

---

 EXPECTATION MAXIMIZATION POUR LES MÉLANGES DE LOIS DE POISSON
 

---

1. Considérons  $\{V_k\}_{1 \leq k \leq K}$  indépendantes et telles que  $V_k \sim \mathcal{P}(\lambda_k)$  pour  $1 \leq k \leq K$  et  $Z$  une variable aléatoire de loi multinomiale de paramètres  $\{\pi_1, \dots, \pi_K\}$  indépendante des  $\{V_k\}_{1 \leq k \leq K}$ . Il suffit alors de poser  $X = V_Z$ .
2. Par définition, pour tout  $j \leq 0$ ,

$$\begin{aligned} \mathbb{P}_\theta(X = j) &= \sum_{k=1}^K \mathbb{P}_\theta(X = j | Z = k) \mathbb{P}_\theta(Z = k) , \\ &= \sum_{k=1}^K e^{-\lambda_k} \frac{\lambda_k^j}{j!} \pi_k . \end{aligned}$$

3. Écrivons la logvraisemblance:

$$\begin{aligned} \log \mathbb{P}_\theta(X_{1:n} = x_{1:n}) &= \sum_{i=1}^n \log \mathbb{P}_\theta(X_i = x_i) , \\ &= \sum_{i=1}^n \log \left( \sum_{k=1}^K e^{-\lambda_k} \frac{\lambda_k^{x_i}}{x_i!} \pi_k \right) , \\ &= \sum_{i=1}^n \left\{ -\log x_i! + \log \left( \sum_{k=1}^K e^{-\lambda_k} \lambda_k^{x_i} \pi_k \right) \right\} . \end{aligned}$$

On remarque ensuite aisément que l'équation  $\nabla_\theta \log \mathbb{P}_\theta(X_{1:n} = x_{1:n}) = 0$  n'admet pas de solution explicite.

4. (a) Pour tout  $\theta$  et tout  $k \in \{1, \dots, K\}$ ,  $j \geq 0$ ,

$$\mathbb{P}_\theta(Z = k | X = j) = \frac{\mathbb{P}_\theta(Z = k; X = j)}{\mathbb{P}_\theta(X = j)} = \frac{\pi_k e^{-\lambda_k} \lambda_k^j / j!}{\sum_{\ell=1}^K \pi_\ell e^{-\lambda_\ell} \lambda_\ell^j / j!} .$$

- (b) La logvraisemblance complète des données est :

$$\mathcal{L}(x, z; \theta) = \sum_{k=1}^K \mathbb{1}_{z=k} \left\{ \log \pi_k + \log \left( e^{-\lambda_k} \frac{\lambda_k^x}{x!} \right) \right\} .$$

Puisque les données sont indépendantes, on obtient,

$$\begin{aligned} \mathcal{L}(x, z; \theta) &= \sum_{i=1}^n \sum_{k=1}^K \mathbb{1}_{z_i=k} \left\{ \log \pi_k + \log \left( e^{-\lambda_k} \frac{\lambda_k^{x_i}}{x_i!} \right) \right\} , \\ &= \sum_{i=1}^n \sum_{k=1}^K \mathbb{1}_{z_i=k} \{ \log \pi_k - \lambda_k + x_i \log(\lambda_k) - \log x_i! \} . \end{aligned}$$

(c) Pour tout  $\theta, \theta'$ ,

$$\begin{aligned} Q(\theta; \theta') &= \mathbb{E}_{\theta'} [\mathcal{L}(X_{1:n}, Z_{1:n}; \theta) | X_{1:n}] , \\ &= \sum_{i=1}^n \sum_{k=1}^K \mathbb{E}_{\theta'} [\mathbb{1}_{Z_i=k} | X_{1:n}] \{ \log \pi_k - \lambda_k + X_i \log(\lambda_k) - \log X_i! \} , \\ &= \sum_{i=1}^n \sum_{k=1}^K \mathbb{P}_{\theta'} (Z_i = k | X_i) \{ \log \pi_k - \lambda_k + X_i \log(\lambda_k) - \log X_i! \} \end{aligned}$$

À l'itération  $p \geq 0$  de l'algorithme, si nous disposons d'un estimateur courant  $\theta^{(p)}$ , nous calculons

$$\omega_{i,k}^{(p)}(X_i) = \mathbb{P}_{\theta^{(p)}} (Z_i = k | X_i) = \frac{\pi_k^{(p)} e^{-\lambda_k^{(p)}} (\lambda_k^{(p)})^j / j!}{\sum_{\ell}^K \pi_{\ell}^{(p)} e^{-\lambda_{\ell}^{(p)}} (\lambda_{\ell}^{(p)})^j / j!}$$

et

$$Q(\theta; \theta^{(p)}) = \sum_{i=1}^n \sum_{k=1}^K \omega_{i,k}^{(p)}(X_i) \{ \log \pi_k - \lambda_k + X_i \log(\lambda_k) - \log X_i! \} .$$

(d) Il est aisé de montrer que la fonction  $\theta \mapsto Q(\theta; \theta^{(p)})$  admet un maximum unique, obtenu en résolvant l'équation  $\nabla_{\theta} Q(\theta; \theta^{(p)}) = 0$ . Pour tout  $1 \leq k \leq K$ ,

$$\partial_{\lambda_k} Q(\theta; \theta^{(p)}) = \sum_{i=1}^n \omega_{i,k}^{(p)}(X_i) \left\{ -1 + \frac{X_i}{\lambda_k} \right\} .$$

On en déduit que

$$\lambda_k^{(p+1)} = \frac{\sum_{i=1}^n \omega_{i,k}^{(p)}(X_i) X_i}{\sum_{i=1}^n \omega_{i,k}^{(p)}(X_i)} .$$

Par ailleurs, pour tout  $1 \leq k \leq K-1$ , en utilisant que  $\pi_K = 1 - \sum_{j=1}^{K-1} \pi_j$ ,

$$\partial_{\pi_k} Q(\theta; \theta^{(p)}) = \sum_{i=1}^n \left\{ \frac{\omega_{i,k}^{(p)}(X_i)}{\pi_k} - \frac{\omega_{i,k}^{(p)}(X_i)}{\pi_K} \right\}$$

et on en déduit que  $k \mapsto \sum_{i=1}^n \omega_{i,k}^{(p)}(X_i) / \pi_k$  est constante. En utilisant par ailleurs que  $\sum_{k=1}^K \pi_k = 1$  et  $\sum_{k=1}^K \omega_{i,k}^{(p)}(X_i) = 1$ , on a

$$\pi_k^{(p+1)} = \frac{1}{n} \sum_{i=1}^n \omega_{i,k}^{(p)}(X_i) .$$

- (e) Pour mettre en place l'algorithme EM, il suffit d'initialiser l'algorithme avec une valeur  $\theta^{(0)}$  puis à chaque itération  $p \geq 0$  d'effectuer l'étape E (i.e. calculer  $\omega_{i,k}^{(p)}$  pour  $1 \leq i \leq n$  et  $1 \leq k \leq K$ ) et de calculer  $\theta^{(p+1)}$  en appliquant les mises à jour de la question précédente.
- (f) La seule garantie que nous avons est que la vraisemblance des observations augmente à chaque itération. En pratique, l'algorithme converge vers un maximum local, et il faut donc analyser les différents points de convergence obtenus si on initialise l'algorithme de différentes façons.