Discrete Mixture Models

Trevor Fiez and Jose Picado

1 Introduction

This project deals with estimating a mixture model for two multimodal distributions for M different categories:

$$P(k_1, \cdots, k_M | p_1, \cdots, p_M) = \frac{n!}{\prod_l k_l!} \prod_{l=1}^{M} p_l^{k_l}, \qquad \sum_l^M p_l = 1 \text{ and } \sum_l^M k_l = n$$

$$P(k_1, \cdots, k_M | q_1, \cdots, q_M) = \frac{n!}{\prod_l k_l!} \prod_{l=1}^{M} q_l^{k_l}, \qquad \sum_l^M q_l = 1 \text{ and } \sum_l^M k_l = n$$

Consider a mixing process of iid realizations from each process $\{k_i(1)^{(p)}\}, \cdots, \{k_i(N)^{(p)}\}$ from p and $\{k_i(1)^{(q)}\}, \cdots, \{k_i(N)^{(q)}\}$ from q. At each time instance $t = 1, \cdots, N$, we observe $\{k_i(t)^{(p)}\}$ with probability α and $\{k_i(t)^{(q)}\}$ with probability $(1-\alpha)$. In this project, we focus on $M = 2$, the binomial case. Because $M = 2$, let $k_2 = n - k_1$, $p_1 = p$, $p_2 = 1 - p$, $q_1 = q$, $q_2 = 1 - q$. Then, the probability mass function for the first bonomial distribution is given by:

$$P(k_1(i)|p, n) = \binom{n}{k_1(i)} p^{k_1(i)} (1-p)^{n-k_1(i)}$$

The probability mass function for the second binomial distribution is given by:

$$P(k_1(i)|q, n) = \binom{n}{k_1(i)} q^{k_1(i)} (1-q)^{n-k_1(i)}$$

Assuming a fixed observation length n at every time instance, the probability mass function for the mixture model is given by:

$$P(k_1(1), \cdots, k_1(N)|\alpha, p, q) = \prod_{i=1}^{N} \left[\alpha \binom{n}{k_1(i)} p^{k_1(i)} (1-p)^{n-k_1(i)} + (1-\alpha) \binom{n}{k_1(i)} q^{k_1(i)} (1-q)^{n-k_1(i)} \right] \tag{1}$$

The parameter vector for this problem is given by:

$$\theta = \begin{bmatrix} \alpha \\ p \\ q \end{bmatrix}$$

2 FIM and CRLB

The derivation for the FIM is included in the appendix. Given equation 1 let the loglikihood be

$$\log L(\theta) = \sum_{i=1}^{N} \log \left[\alpha \binom{n}{k_1(i)} p^{k_1(i)} (1-p)^{n-k_1(i)} + (1-\alpha) \binom{n}{k_1(i)} q^{k_1(i)} (1-q)^{n-k_1(i)} \right]$$

Then it follows that the Fisher Information Matrix and the Cramer-Rao Lower Bound are

$$FIM = -\mathrm{E}\left[\frac{d^2 \log p}{d^2\theta}\right]$$
$$CRLB = FIM^{-1}$$

The derivations for the derivatives are included in the appendix. When doing the derivations we define

$$P(k|p,n) = \binom{n}{k} p^k (1-p)^{n-k}$$
$$P'(k|p,n) = \frac{dP}{dp} = \binom{n}{k}\left[kp^{k-1}(1-p)^{n-k} + (n-k)p^k(1-p)^{n-k-1}\right]$$
$$P\text{"}(k|p,n) = \frac{d^2P}{dp^2} = \binom{n}{k}\left[k(k-1)p^{k-2}(1-p)^{n-k} - (-k+n-1)(n-k)p^k(1-p)^{-k+n-2}\right]$$

The Fisher information matrix is then:

$$-E\left[\frac{\delta^2 \log(L(\theta)}{\delta\theta^2}\right] = -N * E\begin{bmatrix} F_{\alpha\alpha} F_{\alpha p} F_{\alpha q} \\ F_{\alpha p} F_{pp} F_{pq} \\ F_{\alpha q} F_{pq} F_{qq} \end{bmatrix}$$

