2 Solutions

2.1 Exercice 1

1. Plutôt que de considérer ici les temps successifs d'arrivée des clients $T_1, \ldots, T_n$ (suite croissante positive), il est plus aisé de manipuler les inter-temps $\Delta_i = T_i - T_{i-1}$ qui eux sont seulement contraints à être positifs par définition. Le modèle statistique est alors défini via :
   — l'espace des observations : $(\mathbb{R}_+^n, \mathcal{B}(\mathbb{R}_+^n))$: vecteur de n coordonnées positives réelles associée à la tribu des boréliens sur $\mathbb{R}_+^n$.
   — une famille de distributions de probabilités sur $(\mathbb{R}_+^n, \mathcal{B}(\mathbb{R}_+^n))$. On peut considérer un exemple particulier, noté $\mathcal{C}$, de type *paramétrique* ET *de forme produit* :

$$\mathcal{C} = \{\mathbb{P}_\theta, \theta > 0\},$$

   où pour tout $(\delta_1, \ldots, \delta_n) \in \mathbb{R}_+^n$:

$$p_\theta(\delta_1, \ldots, \delta_n) = \prod_{i=1}^{n} \theta e^{-\theta \delta_i}.$$

   Ce dernier modèle statistique traduit ainsi une forme paramétrique exponentielle de chaque inter-temps d'arrivée, associé à l'hypothèse forte d'indépendance entre chaque événement d'arrivée d'un nouveau client. Si l'hypothèse de forme exponentielle est raisonnable, l'indépendance entre les arrivées successives ainsi que le caractère homogène en temps des distributions des inter-temps d'arrivée pourrait être remis en question.
2. On peut simplement proposer l'estimateur suivant basé sur la moyenne empirique des inter-temps

$$\widehat{\Delta}_n := \frac{1}{n} \sum_{i=1}^{n} \Delta_i = \frac{T_n}{n}$$

   Si $\widehat{\Delta}_n$ estime alors le temps moyen d'arrivée d'un nouveau client, on peut imaginer que le nombre de clients arrivant sur un intervalle de temps de longueur h est simplement $h\widehat{\Delta}_n^{-1}$.
3. Dans cette seconde modélisation, cette fois la modélisation porte sur le vecteur de comptage, noté $N = (N_1, \ldots, N_K)$. Par définition, ce vecteur de comptage est un vecteur de K nombres entiers. Le modèle statistique est alors défini via :
   — l'espace des observations : $(\mathbb{N}^K, \mathcal{P}(\mathbb{N}^K))$: vecteur de K entiers positifs, associée à la tribu des parties de $\mathbb{N}^K$.
   — une famille de distributions de probabilités sur $(\mathbb{N}^K, \mathcal{P}(\mathbb{N}^K))$. On peut considérer un exemple particulier, noté $\mathcal{D}$, de type *paramétrique* ET *de forme produit* :

$$\mathcal{D} = \{\mathbb{Q}_\lambda, \lambda > 0\},$$

   où pour tout $(n_1, \ldots, n_K) \in \mathbb{N}^K$:

$$q_\lambda(n_1, \ldots, n_K) = \prod_{j=1}^{K} \frac{e^{-(\lambda h)}(\lambda h)^{n_j}}{n_j!}.$$

Ce dernier modèle statistique traduit ainsi une forme paramétrique Poissonienne du *processus de comptage* N. Là encore, l'hypothèse de forme Poissonienne peut être raisonnable, l'indépendance discutable, et surtout le caractère homogène en temps, *i.e.* la stationarité de la loi des arrivées au cours du temps, très discutable pour une cafétéria où l'affluence doit clairement dépendre du moment de la journée considéré.

Il existe de nombreuses façons de relacher l'hypothèse restrictive de processus de comptage homogène. La plus simple étant de proposer alors un second modèle statistique, dit *non-paramétrique*, où la quantité λ précédente est remplacée par une "intensité" instantanée $t \mapsto \lambda(t)$ et où la famille $\mathcal{D}$ est alors généralisée en $\mathcal{D}' = \{\mathbb{Q}'_\lambda, \lambda \in \mathcal{C}(\mathbb{R}_+, \mathbb{R}_+)\}$, avec

$$\forall (n_1, \dots, n_K) \in \mathbb{N}^K \qquad q_\lambda(n_1, \dots, n_K) = \prod_{j=1}^{K} \frac{e^{-\int_{(j-1)h}^{jh} \lambda(s)\mathrm{d}s)} (\int_{(j-1)h}^{jh} \lambda(s)\mathrm{d}s))^{n_j}}{n_j!}.$$

Ce dernier modèle, beaucoup plus flexible, permet d'appréhender des différences sur les moments de la journée des fréquences d'arrivée de nouveaux clients. Il est bien entendu plus sophistiqué et son étude nécessite des outils statistiques plus élaborés que ceux pour le simple processus de comptage Poissonien homogène.

4. Dans le cas simple du processus homogène Poissonien, on peut alors proposer comme estimateur du nombre de clients arrivant sur un intervalle de temps de longueur h :

$$\widehat{N}_{K,h} = \frac{1}{K} \sum_{j=1}^{K} N_j$$

Aussi, une approximation raisonnable du temps moyen d'arrivée d'un nouveau client est donné par $h\widehat{N}_{K,h}^{-1}$.

5. On peut démontrer que si les $(\Delta_1, \dots, \Delta_n, \dots)$ sont une infinité d' inter-temps d'arrivées exponentielles, *i.e.* les $(\Delta_i)_{i \geq 1}$ sont i.i.d. $\mathcal{E}(\lambda)$, alors le processus $N = (N_h, N_{2h}, \dots, N_{Kh})$ défini par

$$N_{jh} := \sup\{k \geq 0 : T_k < jh\}$$

suit une distribution de comptage i.i.d. Poissonien de paramètre λh. Le caractère "sans mémoire" des lois exponentielles permet l'indépendance des comptages tandis que l'identification de la loi de Poisson se fait via un calcul exact.

2.2 Exercice 2

Posons $p_{n,\theta}(x_1, \cdots, x_n) := \prod_{i=1}^{n} p_\theta(x_i)$. Sous le modèle $p_{n,\theta} \cdot \mathrm{Leb}^{\otimes(n)}$, les v.a. $\{X_i, i \leq n\}$ sont indépendantes.

1. Sous le modèle $p_{n,\theta} \cdot \mathrm{Leb}^{\otimes n}$, les statistiques $\{a_i + b_i X_i, i \leq n\}$ sont indépendantes ; la loi du vecteur est donc le produit des lois de chaque statistique $a_i + b_i X_i$.

Déterminons la loi de $a_i + b_i X_i$. sous p_θ. On fait le changement de variable $y = a_i + b_i x$ pour un réel $b_i \neq 0$ ce qui donne pour toute fonction mesurable positive h

$$\int h(a_i + b_i x)\, p_\theta(x) dx = \int h(y)\, p_\theta\left(\frac{y - a_i}{b_i}\right) \frac{dy}{|b_i|}$$

On en déduit que la densité de $a_i + b_i X_i$ est

$$y \mapsto \frac{1}{|b_i|}\, p_\theta\left(\frac{y - a_i}{b_i}\right).$$

Par suite, sous $p_{n,\theta} \cdot \mathrm{Leb}^{\otimes n}$, la loi du n-uplet est

$$(y_1, \cdots, y_n) \mapsto \prod_{i=1}^{n} |b_i|^{-1}\, p_\theta\left(\frac{y_i - a_i}{b_i}\right).$$

