

**Solutions des exos 8-10 du TD4
et
des exos 4,6,7,9 du TD 5**

TD4

Exo 8.

1. On a

$$E(X) = \frac{2\sqrt{\theta}}{\sqrt{\pi}} \int_0^{\infty} x e^{-\theta x^2} dx = \frac{1}{\sqrt{\pi\theta}} \int_0^{\infty} 2\theta x e^{-\theta x^2} dx = \frac{1}{\sqrt{\pi\theta}} \left[-e^{-\theta x^2} \right]_0^{\infty} = \frac{1}{\sqrt{\pi\theta}}.$$

et

$$\begin{aligned} E(X^2) &= \frac{2\sqrt{\theta}}{\sqrt{\pi}} \int_0^{\infty} x^2 e^{-\theta x^2} dx \\ &= \frac{\sqrt{\theta}}{\sqrt{\pi}} \int_{-\infty}^{\infty} x^2 e^{-\theta x^2} dx \quad (\text{par symétrie}) \\ &= \frac{\sqrt{\theta}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{u^2}{2} e^{-\theta \frac{u^2}{2}} du \quad (\text{par le changement de variable } x = \frac{u}{\sqrt{2}}) \\ &= \frac{1}{2} \frac{1}{\sqrt{2\pi} 1/\sqrt{\theta}} \int_{-\infty}^{\infty} u^2 e^{-\theta \frac{u^2}{2}} du \\ &= \frac{1}{2} \text{Var}(\mathcal{N}(0, 1/\theta)) = \frac{1}{2\theta} \end{aligned}$$

Donc

$$\text{Var}(X) = \frac{\pi - 2}{2\pi\theta}.$$

2. Si on choisit d'estimer $E(X)$ par sa version empirique \bar{X}_n , alors l'estimateur de θ par la méthode des moments est

$$T_n = \frac{1}{\pi \bar{X}_n^2}.$$

En utilisant la loi forte des grands nombres ($E(|X|) = E(X) < +\infty$) et le fait que la fonction $x \rightarrow 1/\pi x^2$ soit continue sur l'ouvert $]0, \infty[$, on montre que T_n converge fortement (ou presque sûrement) vers θ .

3. La log-vraisemblance est donnée par

$$l_n(\theta) = n \log(2/\sqrt{\pi}) + \frac{n}{2} \log(\theta) - \theta \sum_{i=1}^n X_i^2.$$

D'où

$$\frac{\partial l_n(\theta)}{\partial \theta} = \frac{n}{2\theta} - \sum_{i=1}^n X_i^2,$$

où

$$\frac{\partial l_n(\theta)}{\partial \theta} = 0 \iff \theta = \frac{n}{2 \sum_{i=1}^n X_i^2}$$

et

$$\frac{\partial^2 l_n(\theta)}{\partial \theta^2} = -\frac{n}{2\theta^2} < 0 \quad \forall \theta > 0.$$

On en déduit que

$$W_n = \frac{n}{2 \sum_{i=1}^n X_i^2}$$

est l'EMV de θ .

Pour montrer que W_n converge fortement on utilise la loi forte des grands nombres et le fait que la fonction $x \mapsto 1/x$ est continue sur l'ouvert $]0, \infty[$.

Exo 9

1. Pour que le support de la variable X soit égal à $\{-1, 0, 1\}$ il faut et il suffit que $\theta > 0$ et $1 - 2\theta > 0$ ce qui est équivalent à $\theta \in \Theta =]0, 1/2[$.

Il est facile de voir que $E(X) = 0$ et $E(X^2) = V(X) = 2\theta$.

2. L'estimateur de θ par la méthode des moment est donc

$$T_n = \frac{1}{2n} \sum_{i=1}^n X_i^2.$$

Étant donné que $E(X^2) < \infty$, la loi forte des grands nombre implique $T_n \xrightarrow{p.s.} \theta$. En outre, l'estimateur est sans biais et sa variance $Var(T_n) = \frac{\theta}{2n} \rightarrow 0$ donc T_n converge en moyenne quadratique, c-à-d, $E[(T_n - \theta)^2] \rightarrow 0$.

En appliquant le TCL pour les v.a. i.i.d. $Y_i = 1/(2X_i^2)$ ($Var(X^2) < +\infty$ puisque $X^2 \sim Bernoulli(2\theta)$), il vient que

$$\sqrt{n}(T_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{1}{4}Var(X^2)\right)$$

où $Var(X^2) = 2\theta(1 - 2\theta)$, et donc

$$\sqrt{n}(T_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\theta(1 - 2\theta)}{2}\right).$$

2. Soit R le nombre de X_i qui prennent la valeur 0. on peut écrire la vraisemblance en fonction de la variable R . En effet

$$L_n(\theta) = \theta^{n-R}(1 - 2\theta)^R.$$

Ainsi, la log-vraisemblance est donnée par

$$l_n(\theta) = \log L_n(\theta) = (n - R) \log(\theta) + R \log(1 - 2\theta)$$

et

$$\frac{\partial l_n(\theta)}{\partial \theta} = \frac{n - R}{\theta} - \frac{2R}{1 - 2\theta} = \frac{n(1 - 2\theta) - R}{\theta(1 - 2\theta)} = 0 \iff \theta = \frac{n - R}{2n}$$

et

$$\frac{\partial^2 l_n(\theta)}{\partial \theta^2} = -\frac{n - R}{\theta^2} - \frac{4R}{(1 - 2\theta)^2} < 0$$

Il en résulte que l'EMV de θ est

$$W_n = \frac{n - R}{2n}.$$

On a que $R = \sum_{i=1}^n 1_{X_i=0}$ donc égale à la somme de variables i.i.d. de loi de Bernoulli de paramètre $1 - 2\theta$, donc suit une $\mathcal{B}(n, 1 - 2\theta)$. Il s'en suit que

$$Var(W_n) = \frac{1}{4n^2} Var(R) = \frac{n(1 - 2\theta)2\theta}{4n^2} = \frac{\theta(1 - 2\theta)}{2n} \rightarrow 0,$$

aussi $E(W_n) = 1/2 - E(R)/2n = 1/2 - n(1 - 2\theta)/2n = \theta$ donc W_n est sans biais. Donc W_n converge en moyenne quadratique. D'autre part, la loi forte des grands nombres ($E(|1_{X=0}|) = E(1_{X=0}) < +\infty$) implique $R_n/n \xrightarrow{p.s.} E(1_{X=0}) = 1 - 2\theta$ et donc $W_n \xrightarrow{p.s.} \theta$.

