TD/TP : Régression à processus logistique latent

On considère une série temporelle $(Y_1, \ldots, Y_n)$ régie par un processus latent $(Z_1, \ldots, Z_n)$ à deux états où Z_t représente l'état (binaire dans ce cas) à l'instant $t\,(t = 1, \ldots, n)$. $Z_t \in \{0, 1\}$. On dispose d'une série observée $(y_1, \ldots, y_n)$. L'objectif est de prédire et de segmenter la série sur la base du modèle de régression à processus logistique latent définit par :

$$
\begin{aligned}
f(y_t; \boldsymbol{\theta}) &= \pi(t; \boldsymbol{w}) f(y_t; \boldsymbol{\theta}_1) + (1 - \pi(t; \boldsymbol{w})) f(y_t; \boldsymbol{\theta}_0) \\
&= \pi(t; \boldsymbol{w}) \mathcal{N}(\beta_{10} + \beta_{11} t, \sigma_1^2) + (1 - \pi(t; \boldsymbol{w})) \mathcal{N}(\beta_{00} + \beta_{01} t, \sigma_0^2) \qquad (1)
\end{aligned}
$$

où $\pi(t; \boldsymbol{w})$ est le poids logistique de l'état 1 à l'instant t défini par

$$
\pi(t; \boldsymbol{w}) = \mathbb{P}(Z_t = 1; \boldsymbol{w}) = \frac{\exp(w_0 + w_1 t)}{1 + \exp(w_0 + w_1 t)} \qquad (2)
$$

avec $\mathbb{P}(Y_t = 0; \boldsymbol{\theta}) = \frac{1}{1+\exp(w_0 + w_1 t)}$.

Afin d'apprendre ce modèle en estimant les paramètres $\boldsymbol{\theta} = (\boldsymbol{w}, \boldsymbol{\beta}, \sigma_0^2, \sigma_1^2)$ à partir des données et faire la segmentation de la série, on utilise l'algorithme Espérance-Maximisation (EM).
On montre que pour ce modèle l'algorithme EM consiste à partir d'un modèle initial de paramètres $\boldsymbol{\theta}^{(0)}$ et alterner à chaque itération q entre les deux étapes E- et M- suivantes jusqu'à ce qu'il n'y ait plus d'augmentation significative au sens d'un seuil préfixé de la log-vraisemblance du modèle :

1. Étape E : Calculer les probabilités a posteriori :

$$
\tau_{tk}^{(q)} = \mathbb{P}(Z_t = k | y_t, \boldsymbol{\theta}^{(q)}) = \frac{\pi_k(t; \boldsymbol{w}^{(q)}) \mathcal{N}(y_t; \beta_{k0} + \beta_{k1} t, \sigma_k^2)}{f(y_t; \boldsymbol{\theta}^{(q)})} \qquad (3)
$$

2. Étape M : Mettre à jour les paramètres du modèle $\boldsymbol{\theta}^{(t+1)}$:

pour $k \in \{0, 1\}$:

$$
\begin{aligned}
\beta_{k0}^{(q+1)} &= \frac{\sum_{t=1}^{n} \tau_k(t; \boldsymbol{\theta}^{(q)})(y_t - \beta_{k1}^{(q)} t)}{\sum_{t=1}^{n} \tau_k(t; \boldsymbol{\theta}^{(q)})} \qquad (4) \\
\beta_{k1}^{(q+1)} &= \frac{\sum_{t=1}^{n} \tau_k(t; \boldsymbol{\theta}^{(q)})\ (y_t - \beta_{k0}^{(q+1)} t)}{\sum_{t=1}^{n} \tau_k(t; \boldsymbol{\theta}^{(q)})\, t^2} \qquad (5) \\
& \qquad (6)
\end{aligned}
$$

où de façon équivalente

$$
\begin{aligned}
\boldsymbol{\beta}_k^{(q+1)} &= \Big[\sum_{t=1}^{n} \tau_k(t; \boldsymbol{\theta}^{(q)}) \boldsymbol{x}_t \boldsymbol{x}_t^T \Big]^{-1} \sum_{t=1}^{n} \tau_k(t; \boldsymbol{\theta}^{(q)}) y_t \boldsymbol{x}_t \qquad (7) \\
&= (\mathbf{X}^T \mathbf{W}^{(q)} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{(q)} \boldsymbol{y} \qquad (8)
\end{aligned}
$$

