PARTIEL 2019 - 2020

(3 HEURES – DOCUMENTS NON AUTORISÉS)

Exercice 1 (Modèle Poisson-Gamma, 6pts)

On considère des observations $X_1, \dots, X_n$ i.i.d de loi de Poisson $\mathcal{P}(\lambda)$ d'intensité $\lambda > 0$ dont on rappelle la densité $f_\lambda(x) = \lambda^x e^{-\lambda}/x!$ par rapport à la mesure de comptage sur $\mathbb{N}$.

1. (0.5pt) Calculez $\mathbb{E}(X_1)$ et $\mathrm{Var}(X_1)$

   Réponse. On peut utiliser par exemple la fonction génératrice des moments $\phi_X(t) = \mathbb{E}[e^{tX}]$ où $X \sim \mathcal{P}(\lambda)$ car on a $\phi_X'(0) = \mathbb{E}[X]$ et $\phi_X''(0) = \mathbb{E}[X^2]$. On a $\phi_X(t) = \sum_{k \geq 0} e^{tk}\lambda^k e^{-\lambda}/k! = e^{\lambda(e^t-1)}$ et $\phi_X'(t) = \lambda e^t e^{\lambda(e^t-1)}$ et $\phi_X''(t) = \lambda e^t e^{\lambda(e^t-1)} + \lambda^2 e^{2t} e^{\lambda(e^t-1)}$ de sorte que $\mathbb{E}[X] = \phi_X'(0) = \lambda$ et $\mathrm{Var}(X) = \phi_X''(0) - (\phi_X'(0))^2 = \lambda + \lambda^2 - \lambda^2 = \lambda$.

2. (0.5pt) Montrez que l'estimateur $\widehat{\lambda}_n = \overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ est un estimateur sans biais et calculez son risque quadratique. Donnez la loi limite de $\sqrt{n}(\widehat{\lambda}_n - \lambda)$.

   Réponse. On a $\mathbb{E}[\widehat{\lambda}_n] = \mathbb{E}[X_1] = \lambda$ donc $\widehat{\lambda}_n$ est sans biais et $\mathrm{Var}(\widehat{\lambda}_n) = \mathrm{Var}(X_1)/n = \lambda/n$ (car $X_1, \dots, X_n$ i.i.d) et donc la décomposition biais variance donne $R(\widehat{\lambda}_n, \lambda) = \mathrm{Var}(\widehat{\lambda}_n) = \lambda/n$. Le théorème central limite nous donne donc $\sqrt{n}(\widehat{\lambda}_n - \lambda) \rightsquigarrow \mathcal{N}(0, \lambda)$.

3. (1,5pts) Construisez deux intervalles de confiance pour λ asymptotiquement de niveau $1-\alpha$, où $\alpha \in (0,1)$, l'un en utilisant le lemme de Slutsky et l'autre en utilisant la Δ-méthode.

   Réponse. On a $\sqrt{n}(\widehat{\lambda}_n - \lambda) \rightsquigarrow \mathcal{N}(0,\lambda)$ et $\widehat{\lambda}_n \to \lambda$ en probabilités donc d'après Slutsky on obtient $\sqrt{n/\widehat{\lambda}_n}(\widehat{\lambda}_n - \lambda) \rightsquigarrow \mathcal{N}(0,1)$ puisque $x \mapsto 1/\sqrt{x}$ est continue sur $(0, +\infty)$. On en déduit un premier intervalle

   $$\left[\widehat{\lambda}_n - \sqrt{\frac{\widehat{\lambda}_n}{n}}\Phi^{-1}\Big(1 - \frac{\alpha}{2}\Big), \widehat{\lambda}_n + \sqrt{\frac{\widehat{\lambda}_n}{n}}\Phi^{-1}\Big(1 - \frac{\alpha}{2}\Big)\right]$$

   qui est asymptotiquement de niveau $1 - \alpha$, où on rappelle que Φ est la fonction de répartition de $\mathcal{N}(0,1)$ et Φ^{-1} la foncton quantile. On peut obtenir un deuxième intervalle en utilisant la Δ méthode. En effet si g est une fonction C^1 on a $\sqrt{n}(g(\widehat{\lambda}_n) - g(\lambda)) \rightsquigarrow \mathcal{N}(0, \lambda(g'(\lambda))^2)$ de sorte que si on choisit $g(\lambda) = 2\sqrt{\lambda}$ on obtient $2\sqrt{n}(\sqrt{\widehat{\lambda}_n} - \sqrt{\lambda}) \rightsquigarrow \mathcal{N}(0,1)$ ce qui amène à un deuxième intervalle de confiance de niveau asymptotiquement $1 - \alpha$ donné par