Calculons maintenant la borne de Cramér-Rao associée au modèle. On a que

- $\Theta =]0, 1/2[$ ouvert.
- Le support de la loi de $Y = 1_{X=0}$ ne dépend pas de θ car elle appartient à une famille exponentielle puisque sa densité relativement à la mesure de comptage peut s'écrire

$$f(y, \theta) = (1 - 2\theta) \exp\left(\log\left(\frac{2\theta}{1 - 2\theta}\right) y\right), \quad y = 0, 1.$$

- $\theta \mapsto f(\cdot, \theta)$ est $C^1(\Theta)$ puisque $\alpha(\theta) = \log(2\theta/(1-2\theta))$ est C^1 et

$$\frac{\partial}{\partial \theta} \int_{\mathcal{Y}} f(y, \theta) d\nu(y) = \int_{\mathcal{Y}} \frac{\partial f(y, \theta)}{\partial \theta} d\nu(y)$$

où ν est la mesure de comptage (cela est vrai en général car il s'agit d'une famille exponentielle avec $\alpha \in C^1(\Theta)$). Ici, cela est évident car la loi est discrète à valeurs dans l'ensemble fini $\mathcal{Y} = \{0, 1\}$

- Si on pose $W_n = T(Y_1, \dots, Y_n)$, alors la condition

$$\frac{\partial}{\partial \theta} \int_{\mathcal{Y}^n} T(y_1, \dots, y_n) f(y_1, \dots, y_n, \theta) d\nu(y_1) \dots d\nu(y_n) = \int_{\mathcal{Y}^n} T(y_1, \dots, y_n) \frac{\partial f(y_1, \dots, y_n, \theta)}{\partial \theta} d\nu(y_1) \dots d\nu(y_n)$$

est clairement vérifiée car il s'agit d'une loi est discrète à valeurs dans un ensemble fini. En outre, on sait que $E(W_n^2) < +\infty$.

L'information de Fisher apportée par l'échantillon est donnée par

$$\begin{aligned} I_n(\theta) &= nI_X(\theta) \\ &= -nE \left[\frac{\partial^2 \log(f(Y, \theta))}{\partial \theta^2} \right] \end{aligned}$$

(la deuxième égalité est vraie car il s'agit d'une famille exponentielle avec $\alpha \in C^2(\Theta)$) et donc

$$\begin{aligned} I_n(\theta) &= n \left(\frac{4}{(1-2\theta)^2} E(Y) + \frac{1}{\theta^2} E(1-Y) \right), \quad Y = 1_{X=0} \\ &= n \left(\frac{4(1-2\theta)}{(1-2\theta)^2} + \frac{2\theta}{\theta^2} \right) \\ &= \frac{2n}{\theta(1-2\theta)}. \end{aligned}$$

On remarque que $Var(W_n) = 1/I_n(\theta)$, et donc l'estimateur est bien efficace.

3. Notons que $\sum_{i=1}^n X_n^2$ est égale au nombre des variables qui prennent les valeurs -1 ou 1. Donc $\sum_{i=1}^n X_n^2 = n - R$. Par conséquent

$$T_n = \frac{1}{2n} \sum_{i=1}^n X_i^2 = \frac{1}{2} - \frac{1}{2} \frac{R}{n} = W_n.$$

Exo 10

1. Nous devons avoir $a + b = 1$ car $\int_{\mathbb{R}} f(x; a, b) dx = a + b$.

2. $E(X) = a \int_{-1}^0 x dx + (1-a) \int_0^1 x dx = -a/2 + (1-a)/2 = 1/2 - a$.
 $E(X^2) = a \int_{-1}^0 x^2 dx + (1-a) \int_0^1 x^2 dx = a/3 + (1-a)/3 = 1/3$. Donc
 $Var(X) = 1/12 + a - a^2$.

3. L'estimateur de a par la méthode des moments est donné par $T_n = 1/2 - \bar{X}_n$. La loi forte des grands nombres (X est bien intégrable) implique que T_n converge fortement (ou presque sûrement). L'estimateur est sans biais donc son risque quadratique est égal à sa variance. Or $Var(T_n) = Var(\bar{X}_n) = Var(X)/n \rightarrow 0$. Donc W_n est convergent en moyenne quadratique.

4. On peut réécrire la densité de X sous la forme

$$f(x; a, b) = a^{1_{x \in [-1, 0]}} (1-a)^{1-1_{x \in [-1, 0]}} \cdot 1_{x \in [-1, 1]}$$

et donc la log-vraisemblance s'écrit (presque sûrement)

$$l_n(a) = \log(a) \sum_{j=1}^n 1_{X_j \in [-1, 0]} + \log(1-a) \left(n - \sum_{i=1}^n 1_{X_j \in [-1, 0]} \right)$$

avec

$$\frac{\partial l_n}{\partial a} = \frac{R_n}{a} - \frac{n - R_n}{1-a} = 0 \iff a = \frac{R_n}{n}$$

où $R_n = \sum_{i=1}^n 1_{X_i \in [-1, 0]}$, et

$$\frac{\partial^2 l_n}{\partial a^2} = -\frac{R_n}{a^2} - \frac{n - R_n}{(1-a)^2} < 0.$$

On en déduit que $W_n = R_n/n$ est l'EMV de a .

