Université de Strasbourg
M1 - Magistère
Statistique - études de cas

Ségolen Geffray
geffray@math.unistra.fr
Année 2017/2018

Feuille d'exercices: estimation paramétrique

Rappels sur la fonction Gamma:

$$\Gamma(u) = \int_0^\infty e^y y^{u-1} dy, \quad u > 0.$$

La fonction $\Gamma(.)$ est c^∞ sur $\mathbb{R}^{*+}$ et ses dérivées sont données par:

$$\Gamma^{(k)}(u) = \int_0^\infty (\log y)^k \, e^y y^{u-1} dy, \quad u > 0.$$

En voici quelques valeurs remarquables:

$$\begin{aligned} &\Gamma(1) = 1, \qquad &&\Gamma'(1) = -\gamma, \qquad &&\Gamma''(1) = \gamma^2 + \frac{\pi^2}{6}, \\ &\Gamma(2) = 1, \qquad &&\Gamma'(2) = 1 - \gamma, \qquad &&\Gamma''(2) = (1-\gamma)^2 + \frac{\pi^2}{6} - 1, \end{aligned}$$

exprimées au moyen de la constante d'Euler notée γ:

$$\gamma = -\int_0^\infty \log(y) \, e^y dy \approx 0.577\,215.$$

• **Inégalité de Bienaymé-Chebychev**:
Soit X une variable aléatoire admettant un moment d'ordre 2 et soit $a > 0$. On a

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq a) \leq \frac{\mathrm{Var}(X)}{a^2}\,.$$

• **Inégalité de Hoeffding**:
Soit une suite $(X_k)_{1\leq k\leq n}$ de variables aléatoires réelles indépendantes vérifiant, pour deux suites $(a_k)_{1\leq k\leq n}$ et $(b_k)_{1\leq k\leq n}$ de nombres réels tels que $a_k < b_k$ $\forall k$ et $\mathbb{P}(a_k \leq X_k \leq b_k) = 1$, $\forall k$. On pose $S_n = X_1 + X_2 + \cdots + X_n$. Alors, pour tout $t > 0$, on a:

$$\begin{aligned} \mathbb{P}\left(S_n - \mathbb{E}[S_n] \geq t\right) &\leq \exp\left(-\frac{2\,t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right), \\ \mathbb{P}\left(S_n - \mathbb{E}[S_n] \leq -t\right) &\leq \exp\left(-\frac{2\,t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right), \\ \mathbb{P}\left(|S_n - \mathbb{E}[S_n]| \geq t\right) &\leq 2\exp\left(-\frac{2\,t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right). \end{aligned}$$

Exercice 1.
On s'intéresse à la variable de Bernoulli qui représente la capacité qu'a un travailleur salarié de rembourser ou non son emprunt. On suppose que cette capacité s'écrit sous la forme

$$X = I(Z > 1000)$$

où Z désigne la variable aléatoire qui représente le revenu mensuel une fois que l'on en a déduit les charges. On suppose de plus que Z suit une loi $\mathcal{N}(m, \sigma^2)$ de paramètres $m \in \mathbb{R}$ et $\sigma^2 > 0$ inconnus. Définir le modèle statistique pour l'observation de X. Ce modèle est-il identifiable?

Exercice 2.

1. Soit $N \in \mathbb{N}^*$ un entier fixé. Montrer que le modèle $(\mathcal{B}(N,p))_{p\in]0,1[}$ est complet.
2. On observe $\mathbf{X} = (X_1, \ldots, X_n)$ un n-échantillon de loi $\mathcal{B}(p)$ pour un p inconnu dans $]0,1[$. Montrer que $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ est une statistique exhaustive complète pour p.

Exercice 3.
On observe $\mathbf{X} = (X_1, \ldots, X_n)$ un n-échantillon de loi uniforme sur $\left[\theta - \frac{1}{2}, \theta + \frac{1}{2}\right]$ pour $\theta \in \mathbb{R}$.

1. Montrer que la statistique $Y_n = \max(X_1, ..., X_n) - \min(X_1, ..., X_n)$ est libre.
2. Montrer que la statistique $Z_n = (\min(X_1, ..., X_n), \max(X_1, ..., X_n))$ est exhaustive mais pas complète pour θ.
NB: on peut montrer qu'elle est néanmoins minimale.

Exercice 4.
On observe $\mathbf{X} = (X_1, \ldots, X_n)$ un n-échantillon de loi $\mathcal{N}(\theta, \theta^2)$ pour $\theta \in \mathbb{R}$. Montrer que la statistique $\left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2\right)$ est exhaustive mais pas complète pour θ.
NB: on peut montrer qu'elle est néanmoins minimale.