   $$\left[\Big(\Big(\sqrt{\widehat{\lambda}_n} - \frac{\Phi^{-1}(1-\alpha/2)}{2\sqrt{n}}\Big)_+\Big)^2, \Big(\sqrt{\widehat{\lambda}_n} + \frac{\Phi^{-1}(1-\alpha/2)}{2\sqrt{n}}\Big)^2\right]$$

   ou on rappelle que $x_+ = \max(x, 0)$ pour $x \in \mathbb{R}$.

4. (1pt) Construisez un test d'hypothèse $H_0 : \lambda = \lambda_0$ contre $H_1 : \lambda \neq \lambda_0$, où $\lambda_0 > 0$ est fixé, asymptotiquement de niveau $\alpha \in (0,1)$. Calculez la p-valeur de ce test.

   Réponse. On utilise un test de région de rejet

   $$R = \left\{ |\widehat{\lambda}_n - \lambda_0| \geq \sqrt{\frac{\widehat{\lambda}_n}{n}}\Phi^{-1}\Big(1 - \frac{\alpha}{2}\Big) \right\}$$

de sorte que d'après la Question 4. on a par construction $\mathbb{P}_{\lambda=\lambda_0}(R) \to \alpha$ ce qui montre que ce test est de niveau asymptotiquement α. La p-valeur est le niveau α^* tel que $|\widehat{\lambda}_n - \lambda_0| \geq (\widehat{\lambda}_n/n)^{1/2}\Phi^{-1}(1-\alpha/2)$ donc

$$\alpha^* = 2\Big(1 - \Phi\Big(\sqrt{\frac{n}{\widehat{\lambda}_n}}|\widehat{\lambda}_n - \lambda_0|\Big)\Big).$$

5. (1pt) On rappelle que la loi Gamma de paramètres de forme $a > 0$ et d'intensité $\tau > 0$, notée $\text{Gamma}(a,\tau)$, a pour densité $g_{a,\tau}(x) = \frac{\tau^a}{\Gamma(a)}x^{a-1}e^{-\tau x}\mathbf{1}_{x\geq 0}$ contre la densité de Lebesgue sur $\mathbb{R}$. Si G est une variable aléatoire de loi $\text{Gamma}(a,\tau)$, calculez $\mathbb{E}(G)$ et $\text{Var}(G)$.

   Réponse. On utilise le fait que $\int g_{a,\tau}(x)dx = 1$ pour tout $a,\tau > 0$. On a $\mathbb{E}[G] = \int_0^{+\infty}\frac{\tau^a}{\Gamma(a)}x^a e^{-\tau x} = \frac{\Gamma(a+1)}{\tau\Gamma(a)}\int g_{a+1,\tau}(x)dx = a/\tau$ et $\mathbb{E}[G^2] = \frac{\Gamma(a+2)}{\tau^2\Gamma(a)}\int g_{a+2,\tau}(x)dx = a(a+1)/\tau^2$ et donc $\text{Var}[G] = a/\tau^2$.

6. (1pt) On se place maintenant dans un cadre bayesien. On suppose que $X_1,\dots,X_n$ sont i.i.d et de loi conditionnelle à $\boldsymbol{\lambda} = \lambda$ de loi $\mathcal{P}(\lambda)$ et que $\boldsymbol{\lambda}$ suit la loi a priori $\text{Gamma}(a,\tau)$. On introduit $X = (X_1,\dots,X_n)$ et $x = (x_1,\dots,x_n) \in \mathbb{N}^n$. Montrez que la loi a posteriori de $\boldsymbol{\lambda}$ sachant $X = x$ est $\text{Gamma}(\sum_{i=1}^n x_i + a, n+\tau)$. Que dit-on alors concernant les lois Poisson et Gamma ?

