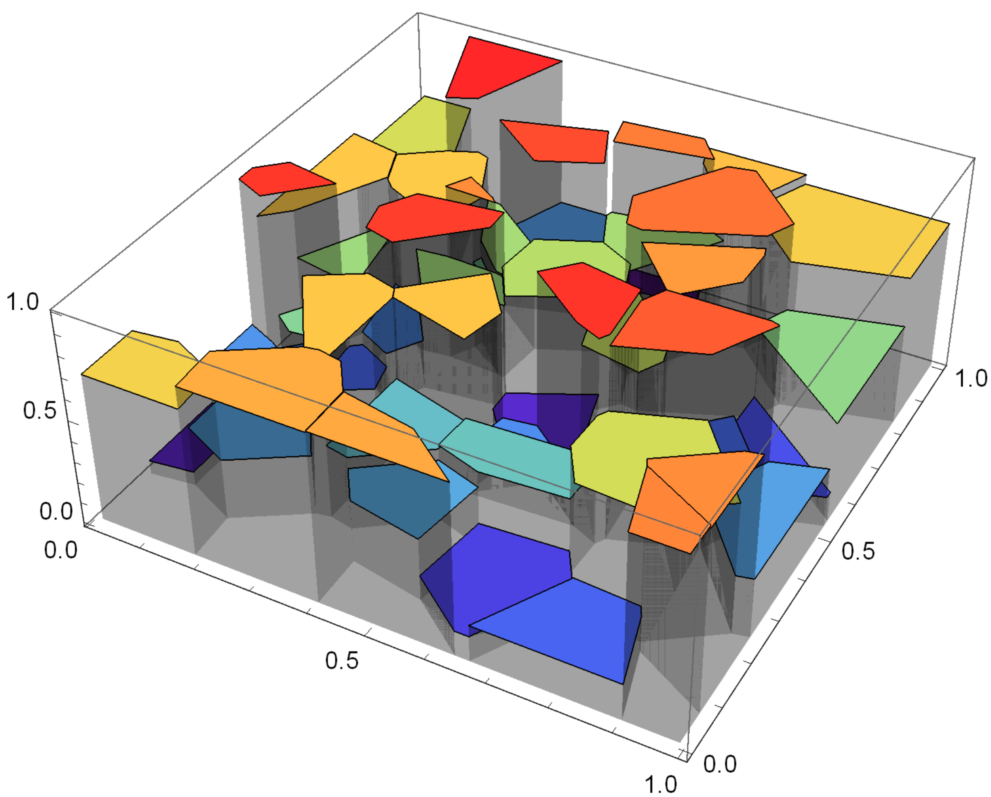
Université de Caen

# Sur l'Estimateur du Maximum de Vraisemblance *(emv)*

Christophe Chesneau

https://chesneau.users.lmno.cnrs.fr/




Caen, le 23 Octobre 2018

# Table des matières



**$\sim$ Note $\sim$**

Ce document présente quelques éléments théoriques et pratiques concernant l'estimateur du maximum de vraisemblance. Je vous invite à me contacter pour tout commentaire : `christophe.chesneau@gmail.com`

# 1 Contexte statistique

### Population et individus

Une population est un ensemble d'objets sur lesquels une étude se porte. Ces objets sont appelés individus.

### Caractère/variable

Toute propriété étudiée chez les individus d'une population est appelée caractère.

Dans ce document, on considèrera un seul caractère noté $X$. Ce qui suit est généralisable à plusieurs caractères étudiées simultanément.

### Nature d'un caractère

Un caractère est dit :

- quantitatif s'il mesure une quantité ou un nombre ; les valeurs que prend $X$ sont numériques (le nombre de personnes dans une salle, le salaire, le nombre d'items dans une liste, le temps de travail. . . ),
- qualitatif/catégoriel s'il mesure une catégorie ; les valeurs que prend $X$ sont des modalités (la couleur des yeux, la marque de téléphone, la présence ou l'absence d'un défaut de fabrication, le stade d'une maladie, la classe d'âges. . . ).

Dans ce document, on suppose que $X$ prend des valeurs numériques : il peut être quantitatif, ou qualitatif avec un codage numérique des modalités (1 pour présence, 0 pour absence. . . ).

### Échantillon d'individus

En pratique, il est souvent difficile de considérer la population dans sa totalité. On choisit alors un groupe d'individus dans la population suivant un processus de sélection précis. Ce groupe est appelé échantillon. Le nombre d'individus dans l'échantillon est noté $n$.

### Processus de sélection

Dans ce document, on suppose qu'un échantillon est constitué de la manière suivante :

- on choisit les $n$ individus un à un, au hasard et avec remise ; il est possible qu'un même individu soit sélectionné plusieurs fois,

- chaque individu a la même probabilité qu'un autre d'être sélectionné.

### Sur la sélection avec remise

Dans de nombreuses situations pratiques, l'idée de faire une sélection avec remise des individus est saugrenue ; une fois que l'on a observé la valeur de $X$ sur un individu, il est souvent logique de ne plus le considérer. Cette sélection avec remise est pourtant une hypothèse implicite classique. Elle sert surtout à simplifier la théorie en rendant les expériences de sélection des individus indépendantes les unes des autres. Le lien entre la pratique et la théorie repose sur le postulat suivant : "lorsque $n$ est faible par rapport au nombre d'individus dans la population, il est raisonnable d'admettre qu'un choix sans remise est assimilable à un choix avec remise".

### Données

Une fois l'échantillon sélectionné, on observe la valeur du caractère $X$ sur chacun des $n$ individus de l'échantillon. Ces valeurs sont appelées données. Elles sont notées $x_1, \ldots, x_n$.

### Estimation paramétrique

L'estimation paramétrique consiste à estimer/évaluer un ou plusieurs paramètres inconnus émanant de $X$ (valeur moyenne, mesures de la variabilité des valeurs...) à partir des données $x_1, \ldots, x_n$. Naturellement, l'estimation du ou des paramètres inconnus doit être aussi précise que possible. Les trois grands types d'estimation paramétriques sont :

- l'estimation ponctuelle ; on estime directement par un ou des réels le ou les paramètres inconnus,
- l'estimation par intervalles de confiance : on détermine des intervalles de réels, les moins étendus possible, qui ont de fortes chances de contenir un paramètre inconnu,
- les tests statistiques : démarches qui consistent à accepter ou non une hypothèse mettant en jeu un ou plusieurs paramètres inconnus, avec un faible risque de se tromper.

### Vers la modélisation mathématique

Deux questions se posent alors :

- Comment faire pour estimer le ou les paramètres inconnus ?
- Comment mesurer la précision de notre estimation ?

Des réponses sont apportées par le biais d'une modélisation mathématique prenant en compte l'aléatoire omniprésent dans ce contexte.

## 2 Modélisation

### Modélisation du caractère $X$

Naturellement, la valeur d'un caractère $X$ peut varier d'un individu à un autre. On peut alors modéliser $X$ par une variable aléatoire réelle (*var*) définie sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$. Cette *var* est également notée $X$ par convention. Dès lors, $X$ est l'application qui, à tout individu choisit au hasard dans la population, associe la valeur observée du caractère.

Les données $x_1, \ldots, x_n$ sont donc des réalisations de la *var* $X$.

### Nature de la *var* $X$

∘ Si les valeurs de $X$ peuvent être énumérées, on considère la *var* $X$ comme discrète.

∘ Si l'énumération des valeurs de $X$ est fastidieuse, on considère la *var* $X$ comme étant à densité.

L'ensemble des valeurs atteignables par la *var* $X$ est appelé support de $X$. Il est noté $X(\Omega)$.

### Loi (de probabilité) d'une *var* $X$

La loi de la *var* $X$ est la probabilité $\mathbb{P}_X$ définie par $\mathbb{P}_X(A) = \mathbb{P}(X \in A)$, pour tout évenement $A$.

### Hypothèses sur la loi de la *var* $X$

A priori, la loi (exacte) de la *var* $X$ est inconnue. Toutefois, on suppose qu'elle est caractérisée par une fonction dont on ne connait pas un ou plusieurs paramètres. Ces paramètres inconnus sont notés/symbolisés par $\theta$. La loi de $X$ est notée $\mathcal{L}(\theta)$. Entre autre, $\theta$ peut-être :

∘ la valeur moyenne de $X$ correspondant à l'espérance de la *var* $X$ notée $\mathbb{E}(X)$,

∘ une mesure de la variabilité des valeurs de $X$ comme l'écart-type de la *var* $X$ noté $\sigma(X)$,

∘ les deux : $\theta = (\mathbb{E}(X), \sigma(X))^t$ (noté sous la forme d'un vecteur colonne).

### Fonction caractérisant la loi de la *var* $X$

La fonction caractérisant la loi $\mathcal{L}(\theta)$ de $X$ est définie par :

$$f(x;\theta) = \begin{cases} \mathbb{P}_\theta(X = x), & \text{la fonction de masse de } X\text{, si } X \text{ est une } \textit{var} \text{ discrète,} \\ f_\theta(x), & \text{une densité de } X\text{, si } X \text{ est une } \textit{var} \text{ à densité.} \end{cases}$$

L'indice $\theta$ précise que $\theta$ apparaît dans les expressions. Il sera omis dans la suite sauf pour clarifier son apparition si besoin est. Le $x$ est pris sur l'ensemble des valeurs atteignables par $X$.

Ainsi, pour tout événement $A$, la loi de $X$ est caractérisée par

$$\mathbb{P}(X \in A) = \begin{cases} \displaystyle\sum_{x \in A} \mathbb{P}_\theta(X = x), & \text{si } X \text{ est une } \textit{var} \text{ discrète,} \\ \displaystyle\int_A f_\theta(x)dx & \text{si } X \text{ est une } \textit{var} \text{ à densité.} \end{cases}$$

### $n$-échantillon

Le processus de sélection des individus étant aléatoire, les données peuvent différer d'un échantillon à l'autre. Cette variabilité peut se modéliser à l'aide de variables aléatoires réelles. Pour tout $i \in \{1, \ldots, n\}$, on considère l'application $X_i$ qui, à tout échantillon obtenu avec ce processus, associe la valeur du caractère $X$ observé sur le $i$-ème individu. Dès lors :

- les données $x_1, \ldots, x_n$ sont des réalisations de $X_1, \ldots, X_n$,
- $X_1, \ldots, X_n$ sont des *var* indépendantes et identiquement distribuées (*iid*),
- la loi commune des *var* $X_1, \ldots, X_n$ est celle de la *var* $X$ : $X_i$ suit la loi $\mathcal{L}(\theta)$.

Le vecteur aléatoire réel $(X_1, \ldots, X_n)$ est appelé $n$-échantillon de $X$.

Ainsi le terme "échantillon" désigne tantôt un groupe d'individus, tantôt $(X_1, \ldots, X_n)$.

### Statistique

On appelle statistique une *var* définie comme étant une fonction d'un $n$-échantillon de $X$.

### Estimateur (aléatoire) de $\theta$

On appelle estimateur (aléatoire) de $\theta$ toute statistique dont les valeurs sont plausiblement proches de $\theta$. On le note $\widehat{\theta}_n$. On peut écrire $\widehat{\theta}_n$ sous la forme $\widehat{\theta}_n = g_n(X_1, \ldots, X_n)$, où $g_n$ désigne une fonction connue dépendant éventuellement de $n$.

### Estimateur ponctuel de $\theta$

Un estimateur ponctuel de $\theta$ est la réalisation d'un estimateur $\widehat{\theta}_n$ de $\theta$ correspondante aux données $x_1, \ldots, x_n$ ; c'est donc un réel ou un vecteur de réels. On le note $\theta^*$.

### Représentation

Partant d'un estimateur $\widehat{\theta}_n = g_n(X_1, \ldots, X_n)$, un estimateur ponctuel de $\theta$ est $\theta^* = g_n(x_1, \ldots, x_n)$ ; on a substitué $X_i$ par $x_i$ pour tout $i \in \{1, \ldots, n\}$.

Réciproquement, si un estimateur ponctuel de $\theta$ est de la forme $\theta^* = g_n(x_1, \ldots, x_n)$ où $g_n$ désigne une fonction connue, l'estimateur sous-jacent est $\widehat{\theta}_n = g_n(X_1, \ldots, X_n)$.

### Estimation paramétrique : enjeu

L'enjeu de l'estimation paramétrique est de

- déterminer un estimateur ponctuel $\theta^*$ qui soit proche de $\theta$,
- évaluer l'écart entre $\theta^*$ et $\theta$ avec le plus de précision possible.

Pour ce faire, la construction d'un "bon estimateur" $\widehat{\theta}_n$ de $\theta$ est nécessaire.

### Notion d'étude asymptotique

On part du postulat suivant : "plus $n$ est grand, plus on a de données, plus on dispose d'information sur les caractéristiques de $X$, plus on doit être en mesure d'estimer précisément $\theta$". L'entier $n$ tel qu'on l'a introduit est fixé. Toutefois, la définition d'un estimateur (aléatoire) $\widehat{\theta}_n$ de $\theta$ nous autorise à considérer $n$ comme une variable entière. Ainsi, on peut introduire la suite de *var* $(\widehat{\theta}_n)_{n\in\mathbb{N}^*}$ et étudier sa précision dans l'estimation de $\theta$ quand $n$ tend vers l'infini. On parle alors d'étude asymptotique.

De manière générale, le mot asymptotique signifie "quand $n$ tend vers l'infini", qui se transpose en pratique par "quand $n$ est assez grand".

### Estimateur performant : première approche

Un estimateur $\widehat{\theta}_n$ de $\theta$ est considéré comme performant si

- la mesure de la *var* $\widehat{\theta}_n - \theta$ sous divers critères probabilistes est faible,
- la suite de *var* $(\widehat{\theta}_n - \theta)_{n\in\mathbb{N}^*}$ considérée sous divers critères probabilistes converge rapidement vers 0 quand $n$ tend vers l'infini,
- la loi de $\widehat{\theta}_n - \theta$ peut être approchée par une loi usuelle quand $n$ est assez grand.

L'erreur d'estimation modélisée par la *var* $\widehat{\theta}_n - \theta$ joue donc un rôle central.

Dans ce document, pour simplifier, on suppose désormais que :

- $(X_1, \ldots, X_n)$ désigne un $n$-échantillon d'une *var* $X$ suivant la loi $\mathcal{L}(\theta)$ caractérisée par la fonction $f(x; \theta)$,
- $\theta$ désigne un seul paramètre inconnu,
- les quantités introduites existent (espérance, variance, écart-type, dérivée...). Si une de ces quantités n'existe pas dans le cadre d'une application, la formule proposée devient absurde ; elle doit être ignorée.

# 3 Performance non-asymptotique d'un estimateur

### Estimateur sans biais

On dit qu'un estimateur $\widehat{\theta}_n$ de $\theta$ est sans biais si et seulement si $\mathbb{E}(\widehat{\theta}_n) = \theta$.

Cela signifie que la valeur moyenne de $\widehat{\theta}_n$ est égale à $\theta$.

### Biais d'un estimateur

Plus généralement, on appelle biais d'un estimateur $\widehat{\theta}_n$ de $\theta$ le réel :

$$B(\widehat{\theta}_n, \theta) = \mathbb{E}(\widehat{\theta}_n - \theta) = \mathbb{E}(\widehat{\theta}_n) - \theta.$$

Ainsi, le biais mesure l'écart moyen entre $\widehat{\theta}_n$ et $\theta$.

### Risque quadratique

On appelle risque quadratique d'un estimateur $\widehat{\theta}_n$ de $\theta$ le réel :

$$R(\widehat{\theta}_n, \theta) = \mathbb{E}((\widehat{\theta}_n - \theta)^2).$$

Le risque quadratique est une mesure de l'erreur moyenne que commet $\widehat{\theta}_n$ dans l'estimation de $\theta$. Il est un critère de performance : plus $R(\widehat{\theta}_n, \theta)$ est petit, plus l'estimateur $\widehat{\theta}_n$ estime bien $\theta$.

Si $\widehat{\theta}_n$ est sans biais, alors on a $R(\widehat{\theta}_n, \theta) = \mathbb{E}((\widehat{\theta}_n - \mathbb{E}(\widehat{\theta}_n))^2) = \mathbb{V}(\widehat{\theta}_n)$.

### Expression pratique du risque quadratique

On peut montrer que

$$R(\widehat{\theta}_n, \theta) = \mathbb{V}(\widehat{\theta}_n) + (B(\widehat{\theta}_n, \theta))^2.$$

### Estimateur préférable

Soient $\widehat{\theta}_n^{(1)}$ et $\widehat{\theta}_n^{(2)}$ deux estimateurs de $\theta$. On dit que $\widehat{\theta}_n^{(1)}$ est préférable à $\widehat{\theta}_n^{(2)}$ si et seulement si

$$R(\widehat{\theta}_n^{(1)}, \theta) \leq R(\widehat{\theta}_n^{(2)}, \theta).$$

### "*Sccri*" = "Sous certaines conditions de régularité et d'intégrabilité"

Dans ce document, on utilisera l'abréviation *Sccri* pour "Sous certaines conditions de régularité et d'intégrabilité" des fonctions mises en jeu. Ces conditions sont purement techniques, difficilement intelligibles et souvent satisfaites dans les applications. Pour ces raisons, on les omet. Toutefois, il est important de noter qu'elles excluent le cas où le support de $f(x;\theta)$ dépend de $\theta$.

### Borne de Cramer-Rao : première approche

*Sccri*, il existe un réel positif $u_n(\theta)$ appelé borne de Cramer-Rao tel que, pour tout estimateur $\widehat{\theta}_n$ de $\theta$ sans biais, on a

$$R(\widehat{\theta}_n, \theta) \geq u_n(\theta),$$

L'estimateur $\widehat{\theta}_n$ étant sans biais, on peut aussi écrire cette inégalité avec $\mathbb{V}(\widehat{\theta}_n)$ à la place de $R(\widehat{\theta}_n, \theta)$.

Cette borne est

- ∘ indépendante de l'expression de l'estimateur,
- ∘ atteignable par un ou plusieurs estimateur a priori.

Elle nous donne donc une indication précise sur la possibilité de bien estimer $\theta$.