Exercice 5.
On observe $\mathbf{X} = (X_1, \ldots, X_n)$ un n-échantillon de loi $\mathcal{N}(\theta, \theta)$ pour $\theta > 0$. On note

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \text{et} \qquad S_n'^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 .$$

1. Comparer $\overline{X}_n$ et $S_n'^2$ au sens du risque quadratique pour l'estimation de θ.
2. Introduisons, pour $0 < \alpha < 1$,

$$T_\alpha = \alpha \overline{X}_n + (1-\alpha) S_n'^2 .$$

Montrer que T_α est sans biais pour θ quelque soit $0 < \alpha < 1$. Déterminer α de sorte que T_α soit de variance minimale.

Exercice 6.
On observe $\mathbf{X} = (X_1, \ldots, X_n)$ un n-échantillon de loi uniforme sur $\{1, 2, \ldots, \theta\}$ pour $\theta \in \mathbb{N}^*$.

1. Déterminer l'estimateur du maximum de vraisemblance de θ noté $\widehat{\theta}_n$.
2. Montrer que $\widehat{\theta}_n = \theta$ à partir d'un certain rang presque-sûrement pour tout $\theta > 0$.

Exercice 7.
Soit $(X_1, ..., X_n)$ un échantillon i.i.d. issu d'une variable parente X de loi $\mathcal{P}(\lambda)$ pour un $\lambda > 0$. Considérons le cas où X_i représente le nombre d'évènements indésirables sévères (EIS) dus à l'absorption d'un médicament l'année i, pour $i = 1, ..., n$. Déterminer un intervalle de confiance bilatéral asymptotique pour le nombre moyen d'EIS par an au niveau de confiance $(1 - \alpha)$. Déterminer un intervalle de confiance bilatéral à distance finie pour le nombre moyen d'EIS par an au niveau de confiance $(1 - \alpha)$ en utilisant l'inégalité de Tchebychev. Commenter.

Exercice 8.

1. Montrer que le modèle $(\mathcal{B}(p))_{p\in]0,1[}$ est complet.
2. Déterminer la matrice d'information de Fisher du modèle, si cela est possible.
3. Soit $(X_1, ..., X_n)$ un échantillon de loi $\mathcal{B}(p)$ pour un $p \in]0, 1[$. Exhiber une statistique exhaustive pour p. Est-elle complète? minimale?
4. Proposer un estimateur de p. Justifier.
5. Etablir son comportement asymptotique.
6. Déterminer un intervalle de confiance asymptotique bilatéral symétrique pour p au niveau de confiance $(1 - \alpha)$.
7. Déterminer un intervalle de confiance à distance finie bilatéral symétrique pour p au niveau de confiance $(1 - \alpha)$ en utilisant l'inégalité de Chebyshev.
8. Déterminer un intervalle de confiance à distance finie bilatéral symétrique pour p au niveau de confiance $(1 - \alpha)$ en utilisant l'inégalité de Hoeffding.

Exercice 9.
Soit $(X_1, ..., X_n)$ un échantillon i.i.d. de loi $\mathcal{P}(\lambda)$ pour un $\lambda > 0$.
Il s'agit d'estimer $\theta = \mathbb{P}(X_1 = 0)$.

1. Déterminer une statistique exhaustive pour θ notée T_n.
2. Déterminer l'estimateur du maximum de vraisemblance de θ noté $\widehat{\theta}_n$.
3. Montrer que $\widehat{\theta}_n$ est biaisé à distance finie.
4. Etablir le comportement asymptotique de $\widehat{\theta}_n$.
5. Déterminer l'estimateur amélioré de Rao-Blackwell de l'estimateur $I(X_1 = 0)$ par conditionnement avec T_n.

Exercice 10.
On observe $\mathbf{X} = (X_1, \ldots, X_n)$ un n-échantillon de loi uniforme sur $[0; \theta]$ pour $\theta > 0$.

