Correction du TD N° 1

Exercice 3 (EMV: loi de Poisson)

1. Si X suit une loi de Poisson de paramètre λ (i.e. $X \sim \mathcal{P}(\lambda)$) alors X est une variable aléatoire discrète avec $X(\Omega) = Val(X) = \{0, 1, 2, \dots\} = \mathbb{N}$ et que la fonction de probabilité de masse s'écrit: $\mathbb{P}[X = k] = \frac{\lambda^k}{k!}e^{-\lambda}, \forall k \in \{0, 1, 2, \dots\}$.
   - L'espérance de X: $\mathbb{E}[X] = \lambda$ et la variance $\mathbb{V}[X] = \lambda$. En effet:

$$\begin{aligned}\mathbb{E}[X] &= \sum_{k\in\mathbb{N}} k\mathbb{P}[X = k] = e^{-\lambda}\sum_{k=0}^{+\infty} k\frac{\lambda^k}{k!} = e^{-\lambda}\sum_{k=1}^{+\infty} k\frac{\lambda^k}{k!} = e^{-\lambda}\sum_{k=1}^{+\infty}\frac{\lambda^k}{(k-1)!} \\ &= e^{-\lambda}\sum_{k'=0}^{+\infty}\frac{\lambda^{k'+1}}{k'!} \quad \text{(Changement d'indice } k' = k-1) \\ &= \lambda e^{-\lambda}\underbrace{\sum_{k'=0}^{+\infty}\frac{\lambda^{k'}}{k'!}}_{e^{\lambda}} = \lambda e^{-\lambda}e^{\lambda} = \lambda.\end{aligned}$$

$$\begin{aligned}\mathbb{V}[X] &= \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \mathbb{E}[X^2] - \lambda^2 = \sum_{k=0}^{+\infty} k^2\mathbb{P}[X = k] - \lambda^2 \\ &= e^{-\lambda}\sum_{k=0}^{+\infty} k^2\frac{\lambda^k}{k!} - \lambda^2 = e^{-\lambda}\sum_{k=1}^{+\infty} k^2\frac{\lambda^k}{k!} - \lambda^2 = e^{-\lambda}\sum_{k=1}^{+\infty} k\frac{\lambda^k}{(k-1)!} - \lambda^2 \\ &= e^{-\lambda}\sum_{k=0}^{+\infty}(k+1)\frac{\lambda^{k+1}}{k!} - \lambda^2 = e^{-\lambda}\Big(\sum_{k=0}^{+\infty}\frac{\lambda^{k+1}}{k!} + \sum_{k=0}^{+\infty} k\frac{\lambda^{k+1}}{k!}\Big) - \lambda^2 \\ &= e^{-\lambda}\Big(\lambda\sum_{k=0}^{+\infty}\frac{\lambda^k}{k!} + \sum_{k=1}^{+\infty} k\frac{\lambda^{k+1}}{k!}\Big) - \lambda^2 = e^{-\lambda}\Big(\lambda\sum_{k=0}^{+\infty}\frac{\lambda^k}{k!} + \sum_{k=1}^{+\infty}\frac{\lambda^{k+1}}{(k-1)!}\Big) - \lambda^2 \\ &= e^{-\lambda}\Big(\lambda\sum_{k=0}^{+\infty}\frac{\lambda^k}{k!} + \sum_{k=0}^{+\infty}\frac{\lambda^{k+2}}{k!}\Big) - \lambda^2 = e^{-\lambda}\Big(\lambda\sum_{k=0}^{+\infty}\frac{\lambda^k}{k!} + \lambda^2\sum_{k=0}^{+\infty}\frac{\lambda^k}{k!}\Big) - \lambda^2 = e^{-\lambda}\Big(\lambda e^{\lambda} + \lambda^2 e^{\lambda}\Big) - \lambda^2 \\ &= \lambda + \lambda^2 - \lambda^2 = \lambda.\end{aligned}$$