2.3 Exercice 3

1. Sous le modèle $p_{n,\theta} \cdot \mathrm{dLeb}^{\otimes n}$, le vecteur $(X_1, \cdots, X_n)$ a une densité de forme produit. Donc les composantes sont indépendantes. De plus, il s'agit d'un produit de la même fonction $u \mapsto g((u-\mu)/\sigma)$ (à une constante multiplicative près) donc ces composantes ont même loi. La constante C_θ est telle que

$$C_\theta \int g\left(\frac{u - \mu}{\sigma}\right) \mathrm{d}u = 1$$

Par suite, Cette loi est donnée par

$$u \mapsto q_\theta(u) := \frac{1}{\sigma}\, g\left(\frac{u - \mu}{\sigma}\right).$$

2. Sous $p_{n,\theta} \cdot \mathrm{dLeb}^{\otimes n}$, les v.a. $\{X_i, 1 \le i \le n\}$ sont i.i.d. de loi $q_\theta \cdot \mathrm{dLeb}$. En appliquant le résultat de l'exercice 2 avec $g \leftarrow q_\theta$, $a_k \leftarrow -\mu/\sigma$ et $b_k \leftarrow 1/\sigma$, on voit que $(X_1 - \mu)/\sigma$ a pour densité g.
3. Sous $p_{n,\theta} \cdot \mathrm{dLeb}^{\otimes n}$, les statistiques $\{(X_i - \mu)/\sigma, i \le n\}$ sont i.i.d. de loi gaussienne centrée réduite : μ est donc l'espérance de X_i et σ^2 est sa variance. Par un argument de loi (faible) des grands nombres, on sait que

$$(\frac{1}{n}S_n, \frac{1}{n}K_n) \overset{\mathbb{P}_{n,\theta}-\mathrm{prob}}{\longrightarrow} (\mu, \sigma^2).$$

Par suite, on peut proposer comme estimateur

$$\begin{aligned}(\hat{\mu}_n, \widehat{\sigma}^2_n) &:= \left(\frac{1}{n}\sum_{i=1}^{n} X_i, \frac{1}{n}\sum_{i=1}^{n} X_i^2 - (\hat{\mu}_n)^2\right) = \left(\frac{1}{n}\sum_{i=1}^{n} X_i, \frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{\mu}_n)^2\right) \\ &= (\frac{1}{n}S_n, \frac{1}{n}K_n).\end{aligned}$$

On aurait aussi pu proposer l'estimateur de la médiane empirique pour μ. La figure 4 compare l'estimateur de μ par la moyenne empirique, et par la médiane empirique. On répète 5000 fois l'algorithme suivant : (i) tirer n réalisations indépendantes de gaussiennes d'espérance $\mu = 2$ et de variance 1 ; (ii) calculer chacun des deux estimateurs à partir de ces n réalisations. Ensuite, on trace le boxplot des 5000 réalisations de chacun des estimateurs. On affiche le résultat dans le cas où $n = 50$ (gauche) et $n = 500$ (droite).

Complément : Estimateur par maximum de vraisemblance (MV) et estimateur par la méthode des moments dans le cas où g est la densité d'une loi gaussienne centrée réduite

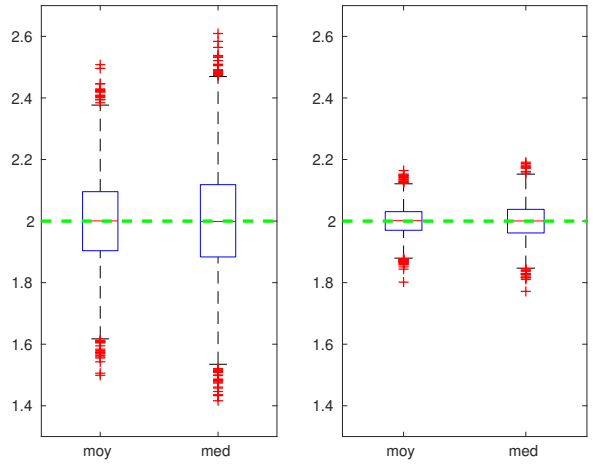

FIGURE 4 – Comparaison des estimateurs : moyenne empirique et médiane empirique, pour la distribution gaussienne d'espérance 2 et de variance 1. Boxplot de 5000 réalisations indépendantes de l'estimateur, dans le cas où $n = 50$ (gauche) et $n = 500$ (droite). On peut observer (a) le rôle de n dans la dispersion des estimateurs ; et (b) que la distribution de l'estimateur moyenne est plus concentrée que celle de l'estimateur médiane.

Les calculs ci-dessous montrent que pour ce modèle gaussien, les deux approches coïncident ; et que l'estimateur MV (et donc celui des moments) est unique.

(a) **Maximum de vraisemblance (MV)** Sous $p_{n,\theta} \cdot \mathrm{dLeb}^{\otimes n}$, la vraisemblance des observations $(X_1, \cdots, X_n)$ s'écrit

$$L_n(\theta) := \frac{1}{(2\pi)^{n/2}\,\sigma^n} \exp\left(-\frac{1}{2}\sum_{k=1}^{n}\frac{(X_k-\mu)^2}{\sigma^2}\right)$$

dont on déduit la log-vraisemblance (à constante C près)

$$\ell_n(\theta) := C - \frac{n}{2}\ln\sigma^2 - \frac{1}{2}\sum_{k=1}^{n}\frac{(X_k-\mu)^2}{\sigma^2}.$$

On peut vérifier aisément que cette quantité possède un unique maximum, donné par la solution des équations de vraisemblance $\nabla\ell_n(\theta) = 0$ et noté $\hat{\theta}_n$. On obtient

$$\hat{\theta}_n = \begin{bmatrix}\hat{\mu}_n \\ \hat{\sigma}_n^2\end{bmatrix} := \begin{bmatrix}\frac{1}{n}\sum_{k=1}^{n} X_k \\ \frac{1}{n}\sum_{k=1}^{n}\left(X_k - \frac{1}{n}\sum_{j=1}^{n} X_j\right)^2\end{bmatrix} = \begin{bmatrix}\frac{1}{n}\sum_{k=1}^{n} X_k \\ \frac{1}{n}\sum_{k=1}^{n} X_k^2 - \hat{\mu}_n^2\end{bmatrix}$$

qui sont resp. la moyenne empirique et la variance empirique.

(b) **Estimateur des moments** On a deux paramètres inconnus ; on sait que

$$\mathbb{E}_\theta[X_1] = \mu \qquad \mathrm{Var}_\theta(X_1) = \sigma^2 \tag{1}$$

dont on déduit les estimateurs

$$\hat{\theta}_n = \begin{bmatrix} \hat{\mu}_n \\ \hat{\sigma}_n^2 \end{bmatrix} := \begin{bmatrix} \frac{1}{n}\sum_{k=1}^n X_k \\ \frac{1}{n}\sum_{k=1}^n \left(X_k - \frac{1}{n}\sum_{j=1}^n X_j\right)^2 \end{bmatrix}$$

Complément: Estimateur MV et estimateur par la méthode des moments dans le cas où g est la densité d'une loi de Laplace

On discute les deux mêmes stratégies que précédemment. Cette fois, les deux approches ne conduisent pas au même estimateur ; de plus, l'estimateur des moments n'est pas unique.

