Kate Lassiter

You are given observations $X = \{x_1, \ldots, x_n\}$ where each $x_i \in \{0, 1, 2, \ldots, 20\}$. You model this as being generated from a Binomial mixture model of the following form:

$$x_i \mid c_i \sim \text{Binomial}(20, \theta_{c_i}), \qquad c_i \overset{iid}{\sim} \text{Discrete}(\pi).$$

In this homework, you will implement two algorithms for learning this mixture model, one using maximum likelihood EM and one using variational inference. Use the data provided.

Problem 1. (35 points)

In this problem, you will derive and implement the EM algorithm for learning maximum likelihood values of π and each $\theta_k \in [0, 1]$ for $k = 1, \ldots, K$. Use c as the model variable integrated out. In your algorithm, initialize π as uniform and each θ_k so that they are evenly spaced between $[0, 1]$.

a) Show your derivation of the E and M steps of the algorithm. You can present your solution for π at the same level of detail as in the notes. Summarize the algorithm in pseudo-code. Since you will need to use the log marginal likelihood in part (b), write that out here.

Problem 1

Recall:

$$binomial \sim p(x = i) = \binom{n}{i} p^i (1 - p)^{n-i}$$

Thus:

$$x_i \sim_{iid} p(x_i|\pi, \boldsymbol{\Theta}) = \sum_{j=1}^{K} \pi_j \, Binomial\,(x_i|20, \theta_j)$$

$$p(x_1, \ldots, x_n|\pi, \boldsymbol{\Theta}) = \prod_{i=1}^{n} (\sum_{j=1}^{K} \pi_j \binom{20}{x_i} \theta_j^{x_i}(1 - \theta_j)^{20-x_i})$$

$$\prod_{i=1}^{n} p(x_i|\pi, \boldsymbol{\Theta}) = \prod_{i=1}^{n} \sum_{j=1}^{K} \pi_j \, Binomial\,(x_i|20, \theta_j)$$

$$\prod_{i=1}^{n} p(x_i, c_i|\pi, \boldsymbol{\Theta}) = \prod_{i=1}^{n} p(x_i|c_i, \boldsymbol{\Theta})p(c_i|\pi)$$

$$\prod_{i=1}^{n} p(x_i, c_i|\pi, \boldsymbol{\Theta}) = \prod_{i=1}^{n} \prod_{j=1}^{K} (\pi_j Binomial\,(x_i|20, \theta_j))^{1(c_i=j)}$$

$$p(x\ |\pi, \boldsymbol{\Theta}) = \sum_{c_1=1}^{K} \ldots \sum_{c_n=1}^{K} \prod_{i=1}^{n} p(x_i, c_i|\pi, \boldsymbol{\Theta}) = \prod_{i=1}^{n} \prod_{j=1}^{K} p(x_i, c_i = j|\pi, \boldsymbol{\Theta})$$

EM equation:

$$\ln p(x|\pi, \boldsymbol{\Theta}) = \sum_{c} q(\mathbf{C}) ln \frac{p(x, \mathbf{C}|\pi, \boldsymbol{\Theta})}{q(\mathbf{C})} + \sum_{c} q(\mathbf{C}) ln \frac{q(\mathbf{C})}{p(\mathbf{C}|x, \pi, \boldsymbol{\Theta})}$$

E – step

Set $q(\mathbf{C}) = p(\mathbf{C}|x, \pi, \boldsymbol{\Theta})$

$$p(\mathbf{C}|x, \pi, \boldsymbol{\Theta}) \propto p(x|\mathbf{C}, \boldsymbol{\Theta})p(\mathbf{C}|\pi)$$

Kate Lassiter

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

$Z = \prod_{i=1}^{n} Z_i \quad \rightarrow$ *Divide by a number Z to normalize the entire function of c by normalizing each individual prior likelihood pair*

$p(\mathbf{C}|x,\pi,\boldsymbol{\theta}) \propto \frac{1}{Z}\prod_{i=1}^{n} p(x_i|c_i, \boldsymbol{\theta})p(c_i|\pi)$

$p(\mathbf{C}|x,\pi,\boldsymbol{\theta}) \propto \frac{\prod_{i=1}^{n} p(x_i|c_i,\boldsymbol{\theta})p(c_i|\pi)}{Z_i}$

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

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

$p(c_i = k|x_i, \pi,\boldsymbol{\theta}) = \frac{\pi_k Binomial\ (x_i|20,\theta_k)}{\sum_{j=1}^{K} \pi_j Binomial\ (x_i|20,\theta_j)}$

$q(\mathbf{C}) = p(\mathbf{C}|x,\pi,\boldsymbol{\theta}) = \prod_{i=1}^{n} p(c_i|x_i, \pi, \boldsymbol{\theta}) = \prod_{i=1}^{n} q(c_i)$

* Let $q(c_i = j) = \phi_i(j)$

$L(\pi, \boldsymbol{\theta}) = \sum_{i=1}^{n} E_{q(c)}[ln\, p(x_i, c_i|\pi, \boldsymbol{\theta})] + constant$

$L(\pi, \boldsymbol{\theta}) = \sum_{i=1}^{n} E_{q(c_i)}[ln\, p(x_i, c_i|\pi, \boldsymbol{\theta})] + constant$

Recall:

$\prod_{i=1}^{n} p(x_i|c_i, \boldsymbol{\theta})p(c_i|\pi)$

Thus:

$L(\pi, \boldsymbol{\theta}) = \prod_{i=1}^{n}\prod_{j=1}^{K} E_{q(c_i)}[ln(\pi_j\, Binomial\ (x_i|20, \theta_j))^{1(c_i=j)}] + Constant$

$L(\pi, \boldsymbol{\theta}) = \prod_{i=1}^{n}\prod_{j=1}^{K} E_{q(c_i)}[\, ln(\pi_j \binom{20}{x_i} \theta_j^{x_i}(1-\theta_j)^{20-x_i})^{1(c_i=j)})]$+ Constant

$L(\pi, \boldsymbol{\theta}) = \sum_{i=1}^{n}\sum_{j=1}^{K} E_{q(c_i)}[1(c_i = j)ln(\pi_j \binom{20}{x_i} \theta_j^{x_i}(1-\theta_j)^{20-x_i} \quad)]$+ Constant

$L(\pi, \boldsymbol{\theta}) = \sum_{i=1}^{n}\sum_{j=1}^{K} E_{q(c_i)}[1(c_i = j)[ln(\pi_j) + ln \binom{20}{x_i} + x_i ln\theta_j + (20 - x_i)ln(1-\theta_j) \quad]$+ Constant

$L(\pi, \boldsymbol{\theta}) = \sum_{i=1}^{n}\sum_{j=1}^{K} E_{q(c_i)}[1(c_i = j)][ln(\pi_j) + x_i ln\theta_j + (20 - x_i)ln(1-\theta_j)\ )]$+ Constant

Recall:

Bernoulli random variable: p(a=0) =1-p p(a=1) =p

E[I] = 1*p(a) +0*p(a^C) =p(a)

Kate Lassiter

Thus:

