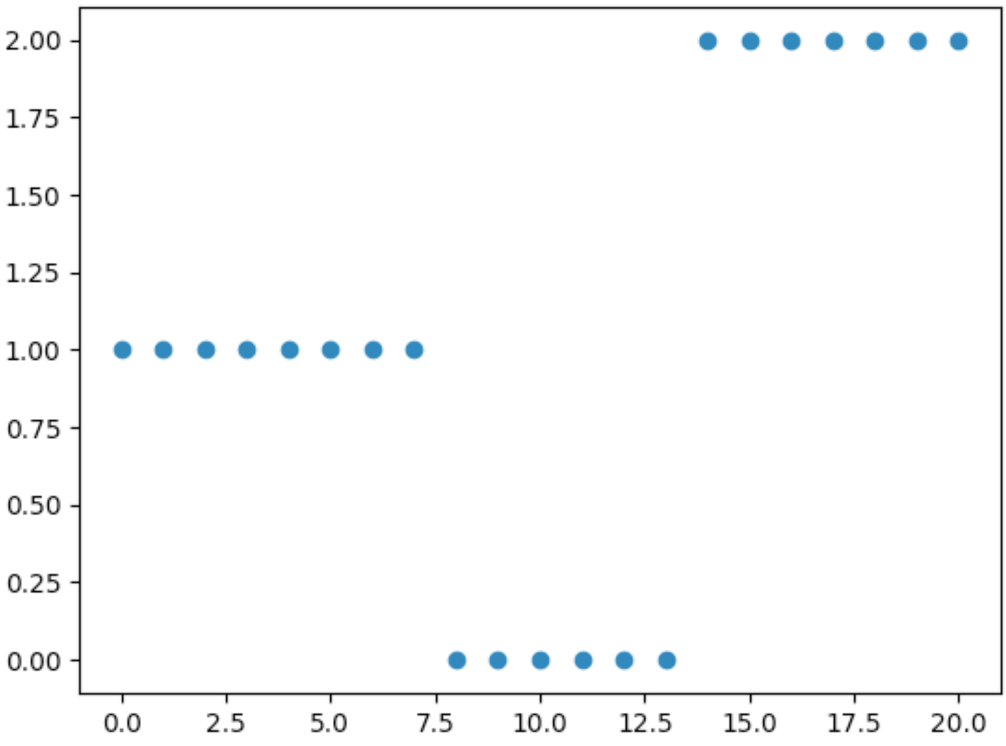

15
14
13
12
11
10
9
8
7
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0

2. VI_Algorithms:

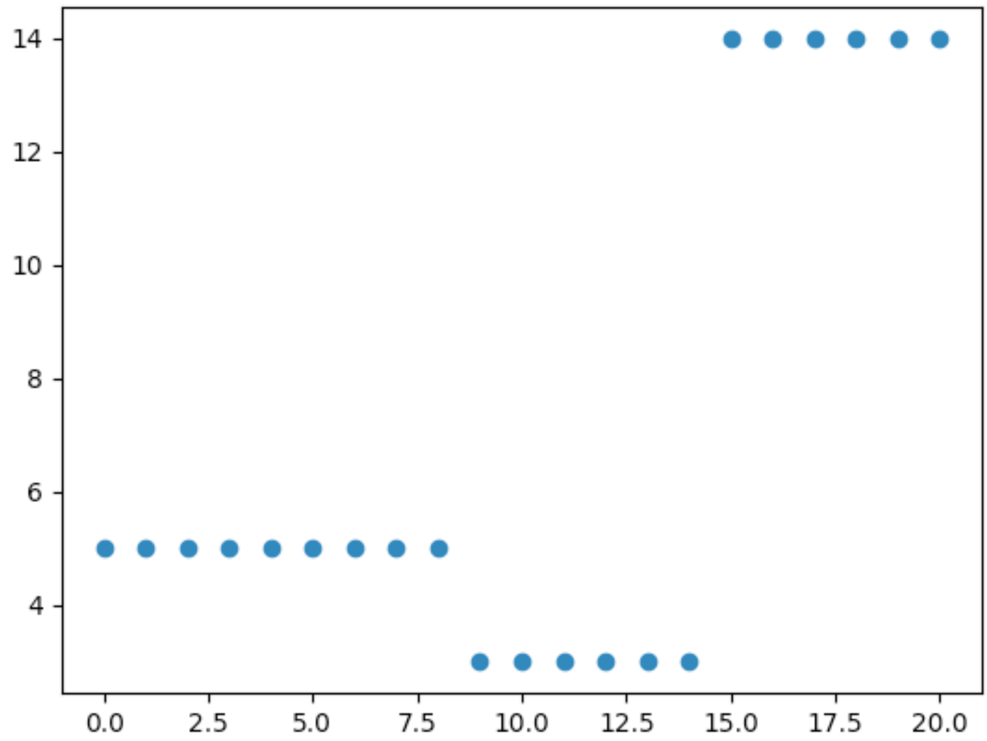
Blue: K = 50; Orange: K = 15; Green: K = 3