### Sur $\theta$ : du paramètre inconnu à la variable

Ne connaissant pas la valeur du paramètre $\theta$, on le considère parfois comme une variable réelle définie sur l'ensemble des ses valeurs possibles. Cela nous permet d'effectuer des opérations usuelles sur des fonctions de $\theta$ (dérivée en $\theta$, maximisation en $\theta\ldots$).

### Information de Fisher

*Sccri*, on appelle information de Fisher/information de Fisher fournit par $(X_1, \ldots, X_n)$ sur $\theta$ le réel :

$$I_n(\theta) = nI_1(\theta),$$

avec

$$I_1(\theta) = \mathbb{E}\left(\left(\frac{\partial}{\partial\theta}\ln(f(X_1;\theta))\right)^2\right).$$

Dans cette expression, $\frac{\partial}{\partial\theta}\ln(f(x;\theta))$ désigne la dérivée partielle de $f(x;\theta)$ en $\theta$. Une fois calculée, on élève $\frac{\partial}{\partial\theta}\ln(f(x;\theta))$ au carré, on remplace $x$ par $X_1$ et on prend l'espérance de la nouvelle *var* obtenue.

L'information de Fisher caractérise la borne de Cramer-Rao ; on a

$$u_n(\theta) = \frac{1}{I_n(\theta)}.$$

Par conséquent,

- plus $I_n(\theta)$ est grand, plus $u_n(\theta)$ est petit, plus le risque $R(\widehat{\theta}_n, \theta)$ peut être petit, plus on peut espérer pourvoir trouver de "bons" estimateurs sans biais de $\theta$,
- plus $I_n(\theta)$ est petit, plus $u_n(\theta)$ est grand, moins le risque $R(\widehat{\theta}_n, \theta)$ peut être petit,moins $\theta$ sera bien estimé, peu importe l'estimateur considéré.

Ainsi, partant d'un $n$-échantillon $(X_1, \ldots, X_n)$ de $X$, l'information de Fisher mesure

- une erreur inhérente à l'hypothèse que $X$ suit la loi $\mathcal{L}(\theta)$,
- la quantité moyenne d'information apportée par $X$ sur $\theta$.

**Autre expression de $I_1(\theta)$**

On peut montrer que

$$I_1(\theta) = \mathbb{V}\left(\frac{\partial}{\partial\theta}\ln(f(X_1;\theta))\right), \qquad I_1(\theta) = -\mathbb{E}\left(\frac{\partial^2}{\partial\theta^2}\ln(f(X_1;\theta))\right).$$

Cette dernière expression est souvent la plus pratique dans les calculs.

**Estimateur efficace**

*Sccri*, on dit qu'un estimateur $\widehat{\theta}_n$ de $\theta$ est efficace si et seulement si

- il est sans biais : $\mathbb{E}(\widehat{\theta}_n) = \theta$,
- on a :

$$\mathbb{V}(\widehat{\theta}_n) = \frac{1}{I_n(\theta)}.$$

Ces deux conditions entraînent que la borne de Cramer-Rao est atteinte : $R(\widehat{\theta}_n, \theta) = u_n(\theta)$. Ainsi, le risque d'un estimateur efficace $\widehat{\theta}_n$ est le plus petit possible ; il est donc préférable à tous les estimateurs sans biais de $\theta$. En ce sens, c'est le "meilleur" estimateur sans biais.

### Borne minimale du risque quadratique d'un estimateur (possiblement biaisé)

*Sccri*, pour tout estimateur $\widehat{\theta}_n$ de $\theta$, on a

$$R(\widehat{\theta}_n, \theta) \geq \frac{1}{I_n(\theta)} \left(1 + \frac{\partial}{\partial \theta} B(\widehat{\theta}_n, \theta)\right)^2 + (B(\widehat{\theta}_n, \theta))^2.$$

La borne inférieure est appelée "borne minimale du risque quadratique de $\widehat{\theta}_n$". Si $\widehat{\theta}_n$ est sans biais, on retrouve la borne de Cramer-Rao.

# 4 Performance asymptotique d'un estimateur

### Estimateur asymptotiquement sans biais

On dit qu'un estimateur $\widehat{\theta}_n$ de $\theta$ est asymptotiquement sans biais si et seulement si

$$\lim_{n\to\infty} \mathbb{E}(\widehat{\theta}_n) = \theta.$$

Cela signifie que la valeur moyenne de $\widehat{\theta}_n$ doit être proche de $\theta$ quand $n$ est assez grand.

Un estimateur sans biais est asymptotiquement sans biais.

### Estimateur faiblement consistant

On dit qu'un estimateur $\widehat{\theta}_n$ de $\theta$ est faiblement consistant si et seulement si la suite de *var* $(\widehat{\theta}_n)_{n\in\mathbb{N}^*}$ converge en probabilité vers $\theta$, *i.e.* pour tout $\epsilon > 0$, on a

$$\lim_{n\to\infty} \mathbb{P}(|\widehat{\theta}_n - \theta| \geq \epsilon) = 0.$$

Cela signifie que $\widehat{\theta}_n$ s'écarte de $\theta$ avec une faible probabilité quand $n$ est assez grand.

### Critère suffisant pour la faible consistance

Si un estimateur $\widehat{\theta}_n$ de $\theta$ vérifie $\lim_{n\to\infty} R(\widehat{\theta}_n, \theta) = 0$, alors il est faiblement consistant ; c'est une conséquence de l'inégalité de Markov.

### Estimateur fortement consistant

On dit qu'un estimateur $\widehat{\theta}_n$ de $\theta$ est fortement consistant de $\theta$ si et seulement si la suite de *var* $(\widehat{\theta}_n)_{n\in\mathbb{N}^*}$ converge presque sûrement vers $\theta$, *i.e.*

$$\mathbb{P}(\lim_{n\to\infty} \widehat{\theta}_n = \theta) = 1.$$

Cela signifie que $\widehat{\theta}_n$ est assurément proche de $\theta$ quand $n$ est assez grand.

La convergence presque sure impliquant la convergence en probabilité, un estimateur fortement consistant est consistant.

### Critère suffisant pour la forte consistance

Si un estimateur $\widehat{\theta}_n$ de $\theta$ vérifie, pour tout $\epsilon > 0$,

$$\sum_{n=1}^{\infty} \mathbb{P}(|\widehat{\theta}_n - \theta| \geq \epsilon) < \infty,$$

alors il est fortement consistant ; c'est une conséquence du lemme de Borel-Cantelli.

En particulier, cette condition est vérifiée si $\sum_{n=1}^{\infty} R(\widehat{\theta}_n, \theta) < \infty$.

### Estimateur asymptotiquement efficace

*Sccri*, on dit qu'un estimateur $\widehat{\theta}_n$ de $\theta$ est asymptotiquement efficace si et seulement si

- il est asymptotiquement sans biais,
- on a
$$\lim_{n\to\infty} I_n(\theta)\mathbb{V}(\widehat{\theta}_n) = 1.$$

Ces deux conditions entraînant que la "borne minimale du risque quadratique de $\widehat{\theta}_n$" est asymptotiquement atteinte. En ce sens, c'est le "meilleur" estimateur asymptotique sans biais.

Un estimateur efficace est asymptotiquement efficace.

### Estimateur asymptotiquement normal

Un estimateur $\widehat{\theta}_n$ de $\theta$ est dit asymptotiquement normal/asymptotiquement distribué selon une loi normale si et seulement si il existe deux suites de réels $(a_n(\theta))_{n\in\mathbb{N}^*}$ et $(b_n(\theta))_{n\in\mathbb{N}^*}$ dépendantes de $\theta$ telles que la suite de *var*

$$\left(\frac{\widehat{\theta}_n - a_n(\theta)}{b_n(\theta)}\right)_{n\in\mathbb{N}^*}$$

converge en loi vers une *var* $Z$ suivant la loi normale $\mathcal{N}(0,1)$, *i.e.* pour tout $x \in \mathbb{R}$,

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{\widehat{\theta}_n - a_n(\theta)}{b_n(\theta)} \leq x\right) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

(cette dernière fonction étant la fonction de répartition de $Z$, laquelle est continue pour tout $x \in \mathbb{R}$).

Cela signifie que $\widehat{\theta}_n$ suit approximativement la loi normale $\mathcal{N}\left(a_n(\theta), b_n^2(\theta)\right)$ quand $n$ est assez grand.

C'est une propriété importante ; la loi normale étant usuelle, on peut l'utiliser pour construire des objets statistiques permettant d'évaluer $\theta$ avec précision (intervalles de confiance, tests statistiques...).

# 5 Estimateurs de la moyenne et de la variance

### Estimateur de $\theta = \mathbb{E}(X)$

On suppose que le paramètre inconnu est $\theta = \mathbb{E}(X)$ et on pose :

$$\widehat{\theta}_n = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

C'est la moyenne des $n$ *var* du $n$-échantillon $(X_1, \ldots, X_n)$. Il est généralement noté $\overline{X}_n$.

Alors $\widehat{\theta}_n$ est un estimateur de $\theta$ qui est

- sans biais : $\mathbb{E}(\widehat{\theta}_n) = \theta$,
- fortement consistant : $(\widehat{\theta}_n)_{n \in \mathbb{N}^*}$ converge presque sûrement vers $\theta$ ; c'est une conséquence de la loi forte des grands nombres,
- asymptotiquement normal : en posant $a_n(\theta) = \theta$ et $b_n(\theta) = \dfrac{\sigma(X)}{\sqrt{n}}$, la suite de *var*

$$\left( \frac{\widehat{\theta}_n - a_n(\theta)}{b_n(\theta)} \right)_{n \in \mathbb{N}^*} = \left( \sqrt{n} \left( \frac{\widehat{\theta}_n - \theta}{\sigma(X)} \right) \right)_{n \in \mathbb{N}^*}$$

  converge en loi vers une *var* $Z$ suivant la loi normale $\mathcal{N}(0,1)$ ; c'est une conséquence du théorème central limite.

Selon la loi $\mathcal{L}(\theta)$, $\widehat{\theta}_n$ peut être efficace ou asymptotiquement efficace ou ni l'un, ni l'autre.

En particulier, si $X$ suit la loi normale $\mathcal{N}(\mu, \sigma^2)$ avec $\mu$ et $\sigma$ inconnus, et $\theta = \mathbb{E}(X) = \mu$, alors $\widehat{\theta}_n$ est un estimateur de $\theta$ sans biais, fortement consistant, asymptotiquement normal et efficace. Il est même "exactement normal" car il suit la loi normale $\mathcal{N}\left(\mu, \dfrac{\sigma^2}{n}\right)$.

Un estimateur ponctuel de $\theta = \mathbb{E}(X)$ est

$$\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i.$$

### Estimateur de $\theta = \mathbb{V}(X)$

On suppose que le paramètre inconnu est $\theta = \mathbb{V}(X)$ et on pose :

$$\widehat{\theta}_n = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2, \qquad \left(\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i\right).$$

C'est $n/(n-1)$ fois la moyenne des carrés des écarts entre la *var* $X_i$ du $n$-échantillon $(X_1, \ldots, X_n)$ et $\overline{X}_n$ pour $i \in \{1, \ldots, n\}$. Il est généralement noté $S_n^2$.

Alors $\widehat{\theta}_n$ est un estimateur de $\theta$ qui est

- sans biais : $\mathbb{E}(\widehat{\theta}_n) = \theta$, c'est même pour cela que l'on divise par $n-1$ et non par $n$,
- fortement consistant : $(\widehat{\theta}_n)_{n\in\mathbb{N}^*}$ converge presque sûrement vers $\theta$ ; c'est une conséquence de la loi forte des grands nombres et du théorème de continuité.

Selon la loi $\mathcal{L}(\theta)$, $\widehat{\theta}_n$ peut être efficace ou asymptotiquement efficace ou ni l'un, ni l'autre.

En particulier, si $X$ suit la loi normale $\mathcal{N}(\mu, \sigma^2)$ avec $\mu$ et $\sigma$ inconnus, et $\theta = \mathbb{V}(X) = \sigma^2$, alors $\widehat{\theta}_n$ est un estimateur de $\theta$ sans biais, fortement consistant, asymptotiquement normal et asymptotiquement efficace. Pour ce dernier point, on peut montrer que

$$\mathbb{V}(\widehat{\theta}_n) = \frac{2\sigma^4}{n-1}, \qquad I_n(\theta) = \frac{n}{2\sigma^4}.$$

Un estimateur ponctuel de $\theta = \mathbb{V}(X)$ est

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2.$$

Aussi, un estimateur ponctuel de $\theta = \sigma(X)$ est

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2}.$$

# 6 Estimateurs du maximum de vraisemblance : première approche

## Méthode du maximum de vraisemblance

La méthode du maximum de vraisemblance est une méthode d'estimation paramétrique qui doit sa popularité à

- la simplicité de son approche,
- sa faculté d'adaptation à une modélisation complexe, *i.e.* une loi $\mathcal{L}(\theta)$ où $\theta$ symbolise une multitude de paramètres inconnus : $\theta = (\theta_1, \theta_2, \theta_3 \ldots)^t$,
- l'aspect numérique accessible grâce à l'application de méthodes d'optimisation connues.

Elle permet de :

- construire des estimateurs performants,
- construire des intervalles de confiances précis,
- mettre en œuvre des tests statistiques "puissants".

La méthode du maximum de vraisemblance fut initialement appelée "le critère absolu" (en 1912), expression lourde de sens, même un siècle plus tard.

## Estimateurs du maximum de vraisemblance : cas $X$ *var* discrète

Dans un premier temps, on suppose que $X$ est une *var* discrète suivant la loi $\mathcal{L}(\theta)$ avec $\theta$ un paramètre inconnu. On rappelle que l'on veut estimer $\theta$ à partir des données $x_1, \ldots, x_n$, le vecteur des données $(x_1, \ldots, x_n)$ étant une réalisation d'un $n$-échantillon $(X_1, \ldots, X_n)$ de $X$.

La méthode du maximum de vraisemblance repose sur l'idée suivante :

- "le fait d'avoir observé les valeurs $x_1, \ldots, x_n$ n'est pas surprenant", soit encore :
- "l'hypothèse d'observer les valeurs $x_1, \ldots, x_n$ plutôt que d'autres était la plus vraisemblable".

Dès lors, on considère $\theta$ comme une variable réelle et on s'intéresse aux valeurs de $\theta$ qui

- "rendent l'observation des valeurs $x_1, \ldots, x_n$ la plus vraisemblable possible", soit encore :
- "maximisent les chances de réalisation de l'événement $\{(X_1, \ldots, X_n) = (x_1, \ldots, x_n)\}$", soit encore :
- "maximisent la probabilité $\mathbb{P}_\theta((X_1, \ldots, X_n) = (x_1, \ldots, x_n))$".

Une telle valeur est donc solution d'un problème d'optimisation visant à maximiser une fonction de $\theta$ caractérisant la vraisemblance d'avoir obtenu $x_1, \ldots, x_n$. Ainsi, $\theta^*$ est un estimateur ponctuel du paramètre inconnu $\theta$ appelé estimateur du maximum de vraisemblance (*emv*).

Aussi, précisons que l'on peut définir la vraisemblance des données $x_1, \ldots, x_n$ par la fonction de $\theta$ :

$$L_n(\theta) = \mathbb{P}_\theta((X_1, \ldots, X_n) = (x_1, \ldots, x_n)).$$

Comme $(X_1, \ldots, X_n)$ est un $n$-échantillon, par l'indépendance et la distribution identique des *var* $X_1, \ldots, X_n$, on peut aussi écrire :

$$L_n(\theta) = \mathbb{P}_\theta\left(\bigcap_{i=1}^{n}\{X_i = x_i\}\right) = \prod_{i=1}^{n} \mathbb{P}_\theta(X_i = x_i) = \prod_{i=1}^{n} \mathbb{P}_\theta(X = x_i).$$

**Méthode du maximum de vraisemblance : cas $X$ *var* à densité**

Si $X$ est une *var* à densité suivant la loi $\mathcal{L}(\theta)$ de densité $f_\theta$, le raisonnement développé précédemment tient toujours. Toutefois, la vraisemblance des données ne peut plus être mesurée par la fonction $L_n(\theta) = \mathbb{P}_\theta((X_1, \ldots, X_n) = (x_1, \ldots, x_n))$ car elle est désormais nulle. Au lieu de $\{(X_1, \ldots, X_n) = (x_1, \ldots, x_n)\}$, une idée est d'introduire un événement non négligeable proche :

$\{(X_1, \ldots, X_n) \in [x_1, x_1 + \epsilon_1[\times \ldots \times [x_n, x_n + \epsilon_n[\}$ avec $\epsilon_1, \ldots, \epsilon_n$ assez petits.

Cela nous amène à mesurer la vraisemblance des données par la fonction de $\theta$ :

$$L_n^{(\epsilon_1,\ldots,\epsilon_n)}(\theta) = \mathbb{P}_\theta((X_1, \ldots, X_n) \in [x_1, x_1 + \epsilon_1[\times \ldots \times [x_n, x_n + \epsilon_n[).$$

En notant $f_\theta(t_1, \ldots, t_n)$ une densité de $(X_1, \ldots, X_n)$, on a

$$L_n^{(\epsilon_1,\ldots,\epsilon_n)}(\theta) = \int_{x_1}^{x_1+\epsilon_1} \ldots \int_{x_n}^{x_n+\epsilon_n} f_\theta(t_1, \ldots, t_n) dt_1 \ldots dt_n,$$

Cependant, déterminer un $\theta$ qui maximise $L_n^{(\epsilon_1,\ldots,\epsilon_n)}(\theta)$ n'est pas chose aisée :

- l'expression analytique d'une fonction intégrale comme $L_n^{(\epsilon_1,\ldots,\epsilon_n)}(\theta)$ n'existe pas toujours,
- il dépend de $\epsilon_1, \ldots, \epsilon_n$ dont la petitesse reste subjective.