$1(c_i = j) = p(c_i) = q(c_i) = \phi_i(j) = p(c_i = k|x_i, \pi, \boldsymbol{\theta}) = \frac{\pi_k Binomial(x_i|20,\theta_k)}{\sum_{j=1}^{K} \pi_j Binomial(x_i|20,\theta_j)} = E[1(c_i = j)]$

$$L(\pi, \boldsymbol{\theta}) = \sum_{i=1}^{n}\sum_{j=1}^{K} E_{q(c_i)}[1(c_i = j)][ln(\pi_j) + x_i ln\theta_j + (20 - x_i)ln(1 - \theta_j)] + \text{Constant}$$

$$L(\pi, \boldsymbol{\theta}) = \sum_{i=1}^{n}\sum_{j=1}^{K} q(c_i = j)[ln(\pi_j) + x_i ln\theta_j + (20 - x_i)ln(1 - \theta_j)) + Constant$$

$$\boxed{L(\pi, \boldsymbol{\theta}) = \sum_{i=1}^{n}\sum_{j=1}^{K} \phi_i(j)\,(ln(\pi_j) + x_i ln\theta_j + (20 - x_i)ln(1 - \theta_j)) + Constant}$$

M – step

Solving for θ_j

$$\frac{d}{d\theta_j}L(\pi, \boldsymbol{\theta}) = \sum_{i=1}^{n} \phi_i(j)\,(ln(\pi_j) + x_i ln\theta_j + (20 - x_i)ln(1 - \theta_j)) + Constant$$

Recall:

$$\frac{d}{dx}[f(g(x))] = f'(g(x))g'(x)$$

$$\frac{d}{d\theta_j}L(\pi, \boldsymbol{\theta}) = \sum_{i=1}^{n} \phi_i(j)\,(\frac{x_i}{\theta_j} - (20 - x_i)\frac{1}{(1-\theta_j)}) + Constant$$

$$\frac{d}{d\theta_j}L(\pi, \boldsymbol{\theta}) = \sum_{i=1}^{n} \phi_i(j)\,(\frac{x_i}{\theta_j} - \frac{(20-x_i)}{(1-\theta_j)}) + Constant$$

$$\frac{d}{d\theta_j}L(\pi, \boldsymbol{\theta}) = 0$$

$$0 = \sum_{i=1}^{n} \phi_i(j)\,(\frac{x_i}{\theta_j} - \frac{(20-x_i)}{(1-\theta_j)})$$

$$0 = \sum_{i=1}^{n} \phi_i(j)\frac{x_i}{\theta_j} - \sum_{i=1}^{n} \phi_i(j)\frac{(20-x_i)}{(1-\theta_j)}$$

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

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

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

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

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

Kate Lassiter

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

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

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

Solving for π_j

$$\frac{d}{d\pi_j}L(\pi,\mathbf{\Theta}) = \sum_{i=1}^{n}\phi_i(j)\,(ln(\pi_j) + x_i ln\theta_j + (20 - x_i)ln(1-\theta_j)) + Constant$$

$$\frac{d}{d\pi_j}L(\pi,\mathbf{\Theta}) = \sum_{i=1}^{n}\phi_i(j)\,(ln(\pi_j)) + Constant$$

$$\frac{d}{d\pi_j}L(\pi,\mathbf{\Theta}) = \sum_{i=1}^{n}\phi_i(j)\,(\frac{1}{\pi_j}) + Constant$$

$$0 = \sum_{i=1}^{n}\phi_i(j)\,(\frac{1}{\pi_j}) + Constant$$

$$0 = \sum_{i=1}^{n}\phi_i(j)\,(\frac{1}{\pi_j}) + Constant$$

subject to $\pi_j \geq 0 \; and \; \sum_{j=1}^{K}\pi_j = 1$

Recall: use Lagrange multipliers

$$\boxed{\pi_j = \frac{\sum_{i=1}^{n}\phi_i(j)}{n}}$$

An Em Algorithm For Binomial Mixture Model

Input: Data $x_1,...x_n$, $x \in R^d$. Number of clusters K.
Output: BMM parameters π, $\mathbf{\Theta}$ and cluster assignment distributions ϕ_i

1. Initialize $\pi^{(0)} \sim Uniform\,(0,1)$ $and\ each$ $\theta_j^{(0)} \sim Uniform\,(0,1)$
2. For iteration t =1, ...,T
   a. E – step: calculate the matrix $\mathbf{E}_{qt}[\phi] = (\mathbf{E}_{qt}[\phi_{1j}],..,\mathbf{E}_{qt}[\phi_{N1j}],..,\mathbf{E}_{qt}[\phi_{i1}],...,\mathbf{E}_{qt}[\phi_{iN2}])$. For i = 1,...n and j = 1,..., K set:
      i. $$\phi_i^{(t)}(j) = \frac{\pi_j^{(t-1)} Binomial\,(x_i|20,\theta_j^{(t-1)})}{\sum_{k=1}^{K}\pi_k^{(t-1)} Binomial\,(x_i|20,\theta_k^{(t-1)})} = \frac{\pi_j^{(t-1)}\theta_{(t-1),j}^{x_i}(1-\theta_{(t-1),j})^{20-x_i}}{\sum_{k=1}^{K}\pi_k^{(t-1)}\theta_{(t-1),k}^{x_i}(1-\theta_{(t-1),k})^{20-x_i}}$$
   b. M – step:

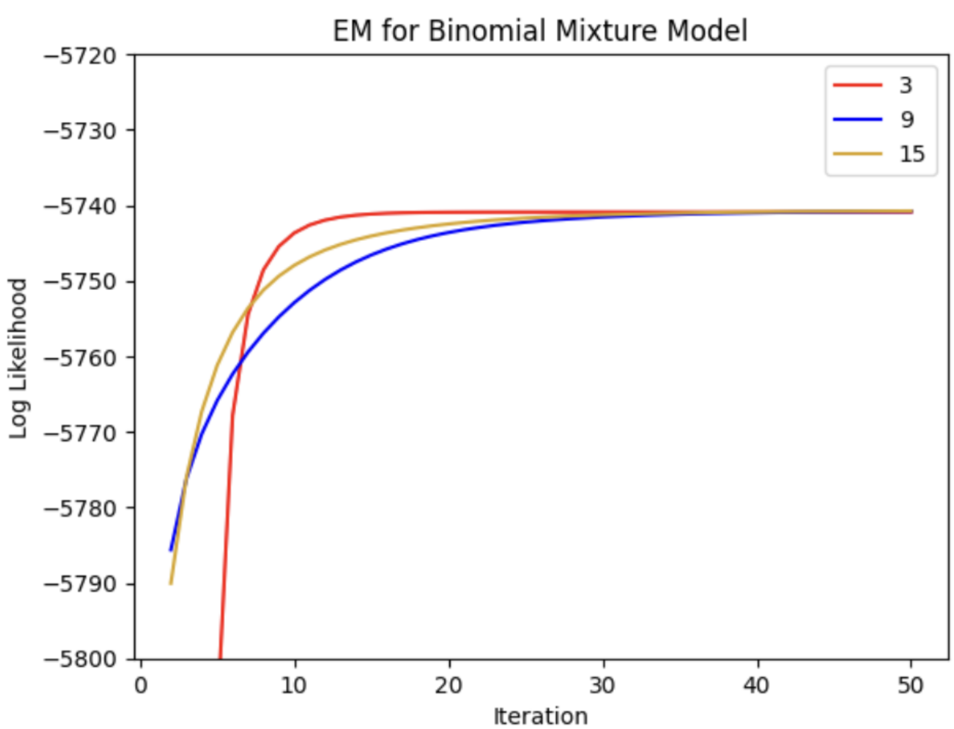
i. $\pi_j^{(t)} = \frac{\sum_{i=1}^{n} \phi_i^{(t)}(j)}{n}$

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

c. Calculate $\ln p(x|\pi^{(t)}, \boldsymbol{\theta}^{(t)})$ using the equation below to assess convergence