1. Le modèle est-il dominé, homogène, exponentiel, régulier?
2. Déterminer une statistique exhaustive pour θ. Est-elle complète? minimale?
3. Déterminer l'estimateur du maximum de vraisemblance de θ noté $\widehat{\theta}_n$.
4. Déterminer l'estimateur de la méthode des moments de θ noté $\widetilde{\theta}_n$.
5. Calculer le biais et le risque quadratique de $\widehat{\theta}_n$ et de $\widetilde{\theta}_n$.
6. Déterminer le comportement asymptotique des différents estimateurs.
7. Quel(s) estimateur(s) préférez-vous?
8. Déterminer un intervalle de confiance asymptotique bilatéral symétrique de niveau $(1-\alpha)$ pour θ en utilisant chacun des estimateurs précédents. Que remarquez-vous?
9. Déterminer un intervalle de confiance à distance finie bilatéral symétrique de niveau $(1-\alpha)$ pour θ en utilisant tour à tour l'inégalité de Tchebychev et l'inégalité de Hoeffding. Que remarquez-vous?

Exercice 11.
Soit $(X_1, ..., X_n)$ un échantillon de variables aléatoires i.i.d. de variable parente X dont la distribution admet la densité suivante par rapport à la mesure de Lebesgue sur $\mathbb{R}$

$$f_\theta(x) = \frac{2x}{\theta^2} I(0 \leq x \leq \theta) \quad \text{avec } \theta > 0.$$

1. Vérifier qu'il s'agit bien d'une densité de probabilité. Le modèle est-il dominé, homogène, exponentiel, régulier?
2. Déterminer la fonction de répartition associée ainsi que l'espérance et la variance de X.
3. Déterminer l'information de Fisher du modèle probabiliste (à une observation).
4. Montrer que la statistique $\max(X_1, ..., X_n)$ est exhaustive minimale pour θ.
5. Déterminer $\widetilde{\theta}_n$ l'estimateur de θ obtenu par la méthode des moments.
6. Déterminer $\widehat{\theta}_n$ l'estimateur de θ obtenu par la méthode du maximum de vraisemblance.
7. Déterminer le biais, la variance et le risque quadratique de $\widetilde{\theta}_n$ et $\widehat{\theta}_n$. Comparer $\widetilde{\theta}_n$ et $\widehat{\theta}_n$.
8. La borne de Fréchet-Darmois-Cramer-Rao s'applique-t-elle ici? Pourquoi?
9. Etudier le comportement asymptotique de $\widetilde{\theta}_n$ et de $\widehat{\theta}_n$.
10. Quel estimateur préférez-vous? Justifier.
11. Déterminer un intervalle de confiance asymptotique bilatéral symétrique de niveau $(1-\alpha)$ pour θ.

Exercice 12.
Un assureur s'intéresse aux sinistres que peuvent subir ses clients afin d'établir sa politique d'assurance. Soit X la variable aléatoire représentant le montant annuel (exprimé en milliers d'euros) des sinistres que subit un client pris au hasard. On suppose que X suit la loi de Pareto dont la distribution admet la densité suivante par rapport à la mesure de Lebesgue sur $\mathbb{R}$

$$f_\theta(x) = \frac{\theta}{x^{\theta+1}} I(x \geq 1).$$

où θ est un paramètre inconnu que l'on veut estimer. On dispose pour cela d'un échantillon i.i.d. $(X_1, ..., X_n)$ distribué comme X.

1. Quel est le support de la loi de X? Quel est l'espace naturel des paramètres noté Θ? Le modèle est-il régulier?
2. Calculer l'espérance de X et en déduire pourquoi on ne peut pas appliquer la méthode des moments pour estimer θ?
3. Déterminer l'information de Fisher du modèle probabiliste (à une observation) lorsque c'est possible.
4. Déterminer $\widehat{\theta}_n$ l'estimateur du maximum de vraisemblance de θ. Est-il biaisé? Est-il fortement consistant? Etudier son comportement asymptotique.
5. Proposer un estimateur de $\mathbb{P}[X > 1000]$. Justifier votre choix.
6. Déterminer un intervalle de confiance unilatéral asymptotique au niveau de confiance 95% pour $\mathbb{P}[X > 1000]$.

Exercice 13.
Soient $X_1, ..., X_n$ des variables aléatoires i.i.d. de variable parente X. On suppose que la distribution de X admet la densité suivante par rapport à la mesure de Lebesgue sur $\mathbb{R}$

$$f_\theta(x) = (1 + \theta)I(0 \leq x \leq 1/2) + (1 - \theta)I(1/2 \leq x \leq 1)$$

où θ est un paramètre inconnu que l'on veut estimer.