Une solution est de considérer une version idéalisée de $L_n^{(\epsilon_1,\ldots,\epsilon_n)}(\theta)$ qui est

$$L_n(\theta) = f_\theta(x_1, \ldots, x_n).$$

Cette expression est une conséquence du résultat suivant : en notant $F_\theta$ la fonction de répartition de $(X_1, \ldots, X_n)$, il vient

$$\lim_{(\epsilon_1,\ldots,\epsilon_n)\to(0,\ldots,0)} \frac{L^{(\epsilon_1,\ldots,\epsilon_n)}(\theta)}{\epsilon_1 \times \ldots \times \epsilon_n} = \frac{\partial^n}{\partial\epsilon_1 \ldots \partial\epsilon_n} F_\theta(x_1, \ldots, x_n) = f_\theta(x_1, \ldots, x_n).$$

En notant $\theta^*$ un maximum de $L_n(\theta)$, $\theta^*$ est un estimateur du maximum de vraisemblance (*emv*) de $\theta$ correspondant aux données.

Comme $(X_1, \ldots, X_n)$ est un $n$-échantillon, par l'indépendance et la distribution identique des *var* $X_1, \ldots, X_n$, on peut aussi écrire :

$$L_n(\theta) = f_\theta(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_\theta(x_i).$$

### Vocabulaires

En anglais,

- "vraisemblance" se dit "likelihood", d'où la notation "L".
- "estimateur du maximum de vraisemblance" (*emv*) se dit "maximum likelihood estimator" (*mle*).

# 7 Estimateurs du maximum de vraisemblance pour $(x_1, \ldots, x_n)$

### Fonction de vraisemblance

Dans ce document, on appelle fonction de vraisemblance pour $(x_1, \ldots, x_n)$ la fonction de $\theta$ :

$$L_n(x_1, \ldots, x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta).$$

On rappelle que $f(x;\theta) = \begin{cases} \mathbb{P}_\theta(X = x), & \text{si } X \text{ est une } var \text{ discrète,} \\ f_\theta(x), & \text{si } X \text{ est une } var \text{ à densité de densité } f_\theta. \end{cases}$

La fonction de vraisemblance n'est intéressante que si $\theta$ et $x_i$ vérifient $f(x_i; \theta) \neq 0$ pour tout $i \in \{1, \ldots, n\}$, sinon on peut d'ores et déjà remettre en cause l'hypothèse que $X$ suit la loi $\mathcal{L}(\theta)$.

### Fonction de vraisemblance : contexte théorique

Dans un contexte théorique, il est possible que seule la modélisation inhérente à $X$ soit décrite, sans mention des données $x_1, \ldots, x_n$. Dès lors, on appelle fonction de vraisemblance la fonction de vraisemblance d'une réalisation quelconque $(x_1, \ldots, x_n)$ de $(X_1, \ldots, X_n)$.

### Estimateurs du maximum de vraisemblance pour $(x_1, \ldots, x_n)$

On appelle estimateur du maximum de vraisemblance de $\theta$ *(emv)* pour $(x_1, \ldots, x_n)$ un réel $\theta^*$ qui maximise la fonction de vraisemblance $L_n(x_1, \ldots, x_n; \theta)$ en $\theta$, *i.e.* pour tout $\theta$,

$$L_n(x_1, \ldots, x_n; \theta) \leq L_n(x_1, \ldots, x_n; \theta^*).$$

Une expression alternative est :

$$\theta^* \in \operatorname*{argmax}_{\theta} L_n(x_1, \ldots, x_n; \theta)$$

où argmax désigne l'argument du maximum qui est l'ensemble des points en lesquels une expression atteint sa valeur maximale.

Puisqu'il dépend de $x_1, \ldots, x_n$, $\theta^*$ est une estimation ponctuelle de $\theta$. Un tel estimateur n'existe pas toujours et peut ne pas être unique.

Dans ce document, pour simplifier, on suppose qu'il existe un unique *emv* $\theta^*$ de $\theta$ pour $(x_1, \ldots, x_n)$.

### Fonction de Log-vraisemblance

On appelle fonction de log-vraisemblance pour $(x_1, \ldots, x_n)$ la fonction de $\theta$ définie par :

$$\ell_n(x_1, \ldots, x_n; \theta) = \ln(L_n(x_1, \ldots, x_n; \theta)).$$

Elle n'a de sens que si $\theta$ vérifie $L_n(x_1, \ldots, x_n; \theta) > 0$.

La fonction logarithme népérien étant croissante, l'*emv* $\theta^*$ de $\theta$ pour $(x_1, \ldots, x_n)$ vérifie :

$$\theta^* \in \underset{\theta}{\operatorname{argmax}}\, L_n(x_1, \ldots, x_n; \theta) = \underset{\theta}{\operatorname{argmax}}\, \ell_n(x_1, \ldots, x_n; \theta).$$

### Équation de vraisemblance

On appelle équation de vraisemblance l'équation en $\theta$ :

$$\frac{\partial}{\partial \theta} \ell_n(x_1, \ldots, x_n; \theta) = 0.$$

### Expression analytique de l'*emv*

Pour envisager d'avoir une expression analytique de l'*emv* $\theta^*$ de $\theta$ pour $(x_1, \ldots, x_n)$, une idée est d'exprimer $L_n(x_1, \ldots, x_n; \theta)$ en fonction de produits de termes exponentiels/puissances, puis de considérer la fonction de log-vraisemblance $\ell_n(x_1, \ldots, x_n; \theta)$. Si cette dernière est dérivable en $\theta$, une condition nécessaire que doit vérifier $\theta^*$ est d'être solution de l'équation de vraisemblance. Il faut ensuite vérifier que $\theta^*$ est bien un maximum pour $\ell_n(x_1, \ldots, x_n; \theta)$ :

- soit en étudiant les variations de $\ell_n(x_1, \ldots, x_n; \theta)$,
- soit en montrant que
$$\frac{\partial^2}{\partial \theta^2} \ell_n(x_1, \ldots, x_n; \theta^*) < 0.$$

### Pratique de l'*emv*

Bien souvent, la résolution de l'équation de vraisemblance n'amène pas une expression analytique pour l'*emv* $\theta^*$ de $\theta$ pour $(x_1, \ldots, x_n)$. En tant que problème d'optimisation, on peut quand même approcher la valeur de $\theta^*$ avec précision à l'aide d'algorithmes itératifs efficaces.

Il y a notamment :

- ○ l'algorithme de Newton-Raphson,
- ○ l'algorithme de Gauss-Newton,
- ○ le score de Fisher.

Par exemple, partant de l'équation de vraisemblance et d'un développement de Taylor au premier ordre de $\frac{\partial}{\partial \theta} \ell_n(x_1, \ldots, x_n; \theta)$, l'algorithme de Newton-Raphson est caractérisé par la $m$-ème itération :

$$\theta^{(m+1)} = \theta^{(m)} - \left( \frac{\partial^2}{\partial \theta^2} \ell_n(x_1, \ldots, x_n; \theta^{(m)}) \right)^{-1} \left( \frac{\partial}{\partial \theta} \ell_n(x_1, \ldots, x_n; \theta^{(m)}) \right).$$

Dès qu'il y a stabilisation pour un $m_*$, *i.e.* $\theta^{(m_*+1)} = \theta^{(m_*)}$, on considère la valeur $\theta^* = \theta^{(m_*)}$.

# 8 Estimateurs du maximum de vraisemblance (aléatoire)

### Estimateur du maximum de vraisemblance (aléatoire)

On appelle estimateur du maximum de vraisemblance (*emv*) un estimateur (aléatoire) $\widehat{\theta}_n$ de $\theta$ défini par

$$\widehat{\theta}_n \in \underset{\theta}{\operatorname{argmax}}\, L_n(X_1, \ldots, X_n; \theta).$$

On a substitué $x_i$ par $X_i$ pour tout $i \in \{1, \ldots, n\}$ dans la définition de $\theta^*$.

Dans ce document, pour simplifier, on suppose qu'il existe un unique *emv* $\widehat{\theta}_n$ de $\theta$.

Si on peut exprimer $\theta^*$ sous la forme $\theta^* = g_n(x_1, \ldots, x_n)$ où $g_n$ désigne une fonction connue, l'*emv* (aléatoire) de $\theta$ est

$$\widehat{\theta}_n = g_n(X_1, \ldots, X_n).$$

C'est la performance de $\widehat{\theta}_n$ qui permet d'évaluer la précision de $\theta^*$ dans l'estimation de $\theta$.

### Propriétés théoriques de l'*emv*

*Sccri*, l'*emv* $\widehat{\theta}_n$ est "au pire" :

- asymptotiquement sans biais,
- fortement consistant,
- asymptotiquement efficace,
- asymptotiquement normal.

C'est donc un estimateur performant.

### Théorème sur l'efficacité

*Sccri*, s'il existe un réel $a_n(\theta)$ dépendant de $\theta$ tels que l'*emv* $\widehat{\theta}_n$ de $\theta$ vérifie la factorisation :

$$\frac{\partial}{\partial \theta} \ell_n(X_1, \ldots, X_n; \theta) = a_n(\theta)(\widehat{\theta}_n - \theta),$$

alors

- $\widehat{\theta}_n$ est un estimateur sans biais : $\mathbb{E}(\widehat{\theta}_n) = \theta$,

- on a
$$\mathbb{V}(\widehat{\theta}_n) = \frac{1}{a_n(\theta)},$$
- on a
$$I_n(\theta) = \frac{1}{\mathbb{V}(\widehat{\theta}_n)}.$$

En particulier, le premier et le troisième points entraînent que $\widehat{\theta}_n$ est un estimateur efficace de $\theta$.

**_Emv_ et efficacité**

S'il existe un estimateur efficace de $\theta$, alors c'est l'*emv* $\widehat{\theta}_n$ de $\theta$ (en revanche, un *emv* n'est pas nécessairement efficace).

**Sur la normalité asymptotique de l'_emv_**

*Sccri*, l'*emv* $\widehat{\theta}_n$ de $\theta$ est asymptotiquement normal ; la suite de *var* :

$$\left(\sqrt{I_n(\theta)}(\widehat{\theta}_n - \theta)\right)_{n\in\mathbb{N}^*}$$

converge en loi vers une *var* $Z$ suivant la loi normale $\mathcal{N}(0,1)$.

**Résultat asymptotique central**

*Sccri*, l'*emv* $\widehat{\theta}_n$ de $\theta$ vérifie : la suite de *var* :

$$\left(\sqrt{I_n(\widehat{\theta}_n)}(\widehat{\theta}_n - \theta)\right)_{n\in\mathbb{N}^*}$$

converge en loi vers une *var* $Z$ suivant la loi normale $\mathcal{N}(0,1)$ ; on a substitué $\theta$ à $\widehat{\theta}_n$ dans la définition de l'information de Fisher du résultat de normalité asymptotique de $\widehat{\theta}_n$.

Ce dernier résultat est central pour la construction d'intervalles de confiance pour $\theta$, de tests statistiques simples ou complexes ...

**Lien entre l'_emv_ et l'information de Fisher**

Le lien existant entre l'*emv* et l'information de Fisher peut s'expliquer ainsi :

- Plus les valeurs de $\left(\frac{\partial}{\partial\theta}\ell_n(x_1,\ldots,x_n;\theta)\right)^2$ sont petites, moins l'*emv* $\theta^*$ de $\theta$ pour $(x_1,\ldots,x_n)$ qui annule cette fonction ne se distingue de ses valeurs voisines. Dans ce contexte, $(x_1,\ldots,x_n)$ ne nous apportent pas assez d'information pour estimer précisément $\theta$.

- A contrario, plus les valeurs de $\left(\frac{\partial}{\partial\theta}\ell_n(x_1,\ldots,x_n;\theta)\right)^2$ sont grandes, plus l'*emv* $\theta^*$ de $\theta$ pour $(x_1,\ldots,x_n)$ qui annule cette fonction se singularise. Dans ce contexte, $(x_1,\ldots,x_n)$ nous apportent suffisamment d'information pour estimer précisément $\theta$.

Ainsi, une mesure de la quantité moyenne d'information apportée par $(X_1,\ldots,X_n)$ sur $\theta$ est $\mathbb{E}\left(\left(\frac{\partial}{\partial\theta}\ell_n(X_1,\ldots,X_n;\theta)\right)^2\right)$. Après calculs, celle-ci correspond à l'information de Fisher $I_n(\theta)$. Cela est en adéquation avec :

- les performances théoriques de l'*emv* $\widehat{\theta}_n$ principalement garanties dans un cadre asymptotique : quand $n$ est assez grand, $I_n(\theta) = nI_1(\theta)$ prend une large étendue de valeurs éloignées de 0,
- la "borne minimale du risque quadratique de $\widehat{\theta}_n$" :

$$R(\widehat{\theta}_n,\theta) \geq \frac{1}{I_n(\theta)}\left(1+\frac{\partial}{\partial\theta}B(\widehat{\theta}_n,\theta)\right)^2 + (B(\widehat{\theta}_n,\theta))^2,$$

qui met en lumière la difficulté d'estimer $\theta$ par $\widehat{\theta}_n$ suivant la grandeur de $I_n(\theta)$ dans un cadre non-asymptotique.

# 9 Intervalles de confiance

### Intervalle de confiance

Soit $\alpha \in ]0,1[$. On appelle intervalle de confiance pour $\theta$ au niveau $100(1-\alpha)\%$ un intervalle de la forme : $\mathcal{IC}_n(X_1,\ldots,X_n) = [a(X_1,\ldots,X_n), b(X_1,\ldots,X_n)]$, où $a(X_1,\ldots,X_n)$ et $b(X_1,\ldots,X_n)$ désignent deux statistiques, tel que

$$\mathbb{P}(\theta \in \mathcal{IC}_n(X_1,\ldots,X_n)) = 1-\alpha.$$

Avec un niveau "au moins de" $100(1-\alpha)\%$, l'intervalle doit vérifier $\mathbb{P}(\theta \in \mathcal{IC}_n(X_1,\ldots,X_n)) \geq 1-\alpha$. Pour que $\mathcal{IC}_n(X_1,\ldots,X_n)$ contienne avec une forte probabilité $\theta$, on prend $\alpha$ petit : $\alpha = 0.05 \ldots$

### Intervalle de confiance obtenu par l'inégalité de Bienaymé-Tchebychev

Soient $\alpha \in ]0,1[$ et $\widehat{\theta}_n$ un estimateur sans biais de $\theta$ tel qu'il existe un réel $a_n$ indépendant de $\theta$ vérifiant $\mathbb{V}(\widehat{\theta}_n) \leq a_n^2$. Alors un intervalle de confiance pour $\theta$ au niveau "au moins de" $100(1-\alpha)\%$ est

$$\mathcal{IC}_n(X_1,\ldots,X_n) = \left[\widehat{\theta}_n - \frac{a_n}{\sqrt{\alpha}}, \widehat{\theta}_n + \frac{a_n}{\sqrt{\alpha}}\right].$$

La preuve repose sur l'inégalité de Bienaymé-Tchebychev.

### Intervalle de confiance asymptotique

Soit $\alpha \in ]0,1[$. On appelle intervalle de confiance asymptotique pour $\theta$ au niveau $100(1-\alpha)\%$ un intervalle de la forme : $\mathcal{IC}_n(X_1,\ldots,X_n) = [a(X_1,\ldots,X_n), b(X_1,\ldots,X_n)]$, où $a(X_1,\ldots,X_n)$ et $b(X_1,\ldots,X_n)$ désignent deux statistiques, tel que

$$\lim_{n\to\infty} \mathbb{P}(\theta \in \mathcal{IC}_n(X_1,\ldots,X_n)) = 1-\alpha.$$

Cela signifie que $\mathcal{IC}_n(X_1,\ldots,X_n)$ est un intervalle de confiance de $\theta$ quand $n$ est assez grand.

Les résultats de convergence en loi permettent la construction de tels intervalles (théorème central limite, méthode Delta...).

### Intervalle de confiance correspondant aux données

Soient $\alpha \in ]0,1[$ et $\mathcal{IC}_n(X_1,\ldots,X_n) = [a(X_1,\ldots,X_n), b(X_1,\ldots,X_n)]$ un intervalle de confiance pour $\theta$ au niveau $100(1-\alpha)\%$. Si l'intervalle de confiance est asymptotique, on suppose que $n$ est assez grand. On appelle intervalle de confiance pour $\theta$ au niveau $100(1-\alpha)\%$ correspondant aux données $(x_1,\ldots,x_n)$ la réalisation de $\mathcal{IC}_n(X_1,\ldots,X_n)$ correspondante aux données :

$$\mathcal{IC}_n(x_1,\ldots,x_n) = [a(x_1,\ldots,x_n), b(x_1,\ldots,x_n)].$$

On obtient alors un intervalle de valeurs dans lequel $\theta$ a $100(1-\alpha)\%$ de chances d'appartenir. Naturellement, plus la longueur de cet intervalle est petite, plus il est précis.

### Intervalle de confiance asymptotique et *emv*

Soient $\alpha \in ]0,1[$ et $\widehat{\theta}_n$ l'*emv* de $\theta$. *Sccri*, un intervalle de confiance asymptotique pour $\theta$ au niveau $100(1-\alpha)\%$ est donné par

$$\mathcal{IC}_n(X_1,\ldots,X_n) = \left[\widehat{\theta}_n - z_\alpha \frac{1}{\sqrt{I_n(\widehat{\theta}_n)}}, \widehat{\theta}_n + z_\alpha \frac{1}{\sqrt{I_n(\widehat{\theta}_n)}}\right],$$

où $z_\alpha$ est le réel tel que $\mathbb{P}(|Z| \geq z_\alpha) = \alpha$ avec $Z$ désignant une *var* suivant la loi normale $\mathcal{N}(0,1)$.

Le réel $z_\alpha$ est aussi le quantile d'ordre $1 - \frac{\alpha}{2}$ de la loi normale $\mathcal{N}(0,1)$, *i.e.* $\mathbb{P}(Z \leq z_\alpha) = 1 - \frac{\alpha}{2}$.

C'est une conséquence du "résultat asymptotique central" de $\widehat{\theta}_n$.

### Intervalle de confiance correspondant aux données et *emv*

Soient $\alpha \in ]0,1[$ et $\theta^*$ l'*emv* de $\theta$ pour $(x_1,\ldots,x_n)$. On suppose que $n$ est assez grand. Alors un intervalle de confiance pour $\theta$ correspondant à $(x_1,\ldots,x_n)$ au niveau $100(1-\alpha)\%$ est donné par

$$\mathcal{IC}_n(x_1,\ldots,x_n) = \left[\theta^* - z_\alpha \frac{1}{\sqrt{I_n(\theta^*)}}, \theta^* + z_\alpha \frac{1}{\sqrt{I_n(\theta^*)}}\right],$$

où $z_\alpha$ est le réel tel que $\mathbb{P}(|Z| \geq z_\alpha) = \alpha$ avec $Z$ désignant une *var* suivant la loi normale $\mathcal{N}(0,1)$.