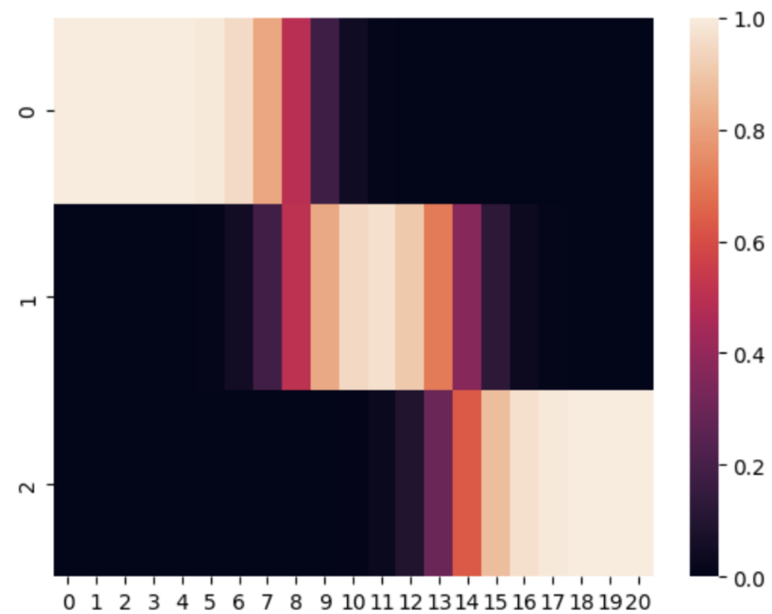
$$L(\pi, \boldsymbol{\theta}) = \sum_{i=1}^{n} \sum_{j=1}^{K} \phi_i^{(t)}(j) \left(\ln(\pi_j^{(t)}) + x_i \ln\theta_j^{(t)} + (20 - x_i) \ln(1 - \theta_j^{(t)}) \right) + constant$$

b) Implement the EM algorithm and run it for 50 iterations for $K = 3, 9, 15$. In one figure, plot the log marginal likelihood for iterations 2 to 50 for each K.

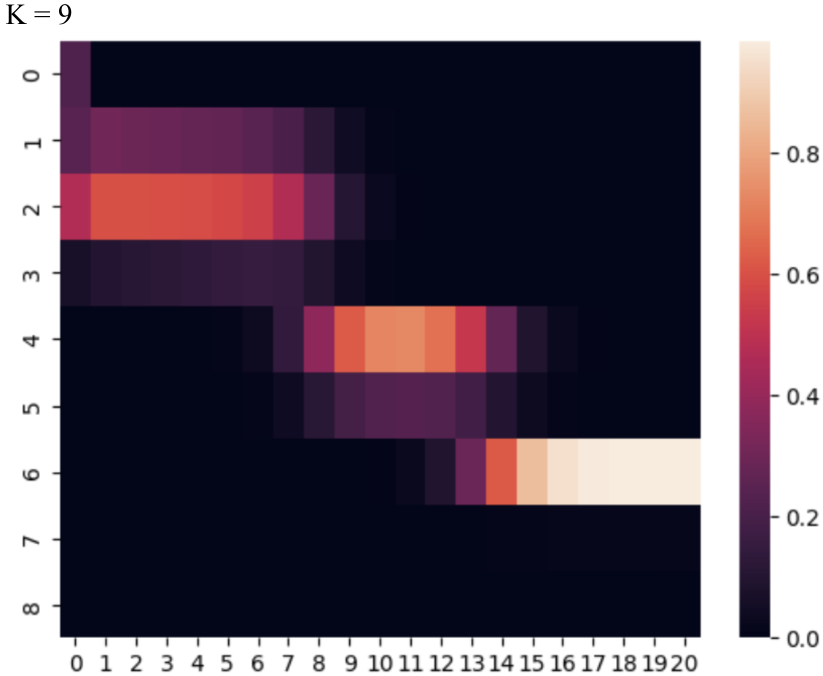


c) For the final iteration of each model show the following: Construct a $K \times 21$ matrix M_K where $M_K(k,i) = q(c=k)$ learned when $x = i-1$. You will use $p(c=k|x=i-1,\pi,\boldsymbol{\theta})$ to find this. Show this as three images with accompanying colorbar legend for $K = 3, 9, 15$. Hint: Since the data set contains instances of all 21 possible values of x, you have already calculated the 21 required q distributions during inference.

K = 3

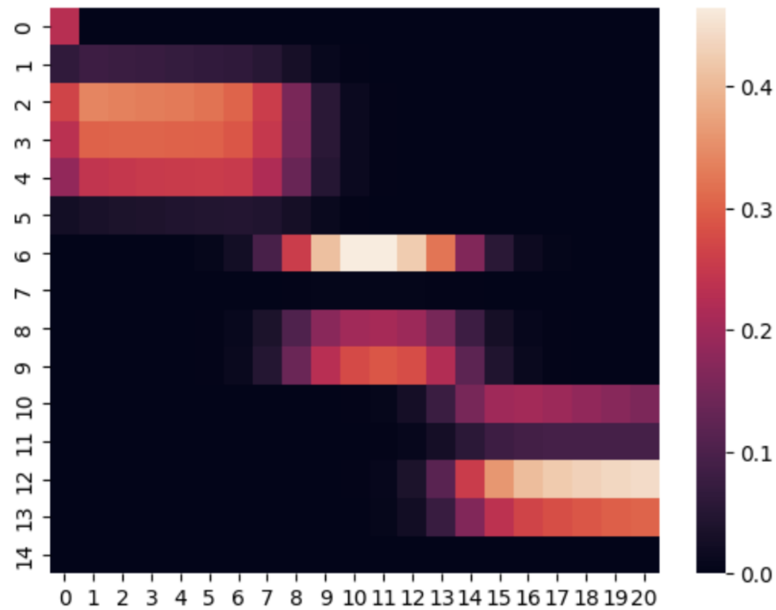


K = 9



Kate Lassiter

K = 15



Problem 2

In this problem, you will implement a variational inference algorithm for approximating the posterior distribution of the Binomial mixture model. As additional priors, let

$$\pi \sim \text{Dirichlet}(\alpha), \qquad \theta_k \overset{iid}{\sim} \text{Beta}(a, b),$$

Set $\alpha = 1/10$, $a = 0.5$ and $b = 0.5$. Approximate the full posterior distribution of π, $\boldsymbol{\theta}$ and $\mathbf{c}$ with the factorized distribution $q(\pi, \boldsymbol{\theta}, \mathbf{c}) = q(\pi)\left[\prod_{k=1}^{K} q(\theta_k)\right]\left[\prod_{i=1}^{n} q(c_i)\right]$ using variational inference.

a) Derive the optimal $q(\pi)$, $q(\theta_k)$ and $q(c_i)$. Also, derive the variational objective function, which will be used in part (b). Summarize the VI algorithm in pseudo-code.

q- Distributions

$q(\pi)$

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

$$p(x\ |\pi, \boldsymbol{\Theta}) = \sum_{c_1=1}^{K} \dots \sum_{c_n=1}^{K} \prod_{i=1}^{n} p(x_i, c_i|\pi, \boldsymbol{\Theta}) = \prod_{i=1}^{n} \prod_{j=1}^{K} \mathrm{p}(x_i, c_i = j|\pi, \boldsymbol{\Theta})$$

$$p(x\ \pi, \boldsymbol{\Theta}, \mathbf{C}) = [\sum_{c_1=1}^{K} \dots \sum_{c_n=1}^{K} \prod_{i=1}^{n} p(x_i|c_i = j, \pi, \boldsymbol{\Theta}) p(c_i = j|\pi)][\sum_{j=1}^{K} p(\theta_j)]p(\pi)$$

$$p(x\ \pi, \boldsymbol{\Theta}, \mathbf{C}) = [\prod_{i=1}^{n} \prod_{j=1}^{K} p(x_i|c_i = j, \pi, \boldsymbol{\Theta}) p(c_i = j|\pi)][\prod_{j=1}^{K} p(\theta_j)]p(\pi)$$

$$p(x\ \pi, \boldsymbol{\Theta}, \mathbf{C}) = [\prod_{i=1}^{n} \prod_{j=1}^{K} p(x_i|c_i = j, \pi_j, \theta_j) p(c_i = j|\pi)][\prod_{j=1}^{K} p(\theta_j|a_j, b_j)][\prod_{j=1}^{K} p(\pi_j|\alpha_j\ ))$$

