STT1000 – Examen Intra 2

(Automne 2021)

Les calculatrices sont autorisées. Les documents sont en revanche interdits, sauf une page d'aide mémoire. L'examen dure 2 heures. Toute sortie avant la fin est autorisée, mais sera définitive.
La feuille propose 7 exercices et un barême approximatif est donné à titre indicatif. Les réponses doivent être reportées sur le cahier joint. Si vous utilisez 2 cahiers, merci de le mentionner, en indiquant 1/2 et 2/2 respectivement.
N'hésitez pas à faire des dessins pour vous aider, mais ne considérez pas un dessin comme une preuve. Si vous utilisez un résultat du cours dans votre preuve, nommez-le aussi précisément que possible.

Si $X \sim \mathcal{N}(0,1)$, on a les probabilités suivantes

$$\begin{cases} \mathbb{P}[X \leq -3] \approx 0.1350\% & \mathbb{P}[X \leq 1] \approx 84.1345\% \\ \mathbb{P}[X \leq -2] \approx 2.2750\% & \mathbb{P}[X \leq 2] \approx 97.7250\% \\ \mathbb{P}[X \leq -1] \approx 15.865\% & \mathbb{P}[X \leq 3] \approx 99.8650\% \\ \mathbb{P}[X \leq 0] = 50.0000\% & \mathbb{P}[X \leq 4] \approx 99.9968\% \end{cases}$$

Si vous pensez que des hypothèses manquent pour répondre à la question, indiquez le dans le cahier. Si vous avez besoin d'introduire des objets mathématiques non définis dans l'énoncé, définissez les clairement.

Exercice 1 – ***(Surface d'un pré)*** [10 points]

Un agriculteur possède un champ carré, dont il veut estimer la superficie. Quand il mesure un côté de son champ, il sait que l'erreur de la mesure est une variable aléatoire d'espérance nulle et de variance σ^2, et que les erreurs sont indépendantes d'une mesure à l'autre. Il réalise une première mesure d'un des côtés et trouve x_1. Il en déduit la superficie $s_1 = x_1^2$. Par précaution, il effectue une seconde mesure de ce même côté, et trouve x_2. Il en déduit la superficie $s_2 = x_2^2$.

Perplexe, il abandonne ses mesures et réfléchit pour savoir quelle est la bonne façon de procéder. Il dispose de deux estimations de la surface de son champ s_1 et s_2, mais considère 3 autres estimations :

$$s_3 = \frac{s_1 + s_2}{2},\ s_4 = \left(\frac{x_1 + x_2}{2}\right)^2 \text{ et } s_5 = x_1 x_2.$$

Parmi ces 5 statistiques, lesquels sont des estimateurs sans biais de la surface ?

Les 5 estimations de la surface du champ sont

$$s_1 = x_1^2,\ s_2 = x_2^2,\ s_3 = \frac{s_1+s_2}{2},\ s_4 = \left(\frac{x_1+x_2}{2}\right)^2 \text{ et } s_5 = x_1x_2.$$

Soit ℓ la **vraie** longueur, inconnue, du champs. Comme le dit l'énoncé, $X_1 = \ell + U_1$ et $X_2 = \ell + U_2$, avec U_1 et U_2 deux variables aléatoires (correspondant aux erreurs de mesure). "*variable aléatoire d'espérance nulle et de variance σ^2, et que les erreurs sont indépendantes d'une mesure à l'autre*", donc $\mathbb{E}[U_1] = \mathbb{E}[U_2] = 0$, $\text{Var}[U_1] = \text{Var}[U_2] = \sigma^2$ et $\text{Cov}[U_1, U_2] = 0$. Comme les variables sont centrées, on a aussi $\mathbb{E}[U_1^2] = \text{Var}[U_1] + \mathbb{E}[U_1]^2 = \sigma^2$, pareil, $\mathbb{E}[U_2^2] = \sigma^2$, et $\mathbb{E}[U_1U_2] = \text{Cov}[U_1, U_2] + \mathbb{E}[U_1]\mathbb{E}[U_2] = 0$. Les 5 estimateurs de la surface du champ sont

$$S_1 = X_1^2,\ S_2 = X_2^2,\ S_3 = \frac{S_1+S_2}{2},\ S_4 = \left(\frac{X_1+X_2}{2}\right)^2 \text{ et } S_5 = X_1X_2.$$