   Réponse. $X_1,\dots,X_n$ sont i.i.d de densité conditionnelle $f_{X|\boldsymbol{\lambda}}(x) = e^{-x}\lambda^x/x!$ pour $x \in \mathbb{N}$ contre la mesure de comptage sur $\mathbb{N}$ et la loi a priori sur $\boldsymbol{\lambda}$ est donnée par $g_{\boldsymbol{\lambda}}(\lambda) = \frac{\tau^a}{\Gamma(a)}\lambda^{a-1}e^{-\tau\lambda}\mathbf{1}_{\lambda>0}$ donc la loi jointe des données et des observations est égale à

$$\begin{aligned} f_{(X_1,\dots,X_n,\boldsymbol{\lambda})}(x_1,\dots,x_n,\lambda) &= f_{(X_1,\dots,X_n)|\boldsymbol{\lambda}=\lambda}(x_1,\dots,x_n)g_{\boldsymbol{\lambda}}(\lambda) \\ &= \frac{\tau^a}{\Gamma(a)\prod_{i=1}^n x_i!}\lambda^{\sum_{i=1}^n x_i + a - 1}e^{-(n+\tau)\lambda}\mathbf{1}_{\lambda>0}. \end{aligned}$$

   On sait que la densité de la loi a posteriori $f_{\boldsymbol{\lambda}|(X_1,\dots,X_n)=(x_1,\dots,x_n)}(\lambda)$ est proportionnelle à la densité jointe $f_{(X_1,\dots,X_n,\boldsymbol{\lambda})}(x_1,\dots,x_n,\lambda)$, donc il s'agit nécessairement de la densité de la loi $\text{Gamma}(\sum_{i=1}^n x_i + a, n+\tau)$. On a donc montré que la loi a priori et la loi a posteriori appartiennent toutes les deux à la famille des loi Gamma lorsque la loi conditionnelle des données est Poisson. On dit alors que les lois Gamma et Poisson sont *conjuguées.*

7. (1,5pts) Montrez que l'estimateur bayesien de λ pour la perte quadratique est égal à $\widehat{\lambda}_n^B = (\sum_{i=1}^n X_i + a)(n+\tau)$. Calculez son risque quadratique et le risque bayesien.

   Réponse. L'estimateur bayesien associé à la perte quadratique est l'espérance de la loi a posteriori, qui vaut ici (cf Question 5.) $\widehat{\lambda}_n^B = (\sum_{i=1}^n X_i + a)(n+\tau)$. On a $\mathbb{E}[\widehat{\lambda}_n^B] = (n\lambda + a)/(n+\tau)$ et donc $\text{biais}(\widehat{\lambda}_n^B,\lambda) = \mathbb{E}[\widehat{\lambda}_n^B] - \lambda = (a - \lambda\tau)/(n+\tau)$. On utilise la décomposition biais-variance pour calculer le risque quadratique

$$\begin{aligned} R(\widehat{\lambda}_n^B,\lambda) &= \text{biais}(\widehat{\lambda}_n^B,\lambda)^2 + \text{Var}(\widehat{\lambda}_n^B) \\ &= \Big(\frac{a-\lambda\tau}{n+\tau}\Big)^2 + \frac{n\lambda}{(n+\tau)^2} \\ &= \frac{\tau^2(\lambda - a/\tau)^2 + n\lambda}{(n+\tau)^2} \end{aligned}$$

   et le risque bayesien de calcule alors en utilisant le fait que $\mathbb{E}_{\lambda\sim\Gamma(a,\tau)}[\lambda] = a/\tau$ et $\mathbb{E}_{\lambda\sim\Gamma(a,\tau)}[(\lambda - a/\tau)^2] = \text{Var}_{\lambda\sim\Gamma(a,\tau)}(\lambda) = a/\tau^2$ de sorte que

$$R_B(\widehat{\lambda}_n^B) = \frac{\tau^2 a/\tau^2 + na/\tau}{(n+\tau)^2} = \frac{a}{\tau(n+\tau)}.$$

Exercice 2 (ANOVA, 4pts)