Kate Lassiter

$$p(x\ \pi, \boldsymbol{\theta}, \mathbf{C}) = [\prod_{i=1}^{n}\prod_{j=1}^{K} (\theta_j^{x_i}(1-\theta_j)^{20-x_i})^{1(c_i=j)} \pi_j^{1(c_i=j)}][\prod_{j=1}^{K} \frac{\Gamma(a_j)+\Gamma(b_j)}{\Gamma(a_j)\Gamma(b_j)} \theta_j^{a_j-1}(1-\theta_j)^{b_j-1}][\frac{\Gamma(\sum_K \alpha_k)}{\prod_{j=1}^{K}\Gamma(\alpha_j)} \prod_{j=1}^{K} \pi_j^{\alpha_j-1}]$$

$$q(\pi) \propto e^{E_{q-\pi}[\sum_{i=1}^{n}\sum_{j=1}^{K} lnp(c_j|\pi)] + \sum_{j=1}^{K} lnp(\pi) + constant\ w.r.t\ \pi]}$$

$$q(\pi) \propto e^{E_{q-\pi}[\sum_{i=1}^{n}\sum_{j=1}^{K} ln\pi_j^{1(c_i=j)} + ln\frac{\Gamma(\sum_K \alpha_k)}{\prod_{j=1}^{K}\Gamma(\alpha_j)}\prod_{j=1}^{K}\pi_j^{\alpha-1} + constant\ w.r.t\ \pi]}$$

$$q(\pi) \propto e^{E_{q-\pi}[\sum_{i=1}^{n}\sum_{j=1}^{K} 1(c_i=j)ln\pi_j + \sum_{j=1}^{K} ln\Gamma(\sum_K \alpha_k) - ln\Gamma(\prod_{j=1}^{K}\alpha_j) + (\alpha-1)ln(\pi_j)]}$$

$$q(\pi) \propto e^{\sum_{i=1}^{n}\sum_{j=1}^{K} E_{q-\pi}[1(c_i=j)]ln\pi_j + \sum_{j=1}^{K}(E_{q-\pi}[ln\Gamma(\sum_K \alpha_k)] - E_{q-\pi}[ln\Gamma(\prod_{j=1}^{K}\alpha_j)] + (\alpha-1)ln(\pi_j))}$$

$$q(\pi) \propto e^{\sum_{i=1}^{n}\sum_{j=1}^{K} E_{q-\pi}[1(c_i=j)]ln\pi_j + \sum_{j=1}^{K}((\alpha-1)ln(\pi_j))}$$

$$q(\pi) \propto e^{\sum_{i=1}^{n}\sum_{j=1}^{K} \phi_i(j)ln\pi_j + \sum_{j=1}^{K}(\alpha-1)ln\pi_j + constant\ w.r.t\ \pi}$$

<u>Recall</u>:

The Power Rule

$\log A^n = n \log A$

$$q(\pi) \propto e^{\sum_{i=1}^{n}\sum_{j=1}^{K} ln\pi_j^{\phi_i(j)} + \sum_{j=1}^{K} ln\pi_j^{(\alpha-1)} + constant\ w.r.t\ \pi}$$

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

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

$$\boxed{q(\pi) \sim Dirichlet(\alpha) \qquad \alpha'_k = \alpha + \sum_{i=1}^{n} E[1(c_i = k)] \qquad \alpha'_k = \alpha + \sum_{i=1}^{n} \phi_i(k)}$$

$q(\theta_j)$

$$p(x\ \pi, \boldsymbol{\theta}, \mathbf{C}) = [\prod_{i=1}^{n}\prod_{j=1}^{K} p(x_i|c_i = j, \pi_j, \theta_j)p(\pi_j)][\prod_{j=1}^{K} p(\theta_j|a_j, b_j)][\prod_{j=1}^{K} p(\pi_j|\alpha_j)]$$

Kate Lassiter

$$p(x\ \pi, \boldsymbol{\Theta}, \mathbf{C}) = [\prod_{i=1}^{n}\prod_{j=1}^{K} (\theta_j^{x_i}(1-\theta_j)^{20-x_i})^{1(c_i=j)} \pi_j^{1(c_i=j)}][\prod_{j=1}^{K} \frac{\Gamma(a_j)+\Gamma(b_j)}{\Gamma(a_j)\Gamma(b_j)} \theta_j^{a_j-1}(1-\theta_j)^{b_j-1}][\frac{\Gamma(\sum_K \alpha_k)}{\prod_{j=1}^{K}\Gamma(\alpha_j)} \prod_{j=1}^{K} \pi_j^{\alpha_j-1}\ ]$$

$$q(\theta_j) = e^{E_{-q}[\sum_{i=1}^{n}\sum_{j=1}^{K} lnp(x_i|c_i=j,\theta_j) + \sum_{j=1}^{K} lnp(\theta_j|a_j,b_j)]}$$

$$q(\theta_j) \propto e^{E_{-q}[\sum_{i=1}^{n}\sum_{j=1}^{K} ln\,(\theta_j^{x_i}(1-\theta_j)^{20-x_i})^{1(c_i=j)}) + \sum_{j=1}^{K} ln\frac{\Gamma(a_j)+\Gamma(b_j)}{\Gamma(a_j)\Gamma(b_j)}\theta_j^{a_j-1}(1-\theta_j)^{b_j-1}]}$$

$$q(\theta_j) \propto e^{E_{-q}[\sum_{i=1}^{n}\sum_{j=1}^{K} 1(c_i=j)ln\,(\theta_j^{x_i}(1-\theta_j)^{20-x_i}) + \sum_{j=1}^{K} ln\Gamma(a_j)+ln\Gamma(b_j)-ln\Gamma(a_j)-ln\Gamma(b_j)+(a_j-1)ln\theta_j + (b_j-1)ln(1-\theta_j)\ ]}$$

$$q(\theta_j) \propto e^{\sum_{i=1}^{n}\sum_{j=1}^{K} E_{-q}[1(c_i=j)]ln\,(\theta_j^{x_i}(1-\theta_j)^{20-x_i}) + \sum_{j=1}^{K} E_{-q}[ln\Gamma(a_j)+ln\Gamma(b_j)-ln\Gamma(a_j)-ln\Gamma(b_j)]+(a_j-1)ln\theta_j + (b_j-1)ln(1-\theta_j)\ ]}$$

$$q(\theta_j) \propto e^{\sum_{i=1}^{n}\sum_{j=1}^{K} E_{-q}[1(c_i=j)]ln\,(\theta_j^{x_i}(1-\theta_j)^{20-x_i}) + \sum_{j=1}^{K} (a_j-1)ln\theta_j + (b_j-1)(1-\theta_j)} + constant\ w.r.t\ \theta_j$$