Where

$$F_{\alpha\alpha} = \frac{\partial^2 L(\theta)}{\partial\alpha^2} = \frac{(P(k|p,n) - P(k|q,n))^2}{(\alpha P(k|p,n) + (1-\alpha)P(k|q,n))^2}$$
$$F_{\alpha p} = \frac{\partial^2 L(\theta)}{\partial\alpha\partial p} = \frac{P'(k|p,n)P(k|q,n)}{(\alpha P(k|p,n) + (1-\alpha)P(k|q,n))^2}$$
$$F_{\alpha q} = \frac{\partial^2 L(\theta)}{\partial\alpha\partial q} = \frac{-P(k|p,n)P'(k|q,n)}{(\alpha P(k|p,n) + (1-\alpha)P(k|q,n))^2}$$
$$F_{pp} = \frac{\partial^2 L(\theta)}{\partial p^2} = \frac{\alpha^2 P\text{"}(k|p,n)P(k|p,n) + \alpha(1-\alpha)P\text{"}(k|p,n)P(k|q,n) - \alpha^2 P(k|p,n)^2}{(\alpha P(k|p,n) + (1-\alpha)P(k|q,n))^2}$$
$$F_{pq} = \frac{\partial^2 L(\theta)}{\partial p\partial q} = \frac{-\alpha(1-\alpha)P'(k|p,n)P'(k|p,n)}{(\alpha P(k|p,n) + (1-\alpha)P(k|q,n))^2}$$
$$F_{qq} = \frac{\partial^2 L(\theta)}{\partial q^2} = \frac{\alpha(1-\alpha)P\text{"}(k|q,n)P(k|p,n) + (1-\alpha)^2 P\text{"}(k|q,n)P(k|q,n) - (1-\alpha)^2 P'(k|q,n)^2}{(\alpha P(k|p,n) + (1-\alpha)P(k|q,n))^2}$$

A plot for the CRLB is shown below for α = 0.1 ... 0.9, p = 0.2, q = 0.4. As you can see, the theoretical upper bound on the mean squared error increases for both b and q when they are sampled at a small rate. Furthermore, you can see that the mean square error increases for α when α is close to 0.4.

3 Maximum Likelihood and Expectation-Maximization

3.1 Maximum Likelihood Equations

Given Equation 1, let the likelihood be

$$L(\theta) = P(k_1(1), \cdots, k_1(N)|\theta)$$

Then, the log-likelihood is

$$\log L(\theta) = \sum_{i=1}^{N} \log\left[\alpha\binom{n}{k_1(i)}p^{k_1(i)}(1-p)^{n-k_1(i)} + (1-\alpha)\binom{n}{k_1(i)}q^{k_1(i)}(1-q)^{n-k_1(i)}\right]$$

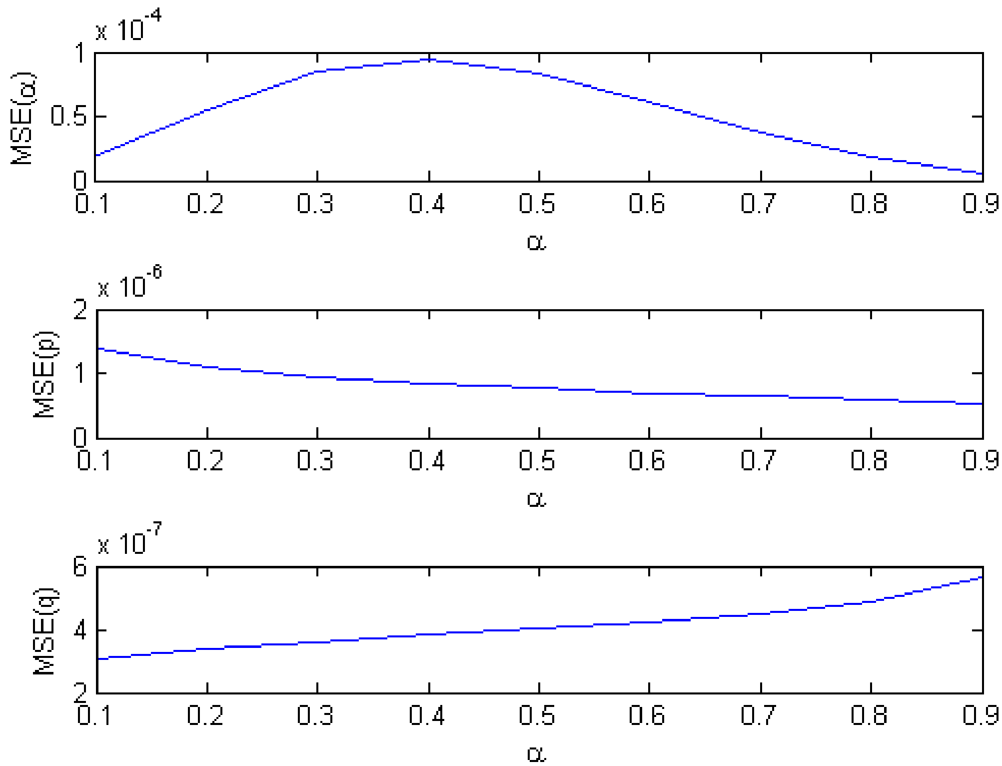

Figure 1: CRLB for different values of α when p = 0.2, q = 0.4

The maximum log-likelihood is given by

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}} \log L(\theta)$$

Let ϕ_1 and ϕ_0 be functions such that

$$\begin{aligned} \phi_1(k_1(i)) &= P(k_1(i)|p, n) \\ \phi_0(k_1(i)) &= P(k_1(i)|q, n) \end{aligned}$$