$$
\begin{aligned}
\sigma_k^{2(q+1)} &= \frac{\sum_{t=1}^{n} \tau_k(y_t; \boldsymbol{\theta}^{(q)})(y_t - (\beta_{k0}^{(q+1)} + \beta_{k1}^{(q+1)} t))^2}{\sum_{t=1}^{n} \tau_k(y_t; \boldsymbol{\theta}^{(q)})} \qquad (9) \\
&= (\boldsymbol{y} - \mathbf{X}^T \boldsymbol{\beta}^{(q+1)}) \mathbf{W}^{(q)} (\boldsymbol{y} - \mathbf{X}^T \boldsymbol{\beta}^{(q+1)}) / \mathrm{trace}(\mathbf{W}^{(q)}) \qquad (10)
\end{aligned}
$$

avec $\mathbf{W}^{(q)} = \mathrm{diag}(\tau_{11}, \ldots, \tau_{n1})$ et $\mathbf{X}$ la matrice de dimensions $n \times (d+1)$ de lignes $(1, t)$ pour $t = 1, \ldots, n$.

On montre que la mise à jour des paramètres des poids logistiques s'effectue par l'algorithme de Nerwton-Raphson qui consiste à partir d'un modèle initial de paramètres $\boldsymbol{w}^{(0)}$ et à mettre à jour, à chaque itération m, les paramètres selon l'équation de mise à jour suivante

$$\begin{aligned} \boldsymbol{w}^{(m+1)} &= \boldsymbol{w}^{(m)} + (\mathbf{X}^T \mathbf{W}^{(m)} \mathbf{X})^{-1} \mathbf{X}^T (\boldsymbol{\tau} - \boldsymbol{p}^{(m)}) \\ &= (\mathbf{X}^T \mathbf{W}^{(m)} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{(m)} \widetilde{\boldsymbol{y}} \end{aligned} \tag{11}$$

où :

- $\mathbf{X}$ est la matrice de dimensions $n \times 2$ de lignes $(1, t)$ pour $t = 1, \ldots, n$;
- $\boldsymbol{\tau}$ est le vecteur colonne de dimension $n \times 1$ des probabilités à posteriori τ_{t1} : $\boldsymbol{\tau} = (\tau_{11}, \ldots, \tau_{n1})^T$;
- $\boldsymbol{p}$ est le vecteur colonne de dimension $n \times 1$ des probabilités : $\boldsymbol{p} = (\pi(\boldsymbol{x}_1; \boldsymbol{\theta}), \ldots, \pi(\boldsymbol{x}_n; \boldsymbol{\theta}))^T$ avec $\pi(\boldsymbol{x}_i; \boldsymbol{\theta}) = \mathbb{P}(Y_i = 1 | \boldsymbol{x}_i; \boldsymbol{\theta}) = \frac{\exp(\boldsymbol{\beta}_k^T \boldsymbol{x}_i)}{1 + \exp(\boldsymbol{\beta}_k^T \boldsymbol{x}_i)}$;
- $\mathbf{W}$ est la matrice diagonale de dimension $n \times n$ d'élements $\pi(\boldsymbol{x}_1; \boldsymbol{\theta})\,(1 - \pi(\boldsymbol{x}_n; \boldsymbol{\theta}))$: $\mathbf{W} = \mathrm{diag}(\boldsymbol{p})$.
- $\widetilde{\boldsymbol{y}} = \mathbf{X} \boldsymbol{\theta}^{(m)} + (\mathbf{W}^{(m)})^{-1} (\boldsymbol{\tau} - \boldsymbol{p}^{(m)})$

De part la forme de (11) qui correspond à celle de l'estimateur des moindres carrés pondérés (avec une matrice de poids $\mathbf{W}$) on appelle l'algorithme Netwton-Raphson dans ce cas l'algorithme IRLS (Iterative Reweighted Least Squares pour moindres carrés pondérés itératifs).

Travail demandé : Implémenter l'algorithme et le tester sur la série temporelle fournie.
On pourra commencer par le cas d'une processus latent à deux états, en le testant sur une série temporelle simulée.