2. Supposons que λ est inconnu et notons que $\theta = \sqrt{\lambda}$. Pour une réalisation $(x_1, \dots, x_n)$ de l'échantillon $(X_1, \dots, X_n)$, La vraisemblance en fonction de θ est définie par:

$$\begin{aligned}L(x_1, \dots, x_n; \theta) &= \prod_{i=1}^{n}\mathbb{P}[X_i = x_i] \\ &= \prod_{i=1}^{n}\frac{\lambda^{x_i}}{(x_i)!}e^{-\lambda} \\ &= \prod_{i=1}^{n}\frac{(\theta^2)^{x_i}}{(x_i)!}e^{-\theta^2} \\ &= (e^{-\theta^2})^n\frac{\prod_{i=1}^{n}(\theta^2)^{x_i}}{\prod_{i=1}^{n}(x_i)!} \\ &= e^{-n\theta^2}\frac{(\theta^2)^{\sum_{i=1}^{n}x_i}}{x_1!\dots x_n!}.\end{aligned}$$

Rappel. La méthode du maximum de vraisemblance consiste à trouver la configuration θ "la plus probable" au vu du modèle et des données. Une autre façon de le dire est que la méthode cherche la valeur de θ qui garantit le meilleur "ajustement" entre les données et le modèle.

Dans le cours nous avons noté $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ la moyenne empirique de l'échantillon $(x_1, \ldots, x_n)$. Ainsi, la vraisemblance s'écrit

$$L(x_1, \ldots, x_n; \theta) = e^{-n\theta^2} \frac{(\theta^2)^{n\bar{x}}}{x_1! \ldots x_n!} = \frac{\{e^{-\theta^2}(\theta^2)^{\bar{x}}\}^n}{x_1! \ldots x_n!}.$$

Maintenant l'estimateur de du maximum de vraisemblance est déterminé par:

$$\hat{\theta} = \arg \max_{\theta \in \Theta =]0,+\infty[} L(x_1, \ldots, x_n; \theta) = \arg \max_{\theta \in \Theta =]0,+\infty[} \log(L(x_1, \ldots, x_n; \theta)).$$

D'abord, on écrit la log de la vraisemblance,

$$\begin{aligned}
\log(L(x_1, \ldots, x_n; \theta) &= \log(\{e^{-\theta^2}(\theta^2)^{\bar{x}}\}^n) - \log(x_1! \ldots x_n!) \\
&= n\big(\log(\{e^{-\theta^2}) + \log((\theta^2)^{\bar{x}}\})\big) - \log(x_1! \ldots x_n!) \\
&= n\big(\log(e^{-\theta^2}) + \log((\theta)^{2\bar{x}}\})\big) - \log(x_1! \ldots x_n!) \\
&= n\big(-\theta^2 + 2\bar{x}\log(\theta)\big) - \log(x_1! \ldots x_n!)
\end{aligned}$$

On calcule la dérivée de $\log(L(x_1, \ldots, x_n; \theta))$ par rapport à θ

$$\begin{aligned}
\frac{\partial}{\partial\theta}(\log(L(x_1, \ldots, x_n; \theta))) &= (\log(L(x_1, \ldots, x_n; \theta)))' \\
&= \Big(n\big(-\theta^2 + 2\bar{x}\log(\theta)\big) - \log(x_1! \ldots x_n!)\Big)' \\
&= -2n\theta + \frac{2n\bar{x}}{\theta}.
\end{aligned}$$

Ainsi,

$$\frac{\partial}{\partial\theta}(\log(L(x_1, \ldots, x_n; \theta))) = 0 \Leftrightarrow -2n\theta + \frac{2\bar{x}n}{\theta} = 0 \Leftrightarrow \theta^2 = \bar{x} \Leftrightarrow \hat{\theta} = \sqrt{\bar{x}}.$$

Alors le point $\hat{\theta} = \sqrt{\bar{x}}$ est un point critique. Pour vérifier qu'il s'agit d'un maximum local, on vérifie que la héssienne est définie négative. Or, la hésienne est juste un scalaire $H(\theta) = \frac{\partial^2}{\partial^2\theta}(\log(L(x_1, \ldots, x_n; \hat{\theta}))) < 0$. En effet,

$$\begin{aligned}
\frac{\partial^2}{\partial^2\theta}(\log(L(x_1, \ldots, x_n; \theta))) &= \frac{\partial}{\partial\theta}(-2n\theta + \frac{2n\bar{x}}{\theta}) \\
&= -2n - \frac{2n\bar{x}}{\theta^2} \\
&= -2n(1 + \frac{1}{\theta^2}).
\end{aligned}$$