Therefore, the maximum likelihood equations are given by:

$$\frac{\delta \log P}{\delta \theta} = \begin{bmatrix} \sum_{i=1}^{N} \frac{\phi_1(k_1(i)) - \phi_0(k_1(i))}{\alpha\phi_1(k_1(i)) + (1-\alpha)\phi_0(k_1(i))} \\ \sum_{i=1}^{N} \frac{\alpha\binom{n}{k_1(i)}\left(k_1(i)p^{k_1(i)-1}(1-p)^{n-k_1(i)} - p^{k_1(i)}(n-k_1(i))(1-p)^{n-k_1(i)-1}\right)}{\alpha\phi_1(k_1(i)) + (1-\alpha)\phi_0(k_1(i))} \\ \sum_{i=1}^{N} \frac{(1-\alpha)\binom{n}{k_1(i)}\left(k_1(i)q^{k_1(i)-1}(1-q)^{n-k_1(i)} - q^{k_1(i)}(n-k_1(i))(1-q)^{n-k_1(i)-1}\right)}{\alpha\phi_1(k_1(i)) + (1-\alpha)\phi_0(k_1(i))} \end{bmatrix} = \mathbf{0}$$

3.2 EM Algorithm

Let $\mathbf{k_1} = k_1(1), \cdots, k_1(N)$ be the observed data. We introduce membership variables $\mathbf{y} = y(i), \cdots, y(N)$ (hidden data) such that $P(y(i) = l) = \alpha_l$. Because we only have two classes, then

$$\begin{aligned} &P(y(i) = 1) = \alpha \\ &P(y(i) = 0) = 1 - P(y(i) = 1) = 1 - \alpha \end{aligned}$$

The joint probability mass function is given by:

$$\begin{aligned} P(\mathbf{k_1}, \mathbf{y}|\theta) &= \prod_{i=1}^{N} P(k_1(i), y(i)|\theta) \\ &= \prod_{i=1}^{N} P(k_1(i)|y(i)) P(y(i)|\theta) \\ &= \prod_{i=1}^{N} \prod_{l=1}^{M} (\phi_l(k_1(i)\alpha_l))^{I(y(i)=l)} \end{aligned}$$

Let Q be an auxiliary function such that

$$\begin{aligned} Q(\theta, \theta') &= \mathrm{E}\left[\log P(\mathbf{k_1}, \mathbf{y}|\theta)|\mathbf{k_1}, \theta'\right] \\ &= \mathrm{E}\left[\sum_{i=1}^{N} \sum_{l=1}^{M} I(y(i) = l)(\log \phi_l(k_1(i)) + \log \alpha_l)|\mathbf{k_1}, \theta'\right] \\ &= \sum_{i=1}^{N} \sum_{l=1}^{M} \mathrm{E}\left[I(y(i) = l)|\mathbf{k_1}, \theta'\right] (\log \phi_l(k_1(i)) + \log \alpha_l) \\ &= \sum_{i=1}^{N} \sum_{l=1}^{M} P(y(i) = l|k_1(i), \theta')(\log \phi_l(k_1(i)) + \log \alpha_l) \end{aligned} \tag{2}$$

In the E-step of the EM algorithm, we compute the term $P(y(i) = l|k_1(i), \theta')$ of Equation 2. In our case of $M = 2$ the E-step is given by:

$$\begin{aligned} P(y(i) = 1|k_1(i), \theta') &= \frac{\alpha\binom{n}{k_1(i)}p^{k_1(i)}(1-p)^{n-k_1(i)}}{\alpha\binom{n}{k_1(i)}p^{k_1(i)}(1-p)^{n-k_1(i)} + (1-\alpha)\binom{n}{k_1(i)}q^{k_1(i)}(1-q)^{n-k_1(i)}} \\ P(y(i) = 0|k_1(i), \theta') &= \frac{(1-\alpha)\binom{n}{k_1(i)}q^{k_1(i)}(1-q)^{n-k_1(i)}}{\alpha\binom{n}{k_1(i)}p^{k_1(i)}(1-p)^{n-k_1(i)} + (1-\alpha)\binom{n}{k_1(i)}q^{k_1(i)}(1-q)^{n-k_1(i)}} \end{aligned}$$