On considère des variables aléatoires $X_{i,j} \sim \mathcal{N}(m_i, \sigma^2)$ pour $i = 1, \ldots, k$ et $j = 1, \ldots, n_i$ toutes indépendantes. Cela revient à observer k échantillons gaussiens de tailles respectives $n_1, n_2, \ldots, n_k$, notés $X_{i,\bullet} = [X_{i,1} \cdots X_{i,n_i}]^\top \in \mathbb{R}^{n_i}$. On suppose que les paramètres $m = [m_1 \cdots m_k]^\top \in \mathbb{R}^k$ et $\sigma^2 > 0$ sont inconnus. On veut construire un test d'hypothèses

$$H_0 : m_1 = m_2 = \cdots = m_k \quad \text{contre} \quad H_1 : \exists i \neq i' \text{ tels que } m_i \neq m_{i'},$$

autrement dit, on veut tester si les moyennes des k échantillons sont égales.

1. (1pt) On considère le vecteur aléatoire $X = (X_{i,j})_{1 \le i \le k, 1 \le j \le n_i} \in \mathbb{R}^n$ avec $n = \sum_{i=1}^k n_i$. Montrez que $\mu = \mathbb{E}(X)$ appartient à un sous-espace vectoriel E de dimension k. Calculez $X_E = \operatorname{proj}_E(X)$ la projection orthogonale de X dans E.

   Réponse. On a
   $$\mu = \mathbb{E}X = \big[m_1 \cdots m_1\, m_2 \cdots m_2 \cdots m_k \cdots m_k\big]^\top \in \mathbb{R}^n$$
   où chaque terme m_i est répété n_i fois pour $i = 1, \ldots, k$. On introduit le vecteur $e_1 \in \mathbb{R}^d$ dont les n_1 premières coordonnées sont 1 et toutes les autres 0, le vecteur e_2 dont les n_1 premières coordonnées sont 0, les n_2 suivantes sont 1, toutes les autres 0, et ainsi de suite pour définir les vecteurs $e_1, \ldots, e_k$ de $\mathbb{R}^n$ qui sont clairement orthogonaux. On a ainsi $\mu = \sum_{i=1}^k m_i e_i$ ce qui montre que $\mu \in V$ où $V = \operatorname{span}(e_1, \ldots, e_k)$ est un sous-espace de dimension k de $\mathbb{R}^n$. Les $e_1, \ldots, e_k$ sont orthogonaux et de normes respectives $\sqrt{n}_1, \ldots, \sqrt{n}_k$ de sorte que $X_E = \operatorname{proj}_E(X) = \sum_{i=1}^k \langle X, \frac{1}{\sqrt{n_i}} e_i \rangle \frac{1}{\sqrt{n_i}} e_i = \sum_{i=1}^k \frac{1}{n_i} \langle X, e_i \rangle e_i = \sum_{i=1}^k \overline{X}_{i,\bullet} e_i$ où on a posé $\overline{X}_{i,\bullet} = \frac{1}{n_i} \langle X, e_i \rangle = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{i,j} =$ la moyenne du i-ème échantillon.

2. (1pt) Montrez que l'hypothèse nulle s'écrit $H_0 : \mu \in F$, où F est un sous-espace vectoriel de E de dimension 1. Calculez X_F.

   Réponse. Sous H_0 toutes les coordonnées de μ sont égales, donc $\mu \in F$ où $F = \operatorname{span}(\mathbf{1})$ avec $\mathbf{1} \in \mathbb{R}^n$ le vecteur dont toutes les coordonnées sont constantes et égales à 1 et évidemment $\dim(F) = 1$ et $X_F = \operatorname{proj}_F(X) = \overline{X}\mathbf{1}$ où $\overline{X} = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^{n_i} X_{i,j}$ est la moyenne de l'union des k échantillons.

3. (2pts) Construisez alors un test de Fisher pour les hypothèses décrites au dessus.