Pour $\theta = \hat{\theta}$,

$$\begin{aligned}
\frac{\partial^2}{\partial^2\theta}(\log(L(x_1, \ldots, x_n; \hat{\theta}))) &= -2n - \frac{2n\bar{x}}{\hat{\theta}^2} \\
&= -2n - \frac{\bar{x}}{\bar{x}^2} = \\
&- 2n - \frac{2n}{\bar{x}} = -2n(1 + \frac{1}{\bar{x}}) < 0 \quad (\bar{x} \geq 0).
\end{aligned}$$

Donc le point critique $\hat{\theta} = \sqrt{\bar{x}}$ est un maximum local. Montrons qu'il s'agit d'un maximum global. En effet, la log de la vraisemblance est une fonction concave, comme somme de deux fonctions concave: $\theta \mapsto -\theta^2$ et $\theta \mapsto \log(\theta)$. L'estimateur du maximum de vraisemblance de l'écart type est donnée par $\hat{\theta} = \sqrt{\bar{x}} = \sqrt{\frac{1}{n}\sum_{i=1}^n x_i}$.

Remarque. Notons que l'estimateur du maximum de vraisemblance est une fonction aléatoire de $(X_1, \ldots, X_n)$, nous écrivons souvent:

$$\hat{\theta}(\omega) = \hat{\theta}(X_1(\omega), \ldots, X_n(\omega)) = \sqrt{\frac{1}{n}\sum_{i=1}^n X_i(\omega)}.$$

Exercice 4 (EMV: loi normale à moyenne et variance inconnus)

1. La fonction densité de probabilité d'une variable $X \sim \mathcal{N}(\mu, \sigma^2)$ est donnée par: $f_X(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{(x-\mu)^2}{2\sigma^2}\Big), \forall x \in \mathbb{R}$.
2. Nous cherchons à estimer un vecteur $\theta = (\mu, \sigma^2)$. L'espace des paramètres $\theta \in \Theta = \mathbb{R} \times]0, +\infty[$. La vraisemblance en fonction de θ est définie par:

$$\begin{aligned} L(x_1, \ldots, x_n; \theta) &= \prod_{i=1}^n f_X(x_i; \mu, \sigma^2) \\ &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{(x_i-\mu)^2}{2\sigma^2}\Big) \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}} \prod_{i=1}^n \exp\Big(-\frac{(x_i-\mu)^2}{2\sigma^2}\Big). \end{aligned}$$

Nous calculons la fonction log de vraisemblance,

$$\begin{aligned} \log(L(x_1, \ldots, x_n; \theta)) &= \log\bigg(\frac{1}{(2\pi\sigma^2)^{n/2}} \prod_{i=1}^n \exp\Big(-\frac{(x_i-\mu)^2}{2\sigma^2}\Big)\bigg) \\ &= -\frac{n}{2}\log(2\pi\sigma^2) + \log\bigg(\prod_{i=1}^n \exp\Big(-\frac{(x_i-\mu)^2}{2\sigma^2}\Big)\bigg) \\ &= -\frac{n}{2}\log(2\pi\sigma^2) - \sum_{i=1}^n \frac{(x_i-\mu)^2}{2\sigma^2} \\ &= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \sum_{i=1}^n \frac{(x_i-\mu)^2}{2\sigma^2} \end{aligned}$$

L'EMV vérifie

$$\hat{\theta} = (\hat{\mu}, \widehat{\sigma^2}) = \arg \max_{\theta \in \mathbb{R} \times]0, +\infty[} \log(L(x_1, \ldots, x_n; \theta)).$$

Pour pouvoir les calculer il suffit de résoudre:

$$\nabla \log(L(x_1, \ldots, x_n; \theta)) = \begin{pmatrix} \frac{\partial}{\partial\mu}(\log(L(x_1, \ldots, x_n; \mu, \sigma^2))) \\ \frac{\partial}{\partial\sigma^2}(\log(L(x_1, \ldots, x_n; \mu, \sigma^2))) \end{pmatrix} = \mathbf{0} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Soit,

$$\begin{cases} \frac{\partial}{\partial\mu}(\log(L(x_1, \ldots, x_n; \mu, \sigma^2))) = 0, \\ \frac{\partial}{\partial\sigma^2}(\log(L(x_1, \ldots, x_n; \mu, \sigma^2))) = 0 \end{cases}$$