In the M-step, we compute

$$\begin{aligned}
\theta^{k+1} &= \underset{\theta}{\operatorname{argmax}}\, Q(\theta, \theta^k) \\
&= \underset{\theta}{\operatorname{argmax}} \sum_{i=1}^{N} \sum_{l=1}^{M} P(y(i) = l | k_1(i), \theta^k)(\log \phi_l(k_1(i)) + \log \alpha_l) \\
&\text{s.t. } \sum_{l=1}^{M} \alpha_l = 1
\end{aligned}$$

In our case of $M = 2$, we have

$$\theta^{k+1} = \underset{\theta}{\operatorname{argmax}} \sum_{i=1}^{N} \left[P(y(i) = 1 | k_1(i), \theta^k)(\log \phi_1(k_1(i)) + \log \alpha) + P(y(i) = 0 | k_1(i), \theta^k)(\log \phi_0(k_1(i)) + \log(1 - \alpha)) \right]$$

To maximize α, we use a Lagrange multiplier λ to define the Lagrangian as

$$L = \sum_{i=1}^{N} \sum_{l=1}^{M} P(y(i) = l | k_1(i), \theta^k)(\log \alpha_l) + \lambda \left(\sum_{l=1}^{M} \alpha_l - 1 \right)$$

Then, taking the derivative and making equal to 0:

$$\begin{aligned}
\frac{\delta L}{\delta \alpha_l} &= \sum_{i=1}^{N} P(y(i) = l | k_1(i), \theta^k) \frac{1}{\alpha_l} + \lambda = 0 \\
\alpha_l &= -\frac{1}{\lambda} \sum_{i=1}^{N} P(y(i) = l | k_1(i), \theta^k)
\end{aligned}$$

To get the value of the Lagrange multiplier:

$$\frac{\delta L}{\delta \lambda} = \sum_{l=1}^{M} \alpha_l - 1 = 0$$

Plugging in the value of α_l:

$$\begin{aligned}
1 &= \sum_{l=1}^{M} \left[-\frac{1}{\lambda} \sum_{i=1}^{N} P(y(i) = l | k_1(i), \theta^k) \right] \\
1 &= -\frac{1}{\lambda} \sum_{i=1}^{N} \sum_{l=1}^{M} P(y(i) = l | k_1(i), \theta^k) \\
\lambda &= -\sum_{i=1}^{N} \sum_{l=1}^{M} P(y(i) = l | k_1(i), \theta^k) \\
\lambda &= -\sum_{i=1}^{N} 1 \\
\lambda &= -N
\end{aligned}$$

Therefore, we get:

$$\alpha^{k+1} = \frac{1}{N} \sum_{i=1}^{N} P(y(i) = 1 | k_1(i), \theta^k)$$

To maximize p, we get:

$$\begin{aligned}
p^{k+1} &= \underset{p}{\operatorname{argmax}} \sum_{i=1}^{N} P(y(i)=1|k_1(i),\theta^k)\log\phi_1(k_1(i)) \\
&= \underset{p}{\operatorname{argmax}} \sum_{i=1}^{N} P(y(i)=1|k_1(i),\theta^k)\log\binom{n}{k_1(i)}p^{k_1(i)}(1-p)^{n-k_1(i)} \\
&= \underset{p}{\operatorname{argmax}} \sum_{i=1}^{N} P(y(i)=1|k_1(i),\theta^k)(\log n! - \log k_1(i)! - \log(n-k_1(i))! + k_1(i)\log p + (n-k_1(i))\log(1-p))
\end{aligned}$$

Taking the derivative and making equal to 0:

$$\begin{aligned}
\frac{\delta}{\delta p} &= \sum_{i=1}^{N} P(y(i)=1|k_1(i),\theta^k)\left(\frac{k_1(i)}{p} + \frac{n-k_1(i)}{1-p}\right) \\
&= \sum_{i=1}^{N} P(y(i)=1|k_1(i),\theta^k)\left(\frac{k_1(i)(1-p)-(n-k_1(i))p}{p(1-p)}\right) \\
&= \sum_{i=1}^{N} P(y(i)=1|k_1(i),\theta^k)\left(\frac{k_1(i)-k_1(i)p-np+k_1(i)p}{p(1-p)}\right) \\
&= \sum_{i=1}^{N} P(y(i)=1|k_1(i),\theta^k)\left(\frac{k_1(i)-np}{p(1-p)}\right) \\
&= \frac{1}{p(1-p)}\sum_{i=1}^{N} P(y(i)=1|k_1(i),\theta^k)k_1(i) - \frac{1}{p(1-p)}\sum_{i=1}^{N} P(y(i)=1|k_1(i),\theta^k)np \\
&= 0
\end{aligned}$$

Therefore, the value that maximizes p is

$$p^{k+1} = \frac{\sum_{i=1}^{N} P(y(i)=1|k_1(i),\theta^k)k_1(i)}{n\sum_{i=1}^{N} P(y(i)=1|k_1(i),\theta^k)}$$