(a) **Estimateur des moments** La méthode reste la même que dans le cas précédent, puisque les relations (1) restent vraies quelle que soit la loi g. On a donc

$$\hat{\theta}_n = \begin{bmatrix} \hat{\mu}_n \\ \hat{\sigma}_n^2 \end{bmatrix} := \begin{bmatrix} \frac{1}{n}\sum_{k=1}^n X_k \\ \frac{1}{n}\sum_{k=1}^n \left(X_k - \frac{1}{n}\sum_{j=1}^n X_j\right)^2 \end{bmatrix} = \begin{bmatrix} \frac{1}{n}\sum_{k=1}^n X_k \\ \frac{1}{n}\sum_{k=1}^n X_k^2 - \hat{\mu}_n^2 \end{bmatrix}$$

(b) **Estimateur MV** Ici, la log-vraisemblance des observations est donnée par

$$\ell_n(\theta) := C - n\ln\sigma - \frac{1}{\sigma}\sum_{k=1}^n |X_k - \mu|\,.$$

Maximiser cette quantité en σ donne

$$\hat{\sigma}_n := \frac{1}{n}\sum_{k=1}^n |X_k - \mu|\,.$$

Maximiser cette quantité en μ est équivalente à minimiser $\mu \mapsto \sum_{k=1}^n |X_k - \mu|$. Le résultat donne (voir ci-dessous une explication)

$$\hat{\mu}_n = \begin{cases} X_{(n+1)/2:n} & \text{lorsque } n \text{ est impair} \\ \text{n'importe quel point de l'intervalle }]X_{n/2:n}, X_{n/2+1:n}[& \text{lorsque } n \text{ est pair} \end{cases}$$

ici, on a noté $X_{1:n}, \cdots, X_{n:n}$ les statistiques d'ordre. On reconnaît la définition de la *médiane*.

Détaillons la minimisation de la fonction $h : \mu \mapsto \sum_{k=1}^n |X_k - \mu|$:

Cette fonction est linéaire par morceaux, continue, dérivable partout sauf en $\mu \in \{X_1, \cdots, X_n\}$:

— Sur $]-\infty, X_{1:n}[$, $h(\mu) = \sum_{k=1}^n X_k - n\mu$, de dérivée $-n$.
— Sur $]X_{n:n}, +\infty[$, $h(\mu) = n\mu + \sum_{k=1}^n X_k$, de dérivée n.
— Sur $]X_{q:n}, X_{q+1:n}[$, $h(\mu) = \sum_{k=1}^q (\mu - X_k) + \sum_{k=q+1}^n (X_k - \mu)$, de dérivée $2q - n$.

On voit donc que $\mu \mapsto h(\mu)$ est d'abord décroissante, puis croissante. Elle atteint son minimum en le point $\mu = X_{q_\star+1:n}$ lorsque $n = 2q_\star + 1$; et son minimum sur l'intervalle $]X_{q_\star:n}, X_{q_\star+1:n}[$ lorsque $n = 2q_\star$.

4. Sous $p_{n,\theta} \cdot \mathrm{dLeb}^{\otimes n}$,
— la loi de X_i est une gaussienne d'espérance μ et de variance σ^2,

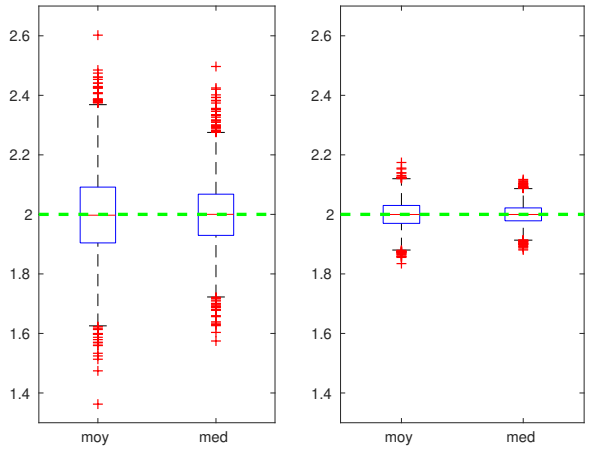

FIGURE 5 – Comparaison des estimateurs de la moyenne empirique et de la médiane empirique pour la distribution Laplace d'espérance 2 et de variance 1. Boxplot de 5000 réalisations indépendantes de chacun des deux estimateurs, dans le cas où $n = 50$ (gauche) et $n = 500$ (droite). On peut observer (a) le rôle de n dans la dispersion des estimateurs ; et (b) que la distribution de l'estimateur de la médiane est plus concentrée que celle de l'estimateur de la moyenne.

- les v.a. $X_1, \cdots, X_n$ sont indépendantes,
- la loi de $S_n := \sum_{i=1}^n X_i$ est une gaussienne d'espérance $n\mu$ et de variance $n\sigma^2$.
- la loi de
$$K_n := \sum_{i=1}^n (X_i - n^{-1}S_n)^2 = \sum_{i=1}^n X_i^2 - n^{-1}S_n^2$$
est une Gamma de paramètres $((n-1)/2, 1/(2\sigma^2))$. Par convention, une loi de Gamma de paramètres (a, b) est une loi sur $\mathbb{R}^+$ dont la densité est proportionnelle à $u \mapsto u^{a-1}\exp(-bu)$.
- les v.a. S_n et K_n sont indépendantes.

La démonstration de ce résultat se trouve dans le polycopié de cours (théorème de Gosset, Théorème IV-5-24). Par suite, le modèle statistique induit est

$$\left(\mathbb{R}\times\mathbb{R}^+, \mathcal{B}(\mathbb{R}\times\mathbb{R}^+), \left\{\rho_{n,\theta}\cdot \mathrm{Leb}^{\otimes 2} \ : \ \theta := (\mu,\sigma^2) \in \mathbb{R}\times\mathbb{R}_\star^+\right\}\right),$$

où

$$\rho_{n,\theta}(z,y) := C_n \ \exp(-(z-n\mu)^2/(2n\sigma^2))\ \exp(-y/(2\sigma^2))y^{(n-3)/2}\ \mathbb{1}_{\mathbb{R}}(z)\mathbb{1}_{\mathbb{R}^+}(y)$$

en ayant posé

$$C_n := \frac{1}{\sqrt{2\pi n\sigma^2}}\left(\frac{1}{2\sigma^2}\right)^{(n-1)/2}\frac{1}{\Gamma((n-1)/2)}\,.$$

Complément : si on veut aller plus loin sur les modèles induits : quel est le modèle induit par $(S_n, \sum_{i=1}^n X_i^2)$? Remarquons tout d'abord que

$$\sum_{i=1}^n X_i^2 = K_n + \frac{S_n^2}{n}\,.$$

Pour identifier le modèle induit, il est nécessaire d'identifier la loi de $(S_n, \sum_{i=1}^n X_i^2)$ sous $p_{n,\theta} \cdot \mathrm{Leb}^{\otimes n}$. Soit h une fonction continue bornée. On a

$$\begin{aligned}
&\mathbb{E}_{n,\theta}\left[h\left(S_n, K_n + n^{-1}S_n^2\right)\right] \\
&= C_n \int_{\mathbb{R}\times\mathbb{R}^+} h\left(z, y + n^{-1}z^2\right)\ \exp(-(z-n\mu)^2/(2n\sigma^2))\ \exp(-y/(2\sigma^2))y^{(n-3)/2}\mathrm{d}z\mathrm{d}y \\
&= C_n \int_{\mathbb{R}} \int_{u^2/n}^{+\infty} h(u,v)\ \exp(-(u-n\mu)^2/(2n\sigma^2))\ \exp(-(v-u^2/n)/(2\sigma^2))\ \left(v - \frac{u^2}{n}\right)^{(n-3)/2} \mathrm{d}u\mathrm{d}v \\
&= C_n \exp(-n\mu^2/(2\sigma^2)) \int_{\mathbb{R}} \int_{u^2/n}^{+\infty} h(u,v)\ \exp(u\mu/\sigma^2 - v/(2\sigma^2))\ \left(v - \frac{u^2}{n}\right)^{(n-3)/2} \mathrm{d}u\mathrm{d}v
\end{aligned}$$

en ayant fait le changement de variable

$$z = u \qquad y = v - u^2/n.$$

On en déduit la densité de la statistique $(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$ sous $p_{n,\theta} \cdot \mathrm{dLeb}^{\otimes n}$; puis la définition du modèle statistique induit :

$$\left(\mathbb{R}\times\mathbb{R}^+, \mathcal{B}(\mathbb{R}\times\mathbb{R}^+), \{\tilde{\rho}_{n,\theta} \cdot \mathrm{dLeb}^2, \theta \in \Theta\}\right)$$

où

$$\tilde{\rho}_{n,\theta}(u,v) \propto \exp(u\mu/\sigma^2 - v/(2\sigma^2))\ \left(v - \frac{u^2}{n}\right)^{(n-3)/2} \mathbb{1}_{\mathbb{R}}(u)\ \mathbb{1}_{[u^2/n, +\infty[}(v).$$

On observera que bien que sous $p_{n,\theta}\cdot\mathrm{Leb}^{\otimes n}$, les statistiques (S_n, K_n) soient indépendantes, il n'en est pas de même pour $(S_n, \sum_{i=1}^n X_i^2)$.