Par un calcul simple, on a
$\frac{\partial}{\partial\mu}(\log(L(x_1,\ldots,x_n;\mu,\sigma^2))) = \sum_{i=1}^n \frac{x_i-\mu}{\sigma^2}$
et $\frac{\partial}{\partial\sigma^2}(\log(L(x_1,\ldots,x_n;\mu,\sigma^2))) = -\frac{n}{2\sigma^2} + \sum_{i=1}^n \frac{(x_i-\mu)^2}{2(\sigma^2)^2}$.
Ainsi,

$$\begin{cases} \frac{\partial}{\partial\mu}(\log(L(x_1,\ldots,x_n;\mu,\sigma^2))) = 0, \\ \frac{\partial}{\partial\sigma^2}(\log(L(x_1,\ldots,x_n;\mu,\sigma^2))) = 0 \end{cases} \Leftrightarrow \begin{cases} \sum_{i=1}^n \frac{x_i-\mu}{\sigma^2} = 0, \\ -\frac{n}{2\sigma^2} + \sum_{i=1}^n \frac{(x_i-\mu)^2}{2\sigma^4} = 0 \end{cases} \Leftrightarrow \begin{cases} \sum_{i=1}^n \frac{x_i-\mu}{\sigma^2} = 0, \\ -\frac{n}{2\sigma^2} + \sum_{i=1}^n \frac{(x_i-\mu)^2}{2\sigma^4} = 0 \end{cases}$$

$$\Leftrightarrow \begin{cases} \sum_{i=1}^n x_i - n\mu = 0, \\ -\frac{n}{2\sigma^2} + \left(\frac{1}{\sigma^2}\right)^2 \left(\sum_{i=1}^n \frac{(x_i-\mu)^2}{2}\right) = 0 \end{cases} \Leftrightarrow \begin{cases} \mu = \frac{1}{n}\sum_{i=1}^n x_i, \\ -\frac{n}{2}\sigma^2 + \sum_{i=1}^n \frac{(x_i-\mu)^2}{2} = 0 \end{cases} \Leftrightarrow \begin{cases} \mu = \frac{1}{n}\sum_{i=1}^n x_i, \\ \sigma^2 = \frac{1}{n}\sum_{i=1}^n (x_i-\mu)^2. \end{cases}$$

On conclut que le vecteur $\hat{\theta} = \begin{pmatrix} \hat{\mu} \\ \widehat{\sigma^2} \end{pmatrix} = \begin{pmatrix} \bar{x} \\ \frac{1}{n}\sum_{i=1}^n (x_i-\bar{x})^2 \end{pmatrix}$ est un point critique pour la fonction de la log de vraisemblance. Pour vérifier qu'il s'agit d'un maximum local, on détermine si la matrice hésienne en $\hat{\theta}$ est définie positive. D'abord on calcule pour tout $\theta = \begin{pmatrix} \mu \\ \sigma^2 \end{pmatrix}$. Soit $\vec{x} = (x_1,\ldots,x_n)$

$$H(\theta) = \begin{pmatrix} \frac{\partial^2}{\partial\mu^2}\log(L(\vec{x};\theta)) & \frac{\partial^2}{\partial\mu\partial\sigma^2}\log(L(\vec{x};\theta)) \\ \frac{\partial^2}{\partial\sigma^2\partial\mu}\log(L(\vec{x};\theta)) & \frac{\partial^2}{\partial^2(\sigma^2)^2}\log(L(\vec{x};\theta)) \end{pmatrix}.$$

Or on a:

$$\begin{cases} \frac{\partial^2}{\partial\mu^2}(\log(L(x_1,\ldots,x_n;\mu,\sigma^2))) = -\frac{n}{\sigma^2}, \\ \frac{\partial^2}{\partial\mu\partial\sigma^2}(\log(L(x_1,\ldots,x_n;\mu,\sigma^2))) = -\frac{\sum_{i=1}^n (x_i-\mu)}{(\sigma^2)^2}, \\ \frac{\partial^2}{\partial(\sigma^2)^2}(\log(L(x_1,\ldots,x_n;\mu,\sigma^2))) = \frac{n}{2(\sigma^2)^2} - \frac{\sum_{i=1}^n (x_i-\mu)^2}{(\sigma^2)^3} \\ \frac{\partial^2}{\partial\sigma^2\partial\mu}(\log(L(x_1,\ldots,x_n;\mu,\sigma^2))) = -\frac{\sum_{i=1}^n (x_i-\mu)}{(\sigma^2)^2}. \end{cases}$$