TD5

Exo 4. Pour les deux questions de l'exercice, notons d'abord qu'il s'agit d'un test d'hypothèses simples de la forme

$$H_0 : \theta = \theta_0 \text{ contre } H_1 : \theta = \theta_1$$

avec $\theta_0 \neq \theta_1$. D'après le Lemme de Neyman-Pearson, la région critique (zone de rejet) du test UPP de niveau α est donnée par

$$\frac{L_n(\theta_1)}{L_n(\theta_0)} > c_\alpha$$

où

$$P_{\theta_0} \left(\frac{L_n(\theta_1)}{L_n(\theta_0)} > c_\alpha \right) = \alpha.$$

La décision est

$$\begin{aligned} & \text{Rejeter } H_0 \text{ si } \frac{L_n(\theta_1)}{L_n(\theta_0)} > c_\alpha \\ & \text{Accepter } H_0 \text{ sinon.} \end{aligned}$$

Revenons maintenant à l'exercice.

1. On se fixe $\alpha = 5\%$. En se basant sur les données collectées (un échantillon de taille $n = 100$ dont les v.a. sont i.i.d. selon $\mathcal{N}(\mu, 49)$) qui ont donné lieu à la réalisation d'un bruit moyen $\bar{x}_n = 79.1$, on veut tester

$$H_0 : \mu = 80 \text{ contre } H_1 : \mu = 78.$$

Posons $\sigma_0^2 = 49$. On a que

$$\begin{aligned} L_n(\mu) &= \prod_{j=1}^n \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left(-\frac{(X_j - \mu)^2}{2\sigma_0^2}\right) \\ &= \frac{1}{(\sqrt{2\pi}\sigma_0)^n} \exp\left(-\frac{\sum_{j=1}^n (X_j - \mu)^2}{2\sigma_0^2}\right) \end{aligned}$$

et donc

$$\begin{aligned} \frac{L_n(\mu_1)}{L_n(\mu_0)} &= \exp\left(-\frac{\sum_{j=1}^n (X_j - \mu_1)^2 - \sum_{j=1}^n (X_j - \mu_0)^2}{2\sigma_0^2}\right) \\ &= \exp\left(-\frac{\sum_{j=1}^n [(X_j - \mu_1)^2 - (X_j - \mu_0)^2]}{2\sigma_0^2}\right) \\ &= \exp\left((\mu_1 - \mu_0) \frac{\sum_{j=1}^n (2X_j - (\mu_0 + \mu_1))}{2\sigma_0^2}\right) \\ &= \exp\left(\frac{n(\mu_1 - \mu_0)}{\sigma_0^2} \left(\bar{X}_n - \frac{\mu_0 + \mu_1}{2}\right)\right). \end{aligned}$$

D'où

$$\begin{aligned} \frac{L_n(\mu_1)}{L_n(\mu_0)} > c_\alpha &\iff \bar{X}_n - \frac{\mu_0 + \mu_1}{2} < d_\alpha, \text{ car } \mu_1 < \mu_0 \\ &\iff \bar{X}_n < k_\alpha. \end{aligned}$$

Sous l'hypothèse H_0 , X_1, \dots, X_n sont i.i.d. $\sim \mathcal{N}(\mu_0, \sigma_0^2)$, ce qui implique que $\bar{X}_n \sim \mathcal{N}(\mu_0, \sigma_0^2/n)$ et donc

$$\begin{aligned} P_{\mu_0} \left(\frac{L_n(\mu_1)}{L_n(\mu_0)} > c_\alpha \right) &= P_{\mu_0}(\bar{X}_n < k_\alpha) \\ &= P_{\mu_0} \left(\frac{\sqrt{n}(\bar{X}_n - \mu_0)}{\sigma_0} < \frac{\sqrt{n}(k_\alpha - \mu_0)}{\sigma_0} \right) \\ &= P(Z < k'_\alpha) \text{ où } Z \sim \mathcal{N}(0, 1) \end{aligned}$$

où

$$k'_\alpha = \frac{\sqrt{n}(k_\alpha - \mu_0)}{\sigma_0}.$$

Pour que le test soit de niveau α , il faut et suffit que

$$P(Z < k'_\alpha) = \alpha$$

c-à-d, que k'_α soit le quantile d'ordre α d'une normale centrée réduite ; $k'_\alpha = z_\alpha$. Le test est donné par

$$\text{Rejeter } H_0 \text{ si } \bar{X}_n < k_\alpha = \mu_0 + \frac{z_\alpha \sigma_0}{\sqrt{n}}$$

Accepter H_0 sinon.

Application numérique : $\alpha = 5\%$, $z_\alpha = z_{0.05} \approx -1.65$, $n = 100$, $\mu_0 = 80$, $\sigma_0 = 7$. On a

$$\bar{x}_n = 79.1 \text{ et } \mu_0 + \frac{z_\alpha \sigma_0}{\sqrt{n}} = 80 - \frac{1.65 \times 7}{10} \approx 78.845$$

donc, on accepte H_0 (l'hypothèse que l'intensité du bruit atteint la limite tolérée 80 dbs) car $79.1 \geq 78.845$.

2. Pour les hypothèses (inversées)

$$H_0 : \mu = 78 \text{ contre } H_1 : \mu = 80$$

le même calcul que précédemment (la seule différence c'est le signe de $\mu_1 - \mu_0$) donne le test (refaites-le pour vous en convaincre !)

$$\text{Rejeter } H_0 \text{ si } \bar{X}_n > \mu_0 + \frac{z_{1-\alpha} \sigma_0}{\sqrt{n}}$$

Accepter H_0 sinon.