$$q(\theta_j) \propto e^{\sum_{i=1}^{n}\sum_{j=1}^{K} \phi_i(j)ln\,(\theta_j^{x_i}(1-\theta_j)^{20-x_i}) + \sum_{j=1}^{K} (a_j-1)ln\theta_j + (b_j-1)ln(1-\theta_j)} + constant\ w.r.t\ \theta_j$$

$$q(\theta_j) \propto e^{\sum_{i=1}^{n}\sum_{j=1}^{K} \phi_i(j)[x_i ln\,\theta_j + (20-x_i)ln(1-\theta_j)\ ] + \sum_{j=1}^{K} (a_j-1)ln\theta_j + (b_j-1)ln(1-\theta_j)} + constant\ w.r.t\ \theta_j$$

$$q(\theta_j) \propto e^{\sum_{i=1}^{n}\sum_{j=1}^{K} \phi_i(j)x_i ln\,\theta_j + \phi_i(j)(20-x_i)ln(1-\theta_j) + \sum_{j=1}^{K} (a_j-1)ln\theta_j + (b_j-1)ln(1-\theta_j)} + constant\ w.r.t\ \theta_j$$

$$q(\theta_j) \propto \theta_j^{\sum_{i=1}^{n}\phi_i(j)x_i + (a_j-1)} (1-\theta_j)^{\sum_{i=1_i}^{n}\phi_i(j)(20-x_i)+(b_j-1)} + constant\ w.r.t\ \theta_j$$

<u>Thus:</u>

$$q(\theta_k) \sim Beta\,(a_k', b_k') \qquad a_k' = a + \sum_{i=1}^{n} \phi_i(k)x_i \qquad b_k' = b + \sum_{i=1_i}^{n} \phi_i(k)(20 - x_i)$$

$$q(\theta_k) \sim Beta(a_k', b_k') \qquad a_k' = a + \sum_{i=1}^{n} E\ [1(c_i = k)]x_i \qquad b_k' = b + \sum_{i=1_i}^{n} E\ [1(c_i = k)](20 - x_i)$$

$$q(c_i = j)$$

$$p(x\ \pi, \boldsymbol{\Theta}, \mathbf{C}) = [\prod_{i=1}^{n}\prod_{j=1}^{K} p(x_i|c_i = j, \pi_j, \theta_j)p(\pi_j)][\prod_{j=1}^{K} p(\theta_j|a_j, b_j)][\prod_{j=1}^{K} p(\pi_j|\alpha_j\ ))$$

$$p(x\ \pi, \boldsymbol{\Theta}, \mathbf{C}) = [\prod_{i=1}^{n}\prod_{j=1}^{K} (\theta_j^{x_i}(1-\theta_j)^{20-x_i})^{1(c_i=j)} \pi_j^{1(c_i=j)}][\prod_{j=1}^{K} \frac{\Gamma(a_j)+\Gamma(b_j)}{\Gamma(a_j)\Gamma(b_j)} \theta_j^{a_j-1}(1-\theta_j)^{b_j-1}][\frac{\Gamma(\sum_K \alpha_k)}{\prod_{j=1}^{K}\Gamma(\alpha_j)} \prod_{j=1}^{K} \pi_j^{\alpha_j-1}\ ]$$

$$q(c_i) = e^{E_{-q}[\sum_{j=1}^{K} lnp(x_i|c_i=j, \theta_j)+lnp(c_i=j|\pi)]]}$$

$$q(c_i) \propto e^{E_{-q}[\sum_{j=1}^{K} ln\,(\theta_j^{x_i}(1-\theta_j)^{20-x_i})^{1(c_i=j)} + ln\pi_j^{1(c_i=j)}]}$$

Kate Lassiter

$$q(c_i) \propto e^{E_{-q}[\sum_{j=1}^{K} 1(c_i=j) ln\,(\theta_j^{x_i}(1-\theta_j)^{20-x_i}) \ + 1(c_i=j) ln\pi_j\,]}$$

$$q(c_i) \propto e^{\sum_{i=1}^{n}\sum_{j=1}^{K} 1(c_i=j)\, E_{-q}[ln\,(\theta_j^{x_i}(1-\theta_j)^{20-x_i})\,] \ + 1(c_i=j)\, E_{-q}[ln\pi_j\,]}$$

* Fix $c_i = j$

$$q(c_i = j) \propto e^{E_{-q}[ln\,(\theta_j^{x_i}(1-\theta_j)^{20-x_i})\,] \ + E_{-q}[ln\pi_j\,]}$$

$$q(c_i = j) \propto exp\{E[ln\pi_j\,] \ + \ ln\binom{20}{x_i} + x_i E[ln\theta_j\,] \ + \ (20 - x_i)\, E\ [ln\,(1-\theta_j\,)]\ \}$$

$$\boxed{q(c_i = j) \propto exp\{E[ln\pi_j\,] + x_i E[ln\theta_j\,] \ + \ (20 - x_i)\, E\ [ln\,(1-\theta_j\,)]\ \}}$$

<u>Thus</u>:

$$\boxed{q(c_i) = \text{Discrete}\,(\phi_i) \qquad \phi_i(j) \ = \frac{exp\{E[ln\pi_j\,]+x_i E[ln\theta_j\,] \ +(20-x_i)\, E\ [ln\,(1-\theta_j\,)]\ \}}{\sum_{k=1}^{K} exp\{E[ln\pi_k\,]+x_i E[ln\theta_k\,] \ +(20-x_i)\, E\ [ln\,(1-\theta_k\,)]\ \}}}$$

<u>Recall</u>:

$$E\ [ln\pi_j\,] = \psi\,(\alpha_j') \ - \ \psi\,(\sum_k \alpha_j') \rightarrow Dirichlet\ r.v$$

$$E\ [ln\theta_j\,] \ = \ \psi\,(a_j') \ - \ \psi\,(a_j' + b_j') \rightarrow Beta\ r.v$$

$$E\ [ln\,(1-\theta_j\,)] \ = \ \psi\,(b_j') \ - \ \psi\,(a_j' + b_j') \rightarrow Beta\ r.v$$

$$\boxed{\phi_i(j) \ = \frac{exp\{([\psi\,(\alpha_j')-\psi\,(\sum_k \alpha_j')\,] + [x_i\ (\psi\,(a_j')-\psi\,(a_j+b_j)) \ \ +(\,20-x_i)(\psi\,(b_j')-\psi\,(a_j+b_j))\ ]\ \}}{\sum_{k=1}^{K} exp\{([\psi\,(\alpha_k')-\psi\,(\sum_m \alpha_m')\,] + [x_i\ (\psi\,(a_k')-\psi\,(a_k+b_k)) \ \ +(\,20-x_i)(\psi\,(b_k')-\psi\,(a_k+b_k))\ ]\ \}}}$$

VI Objective

$$L \ = \ E_q[lnp(c,\pi,\boldsymbol{\Theta}|x)] \ - E_q[lnq(c,\pi,\boldsymbol{\Theta})]$$