> **Remarque:** $\frac{\partial^2}{\partial\mu\partial\sigma^2}(\log(L(x_1,\ldots,x_n;\mu,\sigma^2))) = \frac{\partial^2}{\partial\sigma^2\partial\mu}(\log(L(x_1,\ldots,x_n;\mu,\sigma^2)))$: on peut intervenir l'ordre des dérivées partielles et ceci est dû au théorème de Schwartz. Ce théorème dit que pour les fonctions suffisamment régulières, l'ordre de dérivation par rapport aux variables n'a pas d'importance. Il suffit en fait que les *dérivées partielles existent au voisinage d'un point et soient continues* pour pouvoir intervertir l'ordre de dérivation.

On calcule la hésienne pour $\mu = \hat{\mu} = \bar{x}$ et $\sigma^2 = \widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^n (x_i-\bar{x})^2$.

$$H(\hat{\theta}) = \begin{pmatrix} -\frac{n}{\widehat{\sigma^2}} & -\frac{\sum_{i=1}^n (x_i-\hat{\mu})}{\left(\widehat{\sigma^2}\right)^2} \\ -\frac{\sum_{i=1}^n (x_i-\hat{\mu})}{\left(\widehat{\sigma^2}\right)^2} & \frac{n}{2(\widehat{\sigma^2})^2} - \frac{\sum_{i=1}^n (x_i-\hat{\mu})^2}{(\widehat{\sigma^2})^3} \end{pmatrix}.$$

Remarquons que $\sum_{i=1}^n (x_i-\hat{\mu}) = \sum_{i=1}^n x_i - n\hat{\mu} = n\hat{\mu} - n\hat{\mu} = 0$ et $\frac{\sum_{i=1}^n (x_i-\hat{\mu})^2}{(\widehat{\sigma^2})^3} = \frac{n\widehat{\sigma^2}}{(\widehat{\sigma^2})^3} = \frac{n}{(\widehat{\sigma^2})^2}$. Donc

$$H(\hat{\theta}) = \begin{pmatrix} -\frac{n}{\widehat{\sigma^2}} & 0 \\ 0 & \frac{n}{2(\widehat{\sigma^2})^2} - \frac{n}{(\widehat{\sigma^2})^2} \end{pmatrix} = \begin{pmatrix} -\frac{n}{\widehat{\sigma^2}} & 0 \\ 0 & -\frac{n}{2(\widehat{\sigma^2})^2} \end{pmatrix} = -\begin{pmatrix} \frac{n}{\widehat{\sigma^2}} & 0 \\ 0 & \frac{n}{2(\widehat{\sigma^2})^2} \end{pmatrix}$$

Nous montrons que $H(\hat{\theta})$ est définie négative. $H(\hat{\theta})$ étant diagonale, ses éléments diagonaux sont ses valeurs propres. Comme ses valeurs propres sont négatives, elle est définie négative.

Nous pourrons utiliser une caractérisation avec les déterminants des sous matrices principales. Vous trouverez les détails dans le cours de J. Guérin, N. Lahrichi, S. Le Digabel en cliquant sur le lien suivant: https://www.gerad.ca/Sebastien.Le.Digabel/MTH1007/12_matrices_symetriques_def_positives.pdf (voir les transparents 18, 19 et 20).

Théorème: Une matrice symétrique A de taille $n \times n$ est définie négative si et seulement si les déterminants des sous-matrices principales de A d'ordre impair sont strictement négatives et les déterminants des sous-matrices principales de A d'ordre pair sont strictement positifs.