2.4 Exercice 4

1. Soit $\theta := (\beta, \sigma)$. Sous la loi $p_{n,\theta} \cdot \mathrm{Leb}^{\otimes n}$, le vecteur aléatoire $(Y_1, \cdots, Y_n)$ a la densité
$$(u_1, \cdots, u_n) \mapsto \sigma^{-n} \prod_{i=1}^{n} g\left(\frac{u_i - f(\beta' \mathbf{x}_i)}{\sigma}\right)$$
qui est de forme produit, donc les composantes sont indépendantes. De plus la loi de Y_i est
$$u_i \mapsto g\left(\frac{u_i - f(\beta' \mathbf{x}_i)}{\sigma}\right),$$
à une constante multiplicative près. Un calcul de constante de normalisation montre que cette constante multiplicative vaut $1/\sigma$.
2. En utilisant encore le résultat de l'exercice 2, on montre que sous $p_{n,\theta} \cdot \mathrm{Leb}^{\otimes n}$, la loi de $(Y_i - f(\beta' \mathbf{x}_i))/\sigma$ est g.
3. On sait que sous $p_{n,\theta} \cdot \mathrm{Leb}^{\otimes n}$,
$$Y_i \overset{ind}{\sim} \mathcal{N}\left(\beta_1 + \beta_2 x_i, \sigma^2\right),$$
ce qui entraine que pour tout i, Y_i a même loi que $\beta_1 + \beta_2 x_i + W_i$ où $W_i \sim \mathcal{N}(0, \sigma^2)$. On peut donc comprendre la donnée mesurée y_i comme la mesure de $\beta_1 + \beta_2 x_i$ entachée d'une perturbation modélisée comme la réalisation d'une loi gaussienne centrée de variance σ^2.
► On peut chercher le paramètre qui **maximise la vraisemblance des observations** : la log-vraisemblance est donnée par
$$\theta \mapsto \ell_n(\theta) := C - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{n} \left(Y_i - \beta_1 - \beta_2 \mathbf{x}_i\right)^2,$$
où C collecte tous les termes qui ne dépendent pas de θ. En introduisant les notations
$$\mathbf{X} := \begin{bmatrix} 1 & x_1 \\ \cdots & \cdots \\ 1 & x_n \end{bmatrix} \in \mathbb{R}^{\mathbf{n} \times \mathbf{2}} \qquad \beta := \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \in \mathbb{R}^{\mathbf{2}} \qquad \mathbf{Y} := \begin{bmatrix} Y_1 \\ \cdots \\ Y_n \end{bmatrix} \in \mathbb{R}^{\mathbf{n}}$$
on a
$$\ell_n(\theta) = C - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \|\mathbf{Y} - \mathbf{X}\beta\|^{\mathbf{2}}.$$
Cette fonction possède un unique maximum obtenu en résolvant les équations de vraisemblance $\nabla \ell_n = 0$. Pour β, on doit minimiser la forme quadratique
$$\beta \mapsto \beta' \mathbf{X}' \mathbf{X} \beta - \mathbf{2}\beta' \mathbf{X}' \mathbf{Y} + \mathbf{Y}' \mathbf{Y},$$
et le minimum est atteint en $\hat{\beta}_n$ solution de $\mathbf{X}'\mathbf{X}\beta = \mathbf{X}'\mathbf{Y}$ soit (en supposant la matrice $\mathbf{X}'\mathbf{X}$ inversible)
$$\hat{\beta}_n := \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}'\mathbf{Y}. \tag{2}$$
Noter que $\mathbf{X}\hat{\beta}_{\mathbf{n}}$ est la projection de $\mathbf{Y}$ sur l'espace engendré par les colonnes de $\mathbf{X}$ [2]. Par suite, le coefficient $\hat{\beta}_{n,0}$ est le projeté de $\mathbf{Y}$ sur le vecteur constant $\mathbb{1}_n := (1, \cdots, 1)'$ et il vaut $n^{-1} \sum_{i=1}^{n} y_i$.
Complément: estimation de σ^2

2. posons $H := \mathbf{X}(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}'$. Alors $H^2 = H$, c'est le projecteur sur l'espace vectoriel engendré par les colonnes de $\mathbf{X}$. Donc $H\mathbf{Y}$ est le projeté de $\mathbf{Y}$ sur cet espace.

Pour l'estimateur de σ^2 par la méthode du MV, on obtient

$$\hat{\sigma}_n^2 := \frac{1}{n}\|\mathbf{Y} - \mathbf{X}\hat{\beta}_\mathbf{n}\|^\mathbf{2} = \frac{\mathbf{1}}{\mathbf{n}}\sum_{\mathbf{i=1}}^{\mathbf{n}}\left(\mathbf{Y_i} - \hat{\beta}_{\mathbf{n,1}} - \hat{\beta}_{\mathbf{n,2}}\mathbf{x_i}\right)^\mathbf{2};$$

cet estimateur est la variance des *résidus* i.e. la variance de la composante de $\mathbf{Y}$ orthogonale à l'espace vectoriel engendré par les colonnes de $\mathbf{X}$ (soit en quelque sorte, la "part" de $\mathbf{Y} - \mathbb{E}[\mathbf{Y}]$ non expliquée par les variables explicatives $x_1, \cdots, x_n$).

▶ On peut aussi chercher à minimiser la dispersion du nuage de points $\{(\mathbf{x}_i, y_i), 1 \leq i \leq n\}$ autour de la droite de régression (s'appuyer sur les graphes) ; cela conduit à définir

$$\hat{\beta}_n := \arg\min_\beta \sum_{i=1}^n \left(Y_i - \beta'\mathbf{x}_i\right)^2 = \arg\min_\beta \|\mathbf{Y} - \mathbf{X}\beta\|^\mathbf{2}.$$

On aboutit au même résultat : $\hat{\beta}_n$ est tel que $\mathbf{X}\hat{\beta}_\mathbf{n}$ est le projeté de $\mathbf{Y}$ sur l'espace engendré par les colonnes de $\mathbf{X}$.

4. La figure de gauche représente
   - un nuage de points de coordonnées (x_i, y_i).
   - une droite horizontale : c'est la droite d'équation $y = b_1^{(1)}$ où $b_2^{(1)}$ est l'évaluation de l'estimateur $\hat{\beta}_0^{(1)} := n^{-1}\sum_{i=1}^n Y_i$ en les points mesurés $(y_1, \cdots, y_n)$. On obtient cet estimateur en considérant un modèle de régression avec $f(\beta'\mathbf{x}_i) = \beta_1$.
   - une droite de pente non nulle, d'équation $y = \tilde{b}_1 + \tilde{b}_2 x$ où $(\tilde{b}_1, \tilde{b}_2)$ est l'évaluation de l'estimateur $\hat{\beta}_n$ (voir equation (2)) en les points mesurés $(y_1, \cdots, y_n)$. On obtient cet estimateur en considérant un modèle de régression avec $f(\beta'\mathbf{x}_i) = \beta_1 + \beta_2 x_i$.

