Consignes :
- Sont interdits : Documents, calculettes, téléphones, écouteurs, ordinateurs, tablettes.
- Il est interdit de composer avec un crayon.
- Votre feuille double d'examen doit porter, à l'emplacement réservé, vos nom, prénom, et signature.
- Cette zone réservée doit être cachée par collage.
- Vos feuilles intercalaires doivent être toutes numérotées.
- Le barème est donné à titre indicatif.

On considère une série temporelle $(Y_1, \ldots, Y_n)$, avec $Y_t \in \mathbb{N}$, régie par un processus latent $(Z_1, \ldots, Z_n)$ à K états où Z_t représente l'état latent à l'instant $t\,(t = 1, \ldots, n)$. $Z_t \in \{1, 2, \ldots, K\}$, $K \in \mathbb{N}^\star$. On dispose d'une série observée $(y_1, \ldots, y_n)$. On suppose que pour chaque état k, la variable observée Y_t est modélisé en fonction du temps par le modèle log-linéaire suivant

$$\ln(\lambda_k(t, \boldsymbol{\beta}_k)) = \boldsymbol{\beta}_k^T \boldsymbol{x}_t \tag{1}$$

où $\boldsymbol{x}_t = (1, t)^\top$ et $Y_t | Z_t = k; \boldsymbol{\beta}_k \sim \mathcal{P}\big(\lambda_k(t; \boldsymbol{\beta}_k)\big)$ de loi Poisson

$$\mathbb{P}(Y_t = y_t | Z_t = k; \lambda_k(t; \boldsymbol{\beta}_k)) = \frac{e^{-\lambda_k(t;\boldsymbol{\beta}_k)}(\lambda_k(t; \boldsymbol{\beta}_k))^{y_t}}{y_t!}, \forall y_t \in \mathbb{N}, \tag{2}$$

qu'on notera Poisson$(y_t; \lambda_k(t, \boldsymbol{\beta}_k))$. La probabilité de l'état k à l'instant t définie par le modèle logistique :

$$\mathbb{P}(Z_t = k; \mathbf{w}) = \frac{e^{\boldsymbol{w}_k^\top \boldsymbol{x}_t}}{1 + \sum_{\ell=1}^{K} e^{\boldsymbol{w}_\ell^\top \boldsymbol{x}_t}}, \tag{3}$$

où $\mathbf{w} = (\boldsymbol{w}_1^\top, \ldots, \boldsymbol{w}_{K-1})^\top$ et $\boldsymbol{w}_K = \mathbf{0}$. On notera par $\pi_k(t; \mathbf{w})$ cette probabilité.

1). Montrer que la loi de la variable observée est définie par

$$\mathbb{P}(Y_t = y_t; \boldsymbol{\theta}) \quad = \quad \sum_{k=1}^{K} \pi_k(t; \mathbf{w}) \text{Poisson}(y_t; \lambda_k(t, \boldsymbol{\beta}_k)) \tag{4}$$

$\boldsymbol{\theta}$ étant le vecteur paramètre inconnu du modèle.

L'objectif est de segmenter la série temporelle sur la base du modèle (4). Pour cela, on estime $\boldsymbol{\theta}$ à partir des données en maximisant la log-vraisemblance $L(\boldsymbol{\theta}) = \sum_{t=1}^{n} \ln \mathbb{P}(Y_t = y_t; \boldsymbol{\theta})$ par l'algorithme EM. On montre que celui-ci consiste à partir d'un modèle initial de paramètre $\boldsymbol{\theta}^{(0)}$ et alterner à chaque itération q entre les deux étapes E- et M- suivantes jusqu'à la convergence :

Étape E : Calcul de la fonction $Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(q)})$ définie par :

$$Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(q)}) = \sum_{t=1}^{n} \sum_{k=1}^{K} \tau_k(t; \boldsymbol{\theta}^{(q)}) \ln[\pi_k(t; \mathbf{w}) \text{Poisson}(y_t; \lambda_k(t, \boldsymbol{\beta}_k))] \tag{5}$$

où $\tau_k(t; \boldsymbol{\theta}^{(q)}) = \mathbb{P}(Z_t = k | y_t; \boldsymbol{\theta}^{(q)})$ est la probabilité a posteriori de l'état k à l'instant t.