<u>**Focus first on Decomposing:**</u> $E_q[lnp(c,\pi,\boldsymbol{\Theta}|x)]$

$$p(x\ \pi,\boldsymbol{\Theta},\mathbf{C}) = [\prod_{i=1}^{n}\prod_{j=1}^{K}(\binom{20}{x_i}\ \theta_j^{x_i}(1-\theta_j)^{20-x_i})^{1(c_i=j)}\pi_j^{1(c_i=j)}][\prod_{j=1}^{K}\frac{\Gamma(a_j)+\Gamma(b_j)}{\Gamma(a_j)\Gamma(b_j)}\theta_j^{a_j-1}(1-\theta_j)^{b_j-1}][\frac{\Gamma(\sum_K \alpha_k)}{\prod_{j=1}^{K}\Gamma(\alpha_j)}\prod_{j=1}^{K}\pi_j^{\alpha_j-1}\ ]$$

$$L \ = \ E_q[lnp(x_i|c_i = j,\theta_j\ ) \ + \ lnp(c_j|\pi) \ + \ lnp(\theta_j|a_j,b_j) \ + \ lnp(\pi|\alpha_k) \ +] - E_q[lnq(c,\pi,\boldsymbol{\Theta})]$$

$$L \ = \ E_q[ln(\binom{20}{x_i}\ \theta_j^{x_i}(1-\theta_j)^{20-x_i})^{1(c_i=j)} + \ ln\pi_j^{1(c_i=j)} \ + \ ln\frac{\Gamma(a_j)+\Gamma(b_j)}{\Gamma(a_j)\Gamma(b_j)}\theta_j^{a_j-1}(1-\theta_j)^{b_j-1} \ + \ ln\frac{\Gamma(\sum_K \alpha_k)}{\prod_{j=1}^{K}\Gamma(\alpha_j)}\pi_j^{\alpha_j-1}\ ] - E_q[lnq(c,\pi,\boldsymbol{\Theta})]$$

$$L \ = \ E_q[\sum_{i=1}^{n}\sum_{j=1}^{K} ln(\binom{20}{x_i}\ \theta_j^{x_i}(1-\theta_j)^{20-x_i})^{1(c_i=j)} + \sum_{i=1}^{n}\sum_{j=1}^{K} ln\pi_j^{1(c_i=j)} \ + \sum_{j=1}^{K} ln\frac{\Gamma(a_j)+\Gamma(b_j)}{\Gamma(a_j)\Gamma(b_j)}\theta_j^{a_j-1}(1-\theta_j)^{b_j-1} + \sum_{j=1}^{K} ln\frac{\Gamma(\sum_K \alpha_k)}{\prod_{j=1}^{K}\Gamma(\alpha_j)}\pi_j^{\alpha_j-1}\ ] - E_q[lnq(c,\pi,\boldsymbol{\Theta})]$$

$$L \ = \ E_q[\sum_{i=1}^{n}\sum_{j=1}^{K} 1(c_i = j) ln(\theta_j^{x_i}(1-\theta_j)^{20-x_i}) \ + \sum_{i=1}^{n}\sum_{j=1}^{K} 1(c_i = j) ln\pi_j \ \ + \sum_{j=1}^{K} ln\Gamma(a_j) \ + \ ln\Gamma\,(b_j) \ - \ ln\Gamma\,(a_j \ + \ b_j) + \ (a_j - 1) ln\theta_j \ + \ (b_j - 1) ln(1-\theta_j)$$

$$+ \sum_{j=1}^{K}\ (ln\Gamma\,(\sum_K \alpha_k) \ - \ ln\Gamma\,(\prod_{j=1}^{K}\alpha_j) \ + \ (\alpha_j - 1) ln\pi_j\ ) \ + \ Constant] - E_q[lnq(c,\pi,\boldsymbol{\Theta})$$

Kate Lassiter

$$L = \sum_{i=1}^{n}\sum_{j=1}^{K} E_{ci}[1(c_i = j)]\,(x_i E_\theta[\ln\theta_j] + (20 - x_i)E_\theta[\ln(1-\theta_j)]] + \sum_{i=1}^{n}\sum_{j=1}^{K} E_{ci}[1(c_i = j)]\,E_\pi[\ln\pi_j] + \sum_{j=1}^{K} \ln\Gamma(a_j) + \ln\Gamma(b_j) - \ln\Gamma(a_j + b_j) + (a_j - 1)E_\theta[\ln\theta_j] + (b_j - 1)E_\theta[\ln(1-\theta_j)]]$$
$$+ \sum_{j=1}^{K}(\ln\Gamma(\sum_K \alpha_k) - \ln\Gamma(\prod_{j=1}^{K}\alpha_j) + (\alpha_j - 1)E_\pi[\ln\pi_j]) + Constant] - E_q[\ln q(c,\pi,\mathbf{\Theta})]$$

$$L = \sum_{i=1}^{n}\sum_{j=1}^{K} \phi_i(j)(\,(x_i(\psi(a_j) - \psi(a_j + b_j)) + (20 - x_i)(\psi(b_j) - \psi(a_j + b_j))_i) + \sum_{i=1}^{n}\sum_{j=1}^{K} \phi_i(j)(\psi(\alpha_j) - \psi(\sum_k \alpha_k)) + \sum_{j=1}^{K} \ln\Gamma(a_j) + \ln\Gamma(b_j) - \ln\Gamma(a_j + b_j) + (a_j - 1)(\psi(a_j) - \psi(a_j + b_j)) + (b_j - 1)(\psi(b_j) - \psi(a_j + b_j))$$
$$+ \sum_{j=1}^{K}(\ln\Gamma(\sum_K \alpha_k) - \ln\Gamma(\prod_{j=1}^{K}\alpha_j) + (\alpha_j - 1)(\psi(\alpha_j) - \psi(\sum_k \alpha_k'))) + Constant] - E_q[\ln q(c,\pi,\mathbf{\Theta})]$$

<u>Now decomposing</u> $E_q[\ln q(c,\pi,\mathbf{\Theta})]$

$$L = E_q[\ln p(c,\pi,\mathbf{\Theta}|x)] - E_q[\ln q(c,\pi,\mathbf{\Theta})]$$

$$L = E_q[\ln p(c,\pi,\mathbf{\Theta}|x)] - E_q[\ \ln q(c_i|\theta_j,\pi_j) + \ln q(\pi_j|(\alpha_1,\dots\alpha_k)) + \ln q(\theta_j|a_j,b_j)\ ]$$

$$L = E_q[\ln p(c,\pi,\mathbf{\Theta}|x)] - E_q[\ \sum_{i=1}^{n}\sum_{j=1}^{K} \ln q(c_i = j)^{1(c_i=j)} + \sum_{j=1}^{K} \ln \frac{\Gamma(\sum_K \alpha_k')}{\prod_{j=1}^{K}\Gamma(\alpha_j')}\pi_j^{\alpha_j'-1} + \sum_{j=1}^{K} \ln \frac{\Gamma(a_j') + \Gamma(b_j')}{\Gamma(a_j')\Gamma(b_j')}\theta_j^{a_j'-1}(1-\theta_j)^{b_j'-1}]$$

<u>Recall</u>:

$1(c_i = j) = q(c_i) = \phi_i(j) = E[1(c_i = j)]$

<u>Thus</u>:

$$L = E_q[\ln p(c,\pi,\mathbf{\Theta}|x)] - E_q[\ \sum_{i=1}^{n}\sum_{j=1}^{K} \ln(\phi_{ij})^{1(c_i=j)} + \sum_{j=1}^{K} \ln \frac{\Gamma(\sum_K \alpha_k')}{\prod_{j=1}^{K}\Gamma(\alpha_j')}\pi_j^{\alpha_j'-1} + \sum_{j=1}^{K} \ln \frac{\Gamma(a_j') + \Gamma(b_j')}{\Gamma(a_j')\Gamma(b_j')}\theta_j^{a_j'-1}(1-\theta_j)^{b_j'-1}]$$

$$L = E_q[\ln p(c,\pi,\mathbf{\Theta}|x)] - E_q[\ \sum_{i=1}^{n}\sum_{j=1}^{K} 1(c_i = j)\ln\phi_{ij} + \sum_{j=1}^{K} \ln \frac{\Gamma(\sum_K \alpha_k')}{\prod_{j=1}^{K}\Gamma(\alpha_j')}(\alpha_j' - 1)\ln\pi_j + \sum_{j=1}^{K} \ln \frac{\Gamma(a_j') + \Gamma(b_j')}{\Gamma(a_j')\Gamma(b_j')} + (a_j' - 1)\ln\theta_j + (b_j' - 1)\ln(1-\theta_j)\ ]$$

$$L = E_q[\ln p(c,\pi,\mathbf{\Theta}|x)] - [\ \sum_{i=1}^{n}\sum_{j=1}^{K} E_{ci}[1(c_i = j)]\ln\phi_{ij} + \sum_{j=1}^{K} \ln \frac{\Gamma(\sum_K \alpha_k')}{\prod_{j=1}^{K}\Gamma(\alpha_j')}(\alpha_j' - 1)E_\pi[\ln\pi_j] + \sum_{j=1}^{K} \ln \frac{\Gamma(a_j') + \Gamma(b_j')}{\Gamma(a_j')\Gamma(b_j')} + (a_j' - 1)E_\theta[\ln\theta_j] + (b_j' - 1)E_\theta[\ln(1-\theta_j)]\ ]$$

$$L = E_q[\ln p(c,\pi,\mathbf{\Theta}|x)] - [\ \sum_{i=1}^{n}\sum_{j=1}^{K} E_{ci}[1(c_i = j)]\ln\phi_{ij} + \sum_{j=1}^{K}(\alpha_j' - 1)E_\pi[\ln\pi_j] + \sum_{j=1}^{K}(a_j' - 1)E_\theta[\ln\theta_j] + (b_j' - 1)E_\theta[\ln(1-\theta_j)]\ ] + Constant\ w.r.t\ \theta,\pi,c\ ]$$

$$L = E_q[\ln p(c,\pi,\mathbf{\Theta}|x)] - [\ \sum_{i=1}^{n}\sum_{j=1}^{K} \phi_{ij}\ln\phi_{ij} + \sum_{j=1}^{K}(\alpha_j' - 1)E_\pi[\ln\pi_j] + \sum_{j=1}^{K}(a_j' - 1)E_\theta[\ln\theta_j] + (b_j' - 1)E_\theta[\ln(1-\theta_j)]\ ] + Constant\ w.r.t\ \theta,\pi,c\ ]$$

$$L = E_q[\ln p(c,\pi,\mathbf{\Theta}|x)] - [\ \sum_{i=1}^{n}\sum_{j=1}^{K} \phi_{ij}\ln\phi_{ij} + \sum_{j=1}^{K}(\alpha_j' - 1)(\psi(\alpha_j') - \psi(\sum_k \alpha_k')) + \sum_{j=1}^{K}(a_j' - 1)(\psi(a_j') - \psi(a_j' + b_j')) + (b_j' - 1)(\psi(b_j') - \psi(a_j' + b_j'))\ ] + Constant\ w.r.t\ \theta,\pi,c$$

$$L = E_q[\ln p(c,\pi,\mathbf{\Theta}|x)] - \sum_{i=1}^{n}\sum_{j=1}^{K} \phi_{ij}\ln\phi_{ij} - \sum_{j=1}^{K}(\alpha_j' - 1)(\psi(\alpha_j') - \psi(\sum_k \alpha_k')) - \sum_{j=1}^{K}((a_j' - 1)(\psi(a_j') - \psi(a_j' + b_j')) + (b_j' - 1)(\psi(b_j') - \psi(a_j' + b_j'))) + Constant\ w.r.t\ \theta,\pi,c$$

$$\boxed{\begin{aligned} L = {} & \sum_{i=1}^{n}\sum_{j=1}^{K} E_{ci}[1(c_i = j)]\,(x_i E_\theta[\ln\theta_j] + (20 - x_i)E_\theta[\ln(1-\theta_j)]] + \sum_{i=1}^{n}\sum_{j=1}^{K} E_{ci}[1(c_i = j)]\,E_\pi[\ln\pi_j] + \sum_{j=1}^{K} \ln\Gamma(a_j) + \ln\Gamma(b_j) - \ln\Gamma(a_j + b_j) + (a_j - 1)E_\theta[\ln\theta_j] + (b_j - 1)E_\theta[\ln(1-\theta_j)]] \\ & + \sum_{j=1}^{K}(\ln\Gamma(\sum_K \alpha_k) - \ln\Gamma(\prod_{j=1}^{K}\alpha_j) + (\alpha_j - 1)E_\pi[\ln(\pi_j)]) + Constant] - [\sum_{i=1}^{n}\sum_{j=1}^{K} E_c[1(c_i = j)]\ln\phi_{ij} + \sum_{j=1}^{K}(\alpha_j' - 1)E_\pi[\ln\pi_j] + \sum_{j=1}^{K}(a_j' - 1)E_\theta[\ln\theta_j] + (b_j' - 1)E_\theta[\ln(1-\theta_j)]\ ] + Constant\ ] \end{aligned}}$$

<u>Thus</u>:

$$\boxed{\begin{aligned} L = {} & \sum_{i=1}^{n}\sum_{j=1}^{K} \phi_i(j)(\,(x_i(\psi(a_j') - \psi(a_j + b_j)) + (20 - x_i)(\psi(b_j') - \psi(a_j + b_j))_i) + \sum_{i=1}^{n}\sum_{j=1}^{K} \phi_i(j)(\psi(\alpha_j') - \psi(\sum_k \alpha_k')) + \sum_{j=1}^{K} \ln\Gamma(a_j) + \ln\Gamma(b_j) - \ln\Gamma(a_j + b_j) + (a_j - 1)(\psi(a_j) - \psi(a_j + b_j)) + (b_j - 1)(\psi(b_j) - \psi(a_j + b_j)) \\ & + \sum_{j=1}^{K}(\ln\Gamma(\sum_K \alpha_k) - \ln\Gamma(\prod_{j=1}^{K}\alpha_j) + (\alpha_j - 1)(\psi(\alpha_j) - \psi(\sum_k \alpha_k))) - \sum_{i=1}^{n}\sum_{j=1}^{K} \phi_{ij}\ln\phi_{ij} - \sum_{j=1}^{K}(\alpha_j' - 1)(\psi(\alpha_j') - \psi(\sum_k \alpha_k')) - \sum_{j=1}^{K}((a_j' - 1)(\psi(a_j') - \psi(a_j' + b_j')) + (b_j' - 1)(\psi(b_j') - \psi(a_j + b_j)) + Constant \end{aligned}}$$

A Variational Inference Algorithm For The Binomial Mixture Model

Input: Data $x_1,\dots x_n$, $x \in R^d$. Number of clusters K.
Output: BMM parameters $q(\pi), q(\theta_k), q(c_i)$

1. Initialize $(\alpha_1^{(0)},\dots,\alpha_k^{(0)}) \sim Uniform\ (0,1)\ and\ each\ (a_k^{(0)}, b_k^{(0)}) \sim Uniform\ (10, 100)$
2. For iteration t =1, ...,T
    a. Update $q(c_i)$ for i = 1,...n set:
        i. $$\phi_i^{(t)}(j) = \frac{e^{t_1(j) + t_2(j)}}{\sum_{j=1}^{K} e^{t_1(k) + t_2(k)}}$$
            1. $t_1(j) = \psi(\alpha_j^{'(t-1)}) - \psi(\sum_k \alpha_k^{'(t-1)})$
            2. $t_2(j) = x_i(\psi(a_j^{'(t-1)}) - \psi(a_j^{'(t-1)} + b_j^{'(t-1)})) + (20 - x_i)(\psi(b_j^{'(t-1)}) - \psi(a_j^{'(t-1)} + b_j^{'(t-1)}))$