L'estimateur S sera un estimateur sans biais de la surface du champ si $\mathbb{E}[S] = \ell^2$. Calculons les espérances de nos 5 estimateurs :

$$\mathbb{E}[S_1] = \mathbb{E}[X_1^2] = \mathbb{E}\Big[(\ell+U_1)^2\Big] = \mathbb{E}\Big[\ell^2 + 2\ell U_1 + U_1^2\Big] = \ell^2 + 2\ell\mathbb{E}[U_1] + \mathbb{E}[U_1]^2 = \ell^2 + 0 + \sigma^2 \neq \ell^2,$$

car $\sigma^2 > 0$, donc S_1 est un estimateur biaisé de la surface du champs (et le biais vaut σ^2). De manière parfaitement symmétrique, S_2 est un estimateur biaisé de la surface.

$$\mathbb{E}[S_3] = \mathbb{E}\left[\frac{S_1+S_2}{2}\right] = \frac{\mathbb{E}[S_1]+\mathbb{E}[S_2]}{2} = \frac{(\ell+\sigma^2)+(\ell+\sigma^2)}{2} = \ell + \sigma^2 \neq \ell,$$

autrement dit, S_3 est aussi un estimateur biaisé de la surface du champs. Autrement dit, faire la moyenne des surfaces ne permet pas de construire un estimateur sans biais (ou à réduire le biais).

$$\mathbb{E}[S_4] = \mathbb{E}\left[\left(\frac{X_1+X_2}{2}\right)^2\right] = \frac{1}{4}\mathbb{E}\Big[(\ell+\ell+U_1+U_2)^2\Big] = \frac{1}{4}\mathbb{E}\Big[(2\ell+U_1+U_2)^2\Big]$$

aussi

$$\mathbb{E}[S_4] = \frac{1}{4}\Big[(2\ell)^2 + 2\cdot 2\ell\mathbb{E}[U_1+U_2] + \mathbb{E}[(U_1+U_2)^2]$$

avec $\mathbb{E}[U_1+U_2] = 0$ et

$$\mathbb{E}[(U_1+U_2)^2] = \mathbb{E}[U_1^2] + 2\mathbb{E}[U_1U_2]\mathbb{E}[U_2^2] = \sigma^2 + 0 + \sigma^2 = 2\sigma^2$$

ce qui donne, pour résumer, i

$$\mathbb{E}[S_4] = \frac{1}{4}\Big[4\ell^2 + 2\sigma^2] = \ell + \frac{\sigma^2}{2} \neq \ell$$

autrement dit, S_4 est, là encore, un estimateur biaisé de la surface du champ. Cela dit, prendre le carré dela moyenne des longeur permet de réduire (un peu) le biais (le biais est divisé par 2).

$$\mathbb{E}[S_5] = \mathbb{E}\Big[X_1X_2\Big] = \mathbb{E}\Big[(\ell+U_1)(\ell+U_2)\Big] = \mathbb{E}\Big[\ell^2 + \ell U_1 + \ell U_2 + U_1U_2\Big] = \ell^2 + \ell\mathbb{E}\Big[U_1+U_2\Big] + \mathbb{E}\Big[U_1U_2\Big] = \ell^2$$

c'est à dire que S_5 est un estimateur sans biais de la surface du champ.

Exercice 2 – ***(Moyenne)*** [20 points]

Pour estimer l'espérance μ d'une population, on utilise souvent la moyenne empirique $\widehat{\mu} = \overline{x}$, à partir d'un échantillon $\{x_1, \cdots, x_n\}$, obtenu à partir de variables i.i.d. Soit $a \in [0,1]$. On propose l'estimateur $\widetilde{\mu} = a \cdot \overline{x}$.

1. Calculer biais$(\widehat{\mu})$, Var$(\widehat{\mu})$ et l'erreur quadratique moyenne de $\widehat{\mu}$.
2. Calculer biais$(\widetilde{\mu})$, Var$(\widetilde{\mu})$ et l'erreur quadratique moyenne de $\widetilde{\mu}$.
3. Tracer (sommairement) $a \mapsto$ biais$(\widetilde{\mu})^2$ et $a \mapsto$ Var$(\widetilde{\mu})$ sur $[0,1]$.
4. Existe-t-il a tel que l'erreur quadratique de $\widetilde{\mu}$ soit inférieure à l'erreur quadratique de $\widehat{\mu}$?

