Multinomial distribution [without z_{ij}]

$$P(\vec{x}|\theta) = \sum_{j=1}^{k} P_j \, P(\vec{x_i}|\xi_j)$$

P_j – mixing weight

θ_j – parameters of pdf

$$\sum_{j=1}^{k} P_j = 1$$

$$\theta = \{\theta_1, \theta_2 \cdots \theta_k, P_1 \cdots P_k\}$$

The PDF is a multinomial distribution

$$P(x_i|\theta_j) = \frac{n!}{\prod_{i=1}^{m} n_i!} \prod_{j=1}^{m} P_j^{n_j}$$

The likelihood function is

$$P(\vec{x}|\theta_j) = \prod_{i=1}^{N} \sum_{j=1}^{K} P_j \cdot P(x_i|\theta_j)$$

The log likelihood function

$$\mathcal{L}\phi(X, \theta_j, d_1 \cdots d_k, \Lambda) = \sum_{i=1}^{N} \log \sum_{j=1}^{K} P_j \cdot P(x_i|\theta_j) + \lambda_1 P_1 + \lambda_2 P_2 + \cdots + \lambda_k P_k + \Lambda\left(1 - \sum_{j=1}^{K} P_j\right)$$

To derive the maximum log likelihood, differentiate w.r.t all parameters & equate to zero.

1. $$\frac{\partial \phi (x, \theta_j, d_1 \cdots d_k, \Lambda)}{\partial P_j} = 0$$

② $\frac{\partial \phi}{\partial \theta_j}(x_i, \theta_j, \lambda_1 \cdots \lambda_k, \Lambda) = 0$

③ $\frac{\partial \phi}{\partial \lambda_j}(x_i, \theta_j, \lambda_1 \cdots \lambda_k, \Lambda) = 0$

④ $\frac{\partial \phi}{\partial \Lambda}(x, \theta_j, \lambda_1 \cdots \lambda_k, \Lambda) = 0$

Take ③ & ④

Solve $\frac{\partial \phi}{\partial \lambda_j}(x, \theta, \lambda_1, \cdots \lambda_k, \Lambda) = P_j \neq 0$

As it contradicts constraint

$\lambda_1 = \lambda_2 = \cdots \lambda_k = 0$

$\frac{\partial \phi}{\partial \Lambda}(x, \theta, \lambda_1 \cdots \lambda_k, \Lambda) = 1 - \sum_{j=1}^{k} P_j = 0$

$$\boxed{\sum_{j=1}^{k} P_j = 1} \rightarrow ⑤$$

Take ① $\frac{\partial \phi}{\partial P_j}(x, \theta, \lambda_1 \cdots \lambda_k, \Lambda) = \sum_{i=1}^{N} \frac{1}{\sum_{j=1}^{k} P_j \cdot P(x_i|\theta_j)} \frac{\partial}{\partial P_j}\left\{\sum_{j=1}^{k} P_j \cdot P(x_i|\theta_j)\right\} - \Lambda$

$$\Rightarrow \frac{\sum_{i=1}^{N} P(x_i|\theta_j)}{\sum_{j=1}^{k} P_j \cdot P(x_i|\theta_j)} = 0$$

Multiplying & dividing by P_j

$$\frac{1}{P_j}\left\{\frac{\sum_{i=1}^{N} P_j \cdot P(x_i|\theta_j)}{\sum_{j=1}^{k} P_j \cdot P(x_i|\theta_j)}\right\} - \Lambda = 0$$

By Bayes rule

$$\frac{1}{P_j}\sum_{i=1}^{N} P(j|x_i) - \Lambda = 0$$

$$P_j = \frac{1}{\Lambda}\sum_{i=1}^{N} P(j|x_i) \longrightarrow (6)$$

Sub in (5)

$$\sum_{j=1}^{k}\frac{1}{\Lambda}\sum_{i=1}^{N} P(j|x_i) = 1$$

$$\frac{1}{\Lambda}\sum_{i=1}^{N}\sum_{j=1}^{k} P(j|x_i) = 1$$

$$\frac{1}{\Lambda} N = 1$$

$$\boxed{\Lambda = N} \longrightarrow (7)$$

Sub in (6)