   Réponse. On peut écrire $X = \mu + \varepsilon$ où $\varepsilon \sim \mathcal{N}(0, \sigma^2 I_n)$. On décompose $\mathbb{R}^n = E^\perp \oplus E = E^\perp \oplus F \oplus W$. Le principe du test est d'utiliser le fait que $X_E - X_F = X_W = \operatorname{proj}_W(\mu + \varepsilon) = \operatorname{proj}_W(\mu) + \operatorname{proj}_W(\varepsilon)$ et que *quand H_0 est vraie* $\mu \in F$ donc $\operatorname{proj}_W(\mu) = 0$. On a donc, *sous* H_0, que $\frac{1}{\sigma^2}\|X_E - X_F\|^2 = \|\operatorname{proj}_W(\varepsilon/\sigma)\|^2$. De plus $X - X_E = \operatorname{proj}_{E^\perp}(X) = \operatorname{proj}_{E^\perp}(\mu + \varepsilon) = \operatorname{proj}_{E^\perp}(\varepsilon)$ car $\mu \in E$ et donc $\frac{1}{\sigma^2}\|X - X_E\|^2 = \|\operatorname{proj}_{E^\perp}(\varepsilon/\sigma)\|^2$. Puisque $\varepsilon/\sigma \sim \mathcal{N}(0, I_n)$ on a d'après le théorème de Cochran que $\|\operatorname{proj}_W(\varepsilon/\sigma)\|^2 \sim \chi^2(\dim(W))$, que $\|\operatorname{proj}_{E^\perp}(\varepsilon/\sigma)\|^2 \sim \chi^2(\dim(E^\perp))$ et que $\|\operatorname{proj}_{E^\perp}(\varepsilon/\sigma)\|^2$ et $\|\operatorname{proj}_W(\varepsilon/\sigma)\|^2$ sont indépendantes car $E^\perp \perp W$. Par ailleurs $\dim(E^\perp) = n - k$ et $\dim(W) = k - 1$. Cela entraine d'après la définition de la loi de Fisher $\mathcal{F}(p, q)$ a p, q degrés de libertés que
   $$T = \frac{\|X_E - X_F\|^2/(k-1)}{\|X - X_E\|^2/(n-k)} \sim \mathcal{F}(k-1, n-k)$$
   et qu'on peut donc considérer le test de Fisher de région de rejet $R = \{T \ge q_{k-1,n-k}(1-\alpha)\}$ où $q_{p,q}(1-\alpha)$ est le quantile d'ordre $1-\alpha$ de la loi $\mathcal{F}(k-1, n-k)$. Écrivons T sous une forme un peu plus interprétable. On a d'une part
   $$\|X - X_E\|^2 = \sum_{i=1}^k \sum_{j=1}^{n_i} (X_{i,j} - \overline{X}_{i,\bullet})^2 = \sum_{i=1}^k n_i \frac{1}{n_i} \sum_{j=1}^{n_i} (X_{i,j} - \overline{X}_{i,\bullet})^2 = nV_{\text{intra}}$$

où V_{intra} est la *variance intra-classes* (une *classe* correspondant à un échantillon i) qui correspond à la moyenne des variances $\frac{1}{n_i}\sum_{j=1}^{n_i}(X_{i,j}-\overline{X}_{i,\bullet})^2$ de chaque échantillon $i=1,\dots,k$ et d'autre part

$$\|X_E - X_F\|^2 = \sum_{i=1}^{k} n_i(\overline{X}_{i,\bullet} - \overline{X})^2 = nV_{\text{inter}}$$

où V_{inter} est la *variance inter-classes* qui correspond à la variance des moyennes de chaque échantillon. Cela explique le nom d'ANOVA (ANalysis of VAriance), car le de Fisher utilise dans ce cas particulier la statistique de test

$$T = \frac{V_{\text{inter}}/(k-1)}{V_{\text{intra}}/(n-k)},$$

qui est basée sur le ratio des variances inter et intra V_{inter} et V_{intra}.

Exercice 3. (Sur le range d'un échantillon, 3pts)

Soient $X_1,\dots,X_n$ des variables aléatoires réelles i.i.d. On pose

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{et} \quad \text{range}(X_1,\dots,X_n) = \max_{1\le i\le n} X_i - \min_{1\le i\le n} X_i.$$

1. (1,5pt) On suppose que les $X_1,\dots,X_n$ sont de loi $\mathcal{N}(m,\sigma^2)$. Montrez que $\overline{X}_n$ et range$(X_1,\dots,X_n)$ sont indépendants.