1) Pour $\widehat{\mu} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$,

$$\mathbb{E}(\widehat{\mu}) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X_i) = \frac{1}{n}\sum_{i=1}^{n}\mu = \frac{n}{n}\mu = \mu.$$

$$\text{Var}(\widehat{\mu}) = \text{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\text{Var}(X_i) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{n}{n^2}\sigma^2 = \frac{\sigma^2}{n},$$

où $\sigma^2 = \text{Var}[X_i]$, et

$$\text{EQM}(\widehat{\mu}) = \text{biais}(\widehat{\mu})^2 + \text{Var}(\widehat{\mu}) = 0^2 + \frac{\sigma^2}{n} = \frac{\sigma^2}{n}.$$

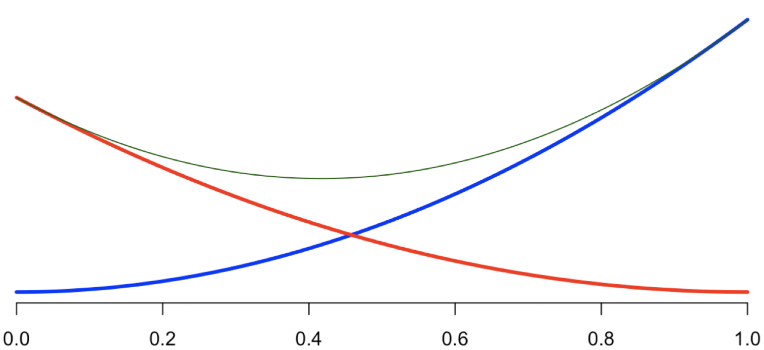
2) Pour $\widetilde{\mu} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$,

$$\mathbb{E}(\widetilde{\mu}) = \mathbb{E}(\alpha\widehat{\mu}) = \alpha\mu.$$

$$\text{Var}(\widetilde{\mu}) = \text{Var}(\alpha\widehat{\mu}) = \alpha^2\text{Var}(\widehat{\mu}) = \alpha^2\frac{\sigma^2}{n},$$

$$\text{EQM}(\widetilde{\mu}) = \text{biais}(\widehat{\mu})^2 + \text{Var}(\widehat{\mu}) = [(1-\alpha)\mu]^2 + \alpha^2\frac{\sigma^2}{n}$$

3) Le dessin ressemble à celui ci-dessous, avec $a \mapsto \text{biais}(\widetilde{\mu})^2 = [(1-\alpha)\mu]^2$ en rouge, et $a \mapsto \text{Var}(\widetilde{\mu}) = \alpha^2\dfrac{\sigma^2}{n}$ en bleu.



La courbe en vert est ici $\alpha \mapsto \text{EQM}(\widetilde{\mu})$, qui semble atteindre son minimum en une valeur entre 0 et 1 (c'est l'idée de la question suivante).
4) soit

$$h : \alpha \mapsto [(1-\alpha)\mu]^2 + \alpha^2\frac{\sigma^2}{n}$$

qui est un polynome de degré 2 en α (continuement différentiable). Le minimum de h est obtenu quand $h'(\alpha) = 0$,

$$h'(\alpha) = -2(1-\alpha)\mu^2 + 2\alpha\frac{\sigma^2}{n}$$

et $h'(\alpha^\star) = 0$ si et seulement si

$$2(1-\alpha^\star)\mu^2 = 2\alpha^\star\frac{\sigma^2}{n}$$

autrement dit

$$\alpha^\star = \frac{n\mu^2}{n\mu^2+\sigma^2}.$$

Exercice 3 – ***(Loi normale censurée)*** *[15 points]*

Considérons un échantillon $\{x_1, \cdots, x_n\}$ tiré suivant une loi $\mathcal{N}(\theta, 1)$. Malheureusement, seules des indicatrices, indiquant que les observations étaient positives, ont été observées, i.e. $\{y_1, \cdots, y_n\}$, avec $y_i = \mathbf{1}(x_i > 0)$.