1. Quel est le support de la loi de X noté $X(\Omega)$? Quel est l'espace naturel des paramètres noté Θ? Le modèle est-il régulier?
2. Déterminer l'espérance et la variance de X.
3. Déterminer l'information de Fisher du modèle probabiliste (à une observation) lorsque c'est possible.
4. Déterminer $\widetilde{\theta}_n$ l'estimateur de θ obtenu par la méthode des moments. Est-il biaisé? Est-il fortement consistant? Etudier son comportement asymptotique.
5. Déterminer $\widehat{\theta}_n$ l'estimateur du maximum de vraisemblance de θ. Est-il biaisé? Est-il fortement consistant? Etudier son comportement asymptotique.
6. Comparer $\widetilde{\theta}_n$ et $\widehat{\theta}_n$.

7. Déterminer un intervalle de confiance bilatéral symétrique de niveau $(1-\alpha)$ pour θ.

8. Proposer un estimateur de $\mathbb{P}[0 \leq X \leq 1/4]$. Justifier votre choix.

9. Déterminer un intervalle de confiance asymptotique au niveau de confiance 95% pour $\mathbb{P}[0 \leq X \leq 1/4]$.

Exercice 14.
Soit $(X_1, ..., X_n)$ un échantillon i.i.d. de variable parente X de loi exponentielle de paramètre $\lambda > 0$. On souhaite estimer le paramètre $\theta = \mathbb{P}(X \leq t_0)$ où $t_0 > 0$ est fixé et connu.

1. Déterminer l'estimateur de θ par la méthode des moments et étudier ses propriétés.

2. Déterminer l'estimateur de λ par la méthode du maximum de vraisemblance et étudier ses propriétés. Vous obtiendrez le comportement asymptotique de l'estimateur par deux méthodes différentes.

3. Déduire de la question précédente l'estimateur de θ par la méthode du maximum de vraisemblance et étudier ses propriétés.

4. Exhiber une statistique exhaustive pour θ.

5. Déterminer l'amélioré de Rao-Blackwell de l'estimateur $\widehat{\eta} = I(X_1 \leq t_0)$ de θ.

   Indication: on admettra que la loi de X_1 sachant que $\displaystyle\sum_{i=1}^{n} X_i = y$ pour $y > 0$ est la loi du minimum d'un échantillon i.i.d. $(U_1, ..., U_{n-1})$ de loi uniforme sur $[0, y]$.

6. Déterminer un intervalle de confiance asymptotique bilatéral symétrique de niveau $(1-\alpha)$ pour θ.

Exercice 15.
Un laboratoire pharmaceutique met en place la production d'un nouveau médicament sous forme de comprimé. Il faut déterminer le dosage du principe actif dans chaque comprimé. Pour cela, plusieurs techniques industrielles sont envisageables. Toutes fournissent un dosage en principe actif (exprimé en μg) qui suit une loi gaussienne de paramètres notés $\mu \in \mathbb{R}$ et $\sigma^2 > 0$. Si la teneur en principe actif est trop importante, le médicament peut devenir dangereux. Le laboratoire considère ainsi que le dosage doit être inférieur à $60\mu g$ avec une probabilité supérieure ou égale à 0.999. Inversement, si la teneur en principe actif est trop faible, le médicament n'est pas efficace. Le laboratoire considère ainsi que le dosage doit être supérieur à $40\mu g$ avec une probabilité supérieure ou égale à 0.95.

1. Que représentent ici les paramètres μ et σ?

2. Traduire les exigences du laboratoire concernant la sécurité et l'efficacité des comprimés.

3. En déduire un système de deux inégalités pour μ et σ.

4. A quel intervalle, dépendant de σ doit alors appartenir μ? En déduire la valeur maximale possible pour le choix de σ notée $\sigma_{\max}$.

5. Plus une technique industrielle de dosage est précise, plus elle est chère à mettre en oeuvre. Pour minimiser les coûts de production, le laboratoire a intérêt à choisir une technique industrielle associée au σ le plus grand possible tout en respectant les exigences de sécurité et d'efficacité. Supposons que le laboratoire choisit une technique industrielle associée à $\sigma_{\max}$. Quelle valeur du dosage moyen devra-t-il choisir pour respecter les exigences? On note μ_{opt} cette valeur.

6. La direction du laboratoire pharmaceutique affirme donc calibrer son procédé industriel de dosage sur μ_{opt} et $\sigma_{\max}$. L'ingénieur d'usine décide un contrôle qualité. Il teste la teneur en principe actif de 40 comprimés. Il obtient une moyenne empirique de $45\mu g$ et un écart-type empirique de $5\mu g$. Que peut-il conclure de ces observations au sujet de la teneur moyenne en principe actif?

Exercice 16.