   Réponse. On a $X=(X_1,\dots,X_n)\sim\mathcal{N}(m\mathbf{1},\sigma^2 I_n)$ et $\varepsilon=(X-m)/\sigma\sim\mathcal{N}(0,I_n)$. On pose $E=\text{span}(\mathbf{1})$ de sorte que $\text{proj}_E(X)=\overline{X}\mathbf{1}$ et $\text{proj}_{E^\perp}(X)=X-\overline{X}\mathbf{1}$ où $\overline{X}=\frac{1}{n}\sum_{i=1}^n X_i$. Mais par ailleurs $\text{proj}_E(X)=\text{proj}_E(m\mathbf{1}+\sigma\varepsilon)=m\mathbf{1}+\sigma\,\text{proj}_E(\varepsilon)$ et $\text{proj}_{E^\perp}(X)=\sigma\,\text{proj}_{E^\perp}(\varepsilon)$ ce qui montre d'après le théorème de Cochran que $\overline{X}\mathbf{1}$ et $X-\overline{X}\mathbf{1}$ sont indépendants et donc en particulier que $\overline{X}$ et range$(X_1,\dots,X_n)$ = range$(X_1-\overline{X},\dots,X_n-\overline{X})$ sont indépendants.

2. (1,5pt) On suppose que les $X_1,\dots,X_n$ sont de loi exponentielle $\mathcal{E}(\lambda)$ d'intensité $\lambda>0$. Déterminez la loi de range$(X_1,\dots,X_n)$.

   Réponse.

Exercice 4. (Loi de Student, 4pts)

Soit X une variable aléatoire de loi $\mathcal{N}(0,1)$ indépendante d'une suite $(Y_n)_{n\ge 1}$ de variables aléatoires où Y_n suit la loi $\chi^2(n)$ du Chi-deux à n degrés de liberté. On rappelle que la loi $\chi^2(n)$ correspond à une loi Gamma$(n/2,1/2)$ dont la densité est rappelée dans l'Exercice 1. On pose $T_n = X/\sqrt{Y_n/n}$.

1. (1,5pts) Montrez que la densité t_n de T_n est donnée par

$$t_n(x) = \frac{1}{\sqrt{\pi}}\frac{\Gamma((n+1)/2)}{\sqrt{n}\,\Gamma(n/2)}\frac{1}{(1+x^2/n)^{(n+1)/2}}.$$

   Réponse. Soit φ continue bornée et calculons puisque $X\sim\mathcal{N}(0,1)$, $Y_n\sim\chi^2(n)=\Gamma(n/2,1/2)$ et comme X et Y_n sont indépendantes:

$$\mathbb{E}[\varphi(T_n,Y_n)] = \mathbb{E}\varphi\Big(\frac{X}{\sqrt{Y_n/n}},Y_n\Big) = \int_{\mathbb{R}}\int_{\mathbb{R}+}\varphi\Big(\frac{x}{\sqrt{y/n}},y\Big)\frac{e^{-x^2/2}}{\sqrt{2\pi}}\frac{e^{-y/2}y^{n/2-1}}{2^{n/2}\Gamma(n/2)}dxdy.$$

La fonction $g : \mathbb{R} \times \mathbb{R}_+^* \to \mathbb{R} \times \mathbb{R}_+^*$ donnée par $g(x,y) = (\frac{x}{\sqrt{y/n}}, y) = (u,v)$ est un C^1-difféomorphisme de jacobien $\begin{vmatrix} \sqrt{n/y} & -\frac{x\sqrt{n}}{2y^{3/2}} \\ 0 & 1 \end{vmatrix} = \sqrt{n/y} = \sqrt{n/v}$, on peut donc procéder au changement de variable $(u,v) = g(x,y)$ pour obtenir

$$\mathbb{E}[\varphi(T_n, Y_n)] = \int_{\mathbb{R}} \int_{\mathbb{R}^+} \varphi(u,v) \frac{e^{-\frac{v}{2}(\frac{u^2}{n}+1)} v^{(n-1)/2}}{2^{n/2}\sqrt{2\pi n}\Gamma(n/2)} dudv$$

ce qui donne la densité jointe de (T_n, Y_n). La loi de T_n s'obtient alors en intégrant par rapport à v et en faisant apparaitre une densité $g_{a,\lambda}$ de la loi Gamma(a, λ):

$$\begin{aligned}\int_{\mathbb{R}^+} \frac{e^{-\frac{v}{2}(\frac{u^2}{n}+1)} v^{(n-1)/2}}{2^{n/2}\sqrt{2\pi n}\Gamma(n/2)} dv &= \frac{\Gamma((n+1)/2)}{\Gamma(n/2)} \frac{1}{\sqrt{\pi n}} \frac{1}{(u^2/n+1)^{(n+1)/2}} \int_{\mathbb{R}^+} g_{(n+1)/2,(u^2/n+1)/2}(x)dx \\ &= \frac{\Gamma((n+1)/2)}{\Gamma(n/2)} \frac{1}{\sqrt{\pi n}} \frac{1}{(u^2/n+1)^{(n+1)/2}},\end{aligned}$$

qui est donc la densité de la loi de student à n degrés de liberté.