1. Quelle est l'estimateur de la méthode des moments de θ (construit à partir des observations y_i) ?
2. Quelle est l'estimateur du maximum de vraisemblance de θ (construit à partir des observations y_i) ?

1) Comme $y_i = \mathbf{1}(x_i > 0)$, y_i prend deux valeurs, 0 ou 1, donc Y_i suit une loi de Bernoulli, $\mathcal{B}(p)$, avec $p = \mathbb{P}[Y_i = 1]$, soit $p = \mathbb{P}[X_i > 0]$. ou $p = \mathbb{P}[X_i - \theta > -\theta]$ avec $X_i - \theta \sim \mathcal{N}(0,1)$. Donc $p = 1 - \mathbb{P}[X_i - \theta \leq -\theta] = 1 - \Phi(-\theta)$ où Φ est (classiquement) la fonction de répartition de la loi $\mathcal{N}(0,1)$. Aussi, $Y_i \sim \mathcal{B}(1 - \Phi(-\theta))$ donc la méthode des moments donne l'égalité

$$\mathbb{E}[Y] = 1 - \Phi(-\theta) \text{ soit } \overline{y} = 1 - \Phi(-\widehat{\theta})$$

i.e.

$$\widehat{\theta} = -\Phi^{-1}\Big(1 - \overline{y}\Big) = \Phi^{-1}\Big(\overline{y}\Big)$$

par symmétrie de la loi $\mathcal{N}(0,1)$.
2.1) Pour l'estimateur du maximum de vraisemblance, la fonction de probabilité est

$$f_p(y) = p^y(1-p)^{1-y} \text{ avec } p = 1 - \Phi(-\theta)$$

dont le logarithme vaut

$$\log f_p(y) = y\log(p) + (1-y)\log(1-p),$$

de telle sorte que la log-vraisemblance s'écrit

$$\log\mathcal{L}(p) = \log(p)\sum_{i=1}^n y_i + \log(1-p)\sum_{i=1}^n(1-y_i).$$

Comme cette fonction est continue et dérivable sur $(0,1)$, on va chrecher le maximum de la (log)-vraisemblance grâce à la condition du premier ordre. La dérivée de la log-vraisemblance s'écrit

$$\frac{\partial\log\mathcal{L}(p)}{\partial p} = \frac{1}{p}\sum_{i=1}^n y_i - \frac{1}{1-p}\sum_{i=1}^n(1-y_i) = \frac{n\overline{y}}{p} - \frac{n(1-\overline{y})}{1-p}$$

qui s'annulle en $\widehat{p} = \overline{y}$, c'est à dire

$$1 - \Phi(-\widehat{\theta}) = \widehat{p} = \overline{y}, \text{ soit } \widehat{\theta} = -\Phi^{-1}\Big(1 - \overline{y}\Big) = \Phi^{-1}\Big(\overline{y}\Big)$$

par symmétrie de la loi $\mathcal{N}(0,1)$. On obtient le même estimateur que la méthode des moments.
2.2) Pour l'estimateur du maximum de vraisemblance, on peut bien entendu faire plus compliqué. En effet, la fonction de probabilité s'écrit aussi, comme $p = 1 - \Phi(-\theta) = \Phi(\theta)$

$$f_\theta(y) = \Phi(\theta)^y(1 - \Phi(\theta))^{1-y}$$

dont le logarithme vaut

$$\log f_\theta(y) = y\log(\Phi(\theta)) + (1-y)\log(1 - \Phi(\theta)),$$

de telle sorte que la log-vraisemblance s'écrit

$$\log \mathcal{L}(\theta) = \log(\Phi(\theta))\sum_{i=1}^n y_i + \log(1 - \Phi(\theta))\sum_{i=1}^n (1 - y_i).$$