Application numérique : $\alpha = 5\%$, $z_{1-\alpha} = z_{0.95} \approx 1.65$, $n = 100$, $\mu_0 = 78$, $\sigma_0 = 7$. On a

$$\bar{x}_n = 79.1 \text{ et } \mu_0 + \frac{z_{1-\alpha} \sigma_0}{\sqrt{n}} = 78 + \frac{1.65 \times 7}{10} \approx 79.155$$

donc, on accepte H_0 (l'hypothèse que l'intensité du bruit est de seulement 78 dbs) car $79.1 \leq 79.155$.

3. On n'arrive pas à rejeter ni l'hypothèse que l'intensité du bruit est de 78 dbs ni qu'elle atteint le seuil toléré 80 dbs. Cela suggère qu'il n'y a pas assez de données pour arriver à discriminer entre les deux hypothèses.

Exo 5. Il s'agit ici d'un test d'hypothèses composites de la forme

$$H_0 : \theta \leq \theta_0 \text{ contre } H_1 : \theta > \theta_0.$$

Comment construire un test de seuil α ? Rappelons que si T est une statistique exhaustive pour θ telle que T admet une densité $g(\cdot, \theta)$ de rapport de vraisemblance monotone croissant c-à-d, $g(\cdot, \theta)$ vérifie la condition

$$t \mapsto \frac{g(t, \theta_1)}{g(t, \theta_2)}$$

est monotone croissante pour tous $\theta_2 > \theta_1$ alors :

- le test (UPP) de niveau α pour les hypothèses

$$H_0 : \theta = \theta_0 \text{ contre } H_1 : \theta > \theta_0.$$

ou

$$H_0 : \theta \leq \theta_0 \text{ contre } H_1 : \theta > \theta_0.$$

est donné par

Rejeter H_0 si $T > t_\alpha$
Accepter H_0 sinon

où t_α est déterminé par $P_{\theta_0}(T > t_\alpha) = \alpha$,

- le test (UPP) de seuil α pour les hypothèses

$$H_0 : \theta = \theta_0 \text{ contre } H_1 : \theta < \theta_0.$$

ou

$$H_0 : \theta \geq \theta_0 \text{ contre } H_1 : \theta < \theta_0.$$

est donné par

Rejeter H_0 si $T < t'_\alpha$
Accepter H_0 sinon

où t'_α est déterminé par $P_{\theta_0}(T < t'_\alpha) = \alpha$.

Les inégalités déterminant les régions critiques doivent être inversées si T a admet une densité de rapport de vraisemblance monotone décroissant.

Revenons à l'exercice.

1. On dispose d'un échantillon X_1, \dots, X_n ($n = 10$) i.i.d. $\sim \mathcal{N}(\mu, \sigma^2)$ où σ est connu (=5). En se basant sur cet échantillon, on veut tester

$$H_0 : \mu \geq 710 \text{ contre } H_1 : \mu < 710.$$

On sait que $T = \bar{X}_n$ est une statistique exhaustive pour le paramètre μ . En plus, $\bar{X}_n \sim \mathcal{N}(\mu, \sigma^2/n)$, donc de densité

$$g(t, \mu) = \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} \exp\left(-\frac{n(t - \mu)^2}{2\sigma^2}\right).$$

Soit $\mu_2 > \mu_1$. On a que

$$\begin{aligned} \frac{g(t, \mu_2)}{g(t, \mu_1)} &= \exp\left(-n \frac{(t - \mu_2)^2 - (t - \mu_1)^2}{2\sigma^2}\right) \\ &= \exp\left(\frac{n(\mu_2 - \mu_1)}{2\sigma^2} (2t - (\mu_1 + \mu_2))\right) \end{aligned}$$

qui est une fonction croissante de t ($\mu_2 - \mu_1 > 0$). Le test UPP de niveau α est donné par

Rejeter H_0 si $\bar{X}_n < t'_\alpha$
Accepter H_0 sinon

avec $P_{\mu_0}(\bar{X}_n < t'_\alpha) = \alpha$ avec $\mu_0 = 710$ et $\alpha = 5\%$. Or, lorsque $\mu = \mu_0$, $\bar{X}_n \sim \mathcal{N}(\mu_0, \sigma^2/n)$, d'où

$$\begin{aligned} P_{\mu_0}(\bar{X}_n < t'_\alpha) &= P_{\mu_0}\left(\frac{\sqrt{n}(\bar{X}_n - \mu_0)}{\sigma} < \frac{\sqrt{n}(t'_\alpha - \mu_0)}{\sigma}\right) \\ &= P\left(Z < \frac{\sqrt{n}(t'_\alpha - \mu_0)}{\sigma}\right), \text{ où } Z \sim \mathcal{N}(0, 1) \\ &= \alpha \end{aligned}$$

impliquant que

$$\frac{\sqrt{n}(t'_\alpha - \mu_0)}{\sigma} = z_\alpha, \text{ le quantile d'ordre } \alpha \text{ d'une normale centrée réduite.}$$

Le test de seuil α est alors donné par

Rejeter H_0 si $\bar{X}_n < \mu_0 + \frac{z_\alpha \sigma}{\sqrt{n}}$
Accepter H_0 sinon

Application numérique. $\mu_0 = 710$, $n = 10$, $\sigma = 5$, $z_{0.05} \approx -1.65$. On a

$$\bar{x}_n = 707 \text{ et } \mu_0 + \frac{z_\alpha \sigma}{\sqrt{n}} \approx 707.4.$$

On rejette l'hypothèse H_0 , c-à-d, l'hypothèse selon laquelle le poids d'un paquet de lessive serait supérieur à 710 g.