2. (1,5pts) Montrez que

$$\lim_{n\to\infty} t_n(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

pour tout $x \in \mathbb{R}$. On pourra tout d'abord montrer que $\Gamma(a+1/2)/\Gamma(a) \sim \sqrt{a}$ quand $a \to \infty$ en remarquant que $\Gamma(a+1/2)/\Gamma(a) = \mathbb{E}[\sqrt{U}]$ où U suit une loi $\Gamma(a,1)$.

Réponse. On commence par remarquer que

$$\frac{\Gamma(a+1/2)}{\Gamma(a)} = \int_{\mathbb{R}_+} \frac{x^{a-1/2}e^{-x}}{\Gamma(a)} dx = \int_{\mathbb{R}_+} \sqrt{x} g_{a,1}(x) dx = \mathbb{E}[\sqrt{U}]$$

où $U \sim$ Gamma$(a,1)$ et on rappelle que $\mathbb{E}U = \mathrm{Var}(U) = a$ et donc $\mathbb{E}[(U-a)^2]/a = 1$. Mais $|\sqrt{U} - \sqrt{a}| = |U-a|/(\sqrt{U}+\sqrt{a}) \le |U-a|/\sqrt{a}$ donc $\mathbb{E}|\sqrt{U}-\sqrt{a}| \le \sqrt{\mathbb{E}[(U-a)^2]}/\sqrt{a} = 1$ d'après l'inégalité de Cauchy-Schwarz et donc $|\mathbb{E}\sqrt{U}/\sqrt{a} - 1| \le \mathbb{E}|\sqrt{U/a} - 1| \le 1/\sqrt{a}$ ce qui montre bien que $\mathbb{E}[\sqrt{U}]/\sqrt{a} \to 1$ quand $a \to +\infty$ et que donc $\Gamma(a+1/2) \sim \sqrt{a}\Gamma(a)$. Cela entraine que $\frac{\Gamma(n/2+1/2)}{\Gamma(n/2)} \sim \sqrt{n/2}$ quand $n \to +\infty$ et le fait que $(u^2/n+1)^{-(n+1)/2} \to e^{-u^2/2}$ permet de conclure.

3. (1pts) En déduire que T_n converge en loi vers $\mathcal{N}(0,1)$.

Réponse. On a vu que $t_n(x) \to \phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ pour tout $x \in \mathbb{R}$. Alors le *lemme de Scheffé* permet de conclure, car on a alors $t_n \to \phi$ dans L^1 et donc en loi, i.e. si $T_n \sim t_n$ alors $T_n \rightsquigarrow \mathcal{N}(0,1)$.

Exercice 5. (Modèle linéaire hétéroscédastique, 3pts)

On observe un vecteur aléatoire $Y = [Y_1 \cdots Y_n]^\top \in \mathbb{R}^n$ de loi $\mathcal{N}(X\beta, \sigma^2 S)$ où

$$X = \begin{bmatrix} x_{1,1} & \dots & x_{1,p} \\ \vdots & \ddots & \vdots \\ x_{n,1} & \dots & x_{n,p} \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} \quad \text{et} \quad S = \begin{bmatrix} f(x_{1,1}) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & f(x_{n,1}) \end{bmatrix},$$

avec $f(x_{i,1}) > 0$ pour tout $i = 1, \dots, n$. On suppose que l'on observe la matrice X, que X est de rang plein et que la fonction f est connue. On souhaite estimer les paramètres inconnus $\beta \in \mathbb{R}^p$ et $\sigma^2 > 0$. Il s'agit d'un modèle linéaire gaussien *hétéroscédastique*, par opposition au modèle linéaire *homoscédastique* étudié dans le cours où $S = I_n$.