Appliquons ce théorème à $H(\hat{\theta})$:
- la première sous-matrice principale est $H_1(\hat{\theta}) = (-\frac{n}{\widehat{\sigma^2}})$ avec un déterminant $det(H_1(\hat{\theta})) = -\frac{n}{\widehat{\sigma^2}} < 0$.
- la deuxième sous-matrice principale est $H_2(\hat{\theta}) = H(\hat{\theta})$ et

$$det(H_2(\hat{\theta})) = det(H(\hat{\theta})) = -\frac{n}{\widehat{\sigma^2}} \times -\frac{n}{2(\widehat{\sigma^2})^2} - 0 = \frac{n^2}{2(\widehat{\sigma^2})^3} > 0.$$

On conclut que $H(\hat{\theta})$ est définie négative. Nous arrivons à montrer que le point $\hat{\theta} = \begin{pmatrix} \hat{\mu} \\ \widehat{\sigma^2} \end{pmatrix} = \begin{pmatrix} \bar{x} \\ \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \end{pmatrix}$ est un maximum local. Montrons qu'il est global. Nous revenons à l'expression de la fonction de la vraisemblance

$$L(x_1, \ldots, x_n; \theta) = \frac{1}{(2\pi\sigma^2)^{n/2}} \prod_{i=1}^{n} \exp\Big(-\frac{(x_i - \mu)^2}{2\sigma^2}\Big) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\Big(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\Big).$$

Vérifions maintenant que pour tout $\mu \neq \hat{\mu} = \bar{x}$ on a $\sum_{i=1}^{n}(x_i - \mu)^2 > \sum_{i=1}^{n}(x_i - \bar{x})^2$. En effet

$$\begin{aligned}
\sum_{i=1}^{n}(x_i - \mu)^2 &= \sum_{i=1}^{n}\big((x_i - \bar{x}\mu) + (\bar{x} - \mu)\big)^2 \\
&= \sum_{i=1}^{n}\Big((x_i - \bar{x})^2 + (\bar{x} - \mu)^2 + 2(x_i - \bar{x})(\bar{x} - \mu)\Big) \\
&= \sum_{i=1}^{n}(x_i - \bar{x})^2 + n(\bar{x} - \mu)^2 + 2(\bar{x} - \mu) \underbrace{\sum_{i=1}^{n}(x_i - \bar{x})}_{\sum_{i=1}^{n} x_i - n\bar{x} = n\bar{x} - n\bar{x} = 0} \\
&= \sum_{i=1}^{n}(x_i - \bar{x})^2 + \underbrace{n(\bar{x} - \mu)^2}_{\text{positif}}
\end{aligned}$$

, Donc $\sum_{i=1}^{n}(x_i - \mu)^2 > \sum_{i=1}^{n}(x_i - \bar{x})^2$ pour tout $\mu \neq \bar{x}$. Ceci implique

$$\frac{1}{(2\pi\sigma^2)^{n/2}} \exp\Big(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\Big) < \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\Big(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \bar{x})^2\Big), \text{pour tout } \sigma^2 > 0.$$

Or l'application $g : \sigma^2 \mapsto \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\Big(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \bar{x})^2\Big)$ atteint son maximum pour $\sigma^2 = \widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ (on peut appliquer étudier comme avant la fonction $\sigma^2 \mapsto \log(\sigma^2)$). Ainsi

$$\frac{1}{(2\pi\sigma^2)^{n/2}} \exp\Big(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \bar{x})^2\Big) < \frac{1}{(2\pi\widehat{\sigma^2})^{n/2}} \exp\Big(-\frac{1}{2\widehat{\sigma^2}}\sum_{i=1}^{n}(x_i - \bar{x})^2\Big).$$

On conclut que le vecteur $\hat{\theta}$ est un maximum global.

3. Le quantile $q_{0,95}$ de la distribution de X est définie par $\mathbb{P}[X < q_{0,95}] = 0,95$

$$\mathbb{P}[X < q_{0,95}] = 0,95 \Leftrightarrow \mathbb{P}\Big[\frac{X-\mu}{\sigma} < \frac{q_{0,95}-\mu}{\sigma}\Big] = 0.95 \Leftrightarrow \mathbb{P}\Big[Z \leq \frac{q_{0,95}-\mu}{\sigma}\Big] = 0.95,$$

où la variable aléatoire $Z := \frac{X-\mu}{\sigma}$ qui suit une loi normale centrée réduite. Ainsi $\frac{q_{0,95}-\mu}{\sigma}$ est le quantile d'ordre $0,95$ de Z qui est $z_{0,95} = 1,645$. On obtient

$$z_{0,95} = \frac{q_{0,95}-\mu}{\sigma} \Leftrightarrow q_{0,95} = z_{0,95}\sigma + \mu$$