   Une fois cette droite de régression estimée (i.e. une fois calculés les coefficients $(\tilde{b}_0, \tilde{b}_1)$, on peut évaluer les *résidus* $e_i := y_i - \hat{y}_i$ i.e. la différence entre la valeur mesurée y_i et la *valeur ajustée* $\hat{y}_i := \tilde{b}_1 + \tilde{b}_2 x_i$. La figure de droite représente le nuage des n points de coordonnées $(\hat{y}_i, e_i)$.

2.5 Exercice 5

`Complément: densité par rapport à la mesure de comptage sur` $\mathbb{N}$

La mesure de comptage μ sur $\mathbb{N}$ est la mesure qui à toute partie A de $\mathbb{N}$ associe son cardinal : $\mu(A) = \mathrm{Card}(A)$.

On dit qu'une loi ξ sur un espace mesurable $(T, \mathcal{T})$ possède une densité f par rapport à la mesure $\check{\nu}$ si pour tout $A \in \mathcal{T}$, $\xi(A) = \int_A f \, \mathrm{d}\check{\nu}$.

Dans le cas où $\check{\nu}$ est la mesure de comptage, cette formule devient

$$\xi(A) = \sum_{a \in A} f(a).$$

Par suite, si X est une v.a. à valeur dans $\mathbb{N}$ dont la loi est donnée par la suite de réels positifs : $p(n) := \mathbb{P}(X = n)$ pour tout n, on dit que "X possède la densité $(p(n))_n$ par rapport à la mesure de comptage μ".

1. Notons $s \mapsto G_X(s)$ la fonction génératrice des moments d'une v.a. X. Lorsque $X \sim \mathcal{P}(\lambda)$, on a pour tout $s \in \mathbb{R}$,

$$G_X(s) := \mathbb{E}\left[s^X\right] = \exp(-\lambda) \sum_{k \geq 0} s^k \frac{\lambda^k}{k!} = \exp(-\lambda(1-s)).$$

2. La loi de Poisson possède des moments de tout ordre ; G_X est donc en particulier deux fois dérivable. On a

$$G'_X(1) = \mathbb{E}[X] \qquad G''_X(1) = \mathbb{E}[X(X-1)]$$

dont on déduit que $\mathbb{E}[X] = \lambda$ et $\mathbb{E}[X(X-1)] = \lambda^2$. Par suite,

$$\mathrm{Var}(X) = \mathbb{E}[X(X-1)] + \mathbb{E}[X] - (\mathbb{E}[X])^2 = \lambda.$$

3. Par indépendance des v.a. X_1, X_2, on a $G_{X_1+X_2} = G_{X_1} G_{X_2}$. On en déduit que

$$G_{X_1+X_2}(s) = \exp(-(\lambda_1 + \lambda_2)(s-1)).$$

La fonction génératrice des moments caractérisant la loi, cette expression prouve que $X_1 + X_2 \sim \mathcal{P}(\lambda_1 + \lambda_2)$.

4. Posons $T_n := \sum_{i=1}^n X_i$. Alors T_n est une v.a. à valeur dans $\mathbb{N}$. D'après la question 3, sous le modèle $p_\lambda \cdot \mathrm{d}\mu$, T_n suit une loi de Poisson de paramètre $n\lambda$. On en déduit que le modèle statistique induit par T_n est le triplet

$$\left(\mathbb{N}, P(\mathbb{N}), \left\{p_{n\lambda} \cdot \mu \, : \, \lambda \in \mathbb{R}_*^+\right\}\right).$$

5. ► Sous $p_\lambda \cdot \mathrm{d}\mu$, T_n est une loi de Poisson de paramètre $n\lambda$. On peut opter pour une approche par maximum de vraisemblance : la log-vraisemblance est donnée par (C désigne une quantité indépendante de λ)

$$\lambda \mapsto \ell_n(\lambda) := C - n\lambda + T_n \ln \lambda,$$

qui atteint son maximum sur $\mathbb{R}_{\star}^{+}$ en un unique point, donné par

$$\hat{\lambda}_n := \frac{1}{n}\sum_{k=1}^{n} X_k. \tag{3}$$

► En exploitant le fait que $(X_1, \cdots, X_n)$ est un n-échantillon d'un modèle statistique de Poisson, on peut proposer un estimateur par la méthode des moments basé sur le moment d'ordre 1 des v.a. X_i. La relation $\mathbb{E}_\lambda[X_1] == \lambda$ définit l'estimateur $\tilde{\lambda}_n$ par la solution de l'équation

$$\frac{1}{n}\sum_{k=1}^{n} X_k = \lambda$$

ce qui donne le même estimateur que celui identifié en (3).

► En exploitant le fait que $(X_1, \cdots, X_n)$ soit un n-échantillon d'un modèle statistique de Poisson, on peut proposer un estimateur par la méthode des moments basé sur le moment d'ordre 2 des v.a. X_i. La relation $\mathbb{E}_\lambda[X_1^2] = \lambda(1+\lambda)$ définit l'estimateur $\check{\lambda}_n$ par la solution de l'équation

$$\lambda > 0 \text{ t.q.} \frac{1}{n}\sum_{k=1}^{n} X_k^2 = \lambda(1+\lambda);$$

ce qui donne l'unique racine strictement positive de

$$\lambda > 0 \text{ t.q.} \lambda^2 + \lambda - \frac{1}{n}\sum_{k=1}^{n} X_k^2 = 0.$$

6. On collecte les données $\ell_1, v_1, \cdots, \ell_n, v_n$ où ℓ_i (resp. v_i) est le nombre de buts marqués par l'équipe locale (resp. visiteuse) lors du i-eme match. Ces données sont une réalisation du vecteur aléatoire $(L_1, V_1, \cdots, L_n, V_n)$ à valeur dans $\mathbb{N}^{2n}$; cet espace est muni de la tribu de ses parties ; on veut munir l'espace mesurable d'une loi telle que sous cette loi, les variables canoniques soient indépendantes et de loi de Poisson de paramètres adequats, à savoir :

$$(\ell_1, v_1, \cdots, \ell_n, v_n) \mapsto \tilde{p}_\theta(\ell_1, v_1, \cdots, \ell_n, v_n) := \prod_{k=1}^{n}\left(\frac{\lambda_1^{\ell_k}}{\ell_k!}\,\frac{\lambda_2^{v_k}}{v_k!}\,\exp(-\lambda_1-\lambda_2)\right), \qquad \theta := (\lambda_1, \lambda_2).$$

Par suite, le modèle statistique est

$$\left(\mathbb{N}^{2n}, \mathcal{P}(\mathbb{N}^{2n}), \{\tilde{p}_\theta \cdot \mu^{\otimes 2}, \theta \in \mathbb{R}_\star^+ \times \mathbb{R}_\star^+\}\right).$$

7. Par un estimateur des moments (qui coincide avec l'estimateur MV), on peut prendre comme estimateur

$$\hat{\lambda}_{1,n} := \frac{1}{n}\sum_{k=1}^{n} L_k \qquad \hat{\lambda}_{2,n} := \frac{1}{n}\sum_{k=1}^{n} V_k.$$

On peut estimer λ_1 et λ_2 resp. par 1.468 et 1.055. On peut discuter le modèle poissonien en observant que la variance des données est supérieure à la moyenne des données - dans les trois cas (les locaux, les visiteurs, le nombre total par match).