Comme cette fonction est continue sur $\mathbb{R}$, on va chrecher le maximum de la (log)-vraisemblance grâce à la condition du premier ordre. La dérivée de la log-vraisemblance s'écrit

$$\frac{\partial \log \mathcal{L}(\theta)}{\partial p} = \frac{\Phi'(\theta)}{\Phi(\theta)}\sum_{i=1}^n y_i - \frac{\Phi'(\theta)}{1 - \Phi(\theta)}\sum_{i=1}^n (1 - y_i) = \Phi'(\theta)\left(\frac{n\overline{y}}{\Phi(\theta)} - \frac{n(1-\overline{y})}{1 - \Phi(\theta)}\right)$$

qui s'annule quand

$$\frac{n\overline{y}}{\Phi(\widehat{\theta})} = \frac{n(1-\overline{y})}{1 - \Phi(\widehat{\theta})} \text{ c'est à dire } \Phi(\widehat{\theta}) = \overline{y}$$

autrement dit $\widehat{\theta} = \Phi^{-1}\Big(\overline{y}\Big)$.

Exercice 4 – [15 points]

On considère des variables independentes $Y_1, \cdots, Y_n$ suivant des lois de Bernoulli $\mathcal{B}(p)$ i.i.d. On veut estimer $\theta = p(1-p)$. Assez naturellement, on pose $\widehat{\theta} = \overline{Y}(1 - \overline{Y})$.

1. Calculer biais$(\widehat{\theta})$.
2. À partir de $\widehat{\theta}$, construire un estimateur sans biais de θ.

1) Pour calculer le biais de $\widehat{\theta}$, commençons par calculer son espérance,

$$\mathbb{E}[\widehat{\theta}] = \mathbb{E}[\overline{Y}(1 - \overline{Y})] = \mathbb{E}[\overline{Y}] - \mathbb{E}\Big[\overline{Y}^2\Big]$$

On peut utiliser ici le fait que $\mathbb{E}[Z^2] = \text{Var}[Z] + \mathbb{E}[Z]^2$, pour écrire

$$\mathbb{E}[\widehat{\theta}] = \mathbb{E}[\overline{Y}] - \Big[\text{Var}[\overline{Y}] + \mathbb{E}[\overline{Y}]^2\Big].$$

Or $\mathbb{E}[\overline{Y}] = p$ et $\text{Var}[\overline{Y}] = \dfrac{p(1-p)}{n}$, et donc

$$\mathbb{E}[\widehat{\theta}] = p - \left[\frac{p(1-p)}{n} + p^2\right] = \frac{n-1}{n}p(1-p) = \frac{n-1}{n}\theta.$$

de telle sorte que le biais s'écrit

$$\text{biais}(\widehat{\theta}) = \mathbb{E}[\widehat{\theta}] - \theta = \frac{n-1}{n}\theta - \theta = -\frac{1}{n}\theta.$$

2) Pour construire un estimateur sans biais de θ, $\widetilde{\theta}$, on peut considérer, assez naturellement

$$\widetilde{\theta} = \frac{n}{n-1}\widehat{\theta}$$

car dans ce cas,

$$\mathbb{E}[\widetilde{\theta}] = \frac{n}{n-1}\mathbb{E}[\widehat{\theta}] = \frac{n}{n-1}\cdot\frac{n-1}{n}\theta = \theta.$$

Exercice 5 – ***(Borne de Cramér-Rao)*** [15 points]

On considère des variables independentes $Y_1, \cdots, Y_n$ de densité

$$f_\theta(y) = 2^\theta \theta y^{-(1+\theta)}\mathbf{1}_{[2,\infty)}(y), \text{ avec } \theta > 1.$$

Calculer la borne de Cramér Rao pour ce modèle.

On commence, comme dans le cours, par définir la fonction score,

$$s(y) = \frac{\partial \log f_\theta(y)}{\partial\theta} = \frac{\partial}{\partial\theta}\Big(\theta\log 2 + \log\theta - (1+\theta)\log(y)\Big)$$

i.e.

$$s(y) = \frac{\partial \log f_\theta(y)}{\partial\theta} = \log 2 + \frac{1}{\theta} - \log(y)$$

et

$$h(y) = \frac{\partial^2 \log f_\theta(y)}{\partial\theta^2} = -\frac{1}{\theta}$$

L'information de Fisher est

$$I_1(\theta) = -\mathbb{E}\Big[h(Y)\Big] = \frac{1}{\theta^2}$$

Par définition de la borne de Cramér Rao,

$$\text{borne de Cramér Rao } = \frac{1}{I_n(\theta)} = \frac{1}{nI_1(\theta)} = \frac{\theta^2}{n}.$$