Pour simplifier, on note $ic_\theta$ cet intervalle ; l'indice $\theta$ précisant que l'intervalle de confiance est pour le paramètre $\theta$.

## 10 Tests de Wald

### Hypothèse statistique

On appelle hypothèse statistique une hypothèse inhérente au contexte de l'étude. À la base d'un test statistique, il y a deux hypothèses complémentaires, notées $H_0$ et $H_1$,

- l'hypothèse $H_0$ formule ce que l'on souhaite rejeter/réfuter,
- l'hypothèse $H_1$ formule ce que l'on souhaite montrer.

### Risque de première espèce

On appelle risque de première espèce le pourcentage de chances de rejeter $H_0$, donc d'accepter $H_1$, alors que $H_0$ est vraie. Il s'écrit sous la forme : $100\alpha\%$, avec

$$\alpha = \mathbb{P}(\text{rejeter } H_0 \text{ alors que } H_0 \text{ est vraie}).$$

Le rejet de $H_0$ est dit "significatif" si elle est rejetée au risque 5%.

### Test statistique

On appelle test statistique une démarche/procédure qui vise à apporter une réponse à la question : est-ce que les données nous permettent de rejeter $H_0$/accepter $H_1$ avec un faible risque de se tromper ?

### Types de test statistique sur un paramètre

Lorsque le test statistique porte sur un paramètre inconnu $\theta$, on dit que le test est

- bilatéral si $H_1$ est de la forme $H_1 : \theta \neq \ldots$
- unilatéral à gauche (sens de $<$) si $H_1$ est de la forme $H_1 : \theta < \ldots$
- unilatéral à droite (sens de $>$) si $H_1$ est de la forme $H_1 : \theta > \ldots$

### p-valeur

On appelle p-valeur le plus petit réel $\alpha \in ]0,1[$ calculé à partir des données tel que l'on puisse se permettre de rejeter $H_0$ au risque $100\alpha\%$. Les logiciels actuels affichent cette p-valeur.

### Degré de significativité

La p-valeur nous donne un degré de significativité du rejet de $H_0$. Le rejet de $H_0$ est dit :

- significatif si p-valeur $\in ]0.01,\ 0.05]$, symbolisé par $\star$,
- très significatif si p-valeur $\in ]0.001,\ 0.01]$, symbolisé par $\star\star$,

- ○ hautement significatif si p-valeur $< 0.001$, symbolisé par $\star\star\star$.

Il y a non rejet de $H_0$ si p-valeur $> 0.05$.

### En cas de non-rejet de $H_0$

S'il y a non-rejet de $H_0$, sauf convention, on ne peut rien conclure du tout (avec le risque considéré). Éventuellement, on peut dire que $H_0$ est plausible (elle "semble pouvoir être admise"). En revanche, peut-être qu'un risque de départ plus élevé ou la disposition de plus de données peuvent conduire à un rejet de $H_0$.

### Tests statistiques et *emv* : test de Wald

Soient $\theta_0$ un réel fixé. On considère les hypothèses :

| Hypothèses | $H_0$ | $H_1$ |
|---|---|---|
| bilatérale | $\theta = \theta_0$ | $\theta \neq \theta_0$ |
| unilatérale à droite | $\theta \leq \theta_0$ | $\theta > \theta_0$ |
| unilatérale à gauche | $\theta \geq \theta_0$ | $\theta < \theta_0$ |

On suppose que $n$ est assez grand. Soient $\theta^*$ l'*emv* de $\theta$ pour $(x_1, \ldots, x_n)$ et

$$z_{obs} = \sqrt{I_n(\theta^*)}(\theta^* - \theta_0).$$

On considère une *var* $Z$ suivant la loi normale $\mathcal{N}(0,1)$. Alors les p-valeurs associées aux hypothèses considérées sont :

| $H_0$ | $H_1$ | p-valeurs | (p-valeurs et fonction de répartition de $Z$) |
|---|---|---|---|
| $\theta = \theta_0$ | $\theta \neq \theta_0$ | $\mathbb{P}(\lvert Z\rvert \geq \lvert z_{obs}\rvert)$ | $2(1 - \mathbb{P}(Z \leq \lvert z_{obs}\rvert))$ |
| $\theta \leq \theta_0$ | $\theta > \theta_0$ | $\mathbb{P}(Z \geq z_{obs})$ | $1 - \mathbb{P}(Z \leq \lvert z_{obs}\rvert)$ |
| $\theta \geq \theta_0$ | $\theta < \theta_0$ | $\mathbb{P}(Z \leq z_{obs})$ | $\mathbb{P}(Z \leq \lvert z_{obs}\rvert)$ |

Les définitions de ces p-valeurs reposent sur le "résultat asymptotique central" de $\widehat{\theta}_n$ et la théorie des tests asymptotiques.

## 11 Complément : Estimation d'une fonction de $\theta$

### Résultat de forte consistance

Soient $\widehat{\theta}_n$ un estimateur de $\theta$ fortement consistant et $h$ une fonction. Alors $h(\widehat{\theta}_n)$ est estimateur de $h(\theta)$ fortement consistant ; c'est une conséquence du théorème de continuité.

### Borne minimale du risque quadratique : généralisation

Soit $h$ une fonction. *Sccri*, pour tout estimateur $\widetilde{\theta}_n$ de $h(\theta)$, on a

$$R(\widetilde{\theta}_n, h(\theta)) \geq \frac{1}{I_n(\theta)} \left( h'(\theta) + \frac{\partial}{\partial \theta} B(\widetilde{\theta}_n, h(\theta)) \right)^2 + (B(\widetilde{\theta}_n, h(\theta)))^2.$$

### Estimateur efficace de $h(\theta)$

Soit $h$ une fonction. *Sccri*, on dit qu'un estimateur $\widetilde{\theta}_n$ de $h(\theta)$ est efficace si et seulement si

- il est sans biais : $\mathbb{E}(\widetilde{\theta}_n) = h(\theta)$,
- on a :
$$\mathbb{V}(\widetilde{\theta}_n) = \frac{(h'(\theta))^2}{I_n(\theta)}.$$

Ainsi, la borne minimale du risque quadratique est atteinte.

### Résultat de normalité asymptotique

Soit $\widehat{\theta}_n$ un estimateur de $\theta$. Supposons qu'il existe une suite de réels $(r_n)_{n \in \mathbb{N}^*}$ vérifiant $\lim_{n \to \infty} r_n = \infty$, et un réel $\sigma > 0$ tels que la suite de *var* $\left( r_n \left( \frac{\widehat{\theta}_n - \theta}{\sigma} \right) \right)_{n \in \mathbb{N}^*}$ converge en loi vers une *var* $Z$ suivant la loi normale $\mathcal{N}(0,1)$. Soit $h$ une fonction. Alors, sous certaines conditions, l'estimateur $h(\widehat{\theta}_n)$ de $h(\theta)$ est asymptotiquement normal.

Plus précisément, en posant $a_n(\theta) = h(\theta)$ et $b_n(\theta) = \frac{\sigma h'(\theta)}{r_n}$, la suite de *var*

$$\left( \frac{h(\widehat{\theta}_n) - a_n(\theta)}{b_n(\theta)} \right)_{n \in \mathbb{N}^*} = \left( r_n \left( \frac{h(\widehat{\theta}_n) - h(\theta)}{\sigma h'(\theta)} \right) \right)_{n \in \mathbb{N}^*}$$

converge en loi vers une *var* $Z$ suivant la loi normale $\mathcal{N}(0,1)$ ; c'est une application de la méthode Delta.

### Théorème général sur l'efficacité

*Sccri*, s'il existe un réel $a_n(\theta)$ dépendant de $\theta$, une fonction $h$ et un estimateur $\widetilde{\theta}_n$ de $h(\theta)$ tels que :

$$\frac{\partial}{\partial\theta}\ell_n(X_1,\ldots,X_n;\theta) = a_n(\theta)(\widetilde{\theta}_n - h(\theta)),$$

alors $\widetilde{\theta}_n$ est un estimateur efficace de $h(\theta)$.

En particulier et plus,

- $\widetilde{\theta}_n$ est un estimateur sans biais : $\mathbb{E}(\widetilde{\theta}_n) = h(\theta)$,
- on a
$$\mathbb{V}(\widetilde{\theta}_n) = \frac{h'(\theta)}{a_n(\theta)}.$$
- on a
$$I_n(\theta) = \frac{(h'(\theta))^2}{\mathbb{V}(\widetilde{\theta}_n)}.$$

### Résultat sur l'*emv*

Soit $h$ une fonction. Si $\widehat{\theta}_n$ est l'*emv* de $\theta$, alors $h(\widehat{\theta}_n)$ est l'*emv* de $h(\theta)$.

# 12 Complément : Cas de multiples paramètres

### Définitions et notations ; ce qui change

On suppose que $\theta$ symbolise $p$ paramètres inconnus : c'est le vecteur colonne $\theta = (\theta_1, \ldots, \theta_p)^t \in \mathbb{R}^p$.

Ce qui ne change pas :

- La définition de la fonction de vraisemblance.
- La définition de la fonction de log-vraisemblance.
- La définition des estimateurs du maximum de vraisemblance.

Ce qui change : Certaines quantités et inégalités s'adaptent avec l'utilisation de matrices. Les plus importantes d'entre elles sont décrites ci-dessous.

### Quelques notations matricielles

Partant d'un vecteur aléatoire colonne à $p$ composantes, on note $\mathbb{E}_p$ son espérance et $\mathbb{C}_{p\times p}$ sa matrice de covariance.

Un estimateur de $\theta = (\theta_1, \ldots, \theta_p)^t$ se note $\widehat{\theta}_n = \left(\widehat{(\theta_1)}_n, \ldots, \widehat{(\theta_p)}_n\right)^t$. Ainsi, pour tout $j \in \{1, \ldots, p\}$, $\widehat{(\theta_j)}_n$ est un estimateur de $\theta_j$.

Le biais de $\widehat{\theta}_n$ devient le vecteur colonne a $p$ composantes :

$B_p(\widehat{\theta}_n, \theta) = \mathbb{E}_p(\widehat{\theta}_n) - \theta = \left(\mathbb{E}(\widehat{(\theta_1)}_n) - \theta_1, \ldots, \mathbb{E}(\widehat{(\theta_p)}_n) - \theta_p\right)^t$.

Le risque de $\widehat{\theta}_n$ devient le réel : $R_p(\widehat{\theta}_n, \theta) = \mathbb{E}(\|\widehat{\theta}_n - \theta\|^2)$, où $\|.\|$ désigne la norme euclidienne : $\|\widehat{\theta}_n - \theta\|^2 = (\widehat{\theta}_n - \theta)^t(\widehat{\theta}_n - \theta)$. On a $R_p(\widehat{\theta}_n, \theta) = \mathrm{trace}(\mathbb{C}_{p\times p}(\widehat{\theta}_n)) + \|B_p(\widehat{\theta}_n, \theta)\|^2$.

La dérivée partielle première de $\ell_n(x_1, \ldots, x_n; \theta)$ en $\theta$ est un vecteur à $p$ composantes :

$$\frac{\partial}{\partial\theta}\ell_n(x_1, \ldots, x_n; \theta) = \left(\frac{\partial}{\partial\theta_1}\ell_n(x_1, \ldots, x_n; \theta), \ldots, \frac{\partial}{\partial\theta_p}\ell_n(x_1, \ldots, x_n; \theta)\right)^t.$$

La dérivée partielle seconde de $\ell_n(x_1, \ldots, x_n; \theta)$ en $\theta$ est une matrice carrée à $p$ lignes et $p$ colonnes :

$$\frac{\partial^2}{\partial\theta^2}\ell_n(x_1, \ldots, x_n; \theta) = \left(\frac{\partial^2}{\partial\theta_j\partial\theta_k}\ell_n(x_1, \ldots, x_n; \theta)\right)_{(j,k)\in\{1,\ldots,p\}^2}.$$

L'information de Fisher est une matrice carrée à $p$ lignes et $p$ colonnes :

$$\begin{aligned} I_n(\theta) &= \mathbb{E}_p\left(\frac{\partial}{\partial\theta}\ell_n(X_1,\ldots,X_n;\theta)\left(\frac{\partial}{\partial\theta}\ell_n(X_1,\ldots,X_n;\theta)\right)^t\right) = \mathbb{C}_{p,p}\left(\frac{\partial}{\partial\theta}\ell_n(X_1,\ldots,X_n;\theta)\right) \\ &= -\mathbb{E}_{p,p}\left(\frac{\partial^2}{\partial\theta^2}\ell_n(X_1,\ldots,X_n;\theta)\right) = \left(-\mathbb{E}\left(\frac{\partial^2}{\partial\theta_j\partial\theta_k}\ell_n(X_1,\ldots,X_n;\theta)\right)\right)_{(j,k)\in\{1,\ldots,p\}^2}. \end{aligned}$$

La borne de Cramer-Rao devient : *Sccri*, pour tout estimateur $\widehat{\theta}_n$ de $\theta$ sans biais, on a

$$\mathbb{C}_{p,p}(\widehat{\theta}_n) \geq I_n^{-1}(\theta),$$

où $I_n^{-1}(\theta)$ désigne l'inverse de la matrice $I_n(\theta)$ et le symbole $\geq$ désigne une inégalité matricielle : pour toutes matrices $A$ et $B$ à $p$ lignes et $p$ colonnes, $A \geq B$ signifie que la matrice $A - B$ est positive, *i.e.* pour tout vecteur colonne $v$ à $p$ composantes réels, $v^t(A-B)v \geq 0$. Ainsi, pour tout vecteur colonne $v$ à $p$ composantes réels, $v^t\left(\mathbb{C}_{p,p}(\widehat{\theta}_n) - I_n^{-1}(\theta)\right)v \geq 0$. En particulier, pour tout $j \in \{1,\ldots,p\}$, on a

$$R\left(\widehat{(\theta_j)}_n, \theta_j\right) \geq [I_n^{-1}(\theta)]_{j,j},$$

où $[I_n^{-1}(\theta)]_{j,j}$ désigne la $j$-ème composante diagonale de $I_n^{-1}(\theta)$.

**Lois approchées de certaines *var* fonction de l'*emv***

Soit $\theta = (\theta_1,\ldots,\theta_p)^t \in \mathbb{R}^p$ et $\widehat{\theta}_n$ l'*emv* de $\theta$. Quand $n$ est assez grand, *sccri*,

- la loi approchée de $\widehat{\theta}_n$ est la loi normale multivariée $\mathcal{N}_p\left(\theta, I_n(\theta)^{-1}\right)$,
- la loi approchée de $(\widehat{\theta}_n - \theta)^t I_n(\theta)(\widehat{\theta}_n - \theta)$ est la loi du Chi-deux $\chi^2(p)$,
- la loi approchée de $-2\left(\ell_n(X_1,\ldots,X_n,\theta) - \ell_n(X_1,\ldots,X_n,\widehat{\theta}_n)\right)$ est la loi du Chi-deux $\chi^2(p)$.

**Tests globaux**

On peut mettre en place des tests d'hypothèses portant sur la comparaison simultanée entre des paramètres inconnus et des valeurs. Soit $\theta = (\theta_1,\ldots,\theta_p)^t \in \mathbb{R}^p$ et $\theta_0$ un vecteur colonne à $p$ composantes fixé. On considère les hypothèses :

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

On suppose que $n$ est assez grand. Soient $\theta^*$ l'*emv* de $\theta$ pour $(x_1,\ldots,x_n)$.

On peut alors décider du rejet de $H_0$ avec :

- le test de Wald : on calcule

$$\chi^2_{obs} = (\theta^* - \theta_0)^t I_n(\theta_0)(\theta^* - \theta_0).$$

  On considère une *var* $K$ suivant la loi du Chi-deux $\chi^2(p)$. Alors la p-valeurs associée est

$$\text{p-valeur} = \mathbb{P}(K \geq \chi^2_{obs}).$$

- le test de rapport de vraisemblance : on calcule

$$\chi^2_{obs} = -2\left(\ell_n(x_1, \ldots, x_n, \theta_0) - \ell_n(x_1, \ldots, x_n, \theta^*)\right).$$

  On considère une *var* $K$ suivant la loi du Chi-deux $\chi^2(p)$. Alors la p-valeurs associée est

$$\text{p-valeur} = \mathbb{P}(K \geq \chi^2_{obs}).$$

Le test de rapport de vraisemblance est très utilisé dans de nombreuses modélisations avec $\theta_0 = (0, \ldots, 0)^t$.

# 13 Complément : Définitions générales

### Fonction de vraisemblance et *emv* : généralité

En toute généralité, pour tout vecteur aléatoire réel $(Y_1, \ldots, Y_n)$ dont la loi dépend d'un paramètre inconnu $\theta$ et $(y_1, \ldots, y_n)$ une réalisation de $(Y_1, \ldots, Y_n)$, on appelle fonction de vraisemblance pour $(y_1, \ldots, y_n)$ la fonction de $\theta$ :

$$L_n(y_1, \ldots, y_n; \theta) = \begin{cases} \mathbb{P}_\theta((Y_1, \ldots, Y_n) = (y_1, \ldots, y_n)), & \text{si } Y_1, \ldots, Y_n \text{ sont discrètes,} \\ f_\theta(y_1, \ldots, y_n), & \text{une densité de } (Y_1, \ldots, Y_n), \text{ si } Y_1, \ldots, Y_n \text{ sont à densité.} \end{cases}$$

Ainsi, les *var* $Y_1, \ldots, Y_n$ ne sont pas nécessairement *iid*. Une version "loi conditionnelle" existe aussi. Ces fonctions de vraisemblance sont utilisées dans le cadre de l'estimation paramétrique pour des modélisations avancées (modèles de régression, séries temporelles. . . ).