To maximize q, we take a similar derivation from the one taken for p. Therefore, the value that maximizes q is

$$q^{k+1} = \frac{\sum_{i=1}^{N} P(y(i)=0|k_1(i),\theta^k)k_1(i)}{n\sum_{i=1}^{N} P(y(i)=0|k_1(i),\theta^k)}$$

4 Method of Moments

For our Method of Moments estimator we used the first, second and third order moments of the value of the number of positive samples to solve for p, q, α. To solve for the values we used the matlab function to solve for a series of unknowns. The moments are given by the following equations.

$$\begin{aligned}
E[K1] &= \alpha np + (1-\alpha)nq = \overline{K1} \\
E[K1^2] &= \alpha(n(n-1)p^2 + np) + (1-\alpha)(n(n-1)q^2 + nq) = \overline{K1^2} \\
E[K1^3] &= \alpha(n(n-1)(n-2)p^3 + 3n(n-1)p^2 + np) + (1-\alpha)(n(n-1)(n-2)q^3 + 3n(n-1)q^2 + nq) = \overline{K1^3}
\end{aligned}$$

Where

$$\overline{K1} = \frac{1}{N}\sum_{i=1}^{N} k_1(i)$$
$$\overline{K1^2} = \frac{1}{N}\sum_{i=1}^{N} k_1(i)^2$$
$$\overline{K1^3} = \frac{1}{N}\sum_{i=1}^{N} k_1(i)^3$$

5 Experiments

We used the Expectation-Maximization (EM) estimator and Method of Moments (MOM) estimator to estimate the parameters of the discrete mixture of models described in Section 1. We compare the estimators with the CRLB derived in Section 2.
The parameters in our model are α, p, and q. We generated data using fixed values for these parameters, and then ran the estimators to find the values for these parameters. We applied the process for $p = 0.2$, $q = 0.4$, and for $\alpha = 0.1, 0.2, \cdots, 0.9$. We set the length of each sequence to be $n = 20$, and generated $N = 200$ i.i.d. sequences. We ran 200 independent Monte-Carlo runs. We ran the EM algorithm for 100 iterations.
We compare the mean squared error (MSE) of each estimator and compare it to the CRLB. The results are shown in Figure 2. As can be seen, the estimators have a bigger MSE when $\alpha = 0.1$ or $\alpha = 0.9$. This is because with $\alpha = 0.1$, the data generator obtained sequences from the distribution $P(k_i(i)|p, n)$ with small probability. The same happens when $\alpha = 0.9$, where the data generator obtained sequences from the distribution $P(k_i(i)|q, n)$ with small probability. The higher the probability that the data generator obtained sequences from a distribution, the easier it is for the estimators to estimate the correct value. The performance of EM and MOM estimators is very similar, with differences mainly with small and large values for α.
As can be seen, the CRLB is very low, compared to both estimators. To visualize the difference more clearly, we plotted the same results with a log-scale on the y axis. The results are shown in Figure 3. As can be seen, the CRLB is always more than 2 orders of magnitude lower than the results obtained by the estimators. We believe this is because the CRLB is not a tight bound. Therefore, in this case, the results obtained by the estimators do not get close to the CRLB.

5.1 EM Initialization

The performance of the EM algorithm may be affected by the initialization of the parameters. To evaluate the impact of initialization, we experimented with three methods of initialization. The first method consists in setting initial values by selecting a random value from a uniform distribution between 0 and 1. This is the easiest method. However, the estimator may get stuck in a local minima. In the second method, we use K-means clustering [1]. This is a more informed method of initialization, without requiring much computational cost. For the third method, we set the initial values to be the exact values that we used to generate the data. Although this is not reasonable in a practical sense, we consider this to be an "empirical lower bound". The results of running the EM algorithm with the described initialization methods are shown in Figure 4. As can be seen, random initialization performs worse when α is low. However, it reaches a similar performance as other initialization methods with greater values of α. K-Means initialization helps the estimator to obtain a lower MSE and get closer to our empirical lower bound.