Et histoire de faire un petit rappel de cours, cela signifie que si $\widehat{\theta}$ est un estimateur sans biais de θ,

$$\text{Var}[\widehat{\theta}] \geq \frac{\theta^2}{n}.$$

Exercice 6 – ***(Estimateurs)*** [20 points]

On considère des variables independentes $Y_1, \cdots, Y_n$ de loi

$$f_\theta(y) = \mathbb{P}[Y_i = y] = (1-\theta)^{y-1}\theta, \text{ pour } y = 1, 2, \cdots \text{ et } \theta \in (0,1).$$

1. Calculer l'estimateur de la méthode des moments de θ.
2. Calculer l'estimateur du maximum de vraisemblance de θ.

1) Pour les amateurs de noms de lois, c'est une loi dite géométrique. L'espérance de cette loi s'écrit

$$\mathbb{E}[Y] = \sum_{y=1}^{\infty} y \cdot f_\theta(y) = \sum_{y=1}^{\infty} y \cdot (1-\theta)^{y-1}\theta = \theta \sum_{y=1}^{\infty} y \cdot (1-\theta)^{y-1} = \theta \sum_{y=0}^{\infty} y \cdot (1-\theta)^{y-1}$$

(la somme peut commencer en 0 ou en 1 car le premier terme de somme vaut 0 quand $y=0$). Une petite ligne de calcul va s'imposer ici: notons

$$g(\theta) = \sum_{y=0}^{\infty} y(1-\theta)^{y-1} = -\sum_{y=0}^{\infty} \frac{\partial}{\partial\theta}(1-\theta)^y = -\frac{\partial}{\partial\theta}\sum_{y=0}^{\infty} \underbrace{(1-\theta)^y}_{r^y} = -\frac{\partial}{\partial\theta}\frac{1}{1-r} = -\frac{\partial}{\partial\theta}\frac{1}{\theta} = -\frac{-1}{\theta^2} = \frac{1}{\theta^2}$$

(on aura reconnu au centre une série géométrique de raison r). Aussi

$$\mathbb{E}[Y] = \theta \cdot \frac{1}{\theta^2} = \frac{1}{\theta}.$$

Aussi, l'estimateur de la méthode des moments va vérifier

$$\overline{y} = \frac{1}{\widehat{\theta}} \text{ soit } \widehat{\theta} = \frac{1}{\overline{y}}$$

2) Pour l'estimateur du maximum de vraisemblance, commençons par écrire la vraisemblance :

$$\mathcal{L}(\theta) = \prod_{i=1}^{n} f_\theta(y_i) = \prod_{i=1}^{n}(1-\theta)^{y-1}\theta = (1-\theta)^{\sum(y_i-1)}\theta^n$$

et la log-vraisemblance

$$\log \mathcal{L}(\theta) = \log(1-\theta)\left(\sum_{i=1}^{n}(y_i-1)\right) + n\log\theta = \log(1-\theta)\Big(n(\overline{y}-1)\Big) + n\log\theta$$

Pour obtenir l'argument du maximum de cette fonction, on passe par la condition du premier ordre

$$\left.\frac{\partial \log \mathcal{L}(\theta)}{\partial\theta}\right|_{\theta=\widehat{\theta}} = 0,$$

avec ici

$$\frac{\partial \log \mathcal{L}(\theta)}{\partial\theta} = -\frac{n(\overline{y}-1)}{1-\theta} + \frac{n}{\theta},$$

et donc $\widehat{\theta} = 1/\overline{y}$. On retrouve (là encore) le même estimateur qu'avec la méthode des moments.

Exercice 7 – ***(Maximum de vraisemblance)*** [20 points]

On considère la fonction de répartition sur $[0,\infty)$,

$$F_\theta(y) = \mathbb{P}[Y \le y] = \left(1 + \frac{1}{y^2}\right)^{-\theta},\ \theta > 0.$$

On dispose d'un échantillon $\{y_1, \cdots, y_n\}$ tiré suivant des variables i.i.d. de loi F_θ.

1. Calculer l'estimateur du maximateur du maximum de vraisemblance $\widehat{\theta}$
2. Calculer l'information de Fisher de θ dans ce modèle.
3. Donner la loi de $\sqrt{n}\left(\widehat{\theta}-\theta\right)$ lorsque $n \to \infty$?