Remarque: modèle identifiable

Noter que si l'on avait uniquement collecté l'information $(x_1 + y_1, \cdots, x_n + y_n)$ - c'est à dire le nombre total de buts marqués par match - on n'aurait pas pu estimer chacun des deux paramètres λ_1, λ_2. En effet, le modèle statistique associé à ces observations aurait été

espace des observations : $\mathbb{N}^n$ muni de la tribu de ses parties.
paramètre $\tilde{\theta} := (\lambda_1, \lambda_2)$ à valeur dans $\tilde{\Theta} := \mathbb{R}_\star^+ \times \mathbb{R}_\star^+$
la famille de lois $\{\check{p}_{\tilde{\theta}} \cdot \mu, \tilde{\theta} \in \tilde{\mathbb{R}}_+^\star \times \mathbb{R}_+^\star\}$ de densité

$$(s_1, \cdots, s_n) \mapsto \exp(-(\lambda_1 + \lambda_2)n) \prod_{k=1}^{n} \frac{(\lambda_1 + \lambda_2)^{s_k}}{s_k!}.$$

La loi ne dépend des paramètres inconnus λ_1, λ_2 qu'à travers la somme $\lambda_1 + \lambda_2$. Ce modèle est dit *non identifiable* puisqu'on peut trouver plusieurs paires (λ_1, λ_2) de réels strictement positifs dont la somme vaut une valeur donnée : il existe $\tilde{\theta} \neq \check{\theta}$ tels que $p_{\tilde{\theta}} = p_{\check{\theta}}$. On pourra donc estimer la somme $\lambda_1 + \lambda_2$ mais on sera incapable d'estimer chacune des quantités λ_1, λ_2.

8. `Complément: Loi mélange`

Soit $f_0 \, \mathrm{d}\nu_0, f_1 \, \mathrm{d}\nu_1$ des lois de probabilités, définies sur un espace mesurable $(T, \mathcal{T})$ avec $T \subseteq \mathbb{R}^d$ - ce contexte inclut par exemple le cas de deux densités par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$; de deux lois discrètes sur $\mathbb{N}$; d'une loi discrète et d'une loi à densité par rapport à Lebesgue.

• Soit $\pi \in]0, 1[$. Une v.a. X de loi *mélange* donnée par

$$(1 - \pi) f_0 \, \mathrm{d}\nu_0 + \pi f_1 \, \mathrm{d}\nu_1$$

a même loi que $(1 - U)\, V + U\, W$ où

U, V, W sont des v.a. indépendantes.
$V \sim f_0 \, \mathrm{d}\nu_0$, $W \sim f_1 \, \mathrm{d}\nu_1$ et U est une v.a. de Bernoulli de paramètre π.

En effet, pour toute fonction h mesurable bornée, on a

$$\begin{aligned}
\mathbb{E}\left[h((1-U)V + UW)\right] &= \mathbb{E}\left[h((1-U)V + UW)\mathbb{1}_{U=0}\right] + \mathbb{E}\left[h((1-U)V + UW)\ \mathbb{1}_{U=1}\right] \\
&= \mathbb{E}\left[h(V)\mathbb{1}_{U=0}\right] + \mathbb{E}\left[h(W)\ \mathbb{1}_{U=1}\right] \\
&= \mathbb{P}(U=0) \int h(v)\, f_0(v) \mathrm{d}\nu_0(v) + \mathbb{P}(U=1) \int h(w)\, f_1(w) \mathrm{d}\nu_1(w) \\
&= (1-\pi) \int h(x) f_0(x) \mathrm{d}\nu_0(x) + \pi \int h(x)\ f_1(x)\ \mathrm{d}\nu_1(x).
\end{aligned}$$

• Cela est équivalent à dire que "avec probabilité $(1-\pi)$, $X = V$; et avec probabilité π, $X = W$".

• Pour obtenir une réalisation d'une v.a. ayant cette loi, il suffit de tirer une v.a. de Bernoulli de paramètre π. Si on obtient 0, on retourne la réalisation d'une v.a. de loi $f_0 \, \mathrm{d}\nu_0$; et si on obtient 1, on retourne une réalisation d'une v.a. de loi $f_1 \, \mathrm{d}\nu_1$.

• On peut généraliser le modèle de mélange à plus de deux composantes :

$$\sum_{i=1}^{I} \alpha_i \, f_i \, \mathrm{d}\nu_i$$

où les poids α_i sont positifs et tels que $\sum_{i=1}^{I} \alpha_i = 1$.

Ici, $f_0 \mathrm{d}\nu_0$ est la loi de Dirac en zero : $V = 0$ avec probabilité 1 ; W est une loi de Poisson de paramètre λ. Par suite, simuler sous cette loi mélange revient à (a) tirer une loi de Bernoulli de paramètre π ; (b) si on obtient 0, retourner 0 et si on obtient 1, retourner le résultat d'un tirage d'une v.a. de Poisson de paramètre λ.

9. Notons X une v.a. ayant cette distribution mélange. Puisque X a même loi que $(1-U)V + UW$ avec U, V, W indépendantes, on a

$$\mathbb{E}[X] = \mathbb{E}[(1-U)V] + \mathbb{E}[UW] = \mathbb{E}[(1-U)]\,\mathbb{E}[V] + \mathbb{E}[U]\ \mathbb{E}[W] = 0 + \pi\lambda.$$

De même, on établit

$$\mathbb{E}\left[X^2\right] = \mathbb{E}\left[U^2 W^2\right] = \mathbb{E}\left[U^2\right]\mathbb{E}\left[W^2\right] = \pi\lambda(1+\lambda).$$

10. On peut déduire de ces relations un estimateur par la méthode des moments pour les paramètres π, λ. Posons

$$S_{1,n} := \frac{1}{n}\sum_{k=1}^{n} X_k \qquad S_{2,n} := \frac{1}{n}\sum_{k=1}^{n} X_k^2.$$

Alors on prend pour estimateur de (π, λ), la solution des conditions

$$S_{1,n} = \pi\lambda, \qquad S_{2,n} = \pi\lambda(1+\lambda) \qquad \pi \in]0,1[\qquad \lambda > 0$$

si elle existe. La résolution des deux premières conditions donne

$$\check{\lambda}_n := \frac{S_{2,n}}{S_{1,n}} - 1 = \frac{S_{2,n} - S_{1,n}}{S_{1,n}} \qquad \hat{\pi}_n := \frac{S_{1,n}^2}{S_{2,n} - S_{1,n}},$$

mais il fait vérifier que ces quantités sont bien resp. à valeur dans $\mathbb{R}_\star^+$ et dans $]0,1[$.
Puisque les v.a. X_k sont à valeur entière, on a toujours $S_{2,n} \geq S_{1,n}$; de plus,

$$\mathbb{P}_{n,(\pi,\lambda)}(S_{2,n} = S_{1,n}) = \mathbb{P}_{n,(\pi,\lambda)}(X_i \leq 1, \forall\, i \leq n) = \left((1-\pi) + \pi\exp(-\lambda)(1+\lambda)\right)^n$$

ce qui entraine que $\sum_n \mathbb{P}_{n,(\pi,\lambda)}(S_{2,n} = S_{1,n}) < \infty$ et donc sous $p_{\pi,\lambda} \cdot \mu$, $S_{2,n} - S_{1,n} > 0$ presque-sûrement pour tout n assez grand.
Par suite, sous ce modèle, et pour tout n assez grand, $\check{\lambda}_n > 0$ et $0 < \hat{\pi}_n < \infty$. En revanche, la condition $\hat{\pi}_n < 1$ est vraie ssi $S_{2,n} - S_{1,n}^2 > S_{1,n}$ i.e. ssi la variance empirique est supérieure à la moyenne empirique. Ce n'est pas toujours le cas, mais ça l'est ici sur les données collectées (c'est justement la raison pour laquelle on a envisagé ce modèle).
L'application numérique donne

$$s_{1,n} = 1.468 \qquad s_{2,n} = 1.617 + 1.468^2 \qquad \check{\lambda}_n = 1.5695 \qquad \hat{\pi}_n = 0.9353.$$

Sur la figure 6, pour $k = 0, \cdots, 4$, on trace (i) la distribution empirique du nombre de buts marqués par l'équipe locale (ii) la probabilité théorique donnée par une loi de Poisson de paramètre $\lambda = 1.468$; et (iii) la probabilité théorique donnée par la loi mélange $(1-\hat{\pi}_n)\delta_0 + \hat{\pi}_n \mathcal{P}(\check{\lambda}_n)$.