Pour chacun des modèles suivants, donner le support de la loi de X et l'espace naturel des paramètres noté Θ. Déterminer si le modèle est dominé (auquel cas exhiber une mesure dominante et donner la dérivée de Radon-Nikodym), homogène, exponentiel, régulier.

1. $F_\theta(x) = \begin{cases} 1 - e^{-x^\theta} & \text{si } x \geq 0 \\ 0 & \text{sinon} \end{cases} \qquad (\theta > 0);$

2. $F_{a,b}(x) = \begin{cases} 0 & \text{si } x < a \\ \dfrac{x-a}{b-a} & \text{si } a \leq x < b \\ 1 & \text{si } b \leq x \end{cases} \qquad (a < b);$

3. $F_\theta(x) = \begin{cases} 0 & \text{si } x < 0 \\ \dfrac{1}{2}(1 - e^{-\theta x}) & \text{si } 0 \leq x < 1/\theta \\ 1 - \dfrac{1}{2}e^{-\theta x} & \text{si } 1/\theta \leq x \end{cases} \qquad (\theta > 0);$

4. $F_{\alpha,\lambda}(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1 - e^{-\lambda x^\alpha} & \text{si } 0 \leq x < 1/\lambda^{1/\alpha} \\ 1 & \text{si } 1/\lambda^{1/\alpha} \leq x \end{cases} \qquad (\alpha > 0,\, \lambda > 0);$

5. $F_\theta(x) = \begin{cases} \displaystyle\sum_{k=0}^{\lfloor x \rfloor} \frac{\theta^k}{k!} e^{-\theta} & \text{si } x \geq 0 \\ 0 & \text{sinon} \end{cases} \qquad (\theta > 0);$

6. $\mathbb{P}_{a,\lambda}[X = x] = e^{-\lambda} \dfrac{\lambda^{x-a}}{(x-a)!}$ pour $x \in \{a, a+1, ...\}$ avec $\lambda > 0$ et $a \in \mathbb{N}$;

7. $F_{\alpha,\lambda}(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1 - e^{-\alpha - \lambda x} & \text{si } x \geq 0 \end{cases}$ avec $\alpha, \lambda > 0$;

8. $F_a(x) = \begin{cases} 0 & \text{si } x < 0 \\ x & \text{si } 0 \leq x < a \\ \dfrac{(a-1)x - a}{a-2} & \text{si } a \leq x < 2 \\ 1 & \text{si } x \geq 2 \end{cases}$ avec $1,5 \leq a < 2$;

9. P_θ est la distribution de $X = (Y - \theta)I(Y > \theta)$ où Y suit la loi normale centrée réduite et où $\theta \in \mathbb{R}$;

10. P_θ est la distribution de $X = I(Y \leq 1) + YI(Y > 1)$ où Y suit la loi uniforme sur $[0;\, 1+\theta]$ pour $\theta > 0$;

11. P_θ avec $\theta = (m, \sigma^2)$ est la distribution de $X = I(Y \leq 1) + YI(Y > 1)$ où Y suit la loi gaussienne de paramètres $\mu \in \mathbb{R}$ et $\sigma^2 > 0$;

12. P_θ est la distribution de la variable aléatoire X définie comme suit avec $\theta = (p, m, \sigma^2)$. Soit $a \in \mathbb{R}$ fixé connu et soit U une variable aléatoire de loi $\mathcal{N}(m, \sigma^2)$ où $m \in \mathbb{R}$ et $\sigma^2 > 0$. On pose $X = a$ avec probabilité p où $0 < p < 1$ et $X = U$ sinon;

13. P_θ est la distribution de $X = cZ + Y(1 - Z)$ où $c \in \mathbb{R}$ est fixé connu, où Y et Z sont deux variables aléatoires indépendantes, où Y suit la loi normale centrée réduite et où Z suit une loi de Bernoulli de paramètre θ pour $0 < \theta < 1$;

14. $P_\theta = \theta\delta_{\{c\}} + (1 - \theta)\mu$ où $\delta_{\{c\}}$ est la mesure de Dirac en c (fixé connu) et où $\forall A \subset \mathcal{B}(\mathbb{R})$, $\mu(A) = \displaystyle\int_A \frac{1}{2} e^{-|x|} dx$;

15. P_θ est absolument continue par rapport à la mesure de Lebesgue sur $\mathbb{R}$ de densité f_θ donnée par
$$f_\theta(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x - \theta^2)^2 \right).$$