On appelle estimateur du maximum de vraisemblance de $\theta$ *(emv)* pour $(y_1, \ldots, y_n)$ un réel $\theta^*$ qui maximise la fonction de vraisemblance $L_n(y_1, \ldots, y_n; \theta)$ en $\theta$, *i.e.*

$$\theta^* \in \operatorname*{argmax}_\theta L_n(y_1, \ldots, y_n; \theta).$$

### Information de Fisher : généralité

En toute généralité, pour tout vecteur aléatoire réel $(Y_1, \ldots, Y_n)$ dont la loi dépend d'un paramètre inconnu $\theta$, on appelle Information de Fisher le réel :

$$I_n(\theta) = \mathbb{E}\left(\left(\frac{\partial}{\partial \theta}\ell_n(Y_1, \ldots, Y_n; \theta)\right)^2\right).$$

Ainsi, les *var* $Y_1, \ldots, Y_n$ ne sont pas nécessairement *iid*.

# 14 Exercices

**Exercice 1.** Soient $n \in \mathbb{N}^*$ et $x_1, \ldots, x_n$ $n$ observations d'un caractère pouvant être modélisé par une *var* $X$ suivant la loi de Poisson $\mathcal{P}(\lambda)$ avec $\lambda > 0$ :

$$\mathbb{P}_\lambda(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}, \qquad k \in \mathbb{N}.$$

Le paramètre $\lambda$ est inconnu. Soit $(X_1, \ldots, X_n)$ un $n$-échantillon de $X$.

1. Déterminer la fonction de vraisemblance pour $(x_1, \ldots, x_n)$.
2. Déterminer la fonction de log-vraisemblance pour $(x_1, \ldots, x_n)$.
3. Déterminer un *emv* $\lambda^*$ de $\lambda$ pour $(x_1, \ldots, x_n)$, puis un *emv* $\widehat{\lambda}_n$ de $\lambda$.
4. Calculer $\mathbb{E}(\widehat{\lambda}_n)$, $\mathbb{V}(\widehat{\lambda}_n)$ et $R(\widehat{\lambda}_n, \lambda)$. Vérifier que $\widehat{\lambda}_n$ est fortement consistant.
5. Calculer l'information de Fisher fournit par $(X_1, \ldots, X_n)$ sur $\lambda$.
6. Est-ce que $\widehat{\lambda}_n$ est efficace ?
7. On suppose désormais que $n = 3500$ (soit "assez grand") et que $\frac{1}{n}\sum_{i=1}^{n} x_i = 3.50$.
   (a) Proposer un intervalle de confiance pour $\lambda$ correspondant à $(x_1, \ldots, x_n)$ au niveau 95%.
   (b) On considère les hypothèses :
   $$H_0 : \ \lambda = 3.42 \qquad \text{contre} \qquad H_1 : \ \lambda \neq 3.42.$$
   Peut-on rejeter $H_0$ au risque 5% ?

**Exercice 2.** Soient $n \in \mathbb{N}^*$ et $x_1, \ldots, x_n$ $n$ observations d'un caractère pouvant être modélisé par une *var* $X$ suivant la loi de Bernoulli $\mathcal{B}(\theta)$ avec $\theta \in ]0, 1[$ :

$$\mathbb{P}_\lambda(X = 0) = 1 - \theta, \qquad \mathbb{P}_\lambda(X = 1) = \theta.$$

Le paramètre $\theta$ est inconnu.

Déterminer un *emv* $\theta^*$ de $\theta$ pour $(x_1, \ldots, x_n)$,

**Exercice 3.** Soient $n \in \mathbb{N}^*$ et $x_1, \ldots, x_n$ $n$ observations d'un caractère pouvant être modélisé par une *var* $X$ de densité :

$$f_\theta(x) = \begin{cases} \dfrac{\theta}{1-\theta} x^{\frac{2\theta-1}{1-\theta}} & \text{si } x \in [0,1], \\ 0 & \text{sinon.} \end{cases}$$

Le paramètre $\theta$ est inconnu avec $\theta \in \left]\frac{1}{2}, 1\right[$.

Déterminer un *emv* $\theta^*$ de $\theta$ pour $(x_1, \ldots, x_n)$,

**Exercice 4.** Soient $n \in \mathbb{N}^*$ et $x_1, \ldots, x_n$ $n$ observations d'un caractère pouvant être modélisé par une *var* $X$ de densité :

$$f_\theta(x) = \begin{cases} \dfrac{\theta+1}{2}(1-|x|)^\theta & \text{si } x \in [-1,1], \\ 0 & \text{sinon.} \end{cases}$$

Le paramètre $\theta$ est inconnu avec $\theta > -1$.

Déterminer un *emv* $\theta^*$ de $\theta$ pour $(x_1, \ldots, x_n)$,

**Exercice 5.** Soient $\lambda > 0$, $a \in \mathbb{R}^*$ et $X$ une *var* suivant la loi de Poisson $\mathcal{P}(\lambda^a)$ :

$$\mathbb{P}_\lambda(X = k) = e^{-\lambda^a} \frac{\lambda^{ak}}{k!}, \qquad k \in \mathbb{N}.$$

Le paramètre $a$ est connu et $\lambda$ est inconnu. Soient $n \in \mathbb{N}^*$ et $(X_1, \ldots, X_n)$ un $n$-échantillon de $X$.

Déterminer l'information de Fisher fournit par $(X_1, \ldots, X_n)$ sur $\lambda$.

**Exercice 6.** Soient $\theta > -1$ et $X$ une *var* de densité :

$$f_\theta(x) = \begin{cases} \dfrac{1+\theta}{(x+\theta)^2} & \text{si } x \geq 1, \\ 0 & \text{sinon.} \end{cases}$$

Le paramètre $\theta$ est inconnu. Soient $n \in \mathbb{N}^*$ et $(X_1, \ldots, X_n)$ un $n$-échantillon de $X$.

Déterminer l'information de Fisher fournit par $(X_1, \ldots, X_n)$ sur $\theta$.

**Exercice 7.** Soient $n \in \mathbb{N}^*$ et $x_1, \ldots, x_n$ $n$ observations d'un caractère pouvant être modélisé par une *var* $X$ suivant la loi exponentielle $\mathcal{E}\left(\frac{1}{\theta}\right)$ avec $\theta > 0$ :

$$f_\theta(x) = \begin{cases} \dfrac{1}{\theta} e^{-\frac{1}{\theta}x} & \text{si } x \geq 0, \\ 0 & \text{sinon.} \end{cases}$$

Le paramètre $\theta$ est inconnu. Soit $(X_1, \ldots, X_n)$ un $n$-échantillon de $X$.

1. Déterminer la fonction de vraisemblance pour $(x_1, \ldots, x_n)$.
2. Déterminer la fonction de log-vraisemblance pour $(x_1, \ldots, x_n)$.
3. Déterminer un *emv* $\theta^*$ de $\theta$ pour $(x_1, \ldots, x_n)$, puis un *emv* $\widehat{\theta}_n$ de $\theta$.
4. Calculer $\mathbb{E}(\widehat{\theta}_n)$, $\mathbb{V}(\widehat{\theta}_n)$ et $R(\widehat{\theta}_n, \theta)$. Vérifier que $\widehat{\theta}_n$ est fortement consistant.
5. Calculer l'information de Fisher fournit par $(X_1, \ldots, X_n)$ sur $\theta$.
6. Est-ce que $\widehat{\theta}_n$ est efficace ?
7. Déterminer :
$$\lim_{n \to \infty} \mathbb{P}\left(\theta \in \left[\widehat{\theta}_n - 1.96\frac{\widehat{\theta}_n}{\sqrt{n}}, \widehat{\theta}_n + 1.96\frac{\widehat{\theta}_n}{\sqrt{n}}\right]\right).$$

**Exercice 8.** Soient $n \in \mathbb{N}^*$ et $x_1, \ldots, x_n$ $n$ observations d'un caractère pouvant être modélisé par une *var* $X$ suivant la loi uniforme $\mathcal{U}([0, \theta])$ avec $\theta > 0$ :

$$f_\theta(x) = \begin{cases} \dfrac{1}{\theta} & \text{si } x \in [0, \theta], \\ 0 & \text{sinon.} \end{cases}$$

Le paramètre $\theta$ est inconnu. Soit $(X_1, \ldots, X_n)$ un $n$-échantillon de $X$.

1. Déterminer un *emv* $\widehat{\theta}_n$ de $\theta$.
2. Déterminer la fonction de répartition de $\widehat{\theta}_n$, puis une densité.
3. Calculer le biais de $\widehat{\theta}_n$. Est-il sans biais ? Est-il asymptotiquement sans biais ?
4. Calculer $R(\widehat{\theta}_n, \theta)$. Vérifier que $\widehat{\theta}_n$ est fortement consistant.

**Exercice 9.** Soient $n \in \mathbb{N}^*$ et $x_1, \ldots, x_n$ $n$ observations d'un caractère pouvant être modélisé par une *var* $X$ de densité :

$$f_\theta(x) = \begin{cases} \dfrac{3}{(x-\theta)^4} & \text{si } x \geq 1+\theta, \\ 0 & \text{sinon.} \end{cases}$$

Le paramètre $\theta$ est inconnu avec $\theta > 0$. Soit $(X_1, \ldots, X_n)$ un $n$-échantillon de $X$.

1. Calculer $\mathbb{E}(X-\theta)$ et $\mathbb{E}((X-\theta)^2)$. En déduire $\mathbb{E}(X)$ et $\mathbb{V}(X)$.
2. Déterminer un *emv* $\widehat{\theta}_n$ de $\theta$.
3. Déterminer la fonction de répartition de $\widehat{\theta}_n - \theta$, puis une densité.
4. Calculer le biais de $\widehat{\theta}_n$. Est-il sans biais ? Est-il asymptotiquement sans biais ?
5. Calculer $R(\widehat{\theta}_n, \theta)$.
6. En utilisant $\widehat{\theta}_n$, construire un estimateur $\widehat{\theta}_n^*$ sans biais de $\theta$. Est-ce que $\widehat{\theta}_n^*$ est préférable à $\widehat{\theta}_n$ ?

**Exercice 10.** Soient $n \in \mathbb{N}^*$ et $x_1, \ldots, x_n$ $n$ observations d'un caractère pouvant être modélisé par une *var* $X$ de densité :

$$f_\theta(x) = \begin{cases} \theta x^{-(\theta+1)} & \text{si } x \geq 1, \\ 0 & \text{sinon.} \end{cases}$$

Le paramètre $\theta$ est inconnu avec $\theta > 1$. Soit $(X_1, \ldots, X_n)$ un $n$-échantillon de $X$. On pose

$$T_n = \frac{1}{n} \sum_{i=1}^{n} \ln(X_i).$$

1. Utiliser le théorème général sur l'efficacité pour montrer que $T_n$ est un estimateur efficace de $\dfrac{1}{\theta}$, puis calculer $\mathbb{E}(T_n)$, $\mathbb{V}(T_n)$ et $I_n(\theta)$.
2. Étudier la convergence en loi de $(\sqrt{n}(\theta T_n - 1))_{n \in \mathbb{N}^*}$.

**Exercice 11.** Soient $n \in \mathbb{N}^*$ et $x_1, \ldots, x_n$ $n$ observations d'un caractère pouvant être modélisé par une *var* $X$ de densité :

$$f_\theta(x) = \begin{cases} \dfrac{\theta+1}{2}(1-|x|)^\theta & \text{si } x \in [-1, 1], \\ 0 & \text{sinon.} \end{cases}$$

Le paramètre $\theta$ est inconnu avec $\theta > 0$. Soit $(X_1, \ldots, X_n)$ un $n$-échantillon de $X$. On pose

$$T_n = \frac{1}{n}\sum_{i=1}^{n} \ln(1 - |X_i|).$$

1. Utiliser le théorème général sur l'efficacité pour montrer que $T_n$ est un estimateur efficace de $-\frac{1}{\theta+1}$, puis calculer $\mathbb{E}(T_n)$, $\mathbb{V}(T_n)$ et $I_n(\theta)$.
2. Étudier la convergence en loi de $\left((\theta+1)\sqrt{n}\left(T_n + \frac{1}{\theta+1}\right)\right)_{n\in\mathbb{N}^*}$.

**Exercice 12.** Soient $n \in \mathbb{N}^*$ et $x_1, \ldots, x_n$ $n$ observations d'un caractère pouvant être modélisé par une *var* $X$ de densité :

$$f_\theta(x) = \begin{cases} e^{-(x-\theta)} & \text{si } x \geq \theta, \\ 0 & \text{sinon.} \end{cases}$$

Le paramètre $\theta$ est inconnu avec $\theta \in \mathbb{R}$.

Déterminer un *emv* $\theta^*$ de $\theta$ pour $(x_1, \ldots, x_n)$.

**Exercice 13.** Soient $n \in \mathbb{N}^*$ et $x_1, \ldots, x_n$ $n$ observations d'un caractère pouvant être modélisé par une *var* $X$ suivant la loi uniforme $\mathcal{U}([\theta, \theta+1])$ avec $\theta \in \mathbb{R}$, *i.e.* de densité :

$$f_\theta(x) = \begin{cases} 1 & \text{si } x \in [\theta, \theta+1], \\ 0 & \text{sinon.} \end{cases}$$

Le paramètre $\theta$ est inconnu.

Montrer que tout point de l'intervalle $[\sup(x_1, \ldots, x_n) - 1, \inf(x_1, \ldots, x_n)]$ est un *emv* $\theta^*$ de $\theta$ pour $(x_1, \ldots, x_n)$.

# 15 Solutions

## Solution 1.

1. La fonction de vraisemblance pour $(x_1, \ldots, x_n)$ est

$$L_n(x_1, \ldots, x_n; \lambda) = \prod_{i=1}^{n} \mathbb{P}_\lambda(X = x_i) = \prod_{i=1}^{n} \left( \frac{\lambda^{x_i}}{x_i!} e^{-\lambda} \right) = e^{-n\lambda} \frac{\lambda^{\sum\limits_{i=1}^{n} x_i}}{\prod\limits_{i=1}^{n} (x_i!)}.$$

2. La fonction de log-vraisemblance pour $(x_1, \ldots, x_n)$ est

$$\begin{aligned} \ell_n(x_1, \ldots, x_n; \lambda) &= \ln(L_n(x_1, \ldots, x_n; \lambda)) = \ln \left( e^{-n\lambda} \frac{\lambda^{\sum\limits_{i=1}^{n} x_i}}{\prod\limits_{i=1}^{n} (x_i!)} \right) \\ &= -n\lambda + \ln(\lambda) \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \ln(x_i!) . \end{aligned}$$

3. Un *emv* $\lambda^*$ de $\lambda$ pour $(x_1, \ldots, x_n)$ est défini par

$$\lambda^* \in \operatorname*{argmax}_{\lambda > 0} L_n(x_1, \ldots, x_n; \lambda) = \operatorname*{argmax}_{\lambda > 0} \ell_n(x_1, \ldots, x_n; \lambda).$$

Par le résultat de la question 2, $\ell_n(x_1, \ldots, x_n; \lambda)$ est dérivable en $\lambda$ avec $\lambda > 0$. Par conséquent, $\lambda^*$ est solution de l'équation de vraisemblance : $\frac{\partial}{\partial \lambda} \ell_n(x_1, \ldots, x_n; \lambda) = 0$. Il vient

$$\frac{\partial}{\partial \lambda} \ell_n(x_1, \ldots, x_n; \lambda^*) = 0 \qquad \Leftrightarrow \qquad -n + \frac{1}{\lambda^*} \sum_{i=1}^{n} x_i = 0 \qquad \Leftrightarrow \qquad \lambda^* = \frac{1}{n} \sum_{i=1}^{n} x_i.$$

De plus, $\lambda^*$ est bien un maximum pour $\ell_n(x_1, \ldots, x_n; \lambda)$ (et aussi pour $L_n(x_1, \ldots, x_n; \lambda)$) car :

- si $\lambda < \lambda^*$, alors $\frac{\partial}{\partial \lambda} \ell_n(x_1, \ldots, x_n; \lambda) > 0$,
- si $\lambda > \lambda^*$, alors $\frac{\partial}{\partial \lambda} \ell_n(x_1, \ldots, x_n; \lambda) < 0$.

Donc $\lambda^* = \frac{1}{n}\sum_{i=1}^{n} x_i$ est bien un *emv* de $\lambda$ pour $(x_1, \ldots, x_n)$. L'*emv* $\widehat{\lambda}_n$ de $\lambda$ associé est

$$\widehat{\lambda}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

4. La loi commune des *var* $X_1, \ldots, X_n$ et l'égalité $\mathbb{E}(X_1) = \lambda$ entraînent

$$\mathbb{E}(\widehat{\lambda}_n) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}(X_i) = \frac{1}{n} n\mathbb{E}(X_1) = \lambda.$$

On en déduit que $\widehat{\lambda}_n$ est un estimateur sans biais de $\lambda$.

L'indépendance et la loi commune des *var* $X_1, \ldots, X_n$, et l'égalité $\mathbb{V}(X_1) = \lambda$ impliquent

$$\mathbb{V}(\widehat{\lambda}_n) = \mathbb{V}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\mathbb{V}\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} \mathbb{V}(X_i) = \frac{1}{n^2} n\mathbb{V}(X_1) = \frac{1}{n}\lambda.$$

Comme $\widehat{\lambda}_n$ est un estimateur sans biais de $\lambda$, on a

$$R(\widehat{\lambda}_n, \lambda) = \mathbb{V}(\widehat{\lambda}_n) = \frac{\lambda}{n}.$$

Comme les *var* $X_1, \ldots, X_n$ sont *iid* et admettent un moment d'ordre 2, la loi forte des grands nombres nous assure que $(\widehat{\lambda}_n)_{n\in\mathbb{N}^*} = \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)_{n\in\mathbb{N}^*}$ converge presque sûrement vers $\mathbb{E}(X_1) = \lambda$. Ainsi, $\widehat{\lambda}_n$ est fortement consistant.

5. L'information de Fisher de $X$ sur $\lambda$ est définie par

$$I_n(\lambda) = nI_1(\lambda), \qquad I_1(\lambda) = -\mathbb{E}\left(\frac{\partial^2}{\partial\lambda^2}\ln(\mathbb{P}_\lambda(X = X_1))\right),$$

l'espérance étant prise par rapport à la *var* $X_1$. Calculons $I_1(\lambda)$.

