Mélange de lois de Poisson

Considérons un mélange de K lois de Poisson. Pour $1 \leq k \leq K$, nous noterons $\lambda_k > 0$ le paramètre de la k-ème composante et $\pi_k \in (0,1)$ son poids. Notons $\theta = (\pi_1, \dots, \pi_K, \lambda_1, \dots, \lambda_K)$ le paramètre inconnu et

$$\Theta = \left\{ \theta = (\pi_1, \dots, \pi_K, \lambda_1, \dots, \lambda_K) \,;\, \forall k \in \{1, \dots, K\}\,,\, \pi_k \in (0,1)\,,\, \lambda_k > 0\,,\, \sum_{k=1}^{K} \pi_k = 1 \right\}.$$

1. Soit $\theta \in \Theta$, expliquer comment construire une variable aléatoire X suivant un mélange de lois de Poisson paramétré par θ.

   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* $X = V_Z = \sum_{k=1}^{K} \mathbb{1}_{Z=k} V_k$.

2. Notons $\mathbb{P}_\theta$ la loi de X. Pour tout $j \geq 0$, calculer $\mathbb{P}_\theta(X = j)$.

   Par définition, pour tout $j \geq 0$,

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

3. Soit $\theta \in \Theta$ et $(x_1, \dots, x_n) \in \mathbb{N}^n$. Calculer $\log \mathbb{P}_\theta(X_{1:n} = x_{1:n})$ où les $(X_i)_{1 \leq i \leq n}$ sont i.i.d de même loi que X.

   Écrivons la logvraisemblance:

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

   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. Puisque nous ne pouvons pas maximiser la logvraisemblance explicitement, nous allons utiliser l'algorithme Expectation Maximization.

   (a) Pour tout $\theta \in \Theta$ et tout $k \in \{1, \dots, K\}$, calculer $\mathbb{P}_\theta(Z = k | X = j)$.

   Pour tout θ *et tout* $k \in \{1, \dots, K\}$, $j \geq 0$,

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

(b) Calculer la logvraisemblance complète des données.

La logvraisemblance complète des données est :

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

Puisque les données sont indépendantes, on obtient,

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

(c) Calculer la quantité intermédiaire de l'algorithme EM.

Pour tout θ, θ',

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

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

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

et

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

(d) En déduire la mise à jour d'une itération de l'algorithme EM.

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) Détailler le fonctionnement complet de l'algorithme EM

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) Cet algorithme converge t'il vers le maximum de vraisemblance ?

La seule garantie que nous avons est que la vraisemblance des observations augmente à chaque itération. On peut montrer sous certaines hypothèses que l'algorithme converge vers un maximum local de la logvraisemblance, et il faut donc analyser les différents points de convergence obtenus si on initialise l'algorithme de différentes façons.

Mélange de régressions linéaires

Considérons des covariables $(x_i)_{i=1}^n$, supposées déterministes, où $x_i \in \mathbb{R}^d$ pour $1 \leq i \leq n$ et le modèle suivant paramétré par

$$\theta = (\pi_1, \ldots, \pi_K, \beta_1, \ldots, \beta_K, \sigma_1^2, \ldots, \sigma_K^2)\,.$$

- Les variables $(Z_i)_{1 \leq i \leq n}$ sont non observées, indépendantes et telles que $Z_i \in \{1, \ldots, K\}$ et $\mathbb{P}_\theta(Z_i = k) = \pi_k$ pour $1 \leq k \leq K$, $1 \leq i \leq n$.
- Conditionnellement aux variables $(Z_i)_{1 \leq i \leq n}$, les observations réelles $(Y_i)_{1 \leq i \leq n}$ sont indépendantes et pout tout $1 \leq i \leq n$, la loi conditionnelle de Y_i est gaussienne, de moyenne $x_i^\top \beta_{Z_i}$ et de variance $\sigma_{Z_i}^2$.

1. Écrire la logvraisemblance complète de $(Z_{1:n}, Y_{1:n})$.