$$\boxed{\hat{P}_j = \frac{1}{N}\sum_{i=1}^{N} P(j|x_i)} \longrightarrow (8)$$

Solve (2)

$$\frac{\partial \phi}{\partial \theta_j}(x, \theta, \lambda_1 \cdots \lambda_k, \Lambda) = \frac{\sum_{i=1}^{N} P_j \frac{\partial}{\partial \theta_j} \sum_{j=1}^{k} P(x_i|\theta_j)}{\sum_{j=1}^{k} P_j \cdot P(x_i|\theta_j)} + \frac{\partial}{\partial \theta_j}\Lambda\left(1 - \sum_{j=1}^{k} P_j\right)$$

$$= \frac{\sum_{i=1}^{N} P_j \cdot P(x_i \mid \theta_j)}{\sum_{j=1}^{K} P_j \cdot P(x_i \mid \theta_j)} \; \frac{\frac{\partial}{\partial \theta_j} \sum_{j=1}^{K} P(x_i \mid \theta_j)}{P(x_i \mid \theta_j)}$$

$$= \sum_{i=1}^{N} P(j \mid x_i) \frac{\partial}{\partial \theta_j} \log P(x_i \mid \theta_j) \longrightarrow (9)$$

$$\log P(x_i \mid \theta_j) = \log \left\{ \frac{n!}{\prod_{i=1}^{m} n_i!} \prod_{j=1}^{m} p_j^{n_j} \right\} = \log n! - \sum_{i=1}^{m} \log n_i! + \sum_{i=1}^{m} n_i$$

Now, to maximise the log likelihood of the PDF adding a new lagrange multiplier (as a property of multinomial distribution)

$$\Lambda(\theta_1 \cdots \theta_m, \lambda) = \log P(x_i|\theta_j) + \lambda\left(1 - \sum_{i=1}^{m} \theta_i\right)$$ [$\because$ in multinomial $\theta_i = P_i$]

Sub in (9)

$$\nabla_{\theta_j,\lambda} \phi(\theta_1 \cdots \theta_m, \lambda) = \nabla_{\theta_j,\lambda}\left\{\sum_{i=1}^{N} P(j|x_i) \log P(x_i|\theta_j)\right\}$$

$$= \sum_{i=1}^{N} P(j|x_i)\left\{\nabla_{\theta_j,\lambda}\left\{\log n! - \sum_{i=1}^{m} n_i! + \sum_{i=1}^{m} n_i \log \theta_i\right\} + \nabla_{\theta_j,\lambda} \lambda\left[-\sum_{i=1}^{m} \theta_i + 1\right]\right\}$$

Deriving w.r.t λ

$$\frac{\partial}{\partial \lambda}\phi(\theta_1, \cdots \theta_m, \lambda) = -1 \cdot \sum_{i=1}^{m} \theta_i + 1$$

$$0 = -\sum_{i=1}^{m} \theta_i + 1$$

$$\boxed{\sum_{i=1}^{m} \theta_i = 1} \longrightarrow (10)$$

$$\frac{\partial}{\partial \theta_j}\phi(\theta_1 \cdots \theta_m, \lambda) = \sum_{i=1}^{N} P(j|x_i) \sum_{i=1}^{m} \frac{n_i}{\theta_i} - \lambda \sum_{i=1}^{N} P(j|x_i)$$

$$0 = \sum_{i=1}^{N} P(j|x_i) \sum_{i=1}^{m} \frac{n_i}{\theta_i} - \lambda \sum_{i=1}^{N} P(j|x_i)$$

$$\lambda \sum_{i=1}^{N} P(j|x_i) = \sum_{i=1}^{N} P(j|x_i) \sum_{i=1}^{m} \frac{n_i}{\theta_i}$$

$$\theta_i = \sum_{i=1}^{m} \frac{n_i \sum_{i=1}^{N} P(j|x_i)}{\lambda \sum_{i=1}^{N} P(j|x_i)}$$

To obtain the maximum θ_i , $\lambda = n$

$$\boxed{\hat{\theta}_i = \sum_{i=1}^{m} \frac{n_i \sum_{i=1}^{N} P(j|x_i)}{n \sum_{i=1}^{N} P(j|x_i)}} \longrightarrow (11)$$

Soln: (11) and (8)