On a

$$\ln(\mathbb{P}_\lambda(X = x)) = \ln\left(e^{-\lambda}\frac{\lambda^x}{x!}\right) = -\lambda + \ln(\lambda)x - \ln(x!),$$

donc

$$\frac{\partial}{\partial\lambda}\ln(\mathbb{P}_\lambda(X = x)) = -1 + \frac{1}{\lambda}x, \qquad \frac{\partial^2}{\partial\lambda^2}\ln(\mathbb{P}_\lambda(X = x)) = -\frac{1}{\lambda^2}x.$$

Comme $\mathbb{E}(X_1) = \lambda$, on obtient

$$I_1(\lambda) = -\mathbb{E}\left(\frac{\partial^2}{\partial \lambda^2} \ln(\mathbb{P}_\lambda(X = X_1))\right) = -\mathbb{E}\left(-\frac{1}{\lambda^2} X_1\right) = \frac{1}{\lambda^2}\mathbb{E}(X_1) = \frac{1}{\lambda}.$$

On en déduit que

$$I_n(\lambda) = nI_1(\lambda) = \frac{n}{\lambda}.$$

6. Par les résultats des questions 4 et 5, $\widehat{\lambda}_n$ est un estimateur de $\lambda$ :
   - sans biais : $\mathbb{E}(\widehat{\lambda}_n) = \lambda$,
   - on a $\mathbb{V}(\widehat{\lambda}_n) = \dfrac{\lambda}{n} = \dfrac{1}{I_n(\lambda)}$.

   On en déduit que $\widehat{\lambda}_n$ est un estimateur efficace de $\lambda$.

7. (a) Nous allons utiliser l'intervalle de confiance asymptotique associé à l'*emv* $\widehat{\lambda}_n$. On a $n$ "assez grand", un *emv* de $\lambda$ pour $(x_1, \ldots, x_n)$ est $\lambda^* = \dfrac{1}{n}\sum_{i=1}^{n} x_i = 3.50$ et, par le résultat de la question 5, $I_n(\lambda^*) = \dfrac{n}{\lambda^*} = \dfrac{3500}{3.50} = 1000$. On a $95\% = 100(1-\alpha)\%$ avec $\alpha = 0.95$ et, en posant $z_\alpha = 1.96$, on a $\mathbb{P}(|Z| \geq z_\alpha) = \alpha = 0.05$ avec $Z$ désignant une *var* suivant la loi normale $\mathcal{N}(0,1)$. Par conséquent, un intervalle de confiance pour $\lambda$ correspondant à $(x_1, \ldots, x_n)$ au niveau 95% est donné par

$$\begin{aligned} ic_\lambda &= \left[\lambda^* - z_\alpha \frac{1}{\sqrt{I_n(\lambda^*)}}, \lambda^* + z_\alpha \frac{1}{\sqrt{I_n(\lambda^*)}}\right] \\ &= \left[3.50 - 1.96 \times \frac{1}{\sqrt{1000}}, 3.50 + 1.96 \times \frac{1}{\sqrt{1000}},\right] = [3.438019, 3.5061981]. \end{aligned}$$

   (b) On considère les hypothèses :

$$H_0 : \ \lambda = \lambda_0 \qquad \text{contre} \qquad H_1 : \ \lambda \neq \lambda_0,$$

   avec $\lambda_0 = 3.42$. Nous allons utiliser le test de Wald associé à l'*emv* $\widehat{\lambda}_n$. On a $n$ "assez grand", un *emv* de $\lambda$ pour $(x_1, \ldots, x_n)$ est $\lambda^* = \dfrac{1}{n}\sum_{i=1}^{n} x_i = 3.50$ et, par le résultat de la question 5, $I_n(\lambda^*) = \dfrac{n}{\lambda^*} = \dfrac{3500}{3.50} = 1000$. On calcule :

$$z_{obs} = \sqrt{I_n(\lambda^*)}(\lambda^* - \lambda_0) = \sqrt{1000}(3.50 - 3.42) = 2.529822.$$

On considère une *var* $Z$ suivant la loi normale $\mathcal{N}(0,1)$. Alors la p-valeur associée est

$$\text{p-valeur} = \mathbb{P}(|Z| \geq |z_{obs}|) = \mathbb{P}(|Z| \geq 2.529822) = 0.01141204.$$

Comme p-valeur $< 0.05$, on rejette $H_0$ au risque 5%. Comme p-valeur $\in [0.01, 0.05[$, le rejet de $H_0$ est (seulement) significatif $\star$.

**Solution 2.** Dans un premier temps, notons que la loi de $X$ peut s'écrire sous la forme analytique : $\mathbb{P}_\theta(X = x) = \theta^x(1-\theta)^{1-x}$, $x \in \{0,1\}$. Un *emv* $\theta^*$ de $\theta$ pour $(x_1, \ldots, x_n)$ est défini par

$$\theta^* \in \underset{\theta \in ]0,1[}{\operatorname{argmax}}\, L_n(x_1, \ldots, x_n; \theta) = \underset{\theta \in ]0,1[}{\operatorname{argmax}}\, \ell_n(x_1, \ldots, x_n; \theta),$$

où $L_n$ désigne la fonction de vraisemblance pour $(x_1, \ldots, x_n)$ et $\ell_n$, la fonction de log-vraisemblance pour $(x_1, \ldots, x_n)$. La fonction de vraisemblance pour $(x_1, \ldots, x_n)$ est

$$L_n(x_1, \ldots, x_n; \theta) = \prod_{i=1}^{n} \mathbb{P}_\theta(X = x_i) = \prod_{i=1}^{n} \left(\theta^{x_i}(1-\theta)^{1-x_i}\right) = \theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{n - \sum_{i=1}^{n} x_i}.$$

La fonction de log-vraisemblance pour $(x_1, \ldots, x_n)$ est

$$\begin{aligned} \ell_n(x_1, \ldots, x_n; \theta) &= \ln(L_n(x_1, \ldots, x_n; \theta)) = \ln\left(\theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{n - \sum_{i=1}^{n} x_i}\right) \\ &= \ln(\theta)\sum_{i=1}^{n} x_i + \ln(1-\theta)\left(n - \sum_{i=1}^{n} x_i\right). \end{aligned}$$

Comme $\ell_n(x_1, \ldots, x_n; \theta)$ est dérivable en $\theta$ avec $\theta \in ]0,1[$, $\theta^*$ est solution de l'équation de vraisemblance : $\frac{\partial}{\partial \theta}\ell_n(x_1, \ldots, x_n; \theta) = 0$. Il vient

$$\frac{\partial}{\partial \theta}\ell_n(x_1, \ldots, x_n; \theta^*) = 0 \quad \Leftrightarrow \quad \frac{1}{\theta^*}\sum_{i=1}^{n} x_i - \frac{1}{1-\theta^*}\left(n - \sum_{i=1}^{n} x_i\right) = 0 \quad \Leftrightarrow$$

$$(1-\theta^*)\sum_{i=1}^{n} x_i - \theta^*\left(n - \sum_{i=1}^{n} x_i\right) = 0 \quad \Leftrightarrow \quad -\theta^* n + \sum_{i=1}^{n} x_i = 0 \quad \Leftrightarrow \quad \theta^* = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

De plus, $\theta^*$ est bien un maximum pour $\ell_n(x_1,\ldots,x_n;\theta)$ (et aussi pour $L_n(x_1,\ldots,x_n;\theta)$) car :

- si $\theta < \theta^*$, alors $\frac{\partial}{\partial\theta}\ell_n(x_1,\ldots,x_n;\theta) > 0$,
- si $\theta > \theta^*$, alors $\frac{\partial}{\partial\theta}\ell_n(x_1,\ldots,x_n;\theta) < 0$.

Donc $\theta^* = \frac{1}{n}\sum_{i=1}^{n} x_i$ est bien un *emv* de $\theta$ pour $(x_1,\ldots,x_n)$.

**Solution 3.** Un *emv* $\theta^*$ de $\theta$ pour $(x_1,\ldots,x_n)$ est défini par

$$\theta^* \in \underset{\theta\in\left]\frac{1}{2},1\right[}{\operatorname{argmax}}\, L_n(x_1,\ldots,x_n;\theta) = \underset{\theta\in\left]\frac{1}{2},1\right[}{\operatorname{argmax}}\, \ell_n(x_1,\ldots,x_n;\theta),$$

où $L_n$ désigne la fonction de vraisemblance pour $(x_1,\ldots,x_n)$ et $\ell_n$, la fonction de log-vraisemblance pour $(x_1,\ldots,x_n)$. La fonction de vraisemblance pour $(x_1,\ldots,x_n)$ est

$$L_n(x_1,\ldots,x_n;\theta) = \prod_{i=1}^{n} f_\theta(x_i) = \prod_{i=1}^{n}\left(\frac{\theta}{1-\theta}x_i^{\frac{2\theta-1}{1-\theta}}\right) = \frac{\theta^n}{(1-\theta)^n}\exp\left(\frac{2\theta-1}{1-\theta}\sum_{i=1}^{n}\ln(x_i)\right).$$

La fonction de log-vraisemblance pour $(x_1,\ldots,x_n)$ est

$$\begin{aligned}\ell_n(x_1,\ldots,x_n;\theta) &= \ln(L_n(x_1,\ldots,x_n;\theta)) = \ln\left(\frac{\theta^n}{(1-\theta)^n}\exp\left(\frac{2\theta-1}{1-\theta}\sum_{i=1}^{n}\ln(x_i)\right)\right)\\ &= n\ln(\theta) - n\ln(1-\theta) + \frac{2\theta-1}{1-\theta}\sum_{i=1}^{n}\ln(x_i).\end{aligned}$$

Comme $\ell_n(x_1,\ldots,x_n;\theta)$ est dérivable en $\theta$ avec $\theta \in \left]\frac{1}{2},1\right[$, $\theta^*$ est solution de l'équation de vraisemblance : $\frac{\partial}{\partial\theta}\ell_n(x_1,\ldots,x_n;\theta) = 0$. Il vient

$$\frac{\partial}{\partial\theta}\ell_n(x_1,\ldots,x_n;\theta^*) = 0 \quad\Leftrightarrow\quad n\frac{1}{\theta^*} - n\frac{1}{1-\theta^*} + \frac{1}{(1-\theta^*)^2}\sum_{i=1}^{n}\ln(x_i) = 0 \quad\Leftrightarrow$$

$$n(1-\theta^*)^2 - n\theta^*(1-\theta^*) + \theta^*\sum_{i=1}^{n}\ln(x_i) = 0 \quad\Leftrightarrow\quad n(1-\theta^*) + \theta^*\sum_{i=1}^{n}\ln(x_i) = 0 \quad\Leftrightarrow$$

$$\theta^* = \frac{n}{n - \sum_{i=1}^{n}\ln(x_i)}.$$

De plus, $\theta^*$ est bien un maximum pour $\ell_n(x_1, \ldots, x_n; \theta)$ (et aussi pour $L_n(x_1, \ldots, x_n; \theta)$) car :

∘ si $\theta < \theta^*$, alors $\frac{\partial}{\partial\theta}\ell_n(x_1, \ldots, x_n; \theta) > 0$,

∘ si $\theta > \theta^*$, alors $\frac{\partial}{\partial\theta}\ell_n(x_1, \ldots, x_n; \theta) < 0$.

Donc $\theta^* = \frac{n}{n - \sum_{i=1}^{n} \ln(x_i)}$ est bien un *emv* de $\theta$ pour $(x_1, \ldots, x_n)$.

**Solution 4.** Un *emv* $\theta^*$ de $\theta$ pour $(x_1, \ldots, x_n)$ est défini par

$$\theta^* \in \operatorname*{argmax}_{\theta > -1} L_n(x_1, \ldots, x_n; \theta) = \operatorname*{argmax}_{\theta > -1} \ell_n(x_1, \ldots, x_n; \theta),$$

où $L_n$ désigne la fonction de vraisemblance pour $(x_1, \ldots, x_n)$ et $\ell_n$, la fonction de log-vraisemblance pour $(x_1, \ldots, x_n)$. La fonction de vraisemblance pour $(x_1, \ldots, x_n)$ est

$$L_n(x_1, \ldots, x_n; \theta) = \prod_{i=1}^{n} f_\theta(x_i) = \prod_{i=1}^{n} \left( \frac{\theta+1}{2}(1-|x_i|)^\theta \right) = \frac{(\theta+1)^n}{2^n} \exp\left( \theta \sum_{i=1}^{n} \ln(1-|x_i|) \right).$$

La fonction de log-vraisemblance pour $(x_1, \ldots, x_n)$ est

$$\begin{aligned} \ell_n(x_1, \ldots, x_n; \theta) &= \ln(L_n(x_1, \ldots, x_n; \theta)) = \ln\left( \frac{(\theta+1)^n}{2^n} \exp\left( \theta \sum_{i=1}^{n} \ln(1-|x_i|) \right) \right) \\ &= n\ln(\theta+1) - n\ln(2) + \theta \sum_{i=1}^{n} \ln(1-|x_i|). \end{aligned}$$

Comme $\ell_n(x_1, \ldots, x_n; \theta)$ est dérivable en $\theta$ avec $\theta > -1$, $\theta^*$ est solution de l'équation de vraisemblance : $\frac{\partial}{\partial\theta}\ell_n(x_1, \ldots, x_n; \theta) = 0$. Il vient

$$\frac{\partial}{\partial\theta}\ell_n(x_1, \ldots, x_n; \theta^*) = 0 \qquad \Leftrightarrow \qquad n\frac{1}{\theta^*+1} + \sum_{i=1}^{n} \ln(x_i) = 0 \qquad \Leftrightarrow$$

$$n + (\theta^*+1)\sum_{i=1}^{n} \ln(1-|x_i|) = 0 \qquad \Leftrightarrow \qquad \theta^* = -\frac{n}{\sum_{i=1}^{n} \ln(1-|x_i|)} - 1.$$

De plus, $\theta^*$ est bien un maximum pour $\ell_n(x_1, \ldots, x_n; \theta)$ (et aussi pour $L_n(x_1, \ldots, x_n; \theta)$) car :

∘ si $\theta < \theta^*$, alors $\frac{\partial}{\partial\theta}\ell_n(x_1, \ldots, x_n; \theta) > 0$,

∘ si $\theta > \theta^*$, alors $\frac{\partial}{\partial\theta}\ell_n(x_1, \ldots, x_n; \theta) < 0$.

Donc $\theta^* = -\frac{n}{\sum\limits_{i=1}^{n}\ln(1-|x_i|)} - 1$ est bien un *emv* de $\theta$ pour $(x_1, \ldots, x_n)$.

**Solution 5.** On a

$$I_n(\lambda) = nI_1(\lambda), \qquad I_1(\lambda) = -\mathbb{E}\left(\frac{\partial^2}{\partial\lambda^2}\ln(\mathbb{P}_\lambda(X = X_1))\right).$$

On a $\ln(\mathbb{P}_\lambda(X = x)) = -\lambda^a + ax\ln(\lambda) - \ln(x!)$, donc

$$\frac{\partial}{\partial\lambda}\ln(\mathbb{P}_\lambda(X = x)) = -a\lambda^{a-1} + \frac{ax}{\lambda}, \qquad \frac{\partial^2}{\partial\lambda^2}\ln(\mathbb{P}_\lambda(X = x)) = -a(a-1)\lambda^{a-2} - \frac{ax}{\lambda^2}.$$

Comme $\mathbb{E}(X_1) = \lambda^a$, on obtient

$$\begin{aligned}\mathbb{E}\left(\frac{\partial^2}{\partial\lambda^2}\ln(\mathbb{P}_\lambda(X = X_1))\right) &= -a(a-1)\lambda^{a-2} - \frac{a}{\lambda^2}\mathbb{E}(X_1) \\ &= -a(a-1)\lambda^{a-2} - \frac{a}{\lambda^2} \times \lambda^a = -a^2\lambda^{a-2}.\end{aligned}$$

Donc

$$I_1(\lambda) = -\mathbb{E}\left(\frac{\partial^2}{\partial\lambda^2}\ln(\mathbb{P}_\lambda(X = X_1))\right) = -\left(-a^2\lambda^{a-2}\right) = a^2\lambda^{a-2}.$$

Au final, il vient

$$I_n(\lambda) = nI_1(\lambda) = n \times a^2\lambda^{a-2} = na^2\lambda^{a-2}.$$

**Solution 6.** On a

$$I_n(\theta) = nI_1(\theta), \qquad I_1(\theta) = -\mathbb{E}\left(\frac{\partial^2}{\partial\theta^2}\ln(f_\theta(X_1))\right).$$

On a $\ln(f_\theta(x)) = \ln\left(\frac{1+\theta}{(x+\theta)^2}\right) = \ln(1+\theta) - 2\ln(x+\theta)$, donc

$$\frac{\partial}{\partial\theta}\ln(f_\theta(x)) = \frac{1}{1+\theta} - \frac{2}{x+\theta}, \qquad \frac{\partial^2}{\partial\theta^2}\ln(f_\theta(x)) = -\frac{1}{(1+\theta)^2} + \frac{2}{(x+\theta)^2}.$$

On en déduit que

$$I_1(\theta) = -\mathbb{E}\left(\frac{\partial^2}{\partial\theta^2}\ln(f_\theta(X_1))\right) = -\mathbb{E}\left(-\frac{1}{(1+\theta)^2} + \frac{2}{(X_1+\theta)^2}\right) = \frac{1}{(1+\theta)^2} - 2\mathbb{E}\left(\frac{1}{(X_1+\theta)^2}\right).$$

On a

$$\begin{aligned}\mathbb{E}\left(\frac{1}{(X_1+\theta)^2}\right) &= \int_{-\infty}^{\infty} \frac{1}{(x+\theta)^2} f_\theta(x)dx = \int_1^{\infty} \frac{1}{(x+\theta)^2} \times \frac{1+\theta}{(x+\theta)^2} dx \\ &= (1+\theta)\int_1^{\infty} \frac{1}{(x+\theta)^4} dx = (1+\theta)\left[-\frac{1}{3(x+\theta)^3}\right]_1^{\infty} \\ &= (1+\theta)\left(\frac{1}{3(1+\theta)^3}\right) = \frac{1}{3(1+\theta)^2}.\end{aligned}$$

D'où

$$I_1(\theta) = \frac{1}{(1+\theta)^2} - 2 \times \frac{1}{3(1+\theta)^2} = \frac{1}{3(1+\theta)^2}.$$

Il s'ensuit

$$I_n(\theta) = nI_1(\theta) = n \times \frac{1}{3(1+\theta)^2} = \frac{n}{3(1+\theta)^2}.$$

**Solution 7.**

1. La fonction de vraisemblance pour $(x_1, \ldots, x_n)$ est

$$L_n(x_1, \ldots, x_n; \theta) = \prod_{i=1}^n f_\theta(x_i) = \prod_{i=1}^n \left(\frac{1}{\theta} e^{-\frac{1}{\theta} x_i}\right) = \frac{1}{\theta^n} \exp\left(-\frac{1}{\theta}\sum_{i=1}^n x_i\right).$$

2. La fonction de log-vraisemblance pour $(x_1, \ldots, x_n)$ est

$$\ell_n(x_1, \ldots, x_n; \theta) = \ln(L_n(x_1, \ldots, x_n; \theta)) = \ln\left(\frac{1}{\theta^n} \exp\left(-\frac{1}{\theta}\sum_{i=1}^n x_i\right)\right) = -n\ln(\theta) - \frac{1}{\theta}\sum_{i=1}^n x_i.$$

3. Un *emv* $\theta^*$ de $\theta$ pour $(x_1, \ldots, x_n)$ est défini par

$$\theta^* \in \underset{\theta>0}{\operatorname{argmax}}\, L_n(x_1, \ldots, x_n; \theta) = \underset{\theta>0}{\operatorname{argmax}}\, \ell_n(x_1, \ldots, x_n; \theta).$$

   Par le résultat de la question 2, $\ell_n(x_1, \ldots, x_n; \theta)$ est dérivable en $\theta$ avec $\theta > 0$. Par conséquent, $\theta^*$ est solution de l'équation de vraisemblance : $\dfrac{\partial}{\partial\theta}\ell_n(x_1, \ldots, x_n; \theta) = 0$.