Exercice 17.
Un système électrique comporte deux diodes montées en série de sorte que le système est en panne lorsque l'une ou l'autre des diodes cesse de fonctionner. La première diode est de type A et on suppose que sa durée de fonctionnement suit une loi exponentielle de paramètre $\lambda_A > 0$. La deuxième diode est de type B et on suppose que sa durée de fonctionnement suit une loi exponentielle de paramètre $\lambda_B > 0$. On suppose enfin que les deux durées de fonctionnement sont indépendantes. On met n systèmes au banc d'essai. Lorsqu'un système tombe en panne, on observe sa durée de fonctionnement modélisée par une variable aléatoire notée Y ainsi que la cause de la panne modélisée par une variable aléatoire notée D.

1. Exprimer Y et D en fonction de la durée de fonctionnement des diodes de type respectif A et B.

2. Montrer que Y et D sont indépendantes.

3. Préciser le modèle statistique pour (Y, D). Est-il dominé? homogène?

4. Donner la fonction de vraisemblance d'un n-échantillon.

5. Déterminer l'estimateur du maximum de vraisemblance de $\theta = (\lambda_A, \lambda_B)^t$ noté $\widehat{\theta}_n$.

6. Etudier le comportement asymptotique de $\widehat{\theta}_n$.

Exercice 18.
Un laboratoire indépendant est chargé par l'Office de protection des consommateurs de vérifier la résistance à l'éclatement (en kg/cm^2) d'un réservoir de carburant d'un certain fabricant.

1. Des essais sont effectués sur un échantillon de 10 réservoirs conduisant à une résistance moyenne à l'éclatement de 219 kg/cm^2 avec un écart-type empirique de 10kg/cm^2. Proposer un intervalle de confiance pour la résistance moyenne à l'éclatement de ce type de réservoir avec un niveau de confiance de 95%. On admet ici que la résistance à l'éclatement est distribuée selon une loi normale.

2. Des essais sont effectués sur un échantillon de 100 réservoirs conduisant à une résistance moyenne à l'éclatement de 219 kg/cm^2 avec un écart-type empirique de 10kg/cm^2. Proposer un intervalle de confiance pour la résistance moyenne à l'éclatement de ce type de réservoir avec un niveau de confiance de 95%. Que constatez-vous?

Exercice 19.
On considère la famille de densités sur $\mathbb{R}^+$ à deux paramètres $u > 0$ et $\alpha > 0$:

$$f_{u,\alpha}(x) = \frac{\alpha^{3u}}{x^{1+u}\Gamma(u)} \exp\left(-\frac{\alpha^3}{x}\right) I_{]0,+\infty[}(x).$$

1. Vérifier qu'il s'agit d'un modèle exponentiel.

2. Utiliser la statistique exhaustive que l'on notera T pour évaluer $\mathrm{E}[\log X]$ et $\mathbb{E}\left[\frac{1}{X}\right]$.

3. Quelle est la loi de $Y = \alpha^3/X$? Evaluer $\mathbb{E}[Y]$ et $\mathrm{Var}(Y)$.

4. Calculer directement (sans passer par le paramétrage canonique) la matrice d'information de Fisher $I(u,\alpha)$ des paramètres u et α en fonction de Γ et de ses dérivées. On notera I^{-1} la matrice inverse sans chercher à la calculer.

5. On dispose de n observations $(X_1, \ldots, X_n)$ i.i.d. de densité $f_{u,\alpha}$. Quel système d'équations vérifient les estimateurs du maximum de vraisemblance $\hat{u}_n$ et $\hat{\alpha}_n$ de u et α? On ne cherchera pas à résoudre ce système d'équations.

6. Quelle est la loi limite (après renormalisation convenable) du vecteur $\begin{pmatrix} \hat{u}_n - u \\ \hat{\alpha}_n - \alpha \end{pmatrix}$?

7. Construire un intervalle de confiance asymptotique bilatéral symétrique pour $\mathbb{E}\left[\frac{1}{X}\right]$ au niveau de confiance $(1-\alpha)$.

Exercice 20.
Soit X une variable aléatoire dont la distribution admet la densité suivante par rapport à la mesure de Lebesgue sur $\mathbb{R}$ pour $\lambda, \mu > 0$:

$$f_{\mu,\lambda}(x) = \frac{\sqrt{\lambda}}{\sqrt{2\pi x^3}} \exp\left(-\lambda \frac{(x-\mu)^2}{2\mu^2 x}\right) I(x > 0).$$

Afin d'estimer les paramètres inconnus μ et λ, on dispose d'un échantillon $(X_1, ..., X_n)$ i.i.d. distribué comme X.