2. On sait que $\bar{X}_n \sim \mathcal{N}(\mu, \sigma^2/n)$, ce qui est équivalent à dire que $\sqrt{n}(\bar{X}_n - \mu)/\sigma \sim \mathcal{N}(0, 1)$. Il s'en suit que

$$P\left(\frac{\sqrt{n}|\bar{X}_n - \mu|}{\sigma} \leq z_{1-\alpha/2}\right) = 1 - \alpha$$

c-à-d

$$P\left(\bar{X}_n - \frac{z_{1-\alpha/2}\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}_n + \frac{z_{1-\alpha/2}\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

et

$$\left[\bar{X}_n - \frac{z_{1-\alpha/2}\sigma}{\sqrt{n}}, \bar{X}_n + \frac{z_{1-\alpha/2}\sigma}{\sqrt{n}}\right]$$

est un intervalle de confiance bilatéral symétrique de niveau $1 - \alpha$ pour μ .
Pour $\alpha = 10\%$, $\bar{x}_n = 707$, $n = 10$, $\sigma = 5$, $z_{1-\alpha/2} = z_{0.95} \approx 1.65$,

$$[707.40, 712.60]$$

est la réalisation d'un intervalle de confiance bilatéral symétrique de niveau 90% pour la vraie moyenne μ .

Exo 11. On nous suggère de construire un test basé sur le rapport de vraisemblance pour tester les hypothèses

$$H_0 : \mu = \mu_0 \quad \text{contre} \quad H_1 : \mu \neq \mu_0$$

(hypothèse simple contre composite).

1. Vraiment facile !

2. Le test du rapport de vraisemblance est donné par

$$\begin{aligned} \text{Rejeter } H_0 \text{ si } & \frac{\sup_{\mu \in \mathbb{R}, \sigma > 0} L_n(\mu, \sigma^2)}{\sup_{\sigma > 0} L_n(\mu_0, \sigma^2)} > c_\alpha \\ \text{Accepter } H_0 \text{ sinon} & \end{aligned}$$

où c_α est déterminé par

$$P\left(\frac{\sup_{\{\mu \leq \mu_0, \sigma > 0\}} L_n(\mu, \sigma)}{\sup_{\{\sigma > 0\}} L_n(\mu_0, \sigma)} > c_\alpha\right) = \alpha.$$

Or,

$$L_n(\mu_0, \sigma) = \frac{1}{(\sqrt{2\pi}\sigma)^n} \exp\left(-\frac{\sum_{j=1}^n (X_j - \mu_0)^2}{2\sigma^2}\right)$$

qu'on peut maximiser par rapport σ en passant à la log-vraisemblance

$$l_n(\mu_0, \sigma) = cte - n \log(\sigma) - \frac{\sum_{j=1}^n (X_j - \mu_0)^2}{2\sigma^2}$$

dont la dérivée est donnée par

$$\frac{\partial l_n(\mu_0, \sigma)}{\partial \sigma} = -\frac{n}{\sigma} + \frac{\sum_{j=1}^n (X_j - \mu_0)^2}{\sigma^3}$$

qui s'annule au point $\hat{\sigma}_0 > 0$ vérifiant

$$\hat{\sigma}_0^2 = \frac{1}{n} \sum_{j=1}^n (X_j - \mu_0)^2.$$

Cette solution est bien le maximum global puisque

$$\begin{aligned} \frac{\partial^2 l_n(\mu_0, \sigma)}{\partial \sigma^2} \Big|_{\sigma=\hat{\sigma}_0} &= \frac{n}{\hat{\sigma}_0^2} - 3 \frac{\sum_{j=1}^n (X_j - \mu_0)^2}{\hat{\sigma}_0^4} \\ &= \frac{1}{\hat{\sigma}_0^4} (n\hat{\sigma}_0^2 - 3n\hat{\sigma}_0^2) = -\frac{2n}{\hat{\sigma}_0^2} < 0 \end{aligned}$$

(cela montre que $\hat{\sigma}_n$ est un maximum local. Pour conclure qu'il est global, notons que si la log-vraisemblance admettait un autre maximum, alors sa dérivée s'annulerait en un autre point puisque la log-vraisemblance devrait "redescendre avant de remonter" et atteindre cet autre maximum, ce qui forcera la dérivée à s'annuler en un minimum local). Or, on a bien trouvé que cette dérivée n'admettait qu'un seul zéro, $\hat{\sigma}_n$).

On en déduit que

$$\begin{aligned} \sup_{\sigma>0} L_n(\mu_0, \sigma^2) &= \frac{1}{(\sqrt{2\pi}\hat{\sigma}_0)^n} \exp\left(-\frac{\sum_{j=1}^n (X_j - \mu_0)^2}{2\hat{\sigma}_0^2}\right) \\ &= \frac{1}{(\sqrt{2\pi}\hat{\sigma}_0)^n} \exp\left(-\frac{n}{2}\right). \end{aligned}$$

Il faut maintenant déterminer le terme du numérateur. On a

$$L_n(\mu, \sigma) = \frac{1}{(\sqrt{2\pi}\sigma)^n} \exp\left(-\frac{\sum_{j=1}^n (X_j - \mu)^2}{2\sigma^2}\right)$$

et

$$l_n(\mu, \sigma) = cte - n \log(\sigma) - \frac{\sum_{j=1}^n (X_j - \mu)^2}{2\sigma^2}$$

qu'on veut maximiser par rapport à $\mu \in]-\infty, \infty[$ et $\sigma > 0$. Les équations

$$\begin{aligned} \frac{\partial l_n(\mu, \sigma)}{\partial \mu} &= \frac{\sum_{j=1}^n (X_j - \mu)}{\sigma^2} = 0 \\ \frac{\partial l_n(\mu, \sigma)}{\partial \sigma} &= -\frac{n}{\sigma} + \frac{\sum_{j=1}^n (X_j - \mu)^2}{\sigma^3} = 0 \end{aligned}$$

sont équivalentes à

$$\mu = \bar{X}_n, \quad \sigma^2 = \frac{1}{n} \sum_{j=1}^n (X_j - \bar{X}_n)^2.$$

