This example assume we know the latent parameter Z.

* We observed a sample of 1000 data points. This sample follow a mixture of 2 Poisson distribution. Let Z_i be the distribution that a data point belong to

$$Z_i \sim \text{Categorical}(\pi_1, 1-\pi_1),$$

$$X_i \mid Z_i = 1 \sim \text{Poisson}(\lambda_1) \text{ and } X_i \mid Z_i = 2 \sim \text{Poisson}(\lambda_2)$$

* We aim to obtain MLE of parameters $\theta = (\pi_1, \lambda_1, \lambda_2)$ using the EM algorithm.

Solution :

E-Step: $\theta^0 = (\pi_1^0, \lambda_1^0, \lambda_2^0)$

computing posteriors:

$$P(Z_i = 1 \mid X, \theta^0) = \frac{P(Z_i = 1, X_i \mid \theta^0)}{P(X_i \mid \theta^0)}$$

$$= \frac{p(X_i \mid Z_i = 1, \theta^0)\, p(Z_i = 1 \mid \theta^0)}{\sum_{k=1}^{2} P(X_i \mid Z_i = k, \theta^0) \cdot P(Z_i = k \mid \theta^0)}$$

$$= \frac{p(X_i \mid Z_i = 1, \theta^0)\, p(Z_i = 1 \mid \theta^0)}{P(X_i \mid Z_i = 1, \theta^0) P(Z_i = 1 \mid \theta^0) + P(X_i \mid Z_i = 2, \theta^0) P(Z_i = 2 \mid \theta^0)}$$

$$= \frac{\frac{(\lambda_1^0)^{x_i} e^{-\lambda_1^0}}{x_i!}\pi_1^0}{\frac{(\lambda_1^0)^{x_i} e^{-\lambda_1^0}}{x_i!}\pi_1^0 + \frac{(\lambda_2^0)^{x_i} e^{-\lambda_1^0}}{x_i!}(1-\pi_1^0)}$$

* Now that we have $P(Z_i = 1 \mid X, \theta^0)$ we can get

$$P(Z_i = 2 \mid X, \theta^0) = 1 - P(Z_i = 1 \mid X, \theta^0)$$

$$= 1 - \frac{\frac{(\lambda_1^0)^{x_i} e^{-\lambda_1^0}}{x_i!}\pi_1^0}{\frac{(\lambda_1^0)^{x_i} e^{-\lambda_1^0}}{x_i!}\pi_1^0 + \frac{(\lambda_2^0)^{x_i} e^{-\lambda_1^0}}{x_i!}(1-\pi_1^0)}$$

M-Step :

* We now have to derive new parameter estimate from the previous posteriors:

$$\hat{\theta} = \text{argmax}_{\theta}\, Q(\theta, \theta^0)$$

where $Q(\theta, \theta^0) = E_{Z|X,\theta^0}\left[\log(P(X, Z \mid \theta))\right]$

for $n = 100$

$$P(X_1, \dots, X_n, Z_1, \dots, Z_n) = \prod_{i=1}^{n} P(X_i | Z_i, \theta) P(Z_i | \theta)$$

$$= \prod_{i=1}^{n} \prod_{K=1}^{2} \left[P(X_i | Z_i = K, \theta) P(Z_i = K | \theta) \right]^{I(Z_i = K)}$$

$$\log\left(P(X_1, \dots, X_n, Z_1, \dots, Z_n)\right)$$

$$= \sum_{i=1}^{n} \sum_{K=1}^{2} I_{(Z_i = K)} \left[\log(P(X_i | Z_i = K, \theta)) + \log(P(Z_i = K | \theta)) \right]$$

$$= \sum_{i=1}^{n} \left[I_{(Z_i = 1)} \left[X_i \log \lambda_1 - \lambda_1 - \log(X_i!) + \log(\pi_1) \right] \right.$$

$$\left. + I_{(Z_i = 2)} \left[X_i \log \lambda_2 - \lambda_2 - \log(X_i!) + \log(1 - \pi_1) \right] \right]$$

$$Q(\theta, \theta^0) = E_{Z|X, \theta^0} \left[\log(P(X, Z | \theta)) \right]$$

$$\boxed{\begin{aligned} = \sum_{i=1}^{n} \Big[& P(Z_i = 1 | X, \theta) \left[X_i \log \lambda_1 - \lambda_1 - \log(X_i!) + \log(\pi_1) \right] \\ + & P(Z_i = 2 | X, \theta) \left[X_i \log \lambda_2 - \lambda_2 - \log(X_i!) + \log(1 - \pi_1) \right] \Big] \end{aligned}}$$

$$\theta = (\pi_1, \lambda_1, \lambda_2)$$

* now we need to get $\hat{\theta} = \text{argmax}_{\theta} Q(\theta, \theta^0)$

$$\frac{dQ(\theta,\theta^0)}{d\pi_1} = \sum_{i=1}^{n}\left[\frac{P(Z_i=1 \mid X,\theta^0)}{\pi_1} - \frac{P(Z_i=2 \mid X,\theta^0)}{1-\pi_1}\right]$$

$$= \sum_{i=1}^{n}\left[\frac{(1-\pi_1)P(Z_i=1 \mid X,\theta^0) - \pi_1 P(Z_i=2 \mid X,\theta^0)}{\pi_1(1-\pi_1)}\right]$$

$$= 0$$

$$\Rightarrow \frac{(1-\pi_1)\sum_{i=1}^{n} P(Z_i=1 \mid X,\theta^0) - \pi_1 \sum_{i=1}^{n}\left(1 - P(Z_i=1 \mid X,\theta^0)\right)}{\pi_1(1-\pi_1)}$$

$$= \frac{(1-\pi_1)\sum_{i=1}^{n} P(Z_i=1 \mid X,\theta^0) - \pi_1 n + \pi_1 \sum_{i=1}^{n} P(Z_i=1 \mid X,\theta^0)}{\pi_1(1-\pi_1)}$$

$$= \frac{\sum_{i=1}^{n} P(Z_i=1 \mid X,\theta^0) - n\pi_1}{\pi_1(1-\pi_1)} = 0$$

$$= \sum_{i=1}^{n} P(Z_i=1 \mid X, \theta^0) - n\pi_1 = 0$$

$$\Rightarrow \hat{\pi}_1 = \sum_{i=1}^{n} P(Z_i=1 \mid X, \theta^0) / n$$

* $P(Z_1=1 \mid X, \theta^0)$ was obtained previously in the E-Step

$$\frac{dQ(\theta, \theta^0)}{d\lambda_1}$$

$$= \frac{\sum_{i=1}^{n} P(Z_i=1 \mid X, \theta^0) X_i}{\lambda_1} - \sum_{i=1}^{n} P(Z_i=1 \mid X, \theta^0) = 0$$

$$\Rightarrow \hat{\lambda}_1 = \frac{\sum_{i=1}^{n} P(Z_i=1 \mid X, \theta^0) X_i}{\sum_{i=1}^{n} P(Z_i=1 \mid X, \theta^0)}$$

Similarly,

$$\frac{dQ(\theta, \theta^0)}{d\lambda_2}$$

$$= \frac{\sum_{i=1}^{n} P(Z_i=2 \mid X, \theta^0) X_i}{\lambda_2} - \sum_{i=1}^{n} P(Z_i=2 \mid X, \theta^0) = 0$$

$$\Rightarrow \hat{\lambda}_2 = \frac{\sum_{i=1}^{n} P(Z_i=2 \mid X, \theta^0) X_i}{\sum_{i=1}^{n} P(Z_i=2 \mid X, \theta^0)}$$

now we have parameter estimate after 1 iteration