6 Conclusion

We derived the FIM and CRLB for estimating the parameters of a mixture model of two binomial distributions with $M = 2$ different categories. Our derivations can easily be extended to multimodal distributions

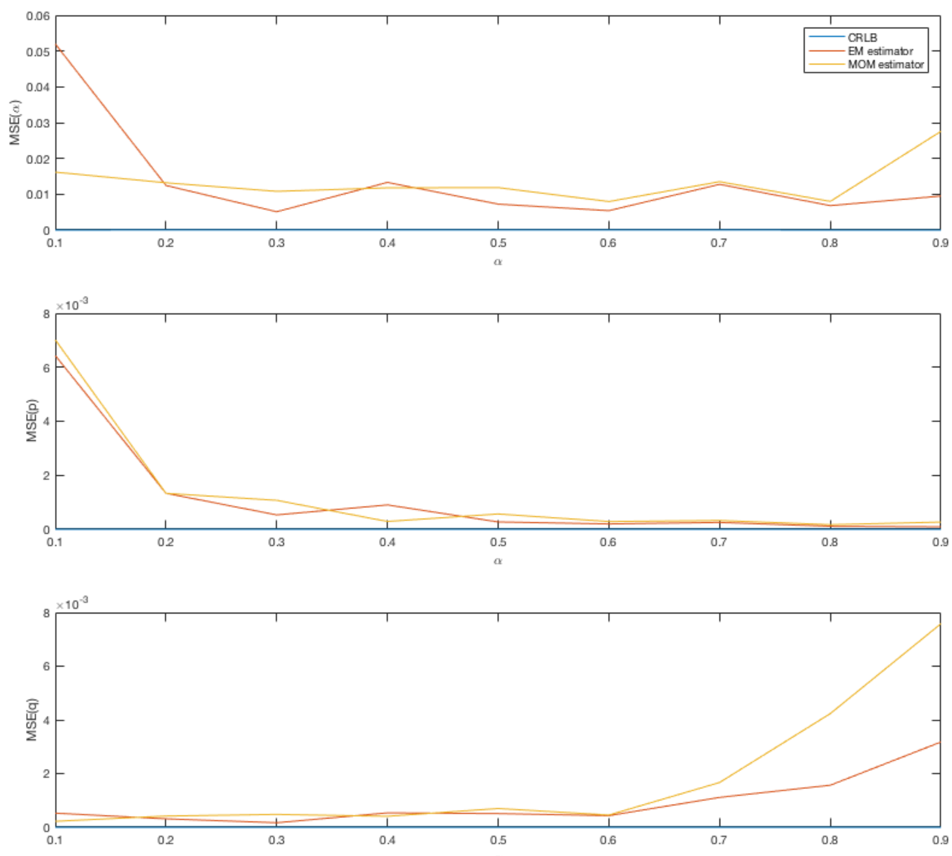

Figure 2: CRLB, EM and MOM results for estimating α (top), p, (middle) and q (bottom).

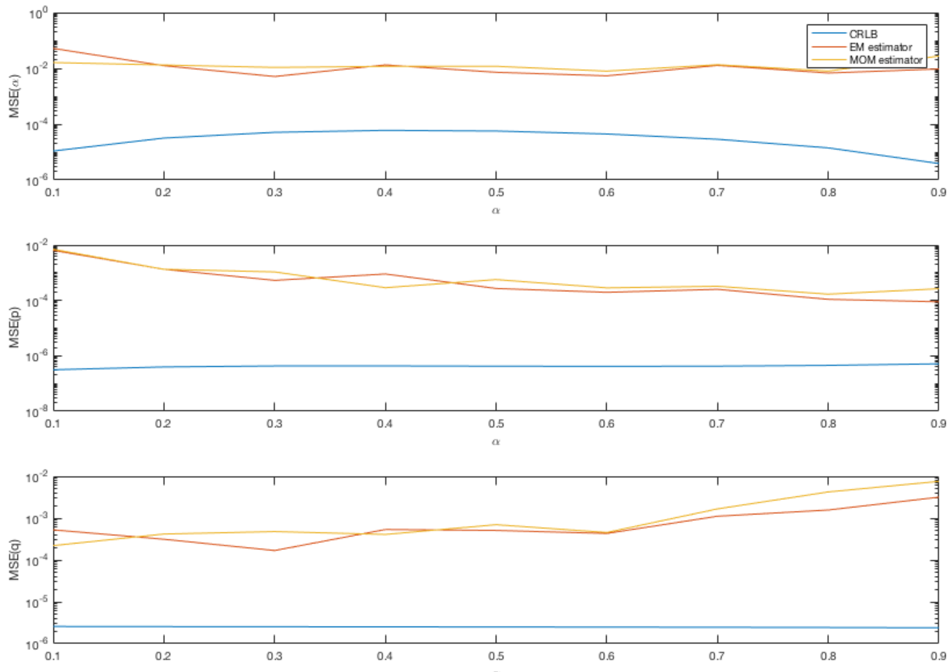

Figure 3: CRLB, EM and MOM results for estimating α (top), p, (middle) and q (bottom) with log-scale on the y axis.