Pour les amateurs de noms de lois, cette loi s'appelle une loi de Dagum, introduite dans les années 1970 par Camilo Dagum, économiste argentin qui a longtemps travaillé à l'Université d'Ottawa.
1) Pour calculer la vraisemblance, il faut d'abord la densité

$$f_\theta(y) = \frac{\partial}{\partial y}F_\theta(y) = \frac{2\theta}{y^3}\left(1+\frac{1}{y^2}\right)^{-\theta-1}$$

de telle sorte que

$$\mathcal{L}(\theta) = \prod_{i=1}^n f_\theta(y_i) = 2^n\theta^n \prod_{i=1}^n \left(y_i^{-3}\left(1+y_i^{-2}\right)^{-(\theta+1)}\right)$$

et

$$\log \mathcal{L}(\theta) = n\log 2 + n\log\theta - \sum_{i=1}^n (3\log y_i) - (\theta+1)\sum_{i=1}^n \log\left(1+y_i^{-2}\right)$$

Comme (presque) toujours, pour obtenir l'argument du maximum de cette fonction, on passe par la condition du premier ordre

$$\left.\frac{\partial \log \mathcal{L}(\theta)}{\partial\theta}\right|_{\theta=\widehat{\theta}} = 0,$$

avec ici

$$\frac{\partial \log \mathcal{L}(\theta)}{\partial\theta} = \frac{n}{\theta} - \sum_{i=1}^n \log\left(1+y_i^{-2}\right)$$

autrement dit

$$\widehat{\theta} = \frac{n}{\displaystyle\sum_{i=1}^n (1+y_i^{-2})}$$

2) Pour calculer l'information de Fisher, on va dériver deux fois $\theta \mapsto \log f_\theta(y)$, car comme on l'a vu en cours c'est plus simple de considérer une seule observation (et donc de travailler avec la densité plutôt que la vraisemblance).

$$\log f_\theta(y) = \log 2 + \log\theta - 3\log y - (\theta+1)\log\left(1+y^{-2}\right)$$

$$\frac{\partial}{\partial\theta}\log f_\theta(y) = \frac{1}{\theta} - \log\left(1+y^{-2}\right)$$

et

$$\frac{\partial^2}{\partial\theta^2}\log f_\theta(y) = -\frac{1}{\theta^2}$$

Et comme

$$I_1(\theta) = -\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\log f_\theta(Y)\right], \text{ avec } Y \sim F_\theta$$

$$I_1(\theta) = \frac{1}{\theta^2}.$$

3) Comme on l'a vu en cours, la loi de $\sqrt{n}\left(\widehat{\theta}-\theta\right)$ lorsque $n\to\infty$ est une loi normale, centrée, de variance l'inverse de l'information de Fisher $I_1(\theta)$, soit

$$\sqrt{n}\left(\widehat{\theta}-\theta\right)\xrightarrow{\mathcal{L}}\mathcal{N}\left(0,\theta^2\right)$$

lorsque $n\to\infty$.
L'exercice s'arrêtait là, mais on peut ensuite utiliser cette propriété pour construire un intervalle de confiance asymptotique, puisque

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

de telle sorte que (si n est asssez grand)

$$\mathbb{P}\left(\Phi^{-1}(\alpha/2)\leq\sqrt{n}\left(\frac{\widehat{\theta}}{\theta}-1\right)\leq\Phi^{-1}(1-\alpha/2)\right)\approx 1-\alpha$$

et donc

$$\mathbb{P}\left(\theta\in\left[\frac{\sqrt{n}\widehat{\theta}}{\sqrt{n}+\Phi^{-1}(1-\alpha/2)};\frac{\sqrt{n}\widehat{\theta}}{\sqrt{n}+\Phi^{-1}(\alpha/2)}\right]\right)\approx 1-\alpha$$

Par exemple si $\alpha=5\%$,

$$\mathbb{P}\left(\theta\in\left[\frac{\sqrt{n}\widehat{\theta}}{\sqrt{n}+2};\frac{\sqrt{n}\widehat{\theta}}{\sqrt{n}-2}\right]\right)\approx 95\%$$