La matrice Hermitienne au point $(\bar{X}_n, 1/n \sum_{j=1}^n (X_j - \bar{X}_n)^2) = (\bar{X}_n, \hat{\sigma}^2)$ est donnée par

$$\begin{pmatrix} -\frac{n}{\hat{\sigma}^2} & 0 \\ 0 & -\frac{2n}{\hat{\sigma}^2} \end{pmatrix}$$

qui est définie négative. Donc, il s'agit (a priori) d'un maximum local. Or, si la log-vraisemblance admettait un autre maximum, son gradient aurait admis un autre zéro (cf. le raisonnement précédent). Il s'en suit que

$$\begin{aligned} \sup_{\{\mu \in \mathbb{R}, \sigma > 0\}} L_n(\mu, \sigma^2) &= \frac{1}{(\sqrt{2\pi}\hat{\sigma})^n} \exp\left(-\frac{\sum_{j=1}^n (X_j - \bar{X}_n)^2}{2\hat{\sigma}^2}\right) \\ &= \frac{1}{(\sqrt{2\pi}\hat{\sigma})^n} \exp\left(-\frac{n}{2}\right). \end{aligned}$$

Le rapport de vraisemblance est donné par

$$\begin{aligned} \frac{\sup_{\mu \in \mathbb{R}, \sigma > 0} L_n(\mu, \sigma^2)}{\sup_{\sigma > 0} L_n(\mu_0, \sigma^2)} &= \left(\frac{\hat{\sigma}_0}{\hat{\sigma}_n}\right)^n \\ &= \left(\frac{\hat{\sigma}_0^2}{\hat{\sigma}_n^2}\right)^{n/2} \end{aligned}$$

et donc dire que

$$\frac{\sup_{\mu \in \mathbb{R}, \sigma > 0} L_n(\mu, \sigma^2)}{\sup_{\sigma > 0} L_n(\mu_0, \sigma^2)}$$

est grand est équivalent à dire que

$$\frac{\hat{\sigma}_0^2}{\hat{\sigma}_n^2}$$

est grand, c-à-d

$$\frac{\sum_{j=1}^n (X_j - \mu_0)^2}{\sum_{j=1}^n (X_j - \bar{X}_n)^2}$$

est grand. Or,

$$\begin{aligned} \sum_{j=1}^n (X_j - \mu_0)^2 &= \sum_{j=1}^n (X_j - \bar{X}_n + \bar{X}_n - \mu_0)^2 \\ &= \sum_{j=1}^n (X_j - \bar{X}_n)^2 + n(\bar{X}_n - \mu_0)^2 \end{aligned}$$

et donc

$$\frac{\sum_{j=1}^n (X_j - \mu_0)^2}{\sum_{j=1}^n (X_j - \bar{X}_n)^2}$$

est grand est équivalent à

$$1 + \frac{n(\bar{X}_n - \mu_0)^2}{\sum_{j=1}^n (X_j - \bar{X}_n)^2}$$

est grand, ou encore à

$$\frac{n(\bar{X}_n - \mu_0)^2}{\sum_{j=1}^n (X_j - \bar{X}_n)^2}$$

est grand ou encore

$$\frac{\sqrt{n}|\bar{X}_n - \mu_0|}{\hat{\sigma}_n}$$

est grand, ou encore

$$\frac{\sqrt{n}|\bar{X}_n - \mu_0|}{\sqrt{\frac{n-1}{n}\hat{\sigma}_{n-1}}}$$

est grand, (Ici $\hat{\sigma}_{n-1}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$). Ce qui est toujours équivalent à

$$\frac{\sqrt{n}|\bar{X}_n - \mu_0|}{\hat{\sigma}_{n-1}}$$

est grand. Ce qu'on a démontré ci-dessus c'est que la zone de rejet est donnée par

$$\frac{\sqrt{n}|\bar{X}_n - \mu_0|}{\hat{\sigma}_{n-1}} > k_\alpha.$$

Or, en se rappelant que si X_1, \dots, X_n sont i.i.d. $\sim \mathcal{N}(\mu_0, \sigma^2)$ alors

$$\frac{\sqrt{n}(\bar{X}_n - \mu_0)}{\hat{\sigma}_{n-1}} \sim \mathcal{T}_{n-1}$$

(une loi de Student de $n - 1$ degrés de liberté), on en déduit que

$$k_\alpha = t_{n-1, 1-\alpha/2}$$

le quantile de \mathcal{T}_{n-1} d'ordre $1 - \alpha/2$.

En résumé, le test du rapport de vraisemblance pour les hypothèses

$$H_0 : \mu = \mu_0 \quad \text{contre} \quad H_1 : \mu \neq \mu_0$$

est donné par

$$\text{Rejeter } H_0 \text{ si } \frac{\sqrt{n}|\bar{X}_n - \mu_0|}{\hat{\sigma}_{n-1}} > t_{n-1, 1-\alpha/2}$$

Accepter H_0 sinon.

Exo 9 (le dernier exo du TD5). Ici, l'espace des paramètres (sans contraintes) est

$$\Theta = \left\{ (p_1, \dots, p_4) \in \mathbb{R}^4 : 0 < p_j < 1; \sum_{j=1}^4 p_j < 1 \right\}.$$

En posant $T_j = \sum_{i=1}^n 1_{\{X_i=j\}}$, la vraisemblance s'écrit

$$L_n(p_1, \dots, p_4) = \prod_{j=1}^5 p_j^{T_j}, \text{ où } p_5 = 1 - p_1 - p_2 - p_3 - p_4.$$

La log-vraisemblance est donnée alors par

$$l_n(p_1, \dots, p_4) = \sum_{j=1}^5 T_j \log(p_j) = \sum_{j=1}^4 T_j \log(p_j) + T_5 \log\left(1 - \sum_{j=1}^4 p_j\right).$$

D'où

$$\frac{\partial l_n}{\partial p_j}(p_1, \dots, p_4) = \frac{T_j}{p_j} - \frac{T_5}{1 - \sum_{j=1}^4 p_j} = \frac{T_j}{p_j} - \frac{T_5}{p_5}$$