Il vient

$$\frac{\partial}{\partial\theta}\ell_n(x_1,\ldots,x_n;\theta^*)=0 \qquad \Leftrightarrow \qquad -n\frac{1}{\theta^*}+\frac{1}{(\theta^*)^2}\sum_{i=1}^n x_i=0 \qquad \Leftrightarrow \qquad \theta^*=\frac{1}{n}\sum_{i=1}^n x_i.$$

De plus, $\theta^*$ est bien un maximum pour $\ell_n(x_1,\ldots,x_n;\theta)$ (et aussi pour $L_n(x_1,\ldots,x_n;\theta)$) car :
- si $\theta<\theta^*$, alors $\frac{\partial}{\partial\theta}\ell_n(x_1,\ldots,x_n;\theta)>0$,
- si $\theta>\theta^*$, alors $\frac{\partial}{\partial\theta}\ell_n(x_1,\ldots,x_n;\theta)<0$.

Donc $\theta^*=\frac{1}{n}\sum_{i=1}^n x_i$ est bien un *emv* de $\theta$ pour $(x_1,\ldots,x_n)$. L'*emv* $\widehat{\theta}_n$ de $\theta$ associé est

$$\widehat{\theta}_n=\frac{1}{n}\sum_{i=1}^n X_i.$$

4. La loi commune des *var* $X_1,\ldots,X_n$ et l'égalité $\mathbb{E}(X_1)=\frac{1}{\frac{1}{\theta}}=\theta$ entraînent

$$\mathbb{E}(\widehat{\theta}_n)=\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^n X_i\right)=\frac{1}{n}\sum_{i=1}^n\mathbb{E}(X_i)=\frac{1}{n}n\mathbb{E}(X_1)=\theta.$$

On en déduit que $\widehat{\theta}_n$ est un estimateur sans biais de $\theta$.

L'indépendance et la loi commune des *var* $X_1,\ldots,X_n$, et l'égalité $\mathbb{V}(X_1)=\frac{1}{\left(\frac{1}{\theta}\right)^2}=\theta^2$ impliquent

$$\mathbb{V}(\widehat{\theta}_n)=\mathbb{V}\left(\frac{1}{n}\sum_{i=1}^n X_i\right)=\frac{1}{n^2}\mathbb{V}\left(\sum_{i=1}^n X_i\right)=\frac{1}{n^2}\sum_{i=1}^n\mathbb{V}(X_i)=\frac{1}{n^2}n\mathbb{V}(X_1)=\frac{1}{n}\theta^2.$$

Comme $\widehat{\theta}_n$ est un estimateur sans biais de $\theta$, on a

$$R(\widehat{\theta}_n,\theta)=\mathbb{V}(\widehat{\theta}_n)=\frac{\theta^2}{n}.$$

Comme les *var* $X_1,\ldots,X_n$ sont *iid* et admettent un moment d'ordre 2, la loi forte des grands nombres nous assure que $(\widehat{\theta}_n)_{n\in\mathbb{N}^*}=\left(\frac{1}{n}\sum_{i=1}^n X_i\right)_{n\in\mathbb{N}^*}$ converge presque sûrement vers $\mathbb{E}(X_1)=\theta$. Ainsi, $\widehat{\theta}_n$ est fortement consistant.

5. L'information de Fisher de $X$ sur $\theta$ est définie par

$$I_n(\theta) = nI_1(\theta), \qquad I_1(\theta) = \mathbb{V}\left(\frac{\partial}{\partial\theta}\ln(f_\theta(X_1))\right),$$

l'espérance étant prise par rapport à la *var* $X_1$. Calculons $I_1(\theta)$. On a

$$\ln(f_\theta(x)) = \ln\left(\frac{1}{\theta}e^{-\frac{1}{\theta}x}\right) = -\ln(\theta) - \frac{1}{\theta}x, \qquad \frac{\partial}{\partial\theta}\ln(f_\theta(x)) = -\frac{1}{\theta} + \frac{1}{\theta^2}x.$$

Comme $\mathbb{V}(X_1) = \theta^2$, on obtient

$$I_1(\theta) = \mathbb{V}\left(\frac{\partial}{\partial\theta}\ln(f_\theta(X_1))\right) = \mathbb{V}\left(-\frac{1}{\theta} + \frac{1}{\theta^2}X_1\right) = \mathbb{V}\left(\frac{1}{\theta^2}X_1\right) = \frac{1}{\theta^4}\mathbb{V}(X_1) = \frac{1}{\theta^2}.$$

On en déduit que

$$I_n(\theta) = nI_1(\theta) = \frac{n}{\theta^2}.$$

6. Par les résultats des questions 4 et 5, $\widehat{\theta}_n$ est un estimateur de $\theta$ :
   - sans biais : $\mathbb{E}(\widehat{\theta}_n) = \theta$,
   - on a $\mathbb{V}(\widehat{\theta}_n) = \dfrac{\theta^2}{n} = \dfrac{1}{I_n(\theta)}$.

   On en déduit que $\widehat{\theta}_n$ est un estimateur efficace de $\theta$.

7. D'une part, en utilisant la question 5, comme $\widehat{\theta}_n > 0$ presque sûrement, on a

$$\begin{aligned}
&\mathbb{P}\left(\theta \in \left[\widehat{\theta}_n - 1.96\frac{\widehat{\theta}_n}{\sqrt{n}}, \widehat{\theta}_n + 1.96\frac{\widehat{\theta}_n}{\sqrt{n}}\right]\right) = \mathbb{P}\left(\widehat{\theta}_n - 1.96\frac{\widehat{\theta}_n}{\sqrt{n}} \leq \theta \leq \widehat{\theta}_n + 1.96\frac{\widehat{\theta}_n}{\sqrt{n}}\right) \\
&= \mathbb{P}\left(-1.96 \leq \frac{\sqrt{n}}{\widehat{\theta}_n}(\theta - \widehat{\theta}_n) \leq 1.96\right) = \mathbb{P}\left(\left|\frac{\sqrt{n}}{\widehat{\theta}_n}(\theta - \widehat{\theta}_n)\right| \leq 1.96\right) \\
&= \mathbb{P}\left(\left|\sqrt{I_n(\widehat{\theta}_n)}(\widehat{\theta}_n - \theta)\right| \leq 1.96\right).
\end{aligned}$$

D'autre part, le résultat asymptotique central nous assure que la suite de *var* :

$$\left(\sqrt{I_n(\widehat{\theta}_n)}(\widehat{\theta}_n - \theta)\right)_{n\in\mathbb{N}^*}$$

converge en loi vers une *var* $Z$ suivant la loi normale $\mathcal{N}(0, 1)$.

On en déduit que

$$\begin{aligned}\lim_{n\to\infty} \mathbb{P}\left(\theta \in \left[\widehat{\theta}_n - 1.96\frac{\widehat{\theta}_n}{\sqrt{n}}, \widehat{\theta}_n + 1.96\frac{\widehat{\theta}_n}{\sqrt{n}}\right]\right) &= \lim_{n\to\infty} \mathbb{P}\left(\left|\sqrt{I_n(\widehat{\theta}_n)}(\widehat{\theta}_n - \theta)\right| \leq 1.96\right) \\ &= \mathbb{P}(|Z| \leq 1.96) = 0.95.\end{aligned}$$

## Solution 8.

1. La fonction de vraisemblance de $\theta$ pour $(x_1, \ldots, x_n)$ est

$$L_n(x_1, \ldots, x_n; \theta) = \prod_{i=1}^{n} f_\theta(x_i).$$

Vu l'expression de $f_\theta(x_i)$, celle-ci est non-nulle si et seulement si, pour tout $i \in \{1, \ldots, n\}$, $x_i \in [0, \theta]$, donc $\sup(x_1, \ldots, x_n) \leq \theta$. Dans ce cas, on a

$$L_n(x_1, \ldots, x_n; \theta) = \frac{1}{\theta^n}.$$

Comme cette fonction est décroissante en $\theta$ quand $\theta > 0$, il vient

$$\theta^* \in \operatorname*{argmax}_{\theta} L_n(x_1, \ldots, x_n; \theta) = \operatorname*{argmax}_{\theta \in [\sup(x_1,\ldots,x_n), \infty[} \frac{1}{\theta^n} = \{\sup(x_1, \ldots, x_n)\}.$$

On en déduit qu'un *emv* $\theta^*$ de $\theta$ pour $(x_1, \ldots, x_n)$ est $\theta^* = \sup(x_1, \ldots, x_n)$. Un *emv* $\widehat{\theta}_n$ de $\theta$ est

$$\widehat{\theta}_n = \sup(X_1, \ldots, X_n).$$

2. Comme $X(\Omega) = [0, \theta]$, on a $\widehat{\theta}_n(\Omega) = (\sup(X_1, \ldots, X_n))(\Omega) = [0, \theta]$. Par conséquent, pour tout $x < 0$, on a $F_{\widehat{\theta}_n}(x) = \mathbb{P}(\widehat{\theta}_n \leq x) = 0$ et pour tout $x > \theta$, on a $F_{\widehat{\theta}_n}(x) = \mathbb{P}(\widehat{\theta}_n \leq x) = 1$. Pour tout $x \in [0, \theta]$, l'indépendance et la loi commune des *var* $X_1, \ldots, X_n$ entraînent

$$\begin{aligned}F_{\widehat{\theta}_n}(x) &= \mathbb{P}(\widehat{\theta}_n \leq x) = \mathbb{P}(\sup(X_1, \ldots, X_n) \leq x) = \mathbb{P}\left(\bigcap_{i=1}^{n} \{X_i \leq x\}\right) = \prod_{i=1}^{n} \mathbb{P}(X_i \leq x) \\ &= (\mathbb{P}(X_1 \leq x))^n = \left(\int_{-\infty}^{x} f_\theta(t)dt\right)^n = \left(\int_0^x \frac{1}{\theta} dt\right)^n = \frac{x^n}{\theta^n}.\end{aligned}$$

Au final, la fonction de répartition de $\widehat{\theta}_n$ est

$$F_{\widehat{\theta}_n}(x) = \begin{cases} \dfrac{x^n}{\theta^n} & \text{si } x \in [0,\theta], \\ 0 & \text{sinon.} \end{cases}$$

On en déduit une densité de $\widehat{\theta}_n$ par dérivation :

$$f_{\widehat{\theta}_n}(x) = \begin{cases} n\dfrac{x^{n-1}}{\theta^n} & \text{si } x \in [0,\theta], \\ 0 & \text{sinon.} \end{cases}$$

3. Le biais de $\widehat{\theta}_n$ est $B(\widehat{\theta}_n,\theta) = \mathbb{E}(\widehat{\theta}_n) - \theta$. En utilisant la densité de $\widehat{\theta}_n$ déterminée au résultat de la question 2, il vient

$$\mathbb{E}(\widehat{\theta}_n) = \int_{-\infty}^{\infty} x f_{\widehat{\theta}_n}(x)dx = \int_0^{\theta} x \times n\frac{x^{n-1}}{\theta^n}dx = \frac{n}{\theta^n}\int_0^{\theta} x^n dx = \frac{n}{\theta^n}\left[\frac{x^{n+1}}{n+1}\right]_0^{\theta} = \frac{n}{n+1}\theta.$$

Donc

$$B(\widehat{\theta}_n,\theta) = \frac{n}{n+1}\theta - \theta = -\frac{1}{n+1}\theta.$$

Comme $B(\widehat{\theta}_n,\theta) \neq 0$, $\widehat{\theta}_n$ n'est pas sans biais. Comme $\lim_{n\to\infty} B(\widehat{\theta}_n,\theta) = \lim_{n\to\infty} -\frac{1}{n+1}\theta = 0$, $\widehat{\theta}_n$ est asymptotiquement sans biais.

4. Le risque quadratique de $\widehat{\theta}_n$ est

$$R(\widehat{\theta}_n,\theta) = \mathbb{E}((\widehat{\theta}_n - \theta)^2) = \mathbb{V}(\widehat{\theta}_n - \theta) + (\mathbb{E}(\widehat{\theta}_n - \theta))^2 = \mathbb{V}(\widehat{\theta}_n) + (B(\widehat{\theta}_n,\theta))^2.$$

On a $\mathbb{V}(\widehat{\theta}_n) = \mathbb{E}(\widehat{\theta}_n^2) - (\mathbb{E}(\widehat{\theta}_n))^2$. En utilisant la densité de $\widehat{\theta}_n$ déterminée au résultat de la question 2, il vient

$$\begin{aligned} \mathbb{E}(\widehat{\theta}_n^2) &= \int_{-\infty}^{\infty} x^2 f_{\widehat{\theta}_n}(x)dx = \int_0^{\theta} x^2 \times n\frac{x^{n-1}}{\theta^n}dx = \frac{n}{\theta^n}\int_0^{\theta} x^{n+1}dx = \frac{n}{\theta^n}\left[\frac{x^{n+2}}{n+2}\right]_0^{\theta} \\ &= \frac{n}{\theta^n} \times \frac{\theta^{n+2}}{n+2} = \frac{n\theta^2}{n+2}. \end{aligned}$$

D'où

$$\mathbb{V}(\widehat{\theta}_n) = \frac{n\theta^2}{n+2} - \left(\frac{n\theta}{n+1}\right)^2 = \frac{n\theta^2}{(n+1)^2(n+2)}.$$

On en déduit que

$$R(\widehat{\theta}_n, \theta) = \frac{n\theta^2}{(n+1)^2(n+2)} + \left(-\frac{\theta}{n+1}\right)^2 = \frac{n\theta^2 + \theta^2(n+2)}{(n+1)^2(n+2)} = \frac{2\theta^2}{(n+1)(n+2)}.$$

5. Par le résultat de la question 4, on a $\sum_{n=1}^{\infty} R(\widehat{\theta}_n, \theta) = \sum_{n=1}^{\infty} \frac{2\theta^2}{(n+1)(n+2)} < \infty$, Cela entraîne la convergence presque sûre de $(\widehat{\theta}_n)_{n\in\mathbb{N}^*}$ vers $\theta$ ; $\widehat{\theta}_n$ est fortement consistant.