1. Quel est le support de la loi de X? Quel est l'espace naturel du paramètre bidimensionnel $\theta = (\mu, \lambda)^t$ noté Θ?
2. Montrer que le modèle appartient à la famille exponentielle. En déduire l'espérance et la variance de X et de $1/X$. Exhiber une statistique exhaustive pour θ. Le modèle est-il régulier?
3. Déterminer l'information de Fisher du modèle probabiliste (à une observation) lorsque c'est possible.
4. Déterminer $\widetilde{\theta}_n$ l'estimateur de θ obtenu par la méthode des moments. Est-il fortement consistant? Etudier son comportement asymptotique.
5. Déterminer $\widehat{\theta}_n$ l'estimateur du maximum de vraisemblance de θ. Est-il fortement consistant? Etudier son comportement asymptotique.
6. En déduire un estimateur consistant de $\mathbb{E}[X]$ de variance minimale. Etudier son comportement asymptotique. En déduire un intervalle de confiance asymptotique au niveau de confiance $(1-\alpha)$ pour $\mathbb{E}[X]$.

Exercice 21.
On considère le modèle admettant la densité suivante par rapport à la mesure de Lebesgue, pour $\mu > 0$ et $\sigma > 0$:

$$f_{\mu,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi x^3}} \exp\left(\frac{-(\mu x - 1)^2}{2\sigma^2 x}\right) I(x > 0).$$

1. Montrer que ce modèle peut se déduire du modèle exponentiel canonique via la transformation $\eta = Q(\mu, \sigma) = (\mu^2/\sigma^2, 1/\sigma^2)^t$.
2. Déterminer la matrice d'information $I^*(\eta)$ associée au modèle canonique puis $I^*(Q(\mu, \sigma))$.
3. En déduire $I(\mu, \sigma)$ la matrice d'information du modèle.
4. On dispose à présent de n observations associées à un n-échantillon de loi $f_{\mu,\sigma}$.
   (a) Déterminer une statistique exhaustive minimale pour $(\mu, \sigma)^t$.
   (b) Déterminer $\widehat{\eta}_n$ l'estimateur du maximum de vraisemblance du paramètre canonique.
   (c) En déduire les estimateurs du maximum de vraisemblance $\widehat{\mu}_n$ et $\widehat{\sigma}^2_n$ de μ et σ^2 (vérifier que $\widehat{\sigma}^2_n$ est toujours positif).
   (d) En déduire l'estimateur du maximum de vraisemblance $\widehat{\sigma}_n$ de σ.
   (e) Déterminer le comportement asymptotique de
   $$\sqrt{n}(\widehat{\mu}_n - \mu, \widehat{\sigma}_n - \sigma)^t\,.$$
   L'estimateur $(\widehat{\mu}_n, \widehat{\sigma}_n)^t$ est-il asymptotiquement efficace?
   (f) On suppose que $\sigma = 1$. L'estimateur $\widehat{\mu}_n$ est-il asymptotiquement efficace? Justifier. Même question pour $\widehat{\sigma}_n$ si $\mu = 3$.

Exercice 22.

1. Soit X une variable aléatoire de loi $P_{\alpha,\lambda}$ admettant la densité $f_{\alpha,\lambda}$ pour $\alpha > 0$ et $\lambda > 0$ par rapport à la mesure de Lebesgue sur $\mathbb{R}^+$ définie par
$$f_{\alpha,\lambda}(x) = \alpha\lambda^\alpha x^{-\alpha-1} I(x \geq \lambda)\,.$$
Le modèle $(P_{\alpha,\lambda})_{(\alpha,\lambda)\in(\mathbb{R}^{*+})^2}$ est-il homogène, exponentiel, régulier?

2. Soit Y une variable aléatoire de loi admettant la densité $f_{\alpha,1}$ par rapport à la mesure de Lebesgue sur $\mathbb{R}^+$. Le modèle $(P_{\alpha,1})_{\alpha\in\mathbb{R}^{*+}}$ est-il homogène, exponentiel, régulier?

3. Déterminer $\mathbb{E}[\log Y]$ et $\mathrm{Var}(\log Y)$.

4. Montrer que si X suit la loi de densité $f_{\alpha,\lambda}$, alors $Y = \dfrac{X}{\lambda}$ suit la loi de densité $f_{\alpha,1}$. En déduire $\mathbb{E}[\log X]$ et $\mathrm{Var}(\log X)$.

