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:

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

$$(1)$$

The parameter vector for this problem is given by:

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

1

2 FIM and CRLB

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

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

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:

$$p(\mathbf{k_1}, \mathbf{y}|\theta) = \prod_{i=1}^{N} p(k_1(i), y(i)|\theta)$$

$$= \prod_{i=1}^{N} p(k_1(i)|y(i))p(y(i)|\theta)$$

$$= \prod_{i=1}^{N} \prod_{l=1}^{M} (\phi_l(k_1(i)\alpha_l))^{I(y(i)=l)}$$

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:

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

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

In the M-step, we compute

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

Then, we get:

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

Stuff im writing

4 Method of Moments