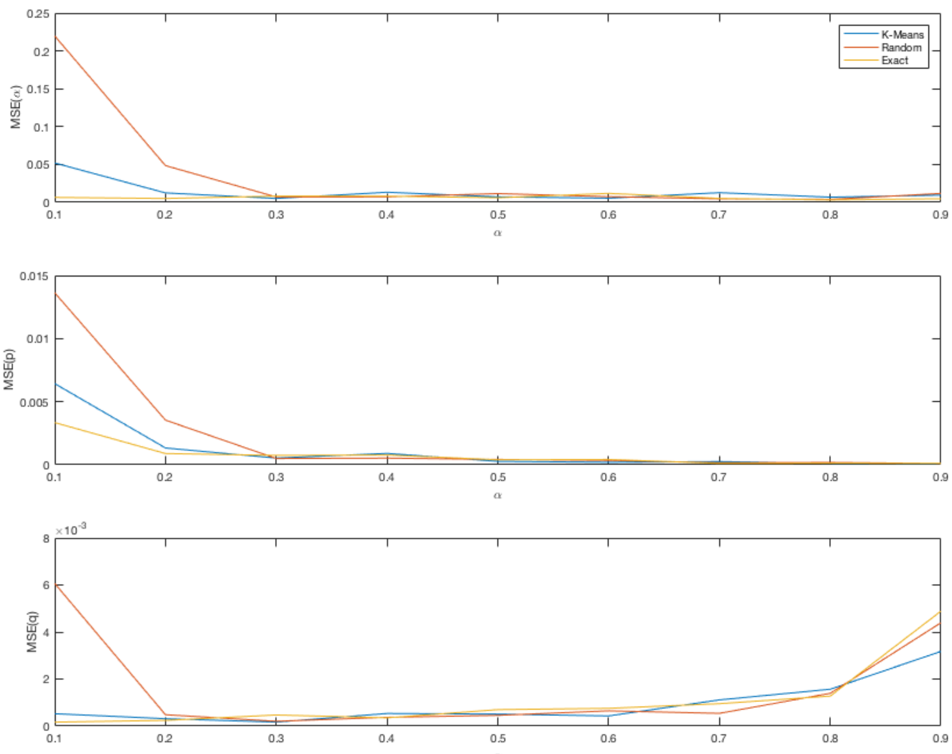

Figure 4: EM results with different initializations for estimating α (top), p, (middle) and q (bottom).

with any number of categories. We showed the Maximum Likelihood equations for our problem. Because the equations do not have a closed-form solution, we used the EM algorithm to obtain an esimator. We derived the equations needed to perform the E-step and M-step. We also showed how to use a MOM estimator, where we simply use the first, second, and third order moments to generate a system of equations. By solving this system, we obtain the estimates for our parameters. We showed in the empirical results that CRLB is a very low bound, as EM and MOM do not get close to it. EM and MOM estimators have a similar performance, with differences only in the edge cases. We also showed the impact of initialization on the EM algorithm.

References

[1] J. Blömer and K. Bujna. Simple methods for initializing the EM algorithm for gaussian mixture models. *CoRR*, abs/1312.5946, 2013.

A Derivations

As a reminder here are the equations for P

$$
\begin{aligned}
P(k|p,n) &= \binom{n}{k} p^k (1-p)^{n-k} \\
P'(k|p,n) &= \frac{dP}{dp} = \binom{n}{k} \left[k p^{k-1} (1-p)^{n-k} + (n-k) p^k (1-p)^{n-k-1} \right] \\
P"(k|p,n) &= \frac{d^2 P}{dp^2} = \binom{n}{k} \left[k(k-1) p^{k-2} (1-p)^{n-k} - (-k+n-1)(n-k) p^k (1-p)^{-k+n-2} \right]
\end{aligned}
$$

Deriving derivatives for the FIM beginning with the first derivative of L(θ)

$$
\frac{\delta \log L}{\delta \theta} = N \begin{bmatrix} \frac{P(k|p,n) - P(k|q,n)}{\alpha P(k|p,n) + (1-\alpha) P(k|q,n)} \\ \frac{\alpha P'(k|p,n)}{\alpha P(k|p,n) + (1-\alpha) P(k|q,n)} \\ \frac{(1-\alpha) P'(k|q,n)}{\alpha P(k|p,n) + (1-\alpha) P(k|q,n)} \end{bmatrix}
$$