1. (1pt) Construisez une matrice M telle que $Z = MY$ suive un modèle linéaire homoscédastique. En déduire un estimateur $\widehat{\beta}$ de β en utilisant la procédure des moindres carrés sur Z.

   Réponse. Posons $M = S^{-1/2}$. On a alors $Z = MY \sim \mathcal{N}(MX\beta, \sigma^2 I_n)$ de sorte que Z suit un modèle linéaire gaussien homoscédastique. L'estimateur des moindres carrés dans ce modèle est donné par $\widehat{\beta} = \arg\min_{b\in\mathbb{R}^p} \|Z - MXb\|^2 = ((MX)^\top MX)^{-1}(MX)^\top Z = (X^\top S^{-1}X)^{-1}X^\top S^{-1}Y$.

2. (2pts) On peut aussi construire l'estimateur des moindres carrés directement sur Y en calculant $\widetilde{\beta} = \arg\min_{b\in\mathbb{R}^p} \|Y - Xb\|^2$. Comparez $\mathbb{E}\big[\|\widehat{\beta} - \beta\|^2\big]$ et $\mathbb{E}\big[\|\widetilde{\beta} - \beta\|^2\big]$ et conclure.

   Réponse. L'estimateur des moindres carrés directement appliqué sur Y est égal à $\widetilde{\beta} = (X^\top X)^{-1}X^\top Y$. Commençons par donner les lois des deux estimateurs. On a $\widehat{\beta} \sim \mathcal{N}(\beta, \sigma^2(X^\top S^{-1}X)^{-1})$ et $\widetilde{\beta} \sim \mathcal{N}(\beta, \sigma^2(X^\top X)^{-1}X^\top SX(X^\top X)^{-1})$. On a $\mathrm{Var}(\widetilde{\beta}) = \mathrm{Var}(\widetilde{\beta} - \widehat{\beta} + \widehat{\beta}) = \mathrm{Var}(\widehat{\beta}) + \mathrm{Var}(\widetilde{\beta} - \widehat{\beta}) + 2\,\mathrm{cov}(\widetilde{\beta} - \widehat{\beta}, \widehat{\beta})$, et on calcule $\mathrm{cov}(\widetilde{\beta} - \widehat{\beta}, \widehat{\beta})$:

$$\begin{aligned}\mathrm{cov}(\widehat{\beta}, \widetilde{\beta} - \widehat{\beta}) &= \mathrm{cov}\Big((X^\top S^{-1}X)^{-1}X^\top S^{-1}Y, \big((X^\top X)^{-1}X^\top - (X^\top S^{-1}X)^{-1}X^\top S^{-1}\big)Y\Big)\\ &= (X^\top S^{-1}X)^{-1}S^{-1}\mathrm{Var}(Y)\big(X(X^\top X)^{-1} - S^{-1}X(X^\top S^{-1}X)^{-1}\big)\\ &= \sigma^2((X^\top S^{-1}X)^{-1} - (X^\top S^{-1}X)^{-1}) = 0_{p\times p}\end{aligned}$$

   ou on a utilisé le fait que $\mathrm{cov}(AY, BY) = A\mathrm{Var}(Y)B^\top$ et $\mathrm{Var}(Y) = \sigma^2 S$. On a donc montré que $\mathrm{Var}(\widetilde{\beta}) = \mathrm{Var}(\widehat{\beta}) + \mathrm{Var}(\widetilde{\beta} - \widehat{\beta})$ et donc que $\mathrm{tr}(\mathrm{Var}(\widetilde{\beta})) \geq \mathrm{tr}(\mathrm{Var}(\widehat{\beta}))$ puisque la matrice $\mathrm{Var}(\widetilde{\beta} - \widehat{\beta})$ est définie positive (c'est une matrice de covariance). Enfin, on a $\mathrm{tr}(\mathrm{Var}(\widehat{\beta})) = \mathrm{tr}(\mathbb{E}[(\widehat{\beta} - \beta)(\widehat{\beta} - \beta)^\top]) = \mathbb{E}[\mathrm{tr}((\widehat{\beta} - \beta)(\widehat{\beta} - \beta)^\top)] = \mathbb{E}[\|\widehat{\beta} - \beta\|^2]$.