Étape M : Mise à jour des paramètres du modèle en calculant $\boldsymbol{\theta}^{(q+1)}$ définie par :

$$\boldsymbol{\theta}^{(q+1)} = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(q)}). \tag{6}$$

2. Donner le vecteur paramètre du modèle, noté $\boldsymbol{\theta}$.
3. Montrer que l'étape E- consiste à calculer les probabilités a posteriori définies par :

$$\tau_k(t; \boldsymbol{\theta}^{(q)}) = \mathbb{P}(Z_t = k|y_t; \boldsymbol{\theta}^{(q)}) = \frac{\pi_k(t; \mathbf{w}^{(q)}) \,\text{Poisson}(y_t; \lambda_k(t, \boldsymbol{\beta}_k^{(q)})))}{\mathbb{P}(Y_t = y_t; \boldsymbol{\theta}^{(q)})}, \quad \forall t = 1, \ldots, n \tag{7}$$

4. Afin de maximiser (5) par rapport aux vecteurs $\boldsymbol{\beta}_k$, montrer que

$$\frac{\partial Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(q)})}{\partial \boldsymbol{\beta}_k} = \sum_{t=1}^{n} \tau_k(t; \boldsymbol{\theta}^{(q)}) \boldsymbol{x}_t \left(y_t - e^{\boldsymbol{\beta}_k^\top \boldsymbol{x}_t} \right). \tag{8}$$

5. Cette dérivée n'admet pas de racine analytique ; Ainsi, pour cette maximisation, on utilise l'algorithme Newton-Raphson pour la mise à jour des β_k. Montrer que à chaque itération m de cet algorithme, la mise à jour est donnée par l'équation suivante :

$$\boldsymbol{\beta}_k^{(m+1)} = \boldsymbol{\beta}_k^{(m)} + \Big[\sum_{t=1}^{n} \tau_k(t; \boldsymbol{\theta}^{(q)}) e^{\boldsymbol{\beta}_k^\top \boldsymbol{x}_t} \boldsymbol{x}_t \boldsymbol{x}_t^\top \Big]^{-1}_{\boldsymbol{\beta}_k = \boldsymbol{\beta}_k^{(m)}} \sum_{t=1}^{n} \tau_k(t; \boldsymbol{\theta}^{(q)}) \boldsymbol{x}_t \left(y_t - e^{\boldsymbol{\beta}_k^\top \boldsymbol{x}_t} \right)_{\boldsymbol{\beta}_k = \boldsymbol{\beta}_k^{(m)}} \tag{9}$$

6. Donner une formulation vectorielle de cette mise à jour
7. En déduire la formulation sous forme de moindres cassés itératifs pondérés (IRLS) de cette mise à jour
8. Donner la forme générale de l'algorithme de mise à jour du vecteur $\mathbf{w}$.
9. Soit $\widehat{\boldsymbol{\theta}}$ le vecteur paramètre estimé par l'algorithme EM. En déduire la séquence optimale des états $(\widehat{Z}_1, \ldots, \widehat{Z}_n)$
10. Soit $K = 2$. Montrer que la mise à jour itérative du vecteur $\boldsymbol{w}$ à chaque itération s de l'algorithme Newton-Raphson est donnée par :

$$\boldsymbol{w}^{(s+1)} = \boldsymbol{w}^{(s)} + \Big[\sum_{t=1}^{n} \boldsymbol{x}_t \boldsymbol{x}_t^\top \pi(t; \boldsymbol{w})(1 - \pi(t; \boldsymbol{w})) \Big]^{-1}_{\boldsymbol{w} = \boldsymbol{w}^{(s)}} \sum_{t=1}^{n} \boldsymbol{x}_t \left(\tau_t(\boldsymbol{\theta}^{(q)}) - \pi(t; \boldsymbol{w}) \right)_{\boldsymbol{w} = \boldsymbol{w}^{(s)}} \tag{10}$$

où $\pi(t; \boldsymbol{w}) = \frac{e^{\boldsymbol{w}^\top \boldsymbol{x}_t}}{1 + e^{\boldsymbol{w}^\top \boldsymbol{x}_t}}$ est la probabilité de l'un des deux états.

11. Donner une formulation vectorielle de cette mise à jour
12. En déduire la formulation IRLS de cette mise à jour