**Solution 9.**

1. On a

$$\begin{aligned}\mathbb{E}(X-\theta) &= \int_{-\infty}^{\infty} (x-\theta) f_\theta(x) dx = \int_{1+\theta}^{\infty} (x-\theta) \times \frac{3}{(x-\theta)^4} dx \\ &= \int_{1+\theta}^{\infty} \frac{3}{(x-\theta)^3} dx = 3\left[-\frac{1}{2(x-\theta)^2}\right]_{1+\theta}^{\infty} = \frac{3}{2}.\end{aligned}$$

On a

$$\begin{aligned}\mathbb{E}((X-\theta)^2) &= \int_{-\infty}^{\infty} (x-\theta)^2 f_\theta(x) dx = \int_{1+\theta}^{\infty} (x-\theta)^2 \times \frac{3}{(x-\theta)^4} dx \\ &= \int_{1+\theta}^{\infty} \frac{3}{(x-\theta)^2} dx = 3\left[-\frac{1}{x-\theta}\right]_{1+\theta}^{\infty} = 3.\end{aligned}$$

On en déduit que

$$\mathbb{E}(X) = \mathbb{E}(X-\theta) + \theta = \frac{3}{2} + \theta.$$

On a

$$\mathbb{V}(X) = \mathbb{V}(X-\theta) = \mathbb{E}((X-\theta)^2) - (\mathbb{E}(X-\theta))^2 = 3 - \left(\frac{3}{2}\right)^2 = \frac{3}{4}.$$

2. La fonction de vraisemblance de $\theta$ pour $(x_1, \ldots, x_n)$ est

$$L_n(x_1, \ldots, x_n; \theta) = \prod_{i=1}^{n} f_\theta(x_i).$$

Vu l'expression de $f_\theta(x_i)$, celle-ci est non-nulle si et seulement si, pour tout $i \in \{1, \ldots, n\}$, $x_i \geq 1 + \theta$, donc $\inf(x_1, \ldots, x_n) \geq 1 + \theta$, soit $\theta \leq \inf(x_1, \ldots, x_n) - 1$. Dans ce cas, on a

$$L_n(x_1, \ldots, x_n; \theta) = \prod_{i=1}^{n} \frac{3}{(x_i - \theta)^4}.$$

Comme cette fonction est croissante en $\theta$ quand $\theta > 0$, il vient

$$\theta^* \in \underset{\theta}{\operatorname{argmax}}\, L_n(x_1, \ldots, x_n; \theta) = \underset{\theta \in ]0, \inf(x_1, \ldots, x_n) - 1]}{\operatorname{argmax}} \prod_{i=1}^{n} \frac{3}{(x_i - \theta)^4} = \{\inf(x_1, \ldots, x_n) - 1\}.$$

On en déduit qu'un *emv* $\theta^*$ de $\theta$ pour $(x_1, \ldots, x_n)$ est $\theta^* = \inf(x_1, \ldots, x_n) - 1$. Un *emv* $\widehat{\theta}_n$ de $\theta$ est

$$\widehat{\theta}_n = \inf(X_1, \ldots, X_n) - 1.$$

3. Comme $X(\Omega) = [1 + \theta, \infty[$, on a $(\widehat{\theta}_n - \theta)(\Omega) = (\inf(X_1, \ldots, X_n) - 1 - \theta)(\Omega) = [0, \infty[$. Par conséquent, pour tout $x < 0$, on a $F_{\widehat{\theta}_n - \theta}(x) = \mathbb{P}(\widehat{\theta}_n - \theta \leq x) = 0$. Pour tout $x \geq 0$, l'indépendance et la loi commune des *var* $X_1, \ldots, X_n$ entraînent

$$\begin{aligned}
F_{\widehat{\theta}_n - \theta}(x) &= \mathbb{P}(\widehat{\theta}_n - \theta \leq x) = \mathbb{P}\left(\inf(X_1, \ldots, X_n) - 1 - \theta \leq x\right) \\
&= 1 - \mathbb{P}\left(\inf(X_1, \ldots, X_n) \geq x + \theta + 1\right) = 1 - \mathbb{P}\left(\bigcap_{i=1}^{n} \{X_i \geq x + \theta + 1\}\right) \\
&= 1 - \prod_{i=1}^{n} \mathbb{P}\left(X_i \geq x + \theta + 1\right) = 1 - \left(\mathbb{P}\left(X_1 \geq x + \theta + 1\right)\right)^n = 1 - \left(\int_{x+\theta+1}^{\infty} f_\theta(y) dy\right)^n \\
&= 1 - \left(\int_{x+\theta+1}^{\infty} \frac{3}{(y - \theta)^4} dy\right)^n = 1 - \left(\left[-\frac{1}{(y - \theta)^3}\right]_{x+\theta+1}^{\infty}\right)^n = 1 - \frac{1}{(x + 1)^{3n}}.
\end{aligned}$$

Au final, la fonction de répartition de $\widehat{\theta}_n - \theta$ est

$$F_{\widehat{\theta}_n - \theta}(x) = \begin{cases} 1 - \dfrac{1}{(x + 1)^{3n}} & \text{si } x \geq 0, \\ 0 & \text{sinon.} \end{cases}$$

On en déduit une densité de $\widehat{\theta}_n - \theta$ par dérivation :

$$f_{\widehat{\theta}_n - \theta}(x) = \begin{cases} \dfrac{3n}{(x+1)^{3n+1}} & \text{si } x \geq 0, \\ 0 & \text{sinon.} \end{cases}$$

4. En utilisant la densité de $\widehat{\theta}_n - \theta$ déterminée au résultat de la question 3, le biais de $\widehat{\theta}_n$ est

$$\begin{aligned} B(\widehat{\theta}_n, \theta) &= \mathbb{E}(\widehat{\theta}_n - \theta) = \int_{-\infty}^{\infty} x f_{\widehat{\theta}_n - \theta}(x) dx = \int_0^{\infty} x \times \frac{3n}{(x+1)^{3n+1}} dx \\ &= 3n \int_0^{\infty} \frac{x+1-1}{(x+1)^{3n+1}} dx = 3n \int_0^{\infty} \frac{1}{(x+1)^{3n}} dx - 3n \int_0^{\infty} \frac{1}{(x+1)^{3n+1}} dx \\ &= 3n \left[ -\frac{1}{(3n-1)(x+1)^{3n-1}} \right]_0^{\infty} - 3n \left[ -\frac{1}{3n(x+1)^{3n}} \right]_0^{\infty} \\ &= \frac{3n}{3n-1} - 3n \times \frac{1}{3n} = \frac{1}{3n-1}. \end{aligned}$$

Comme $B(\widehat{\theta}_n, \theta) \neq 0$, $\widehat{\theta}_n$ n'est pas sans biais. Comme $\lim_{n\to\infty} B(\widehat{\theta}_n, \theta) = \lim_{n\to\infty} \dfrac{1}{3n-1} = 0$, $\widehat{\theta}_n$ est asymptotiquement sans biais.

5. En utilisant la densité de $\widehat{\theta}_n - \theta$ déterminée au résultat de la question 3, le risque quadratique de $\widehat{\theta}_n$ est

$$\begin{aligned} R(\widehat{\theta}_n, \theta) &= \mathbb{E}((\widehat{\theta}_n - \theta)^2) = \int_{-\infty}^{\infty} x^2 f_{\widehat{\theta}_n - \theta}(x) dx = \int_0^{\infty} x^2 \times \frac{3n}{(x+1)^{3n+1}} dx \\ &= 3n \int_0^{\infty} \frac{x^2}{(x+1)^{3n+1}} = 3n \int_0^{\infty} \frac{(x+1)^2 - 2(x+1) + 1}{(x+1)^{3n+1}} dx \\ &= 3n \int_0^{\infty} \frac{1}{(x+1)^{3n-1}} dx - 6n \int_0^{\infty} \frac{1}{(x+1)^{3n}} dx + 3n \int_0^{\infty} \frac{1}{(x+1)^{3n+1}} dx \\ &= 3n \left[ -\frac{1}{(3n-2)(x+1)^{3n-2}} \right]_0^{\infty} - 6n \left[ -\frac{1}{(3n-1)(x+1)^{3n-1}} \right]_0^{\infty} \\ &+ 3n \left[ -\frac{1}{3n(x+1)^{3n}} \right]_0^{\infty} = 3n \times \frac{1}{3n-2} - 6n \times \frac{1}{3n-1} + 3n \times \frac{1}{3n} \\ &= \frac{3n(3n-1) - 6n(3n-2) + (3n-1)(3n-2)}{(3n-1)(3n-2)} = \frac{2}{(3n-1)(3n-2)}. \end{aligned}$$

Ainsi, on a

$$R(\widehat{\theta}_n, \theta) = \frac{2}{(3n-1)(3n-2)}.$$

6. Par le résultat de la question 4, on a $B(\widehat{\theta}_n, \theta) = \dfrac{1}{3n-1}$. Par conséquent, un estimateur sans biais de $\theta$ est

$$\widehat{\theta}_n^* = \widehat{\theta}_n - \frac{1}{3n-1}.$$

On a

$$\begin{aligned} R(\widehat{\theta}_n, \theta) &= \mathbb{V}(\widehat{\theta}_n) + (B(\widehat{\theta}_n, \theta))^2 = \mathbb{V}\left(\widehat{\theta}_n^* + \frac{1}{3n-1}\right) + (B(\widehat{\theta}_n, \theta))^2 \\ &= \mathbb{V}(\widehat{\theta}_n^*) + (B(\widehat{\theta}_n, \theta))^2 = R(\widehat{\theta}_n^*, \theta) + (B(\widehat{\theta}_n, \theta))^2 > R(\widehat{\theta}_n^*, \theta). \end{aligned}$$

On en déduit que $\widehat{\theta}_n^*$ est préférable à $\widehat{\theta}_n$.

## Solution 10.

1. La fonction de vraisemblance pour $(x_1, \ldots, x_n)$ est

$$L_n(x_1, \ldots, x_n; \theta) = \prod_{i=1}^{n} f_\theta(x_i) = \prod_{i=1}^{n} \left(\theta x_i^{-(\theta+1)}\right) = \theta^n \exp\left(-(\theta+1)\sum_{i=1}^{n} \ln(x_i)\right).$$

La fonction de log-vraisemblance pour $(x_1, \ldots, x_n)$ est

$$\begin{aligned} \ell_n(x_1, \ldots, x_n; \theta) &= \ln(L_n(x_1, \ldots, x_n; \theta)) = \ln\left(\theta^n \exp\left(-(\theta+1)\sum_{i=1}^{n} \ln(x_i)\right)\right) \\ &= n\ln(\theta) - (\theta+1)\sum_{i=1}^{n} \ln(x_i). \end{aligned}$$

On a $\dfrac{\partial}{\partial\theta}\ell_n(x_1, \ldots, x_n; \theta) = n\dfrac{1}{\theta} - \sum_{i=1}^{n} \ln(x_i)$. Donc, en substituant $x_i$ par $X_i$ pour tout $i \in \{1, \ldots, n\}$, on peut écrire

$$\frac{\partial}{\partial\theta}\ell_n(X_1, \ldots, X_n; \theta) = -n\left(\frac{1}{n}\sum_{i=1}^{n} \ln(X_i) - \frac{1}{\theta}\right) = a_n(\theta)\left(T_n - h(\theta)\right),$$

avec

$$T_n = \frac{1}{n}\sum_{i=1}^{n} \ln(X_i), \qquad a_n(\theta) = -n, \qquad h(\theta) = \frac{1}{\theta}.$$

Le théorème général sur l'efficacité nous assure que $T_n$ est un estimateur efficace de $h(\theta) = \frac{1}{\theta}$.

En particulier et plus, on a

- $T_n$ est sans biais de $h(\theta) = \frac{1}{\theta}$ : $\mathbb{E}(T_n) = h(\theta) = \frac{1}{\theta}$,
- $\mathbb{V}(T_n) = \frac{h'(\theta)}{a_n(\theta)} = \frac{-\frac{1}{\theta^2}}{-n} = \frac{1}{\theta^2 n}$,
- $I_n(\theta) = \frac{(h'(\theta))^2}{\mathbb{V}(T_n)} = \frac{\left(\frac{1}{\theta^2}\right)^2}{\frac{1}{\theta^2 n}} = \frac{n}{\theta^2}$.

2. En utilisant le résultat de la question 1, on a

$$\sqrt{n}\,(\theta T_n - 1) = \sqrt{n}\left(\frac{T_n - \frac{1}{\theta}}{\frac{1}{\theta}}\right) = \sqrt{n}\left(\frac{\frac{1}{n}\sum_{i=1}^{n}\ln(X_i) - \mathbb{E}(\ln(X_1))}{\sigma(\ln(X_1))}\right).$$

Comme les *var* $(\ln(X_n))_{n\in\mathbb{N}^*}$ sont *iid* et admettent un moment d'ordre 2, le théorème central limite nous assure que $(\sqrt{n}\,(\theta T_n - 1))_{n\in\mathbb{N}^*} = \left(\sqrt{n}\left(\frac{\frac{1}{n}\sum_{i=1}^{n}\ln(X_i) - \mathbb{E}(\ln(X_1))}{\sigma(\ln(X_1))}\right)\right)_{n\in\mathbb{N}^*}$ converge en loi vers une *var* $Z$ suivant la loi normale $\mathcal{N}(0,1)$.

## Solution 11.

1. La fonction de vraisemblance pour $(x_1, \ldots, x_n)$ est

$$L_n(x_1, \ldots, x_n; \theta) = \prod_{i=1}^{n} f_\theta(x_i) = \prod_{i=1}^{n}\left(\frac{\theta+1}{2}(1-|x_i|)^\theta\right) = \frac{(\theta+1)^n}{2^n}\exp\left(\theta\sum_{i=1}^{n}\ln(1-|x_i|)\right).$$

La fonction de log-vraisemblance pour $(x_1, \ldots, x_n)$ est

$$\begin{aligned}\ell_n(x_1, \ldots, x_n; \theta) &= \ln(L_n(x_1, \ldots, x_n; \theta)) = \ln\left(\frac{(\theta+1)^n}{2^n}\exp\left(\theta\sum_{i=1}^{n}\ln(1-|x_i|)\right)\right) \\ &= n\ln(\theta+1) - n\ln(2) + \theta\sum_{i=1}^{n}\ln(1-|x_i|).\end{aligned}$$

On a $\frac{\partial}{\partial\theta}\ell_n(x_1, \ldots, x_n; \theta) = n\frac{1}{\theta+1} + \sum_{i=1}^{n}\ln(1-|x_i|)$.

Donc, en substituant $x_i$ par $X_i$ pour tout $i \in \{1,\ldots,n\}$, on peut écrire

$$\frac{\partial}{\partial\theta}\ell_n(X_1,\ldots,X_n;\theta) = n\left(\frac{1}{n}\sum_{i=1}^{n}\ln(1-|X_i|) - \left(-\frac{1}{\theta+1}\right)\right) = a_n(\theta)\left(T_n - h\left(\theta\right)\right),$$

avec

$$T_n = \frac{1}{n}\sum_{i=1}^{n}\ln(1-|X_i|), \qquad a_n(\theta) = n, \qquad h\left(\theta\right) = -\frac{1}{\theta+1}.$$

Le théorème général sur l'efficacité nous assure que $T_n$ est un estimateur efficace de $h(\theta) = -\frac{1}{\theta+1}$. En particulier et plus, on a

- $T_n$ est sans biais de $h(\theta) = -\frac{1}{\theta+1}$ : $\mathbb{E}\left(T_n\right) = h(\theta) = -\frac{1}{\theta+1}$,
- $\mathbb{V}\left(T_n\right) = \frac{h'\left(\theta\right)}{a_n(\theta)} = \frac{\frac{1}{(\theta+1)^2}}{n} = \frac{1}{(\theta+1)^2 n}$,
- $I_n(\theta) = \frac{\left(h'\left(\theta\right)\right)^2}{\mathbb{V}\left(T_n\right)} = \frac{\left(\frac{1}{(\theta+1)^2}\right)^2}{\frac{1}{(\theta+1)^2 n}} = \frac{n}{(\theta+1)^2}$.

2. En utilisant le résultat de la question 1, on a

$$\begin{aligned}(\theta+1)\sqrt{n}\left(T_n + \frac{1}{\theta+1}\right) &= \sqrt{n}\left(\frac{T_n - \left(-\frac{1}{\theta+1}\right)}{\frac{1}{\theta+1}}\right) \\ &= \sqrt{n}\left(\frac{\frac{1}{n}\sum\limits_{i=1}^{n}\ln(1-|X_i|) - \mathbb{E}\left(\ln(1-|X_1|)\right)}{\sigma\left(\ln(1-|X_1|)\right)}\right).\end{aligned}$$

Comme les *var* $\left(\ln(1-|X_n|)\right)_{n\in\mathbb{N}^*}$ sont *iid* et admettent un moment d'ordre 2, le théorème central limite nous assure que $\left((\theta+1)\sqrt{n}\left(T_n + \frac{1}{\theta+1}\right)\right)_{n\in\mathbb{N}^*} = \left(\sqrt{n}\left(\frac{\frac{1}{n}\sum\limits_{i=1}^{n}\ln(1-|X_i|)-\mathbb{E}(\ln(1-|X_1|))}{\sigma(\ln(1-|X_1|))}\right)\right)_{n\in\mathbb{N}^*}$ converge en loi vers une *var* $Z$ suivant la loi normale $\mathcal{N}(0,1)$.

**Solution 12.** La fonction de vraisemblance pour $(x_1,\ldots,x_n)$ est

$$L_n(x_1,\ldots,x_n;\theta) = \prod_{i=1}^{n} f_\theta(x_i).$$

Vu l'expression de $f_\theta(x_i)$, celle-ci est non-nulle si et seulement si, pour tout $i \in \{1, \ldots, n\}$, $x_i \geq \theta$, soit $\inf(x_1, \ldots, x_n) \geq \theta$. Dans ce cas, on a

$$L_n(x_1, \ldots, x_n; \theta) = \prod_{i=1}^{n} e^{-(x_i - \theta)} = e^{n\theta - \sum\limits_{i=1}^{n} x_i}.$$

Comme cette fonction est croissante en $\theta$, il vient

$$\theta^* \in \underset{\theta}{\operatorname{argmax}}\, L_n(x_1, \ldots, x_n; \theta) = \underset{\theta \in ]0, \inf(x_1, \ldots, x_n)]}{\operatorname{argmax}} e^{n\theta - \sum\limits_{i=1}^{n} x_i} = \{\inf(x_1, \ldots, x_n)\}.$$

On en déduit qu'un/que l' *emv* $\theta^*$ de $\theta$ pour $(x_1, \ldots, x_n)$ est $\theta^* = \inf(x_1, \ldots, x_n)$.

**Solution 13.** La fonction de vraisemblance pour $(x_1, \ldots, x_n)$ est

$$L_n(x_1, \ldots, x_n; \theta) = \prod_{i=1}^{n} f_\theta(x_i).$$

Vu l'expression de $f_\theta(x_i)$, celle-ci est non-nulle si et seulement si, pour tout $i \in \{1, \ldots, n\}$,

- $x_i \geq \theta$, soit $\inf(x_1, \ldots, x_n) \geq \theta$,
- $x_i \leq \theta + 1$, soit $\sup(x_1, \ldots, x_n) \leq \theta + 1 \quad \Leftrightarrow \quad \theta \geq \sup(x_1, \ldots, x_n) - 1$,

donc $\theta \in [\sup(x_1, \ldots, x_n) - 1, \inf(x_1, \ldots, x_n)]$. Dans ce cas, on a

$$L_n(x_1, \ldots, x_n; \theta) = 1.$$

Comme cette fonction atteint sa valeur maximale sur cet intervalle, on a

$$\theta^* \in \underset{\theta}{\operatorname{argmax}}\, L_n(x_1, \ldots, x_n; \theta) = \underset{\theta \in [\sup(x_1, \ldots, x_n) - 1, \inf(x_1, \ldots, x_n)]}{\operatorname{argmax}} 1 = [\sup(x_1, \ldots, x_n) - 1, \inf(x_1, \ldots, x_n)].$$

On en déduit que tout point de l'intervalle $[\sup(x_1, \ldots, x_n) - 1, \inf(x_1, \ldots, x_n)]$ est un *emv* $\theta^*$ de $\theta$ pour $(x_1, \ldots, x_n)$.