Pour tout $1 \leq i \leq n$, $y_i \in \mathbb{R}$, $z_i \in \{1, \ldots K\}$,

$$p_\theta(y_i, z_i \mid x_i) = \prod_{k=1}^K \left(\pi_k \, \varphi_{x_i^\top \beta_k, \sigma_k^2}(y_i)\right)^{\mathbb{1}_{Z_i = k}}\,,$$

où φ_{μ,σ^2} *est la densité de la loi gaussienne de moyenne* μ *et de variance* σ^2. *Ainsi,*

$$\log p_\theta(y_{1:n}, z_{1:n} \mid x_{1:n}) = \sum_{i=1}^n \sum_{k=1}^K \mathbb{1}_{Z_i = k} \log \left(\pi_k \, \varphi_{x_i^\top \beta_k, \sigma_k^2}(y_i)\right)\,.$$

Finalement,

$$\log p_\theta(y_{1:n}, z_{1:n} \mid x_{1:n}) = \sum_{i=1}^n \sum_{k=1}^K \mathbb{1}_{Z_i = k} \left\{\log \pi_k - \frac{1}{2}\log(2\pi\sigma_k^2) - \frac{1}{2\sigma_k^2}\left(y_i - x_i^\top \beta_k\right)^2\right\}\,.$$

2. Soit $\theta^{(p)}$ l'estimateur de θ à l'itération $p \geq 0$ de l'algorithme EM. Donner la quantité intermédiaire $\theta \mapsto Q(\theta, \theta^{(p)})$.

Dans la suite, nous notons $v_k = \sigma_k^2$, *pour* $1 \leq k \leq K$. *Pour tout* $1 \leq i \leq n$, $1 \leq k \leq K$, *par la formule de Bayes,*

$$\omega_{ik}^{(p)} = \mathbb{P}_{\theta^{(p)}}(Z_i = k \mid x_i, y_i) = \frac{\pi_k^{(p)} \, \varphi_{x_i^\top \beta_k^{(p)}, v_k^{(p)}}(y_i)}{\sum_{j=1}^K \pi_j^{(p)} \, \varphi_{x_i^\top \beta_k^{(p)}, v_k^{(p)}}(y_i)}.$$

Ainsi, pour tout $\theta \in \Theta$,

$$Q(\theta, \theta^{(p)}) = \sum_{i=1}^n \sum_{k=1}^K \omega_{ik}^{(p)} \left\{\log \pi_k - \frac{1}{2}\log(2\pi v_k) - \frac{1}{2v_k}\left(y_i - x_i^\top \beta_k\right)^2\right\}\,.$$

3. Donner la mise à jour des poids π_k, $1 \leq k \leq K$.

   En utilisant le fait que $\pi_K = 1 - \sum_{k=1}^{K-1} \pi_k$, *nous obtenons, pour tout* $1 \leq k \leq K-1$,

$$\frac{\sum_{i=1}^n \omega_{ik}^{(p)}}{\pi_k} = \frac{\sum_{i=1}^n \omega_{iK}^{(p)}}{\pi_K}.$$

   Ainsi la fonction $k \mapsto \sum_{i=1}^n \omega_{ik}^{(p)} / \pi_k$ *est constante sur* $\{1, \ldots, K\}$. *On en déduit que, pour tout* $1 \leq k \leq K$

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

4. Donner la mise à jour des β_k, $1 \leq k \leq K$.

   On remarque que la maximisation de la quantité intermédiare par rapport aux coefficients de régression revient à la résolution d'un problème de moindres carrés pondérés. En introduisant X *la matrice de* $\mathbb{R}^{n \times d}$ *telle que la i-ème ligne de* X *est* $x_i^\top$, $y = (y_1, \ldots, y_n)^\top$ *et* $W_k = \mathrm{diag}(\omega_{1k}^{(p)}, \ldots, \omega_{nk}^{(p)})$, *on souhaite résoudre*

$$(X^\top W_k X)\beta_k = X^\top W_k y,$$

   ce qui conduit à

$$\beta_k^{(p+1)} = (X^\top W_k X)^{-1} X^\top W_k y.$$

5. Donner la mise à jour des variances σ_k^2, $1 \leq k \leq K$.

   Pour tout $1 \leq k \leq K$, *il suffit de résoudre*

$$\partial_{v_k} Q(\theta, \theta^{(p)}) = 0,$$

   ce qui conduit à

$$v_k^{(p+1)} = \frac{\sum_{i=1}^n \omega_{ik}^{(p)} \left(y_i - x_i^\top \beta_k^{(p+1)} \right)^2}{\sum_{i=1}^n \omega_{ik}^{(p)}}.$$