The second derivatives were derived as follows

$$
\begin{aligned}
F_{\alpha\alpha} &= \frac{\partial^2 \log P(\theta)}{\partial \alpha^2} \\
F_{\alpha\alpha} &= \frac{\partial}{\partial \alpha} \frac{P(k|p,n) - P(k|q,n)}{\alpha P(k|p,n) + (1-\alpha) P(k|q,n)} \\
F_{\alpha\alpha} &= \frac{-(P(k|p,n) - P(k|q,n))^2}{(\alpha P(k|p,n) + (1-\alpha) P(k|q,n))^2}
\end{aligned}
$$

$$F_{\alpha p} = \frac{\partial^2 \log L(\theta)}{\partial \alpha \partial p}$$
$$F_{\alpha p} = \frac{\partial}{\partial p} \frac{P(k|p,n) - P(k|q,n)}{\alpha P(k|p,n) + (1-\alpha)P(k|q,n)}$$
$$F_{\alpha p} = \frac{P'(k|p,n)}{\alpha P(k|p,n) + (1-\alpha)P(k|q,n)} - \frac{(P(k|p,n) - P(k|q,n))\alpha P'(k|p,n)}{(\alpha P(k|p,n) + (1-\alpha)P(k|q,n))^2}$$
$$F_{\alpha p} = \frac{\alpha P'(k|p,n)P(k|p.n) + (1-\alpha)P(k|p,n)P(k|q,n) - \alpha P'(k|p,n)P(k|p,n) + \alpha P'(k|p,n)P(k|q,n)}{(\alpha P(k|p,n) + (1-\alpha)P(k|q,n))^2}$$
$$F_{\alpha p} = \frac{P'(k|p,n)P(k|q,n)}{(\alpha P(k|p,n) + (1-\alpha)P(k|q,n))^2}$$

$$F_{\alpha q} = \frac{\partial^2 \log L(\theta)}{\partial \alpha \partial q}$$
$$F_{\alpha q} = \frac{\partial}{\partial q} \frac{P(k|p,n) - P(k|q,n)}{\alpha P(k|p,n) + (1-\alpha)P(k|q,n)}$$
$$F_{\alpha p} = \frac{-P'(k|q,n)}{\alpha P(k|p,n) + (1-\alpha)P(k|q,n)} - \frac{(P(k|p,n) - P(k|q,n))(1-\alpha)P'(k|q,n)}{(\alpha P(k|p,n) + (1-\alpha)P(k|q,n))^2}$$
$$F_{\alpha p} = \frac{-\alpha P'(k|q,n)P(k|p,n) - (1-\alpha)P'(k|q,n)P(k|q,n) - (1-\alpha)P'(k|q,n)P(k|p,n) + (1-\alpha)P(k|q,n)P(k|q,n)}{(\alpha P(k|p,n) + (1-\alpha)P(k|q,n))^2}$$
$$F_{\alpha p} = \frac{-P'(k|q,n)P(k|p,n)}{(\alpha P(k|p,n) + (1-\alpha)P(k|q,n))^2}$$

$$F_{pp} = \frac{\partial^2 \log L(\theta)}{\partial p^2}$$
$$F_{pp} = \frac{\partial}{\partial p} \frac{\alpha P'(k|p,n)}{\alpha P(k|p,n) + (1-\alpha)P(k|q,n)}$$
$$F_{pp} = \frac{\alpha P"(k|p,n)}{\alpha P(k|p,n) + (1-\alpha)P(k|q,n)} - \frac{(\alpha P'(k,|p,n))^2)}{(\alpha P(k|p,n) + (1-\alpha)P(k|q,n))^2}$$
$$F_{pp} = \frac{\alpha^2 P(k|p,n)P(k|p,n) + \alpha(1-\alpha)P"(k|p,n)P(k|q,n) - \alpha^2 P'(k|p,n)^2}{(\alpha P(k|p,n) + (1-\alpha)P(k|q,n))^2}$$

$$F_{pq} = \frac{\partial^2 \log L(\theta)}{\partial p \partial q}$$
$$F_{pq} = \frac{\partial}{\partial q} \frac{\alpha P'(k|p,n)}{\alpha P(k|p,n) + (1-\alpha)P(k|q,n)}$$
$$F_{pq} = \frac{-\alpha(1-\alpha)P'(k|p,n)P'(k|q,n)}{(\alpha P(k|p,n) + (1-\alpha)P(k|q,n))^2}$$

$$F_{qq} = \frac{\partial^2 \log L(\theta)}{\partial q^2}$$
$$F_{qq} = \frac{\partial}{\partial q} \frac{(1-\alpha)P'(k|q,n)}{\alpha P(k|p,n) + (1-\alpha)P(k|q,n)}$$
$$F_{qq} = \frac{(1-\alpha)P"(k|q,n)}{\alpha P(k|p,n) + (1-\alpha)P(k|q,n)} - \frac{((1-\alpha)P'(k|q,n))^2}{(\alpha P(k|p,n) + (1-\alpha)P(k|q,n))^2}$$
$$F_{qq} = \frac{\alpha(1-\alpha)P"(k|q,n)P(k|p,n) + (1-\alpha)^2 P"(k|q,n)P(k|q,n) - (1-\alpha)^2 P"(k|q,n)^2}{(\alpha P(k|p,n) + (1-\alpha)P(k|q,n))^2}$$