On en déduit donc que, $\nabla l_n(p_1, \dots, p_4) = 0$ est équivalent à

$$p_j = \frac{p_5}{T_5} T_j \quad \forall j = 1, \dots, 4$$

On déduit des relations $\sum_{j=1}^5 p_j = 1$ et $\sum_{j=1}^5 T_j = n$ que $\nabla l_n(p_1, \dots, p_4) = 0$ admet la solution unique

$$\hat{p}_j = \frac{T_j}{n}, \quad j = 1, \dots, 4.$$

La matrice Hessienne, $H = H(p_1, p_2, p_3, p_4)$ de la log-vraisemblance est donnée

$$\begin{aligned} & \begin{pmatrix} -\frac{T_1}{p_1^2} - \frac{T_5}{(1-\sum_{j=1}^4 p_j)^2} & -\frac{T_5}{(1-\sum_{j=1}^4 p_j)^2} & -\frac{T_5}{(1-\sum_{j=1}^4 p_j)^2} & -\frac{T_5}{(1-\sum_{j=1}^4 p_j)^2} \\ -\frac{T_5}{(1-\sum_{j=1}^4 p_j)^2} & -\frac{T_2}{p_2^2} - \frac{T_5}{(1-\sum_{j=1}^4 p_j)^2} & -\frac{T_5}{(1-\sum_{j=1}^4 p_j)^2} & -\frac{T_5}{(1-\sum_{j=1}^4 p_j)^2} \\ -\frac{T_5}{(1-\sum_{j=1}^4 p_j)^2} & -\frac{T_5}{(1-\sum_{j=1}^4 p_j)^2} & -\frac{T_3}{p_3^2} - \frac{T_5}{(1-\sum_{j=1}^4 p_j)^2} & -\frac{T_5}{(1-\sum_{j=1}^4 p_j)^2} \\ -\frac{T_5}{(1-\sum_{j=1}^4 p_j)^2} & -\frac{T_5}{(1-\sum_{j=1}^4 p_j)^2} & -\frac{T_5}{(1-\sum_{j=1}^4 p_j)^2} & -\frac{T_4}{p_4^2} - \frac{T_5}{(1-\sum_{j=1}^4 p_j)^2} \end{pmatrix} \\ & = \begin{pmatrix} -\frac{T_1}{p_1^2} - \frac{T_5}{p_5^2} & -\frac{T_5}{p_5^2} & -\frac{T_5}{p_5^2} & -\frac{T_5}{p_5^2} \\ -\frac{T_5}{p_5^2} & -\frac{T_2}{p_2^2} - \frac{T_5}{p_5^2} & -\frac{T_5}{p_5^2} & -\frac{T_5}{p_5^2} \\ -\frac{T_5}{p_5^2} & -\frac{T_5}{p_5^2} & -\frac{T_3}{p_3^2} - \frac{T_5}{p_5^2} & -\frac{T_5}{p_5^2} \\ -\frac{T_5}{p_5^2} & -\frac{T_5}{p_5^2} & -\frac{T_5}{p_5^2} & -\frac{T_4}{p_4^2} - \frac{T_5}{p_5^2} \end{pmatrix} \end{aligned}$$

Pour montrer que la solution obtenue est le maximum global, il suffit de montrer que $-H$ est définie positive, c-à-d, le produit scalaire

$$(u_1, u_2, u_3, u_4)(-H)(u_1, u_2, u_3, u_4)^T > 0$$

pour tout $(u_1, u_2, u_3, u_4) \in \mathbb{R}^4 \setminus \{0\}$. Notons $\underline{u} = (u_1, \dots, u_4)$ et posons $\lambda_j = T_j/p_j^2$ pour $j = 1, \dots, 4$ et $r = T_5/p_5^2$ où $p_5 = 1 - \sum_{j=1}^4 p_j$. On a que

$$-H = \begin{pmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & \lambda_3 & 0 \\ 0 & 0 & 0 & \lambda_4 \end{pmatrix} + r \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}.$$

Or, il existe P une matrice orthogonale telle que

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix} = P^T \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix} P.$$

Il vient que

$$\underline{u}(-H)\underline{u}^T = \sum_{j=1}^4 \lambda_j u_j^2 + 4r([P\underline{u}^T]_4)^2, \quad \text{où } [P\underline{u}^T]_4 \text{ est la 4ème composante du vecteur } P\underline{u}^T$$

et

$$\underline{u}(-H)\underline{u}^T > 0, \quad \forall \underline{u} \in \mathbb{R}^4 \setminus \{0\}.$$

On en déduit que $(-H)$ est définie positive et que l'EMV (sans contraintes) de $\theta = (p_1, \dots, p_4)$ est

$$\hat{\theta}_n = (\hat{p}_1, \dots, \hat{p}_4) = \left(\frac{T_1}{n}, \dots, \frac{T_4}{n} \right).$$

Maintenant sous la contrainte que $p_1 = p_2 = p_3$ et $p_4 = p_5$, on a clairement $p_4 = p_5 = (1 - 3p_1)/2$. Pour que le support de X soit égal $\{1, 2, \dots, 5\}$ il faut et suffit que $p = p_1 \in]0, 1/3[$, et le nouvel espace des paramètres se réduit donc à

$$\Theta_0 = \{p : p \in]0, 1/3[\} =]0, 1/3[.$$

Sous l'hypothèse H_0 , la log-vraisemblance (qui ne dépend que de p) est égal à

$$\begin{aligned} l_n(p) &= (T_1 + T_2 + T_3) \log(p) + (T_4 + T_5) \log\left(\frac{1-3p}{2}\right) \\ &= T^* \log(p) + (n - T^*) \log\left(\frac{1-3p}{2}\right) \end{aligned}$$

et donc

$$\frac{\partial l_n(p)}{\partial p} = \frac{T^*}{p} - \frac{3(n - T^*)}{1 - 3p}$$

et donc

$$\frac{\partial l_n(p)}{\partial p} = 0 \iff p = \frac{T^*}{3n}.$$

D'autre part,

$$\frac{\partial^2 l_n(p)}{\partial p^2} = -\frac{T^*}{p^2} - \frac{9(n - T^*)}{(1 - 3p)^2} < 0.$$