Donc l'estimateur EMV de $q_{0,95}$ est donné par $\widehat{q_{0,95}} = z_{0,95}\hat{\sigma} + \hat{\mu} = z_{0,95}\sqrt{\widehat{\sigma^2}} + \hat{\mu}$. Soit

$$\widehat{q_{0,95}} = 1,645\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} + \bar{x}.$$

Exercice 5 (Théorème de Bayes: estimation bayésienne)

1. Notons $\boldsymbol{x} = (x_1, \ldots, x_n)$ une réalisation de l'échantillon $\mathcal{D}_n$. Par le théorème de Bayes, la distribution a posteriori de Λ s'écrit:

$$\begin{aligned} f_{\Lambda|\mathcal{D}_n}(\lambda|\boldsymbol{x}) &= \frac{f_{\mathcal{D}_n|\Lambda}(\boldsymbol{x}|\lambda) f_\Lambda(\lambda)}{f_{\mathcal{D}_n}(\boldsymbol{x})} \\ &\propto f_{\mathcal{D}_n|\Lambda}(\boldsymbol{x}|\lambda) f_\Lambda(\lambda). \end{aligned}$$

La distribution conditionnelle $f_{\mathcal{D}_n|\Lambda}(\boldsymbol{x}|\lambda)$ est donnée par la vraisemblance de l'échantillon $\mathcal{D}_n$ sachant $\Lambda = \lambda$, c'est à dire

$$f_{\mathcal{D}_n|\Lambda}(\boldsymbol{x}|\lambda) = \prod_{i=1}^{n} P[X_i = x_i] = \prod_{i=1}^{n} e^{-\lambda}\frac{\lambda^{x_i}}{x_i!} = e^{-n\lambda}\frac{\lambda^{\sum_{i=1}^{n} x_i}}{x_1! \ldots x_n!},$$

et pour tout $\lambda > 0$, la distribution de masse de probabilité de la variable aléatoire Λ est $f_\Lambda(\lambda) = \frac{\alpha^p}{\Gamma(p)}\lambda^{p-1}e^{-\alpha\lambda}$. Soit

$$f_{\Lambda|\mathcal{D}_n}(\lambda|\boldsymbol{x}) \propto e^{-n\lambda}\frac{\lambda^{\sum_{i=1}^{n} x_i}}{x_1! \ldots x_n!}\frac{\alpha^p}{\Gamma(p)}\lambda^{p-1}e^{-\alpha\lambda} \propto e^{-(n+\alpha)\lambda}\lambda^{\sum_{i=1}^{n} x_i + p - 1} \propto \text{Gamma}(\sum_{i=1}^{n} x_i + p, n + \alpha),$$

La distribution a priori de Λ est proportionnelle à une loi Gamma de paramètre $(\sum_{i=1}^{n} x_i + p, n + p)$.

2. Puisque la fonction log est croissante, donc on peut calculer le maximum a posteriori (MAP) de Λ par

$$\begin{aligned} \Lambda_{\text{MAP}} &= \arg\max_\lambda \Big\{ \log(f_{\Lambda|\mathcal{D}_n}(\lambda|\boldsymbol{x})) \Big\} \\ &\propto \arg\max_\lambda \Big\{ \log\Big(e^{-(n+\alpha)\lambda}\lambda^{\sum_{i=1}^{n} x_i + p - 1}\Big) \Big\} \\ &= \arg\max_\lambda \Big\{ -(n+\alpha)\lambda + \big(\sum_{i=1}^{n} x_i + p - 1\big)\log(\lambda) + C \Big\}, \end{aligned}$$

où C est une constante par rapport à λ. Prenons la dérivée par rapport à λ

$$\begin{aligned} &\Big\{ -(n+\alpha)\lambda + \big(\sum_{i=1}^{n} x_i + p - 1\big)\log(\lambda) + C \Big\}' = 0 \\ &\Leftrightarrow -(n+\alpha) + \frac{\sum_{i=1}^{n} x_i + p - 1}{\lambda} = 0 \Leftrightarrow \lambda = \frac{\sum_{i=1}^{n} x_i + p - 1}{n+\alpha}. \end{aligned}$$

On conclut que si la distribution a priori de $\Lambda \sim \text{Gamma}(p, \alpha)$ alors le maximum a posteriori est

$$\Lambda_{\text{MAP}} = \frac{\sum_{i=1}^{n} X_i + p - 1}{n+\alpha}.$$