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

e. Update $\pi_j^{(t)}$ for j = 1,..., K by setting

i. $\alpha_j^{'(t)} = \alpha + n_j^{(t)}$

f. Update $q(\theta_j^{(t)})$ for j = 1,..., K by setting

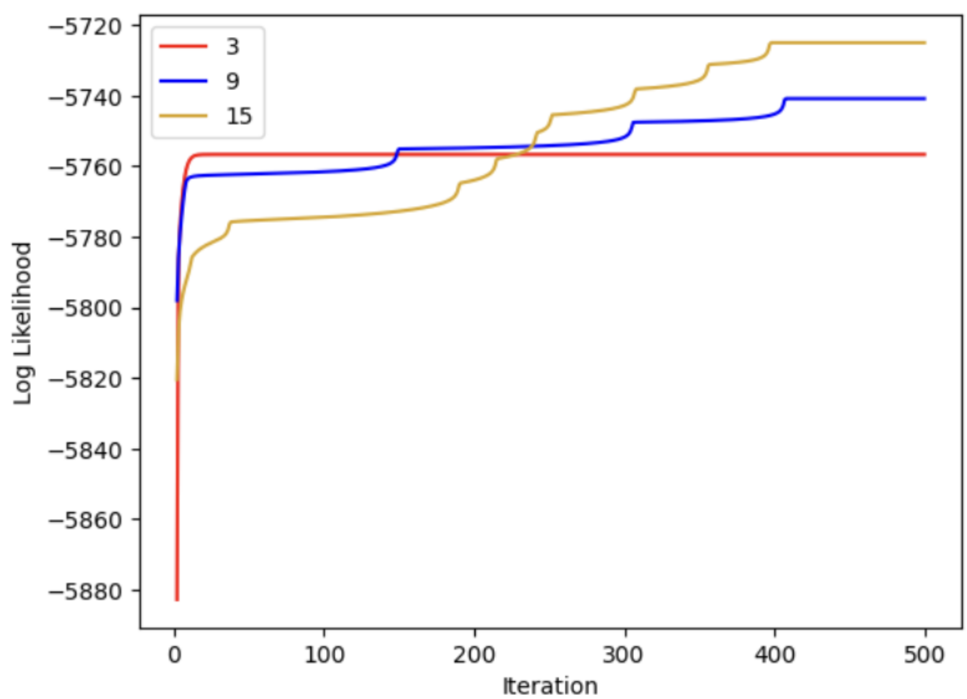
i. $a_j^{(t)'} = a + \sum_{i=1}^{n} x_i \phi_i^{(t)}(j)$

ii. $b_j^{(t)'} = b + \sum_{i=1}^{n} (20 - x_i)\phi_i^{(t)}(j)$

g. Calculate variational objective function to assess convergence as a function of iteration

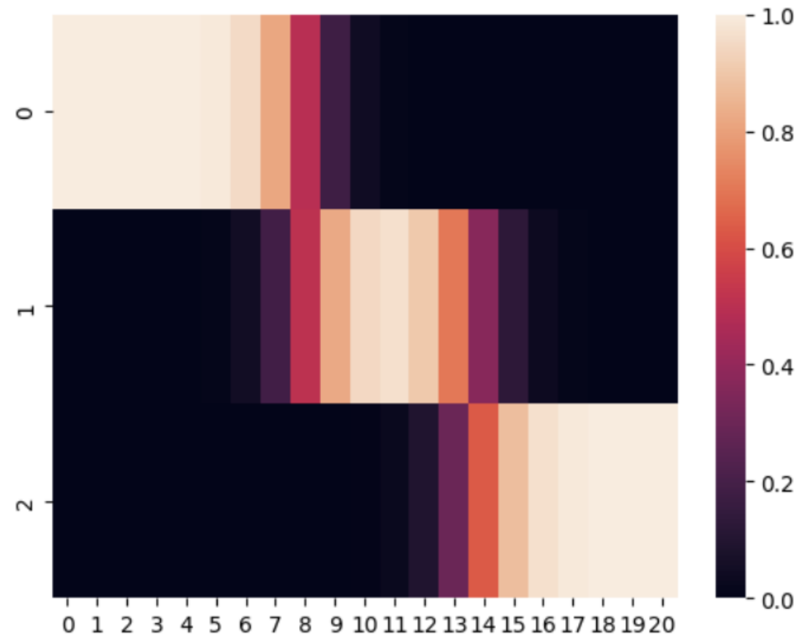
$$\boxed{L(\pi, \boldsymbol{\Theta}) = E[\ln p(x,c,\pi,\boldsymbol{\Theta})] - E[\ln q]}$$

b) Implement this VI algorithm. Initialize your algorithm so that $q(\pi)$ encourages a uniform distribution and the variational parameters of all $q(\theta)$ are each initialized randomly between $[10, 100]$. Update $q(c)$ first. For $K = 3, 9, 15$ run your algorithm 10 times for 500 iterations each time and save the run with the highest variational objective function. You will therefore store one best run for each value of K. In one figure, plot the variational objective function for iterations 2 to 500 using the best run for each value of K.

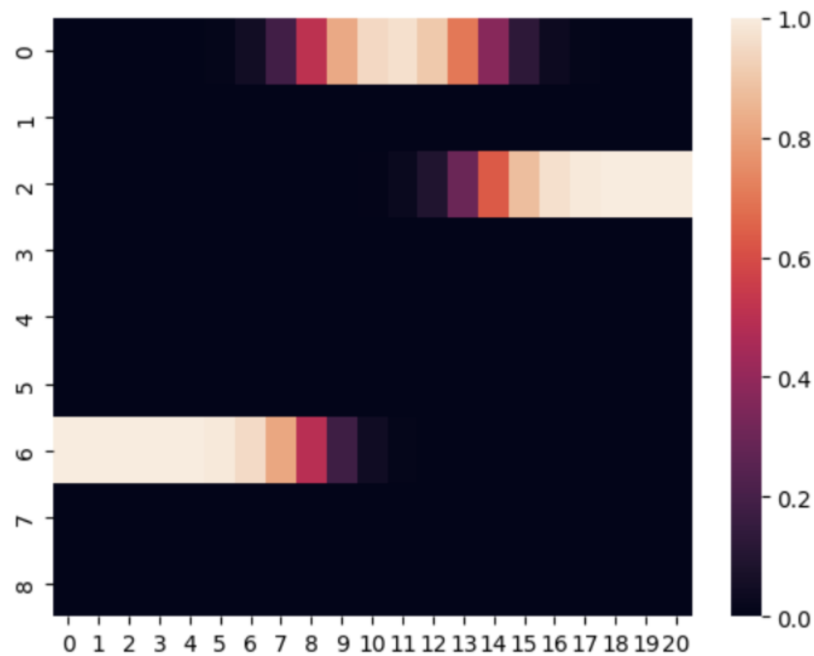


c) This part is similar to Problem 2c. For the final iteration of each saved model show the following: Construct a $K \times 21$ matrix M_K where $M_K(k, i) = q(c = k)$ learned when $x = i - 1$. Note that this q is learned differently from the EM algorithm. Show this as three images with accompanying colorbar legend for $K = 3, 9, 15$ using the best runs you saved from part (b). Hint: Since the data set contains instances of all 21 possible values of x, you have already calculated the 21 required q distributions during inference.

K =3



K = 9



Kate Lassiter

K = 15