On en déduit que l'EMV de $p \in \Theta_0$ est donné par

$$\hat{p}_n = \frac{T^*}{3n}.$$

Le test du rapport de vraisemblance de niveau asymptotique α a pour zone de rejet

$$R_n = \{\xi_n^{RV} > \chi_{m,1-\alpha}^2\}$$

c-à-d,

$$\lim_{n \rightarrow \infty} \sup_{\theta \in \Theta_0} P_\theta(R_n) = \alpha$$

avec

$$\begin{aligned} \xi_n^{RV} &= 2(l_n(\hat{p}_1, \dots, \hat{p}_4) - l_n(\hat{p}_n)) \\ &= 2 \left(\sum_{j=1}^5 T_j \log \left(\frac{T_j}{n} \right) - T^* \log \left(\frac{T^*}{3n} \right) - (n - T^*) \log \left(\frac{n - T^*}{2n} \right) \right) \\ &= 2 \sum_{j=1}^3 T_j \log \left(\frac{3T_j}{T^*} \right) + 2 \sum_{j=4}^5 T_j \log \left(\frac{2T_j}{n - T^*} \right) \end{aligned}$$

et $\chi_{m,1-\alpha}^2$ est le quantile d'ordre $1 - \alpha$ de la loi χ^2 à m degrés de liberté ici $m = \dim(\Theta) - \dim(\Theta_0) = 4 - 1 = 3$.

4. (Test de Wald) Dans cette partie, comme on a des produits matriciels à calculer, on considère un paramètre ainsi qu'un estimateur quelconque de celui-ci comme un vecteur de \mathbb{R}^4 (donc d'une colonne et de 4 lignes). Observons que l'EMV

$$\hat{\theta} = \begin{bmatrix} \frac{T_1}{n} \\ \frac{T_2}{n} \\ \frac{T_3}{n} \\ \frac{T_4}{n} \end{bmatrix} = \frac{1}{n} \sum_{i=1}^n \begin{bmatrix} 1_{X_i=1} \\ 1_{X_i=2} \\ 1_{X_i=3} \\ 1_{X_i=4} \end{bmatrix}$$

donc pour $\theta = (p_1, p_2, p_3, p_4)^T$ on a bien $E(\hat{\theta}) = (p_1, p_2, p_3, p_4)^T$ (sans biais)
et

$$\Sigma(\theta) = \text{Var} \begin{bmatrix} 1_{X_1=1} \\ 1_{X_1=2} \\ 1_{X_1=3} \\ 1_{X_1=4} \end{bmatrix} = \begin{bmatrix} p_1(1-p_1) & -p_1p_2 & -p_1p_3 & -p_1p_4 \\ -p_1p_2 & p_2(1-p_2) & -p_2p_3 & -p_2p_4 \\ -p_3p_1 & -p_3p_2 & p_3(1-p_3) & -p_3p_4 \\ -p_4p_1 & -p_4p_2 & -p_4p_3 & p_4(1-p_4) \end{bmatrix}$$

par conséquent

$$\text{Var}(\hat{\theta}) = \frac{\Sigma(\theta)}{n}.$$

L'hypothèse H_0 peut être écrite sous la forme $\theta : g(\theta) = 0$ en considérant l'application $g : \mathbb{R}^4 \mapsto \mathbb{R}^3$, $g(p_1, p_2, p_3, p_4) = (p_1 - p_2, p_1 - p_3, 2p_4 + 3p_1 - 1)^T$.
Le noyau de cette application affine est de dimension 1 donc son rang est de dimension 3 et sa matrice jacobienne est

$$J_g = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 3 & 0 & 0 & 2 \end{bmatrix}$$

Le test de Wald de niveau asymptotique α est le test dont la zone de rejet est donnée par $R_n = \{\xi_n^W > \chi_{3,1-\alpha}^2\}$ ($k = 3$ étant le rang de l'application g)
où

$$\xi_n^W = g(\hat{\theta})^T \cdot \left(J_g(\hat{\theta}) \cdot I_n^{-1}(\hat{\theta}) \cdot J_g(\hat{\theta})^T \right)^{-1} \cdot g(\hat{\theta})$$

où $I_n^{-1}(\theta) = \frac{1}{n}\Sigma(\theta)$, et donc

$$I_n^{-1}(\hat{\theta}) = \frac{1}{n} \begin{bmatrix} \hat{p}_1(1-\hat{p}_1) & -\hat{p}_1\hat{p}_2 & -\hat{p}_1\hat{p}_3 & -\hat{p}_1\hat{p}_4 \\ -\hat{p}_1\hat{p}_2 & \hat{p}_2(1-\hat{p}_2) & -\hat{p}_2\hat{p}_3 & -\hat{p}_2\hat{p}_4 \\ -\hat{p}_1\hat{p}_3 & -\hat{p}_2\hat{p}_3 & \hat{p}_3(1-\hat{p}_3) & -\hat{p}_3\hat{p}_4 \\ -\hat{p}_1\hat{p}_4 & -\hat{p}_2\hat{p}_4 & -\hat{p}_3\hat{p}_4 & \hat{p}_4(1-\hat{p}_4) \end{bmatrix}.$$

D'autre part

$$g(\hat{\theta}) = (\hat{p}_1 - \hat{p}_2, \hat{p}_1 - \hat{p}_3, 3\hat{p}_1 + 2\hat{p}_4 - 1)^T$$

et

$$J_g = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 3 & 0 & 0 & 2 \end{bmatrix}.$$

Pour le calcul matriciel

$$\left(J_g(\hat{\theta}) \cdot I_n^{-1}(\hat{\theta}) \cdot J_g(\hat{\theta})^T \right)^{-1}$$

on vous souhaite bon courage...