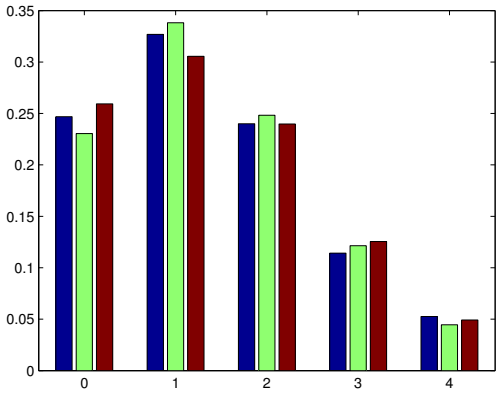

FIGURE 6 – Exercice 5 Distribution empirique (gauche), Loi de Poisson de paramètre 1.468 (centre) et loi mélange (droite) pour $k = 0, 1, \cdots, 4$.

2.6 Exercice 6

`Complément: cumulants, kurtosis, moments`

Lorsque le coefficient d'asymétrie est positif, les queues de distribution sont plus étalées à droite.

Les cumulants $(\kappa_k)_k$ sont les coefficients dans le développement du log de la transformée de Laplace (appelée fonction génératrice des cumulants) :

$$\ln \mathbb{E}\left[\exp(tX)\right] = \sum_{k\geq 0} \frac{\kappa_k}{k!} t^k = \sum_{k\geq 1} \frac{\kappa_k}{k!} t^k.$$

Par dérivation de cette fonction puis évaluation en $t = 0$, on obtient des relations entre cumulants et moments centrés ; par exemple

$$\mu_1 (= \mu) = \kappa_1; \quad \mu_2 (= \sigma^2) = \kappa_2; \quad \mu_3 = \kappa_3; \quad \mu_4 = \kappa_4 + 3\kappa_2^2 = \kappa_4 + 3\mu_2^2.$$

On a donc une autre expression de l'excès de kurtosis

$$\gamma_2 = \frac{\mu_4}{\mu_2^2} - 3 = \frac{\kappa_4}{\kappa_2^2}.$$

Noter que par Cauchy-Schwarz, $\mu_4 > \mu_2^2$ donc $\gamma_2 > -2$. Lorsque $\gamma_2 = 0$ (c'est le cas de la loi gaussienne, voir ci-dessous), la loi de la v.a. X est dite "à queues plates". Lorsque $\gamma_2 > 0$, on parle de "distribution leptokurtique" / lois à queues épaisses. Lorsque $\gamma_2 < 0$, on parle de "distribution platikurtique" / lois à queues légères.

1. Les données collectées $(x_1, \cdots, x_n)$ sont une réalisation du vecteur aléatoire $Z := (X_1, \cdots, X_n)$, à valeur dans $\mathbb{R}^n$. On munit cet espace des observations de la tribu borélienne $\mathcal{B}(\mathbb{R}^n)$ et d'une famille de probabilités $\{p_\theta \cdot \mathrm{Leb}^{\otimes n}, \theta \in \Theta\}$ avec

   $$\theta := (\mu, \sigma^2) \in \Theta := \mathbb{R} \times \mathbb{R}^+$$

   $$(x_1, \cdots, x_n) \mapsto p_\theta(x_1, \cdots, x_n) := \prod_{k=1}^n \left(\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2} \sum_{k=1}^n \frac{(x_k - \mu)^2}{\sigma^2} \right) \right).$$

2. Dans ce modèle statistique, sous $p_\theta \cdot \mathrm{Leb}^{\otimes n}$, les variables canoniques $X_1, \cdots, X_n$ sont i.i.d. de loi gaussienne de paramètres (μ, σ^2). Comme dans l'exercice 3, on peut choisir un estimateur des moments basé sur le premier et second moment ; ou un estimateur par maximum de vraisemblance. Les deux approches conduisent au même estimateur

   $$\hat{\theta}_n = \begin{bmatrix} \hat{\mu}_n \\ \hat{\sigma}_n^2 \end{bmatrix} := \begin{bmatrix} n^{-1} \sum_{k=1}^n X_k \\ n^{-1} \sum_{k=1}^n (X_k - \hat{\mu}_n)^2 \end{bmatrix}$$

3. Ce modèle gaussien est discutable : on observe en effet un histogramme avec des queues plus lourdes que la loi gaussienne. La symétrie de l'histogramme n'est pas non plus convaincante.

4. Si $X \sim \mathcal{N}(\mu, \sigma^2)$ alors $(X-\mu)/\sigma \sim \mathcal{N}(0,1)$. Par symétrie de la loi gaussienne centrée (qui admet des moments d'ordre quelconque), on a

$$\gamma_1 \propto \int_{\mathbb{R}} x^3 \exp(-x^2) dx = 0.$$

La transformée de Laplace de la loi gaussienne $\mathcal{N}(\mu, \sigma^2)$ est donnée par

$$t \mapsto \mathbb{E}\left[\exp(tX)\right] = \exp(t\mu + 0.5t^2\sigma^2)$$

ce qui donne $\ln \mathbb{E}\left[\exp(tX)\right] = t\mu + 0.5t^2\sigma^2$ dont on déduit les cumulants

$$\kappa_1 = \mu \qquad \kappa_2 = \sigma^2 \qquad \kappa_q = 0, q \geq 3.$$

Il vient que l'excès de kurtosis est nul.

5. On remplace les moments exacts par les moments empiriques ; on obtient pour estimateur du coefficient d'asymétrie

$$\hat{\gamma}_{1,n} := \left(\frac{1}{n}\sum_{k=1}^{n}(X_k - \hat{\mu}_n)^2\right)^{-3/2} \left(\frac{1}{n}\sum_{k=1}^{n}(X_k - \hat{\mu}_n)^3\right), \qquad \text{avec } \hat{\mu}_n := n^{-1}\sum_{k=1}^{n} X_k$$

et pour estimateur de l'excès de kurtosis

$$\hat{\gamma}_{2,n} := \left(\frac{1}{n}\sum_{k=1}^{n}(X_k - \hat{\mu}_n)^2\right)^{-2} \left(\frac{1}{n}\sum_{k=1}^{n}(X_k - \hat{\mu}_n)^4\right) - 3.$$

6. L'excès de kurtosis semble signicativement différent de 0, et il est positif ; cela confirme l'observation que la distribution empirique a des queues plus lourdes que la loi gaussienne. Il est donc conseillé d'envisager un autre modèle.

7. Les données collectées $(x_1, \cdots, x_n)$ sont une réalisation du vecteur aléatoire $Z = (X_1, \cdots, X_n)$ à valeur dans $\mathbb{R}^n$. On munit cet espace de sa tribu borélienne, et d'une famille de probabilités $\{\tilde{p}_\theta \cdot \mathrm{Leb}^{\otimes n}, \theta \in \Theta\}$ avec

$\theta := (\alpha_0, \alpha_1)$; $\Theta := \mathbb{R}_+^\star \times \mathbb{R}_+^\star$.

$$(x_1, \cdots, x_n) \mapsto \tilde{p}_\theta(x_1, \cdots, x_n) := \prod_{k=1}^{n}\left(\frac{1}{\sqrt{2\pi}\sqrt{\alpha_0 + \alpha_1 x_{k-1}^2}} \exp\left(-\frac{1}{2}\frac{x_k^2}{\alpha_0 + \alpha_1 x_{k-1}^2}\right)\right).$$

Pour l'expression de la densité, on a utilisé le fait que

— les v.a. $(X_k)_k$ ne sont pas indépendantes ; par définition de la loi conditionnelle on a "la loi de $(X_1, \cdots, X_n)$ est le produit de la loi de $(X_1, \cdots, X_{n-1})$ et de la loi conditionnelle de X_n sachant $(X_1, \cdots, X_{n-1})$" ce qui par récurrence donne (écriture sans rigueur)

$$\prod_{k=1}^{n} \text{"loi conditionnelle de } X_k \text{ sachant } (X_1, \cdots, X_{k-1})\text{"}$$

— Les v.a. $(Z_k)_k$ sont indépendantes et par récurrence, on voit que X_{k-1} est une fonction de $Z_1, \cdots, Z_{k-1}$. Donc Z_k est indépendant de X_{k-1}.