K = 3:

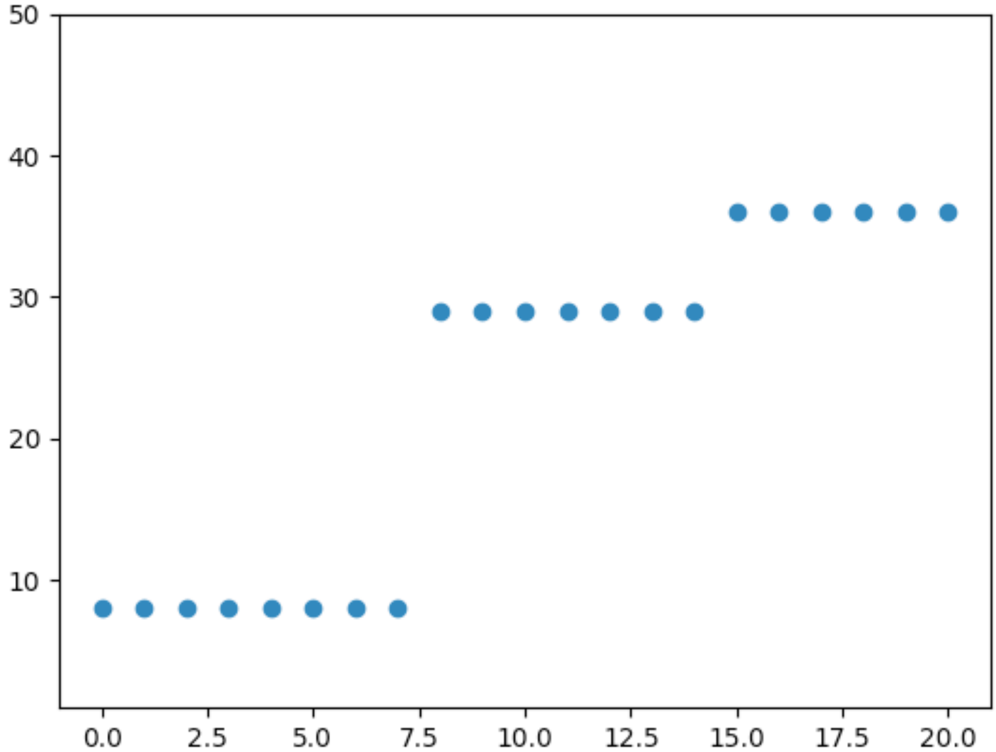


K = 15:



14
12
10
8
6
4
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0

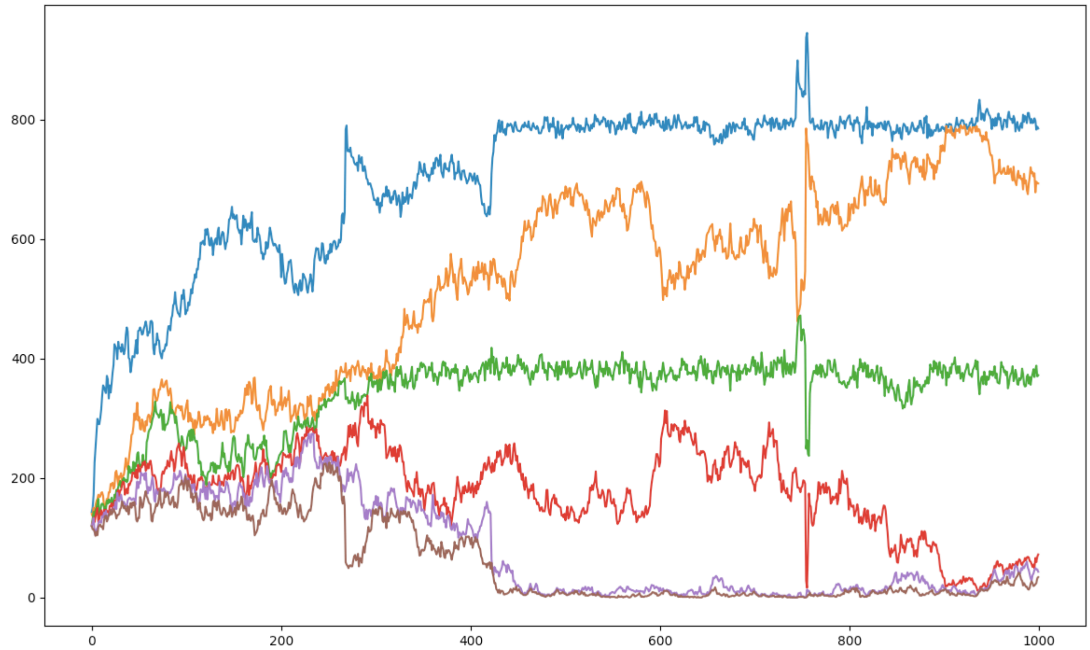
K = 50:



50
40
30
20
10
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0

3. Gibbs Sampling:

six most probable clusters



Total number of clusters for each iteration