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:

$$\phi_1(k_1(i)) = p(k_1(i)|p) = \binom{n}{k_1(i)} p^{k_1(i)} (1-p)^{n-k_1(i)}$$

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

$$\phi_0(k_1(i)) = p(k_1(i)|q) = \binom{n}{k_1(i)} q^{k_1(i)} (1-q)^{n-k_1(i)}$$

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 log-likihood 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

$$P(k|p,n) = \binom{n}{k} p^k (1-p)^{n-k}$$
$$P'(k|p,n) = \frac{dP}{dp} = P"(k|p,n) \qquad = \frac{d^2P}{d^2p} =$$

The Fisher information matrix is then:

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

Where

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

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

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

The maximum log-likelihood is given by

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

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 $p(y(i) = c) = \alpha_c$. Because we only have two classes, then

$$p(y(i) = 1) = \alpha$$
$$p(y(i) = 0) = 1 - p(y(i) = 1) = 1 - \alpha$$

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

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

In the E-step of the EM algorithm, we compute $p(y(i) = c|k_1(i), \theta')$. 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

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

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

$$\begin{aligned}
\theta^{k+1} = \underset{\theta}{\operatorname{argmax}} \sum_{i=1}^{N} [&p(y(i)=1|k_1(i), \theta^k)(\log \phi_1(k_1(i)) + \log \alpha) + \\
&p(y(i)=0|k_1(i), \theta^k)(\log \phi_0(k_1(i)) + \log(1-\alpha))]
\end{aligned}$$

We first maximize α. Using Lagrangiang, we get

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

Then, taking the derivative

$$\frac{\delta L}{\delta \alpha} = \sum_{i=1}^{N}\sum_{l=1}^{M} p(y(i)=c|k_1(i), \theta^k)\frac{1}{\alpha_l} + \lambda \qquad = 0 \alpha_l = -\frac{1}{\lambda}$$

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

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

Then, we get:

$$\alpha_1^{k+1} = \alpha^{k+1} = \frac{1}{N}\sum_{i=1}^{N} p(y(i) = 1|k_1(i), \theta^k)$$
$$\alpha_0^{k+1} = 1 - \alpha^{k+1}$$
$$p^{k+1} = \frac{\sum_{i=1}^{N} p(y(i) = 1|k_1(i), \theta^k)k_1(i)}{n\sum_{i=1}^{N} p(y(i) = 1|k_1(i), \theta^k)}$$
$$q^{k+1} = \frac{\sum_{i=1}^{N} p(y(i) = 0|k_1(i), \theta^k)k_1(i)}{n\sum_{i=1}^{N} p(y(i) = 0|k_1(i), \theta^k)}$$

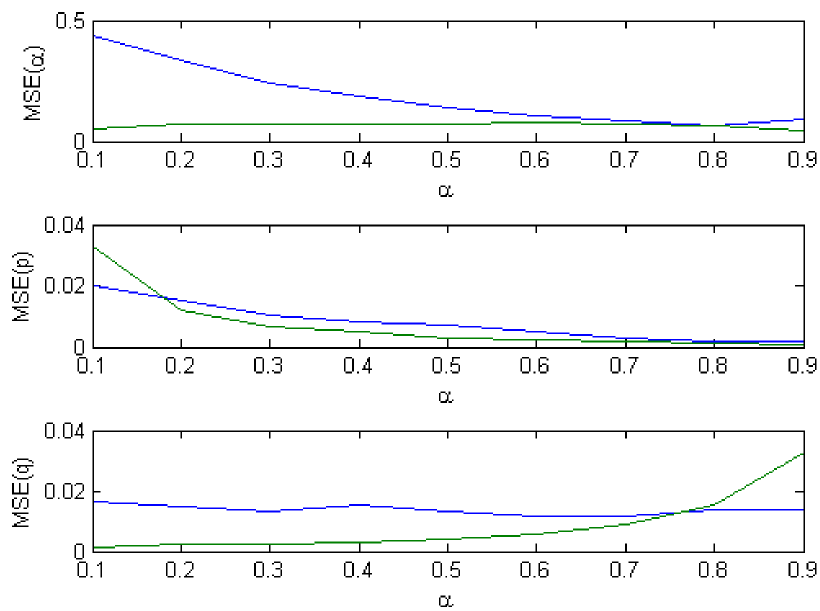

Figure 1: CRLB (green) vs EM (blue).

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.

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

Where

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