— Vue la relation liant X_k, Z_k, X_{k-1}, et puisque Z_k, X_{k-1} sont indépendants, on voit que la loi conditionnelle de X_k sachant $(X_1, \cdots, X_{k-1})$ est la loi conditionnelle de X_k sachant X_{k-1} ; et cette loi est une loi gaussienne centrée de variance $\alpha_0 + \alpha_1 X_{k-1}^2$.

2.7 Exercice 7

1. On considère une famille de scalaires $\beta = (\beta_1, \dots, \beta_K)$ tels que

$$\forall t \in \mathbb{R} \qquad \sum_{k=1}^{K} \beta_k e^{i\mu_k t} = 0. \tag{4}$$

On définit la fonction $f : t \longmapsto \sum_{k=1}^{K} \beta_k e^{i\mu_k t}$. On calcule les dérivées successives de f en remarquant que

$$\forall j \in \mathbb{N} \qquad f^{(j)}(t) = \sum_{k=1}^{K} \beta_k (i\mu_k)^j e^{i\mu_k t} = i^j \sum_{k=1}^{K} \beta_k (\mu_k)^j e^{i\mu_k t}.$$

Puisque f est la fonction nulle (voir (4)), ses dérivés successives sont nulles ce qui entraine en particulier que

$$\forall j \in \{0, \dots, K-1\} \qquad \sum_{k=1}^{K} \beta_k \{\mu_k\}^j = 0.$$

On reconnaît ainsi un système linéaire en les $\{\beta_k, 1 \le k \le K\}$ de Vandermonde $V\beta = 0$ avec

$$V := \begin{pmatrix} 1 & \dots & 1 \\ \mu_1 & \dots & \mu_K \\ \vdots & \dots & \vdots \\ \{\mu_1\}^{K-1} & \dots & \{\mu_K\}^{K-1} \end{pmatrix}.$$

La matrice V est inversible dès que les coefficients $\{\mu_k, 1 \le k \le K\}$ sont deux à deux distincts ; puisque $\mu \in \mathsf{M}$, V est inversible et β est le vecteur nul. La condition (4) implique donc la famille de fonctions est libre.

2. Dans ce modèle statistique, les variables canoniques sont indépendantes, de loi gaussienne d'espérance μ_i et de variance 1. X suit une loi mélange gaussien, alors

$$\mathbb{E}_{\alpha,\mu}\left[\exp(itX)\right] = \sum_{k=1}^{K} \alpha_k \, \mathbb{E}\left[\exp(itY_k)\right] = \sum_{k=1}^{K} \alpha_k \, \exp(it\mu_k - 0.5t^2)$$

en ayant noté Y_i une v.a. de loi $\mathcal{N}(\mu_i, 1)$. Si deux lois sont identiques, alors elles ont même fonction caractéristique. Considérons deux mélanges gaussiens, de paramètres respectifs (α, μ) et $(\tilde{\alpha}, \tilde{\mu})$; et supposons que $f_{\alpha,\mu} = f_{\tilde{\alpha},\tilde{\mu}}$. On obtient donc en passant aux fonctions caractéristiques que

$$\sum_{k=1}^{K} \alpha_k \exp(it\mu_k - 0.5t^2) = \sum_{k=1}^{K} \tilde{\alpha}_k \exp(it\tilde{\mu}_k - 0.5t^2)$$

En simplifiant par $e^{-0.5t^2}$, on obtient que

$$\sum_{k=1}^{K} \alpha_k \exp(it\mu_k) - \sum_{k=1}^{K} \tilde{\alpha}_k \exp(it\tilde{\mu}_k) = 0.$$

En utilisant la question précédente, on voit que

— s'il existe $1 \leq i, j \leq K$ tels que $\mu_i = \tilde{\mu}_j$, alors $\alpha_i = \tilde{\alpha}_j$
— pour tout i tel que $\forall j,\ \mu_i \neq \tilde{\mu}_j$: $\alpha_i = 0$ - ce qui contredit la condition "les composantes du vecteur α sont strictement positives".

On en déduit que seul le premier cas se produit, ce qui entraine : $\alpha = \tilde{\alpha}$ et $\mu = \tilde{\mu}$.

3. On vérifie la formule de la fonction caractéristique d'une loi de Poisson de paramètre μ par un calcul direct : si X suit une loi de Poisson $\mathcal{P}(\mu)$, alors :

$$\mathbb{E}[e^{itX}] = e^{-\mu} \sum_{k=0}^{+\infty} e^{-\mu} \frac{\mu^k}{k!} e^{itk} = \sum_{k=0}^{+\infty} \frac{(\mu e^{it})^k}{k!} = e^{\mu[e^{it}-1]}.$$

On considère ensuite deux mélanges de Poisson correspondant au même modèle statistique :

$$q_{\alpha,\mu} = q_{\tilde{\alpha},\tilde{\mu}},$$

pour deux familles de paramètres (α, μ) et $(\tilde{\alpha}, \tilde{\mu})$ (avec le même nombre de composantes K, et des composantes des vecteurs $\alpha, \tilde{\alpha}$ non nulles). Si les deux lois sont identiques, elles ont alors mêmes fonctions caractéristiques, et on obtient donc que

$$\forall t \in \mathbb{R} \qquad \sum_{k=1}^{K} \alpha_k e^{\mu_k[e^{it}-1]} = \sum_{k=1}^{K} \tilde{\alpha}_k e^{\tilde{\mu}_k[e^{it}-1]}$$

On montre alors comme précédemment que la famille de fonctions $(e^{\mu[e^{it}-1]})_{\mu\in\mathbb{R}}$ est une famille libre. Pour ce faire, on considère un nombre fini de paramètres $\mu_1 < \ldots < \mu_s$ ainsi qu'un nombre de paramètres $\beta = (\beta_1, \ldots, \beta_s)$ tels que la fonction f définie par

$$f : t \longmapsto \sum_{k=1}^{s} \beta_k e^{\mu_k[e^{it}-1]},$$

est nulle pour tout t. On remarque que $f(0) = \sum_{k=1}^{s} \beta_k$, de sorte que $\sum_{k=1}^{s} \beta_k = 0$.
De même, $f'(t) = \sum_{k=1}^{s} \beta_k \mu_k e^{it} e^{\mu_k[e^{it}1]}$ et $f'(0) = \sum_{k=1}^{s} \beta_k \mu_k$; et donc $\sum_{k=1}^{s} \beta_k \mu_k = 0$. Puis

$$f^{(2)}(t) = if'(t) + \sum_{k=1}^{s} \beta_k \{\mu_k\}^2 e^{2it} e^{\mu_k[e^i t-1]},$$

de sorte que $f^{(2)}(0) = \sum_{k=1}^{s} \beta_k \{\mu_k\}^2$.
On démontre ainsi par récurrence sur j que

$$\forall j \geq 1 \qquad f^{(j)}(0) = \sum_{k=1}^{s} \beta_k \{\mu_k\}^j.$$

Encore une fois, on obtient un système linéaire en les coefficients β de matrice, une matrice de Vandermonde. On déduit donc que nécessairement tous les coefficients $(\beta_1, \ldots, \beta_s)$ sont nuls.
Cela implique l'identifiabilité du modèle statistique.