5. Calculer si possible la matrice d'information de Fisher du modèle $(P_{\alpha,\lambda})_{(\alpha,\lambda)\in(\mathbb{R}^{*+})^2}$.

6. Soit $(X_1, ..., X_n)$ un échantillon i.i.d. de variable parente X. Déterminer l'estimateur du maximum de vraisemblance de (α, λ) noté $(\widehat{\alpha}_n, \widehat{\lambda}_n)$.

7. Etablir le comportement asymptotique de $\widehat{\lambda}_n$.

8. Etablir le comportement asymptotique de $1/\widehat{\alpha}_n$. En déduire celui de $\widehat{\alpha}_n$.

9. Pourquoi ne peut-on pas utiliser la méthode des moments?

Exercice 23.

Soit $(X_1, ..., X_n)$ un échantillon i.i.d. de variable parente X de loi de Gumbel définie par sa fonction de répartition $F_{\alpha,\beta}$, paramétrée par $\theta = \begin{pmatrix}\alpha\\ \beta\end{pmatrix}$ avec $\alpha \in \mathbb{R}$ et $\beta > 0$, de la façon suivante:
$$F_{\alpha,\beta}(x) = \exp\left(-\exp\left(-\frac{x-\alpha}{\beta}\right)\right), \quad x \in \mathbb{R}.$$
La loi de Gumbel est très fréquemment utilisée en hydrologie, par exemple, lorsque X modélise le niveau maximal annuel de crue d'un fleuve.

1. Vérifier qu'il s'agit bien d'une fonction de répartition. Justifier l'existence d'une densité. Déterminer l'expression de cette densité. Le modèle est-il homogène? Appartient-il à la famille exponentielle?

2. Calculer les quantités suivantes:
$$\mathbb{E}_\theta\left[\left(\frac{X-\alpha}{\beta}\right)^k\right] \qquad \text{pour } k = 1, 2, 3, 4 \quad \text{ et } \quad \mathrm{Var}_\theta\left(\frac{X-\alpha}{\beta}\right).$$

3. Déterminer l'estimateur de $\theta = \begin{pmatrix}\alpha\\ \beta\end{pmatrix}$ par la méthode des moments. On notera $\widetilde{\theta}_n = \begin{pmatrix}\widetilde{\alpha}_n\\ \widetilde{\beta}_n\end{pmatrix}$ cet estimateur.

4. Etudier son comportement asymptotique.
5. Montrer que l'estimateur du maximum de vraisemblance de β satisfait l'équation suivante:

$$\sum_{i=1}^{n}(X_i - \overline{X}_n + \beta)\exp\left(-\frac{X_i}{\beta}\right) = 0$$

en notant $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Quelle est l'équation satisfaite par l'estimateur du maximum de vraisemblance de α?

On notera $\widehat{\theta}_n = \begin{pmatrix}\widehat{\alpha}_n \\ \widehat{\beta}_n\end{pmatrix}$ l'estimateur du maximum de vraisemblance de θ.

6. Calculer pour $k = 0, 1, 2$:

$$\mathbb{E}_\theta\left[\left(\frac{X-\alpha}{\beta}\right)^k \exp\left(-\frac{X-\alpha}{\beta}\right)\right].$$

7. En admettant que le modèle est régulier, calculer la matrice d'information de Fisher (dans le modèle probabiliste à une observation).
8. Que peut-on dire du comportement asymptotique de $\widehat{\theta}_n$ si l'on sait trouver une solution au problème d'optimisation du maximum de vraisemblance qui soit consistante pour θ (toujours en admettant que le modèle est régulier)?
9. On se donne une probabilité $p \in]0, 1[$ (par exemple 10^{-4}). Proposer un estimateur de x_p défini par la relation $\mathbb{P}(X > x_p) = p$. Justifier. Construire un intervalle de confiance unilatéral pour x_p au niveau de confiance $(1 - \alpha)$.
10. Soit $x_0 > 0$ fixé quelconque. Proposer un estimateur de $q = \mathbb{P}(X > x_0)$. Justifier. Construire un intervalle de confiance unilatéral pour q au niveau de confiance $(1 - \alpha)$.
11. On considère dans cette question le cas particulier où $\beta = 1$.
    (a) Le modèle est-il exponentiel?
    (b) Déterminer une statistique exhaustive pour α dans le modèle à n observations. Est-elle complète? minimale?
    (c) Déterminer l'estimateur du maximum de vraisemblance de α noté $\widehat{\alpha}_n$.
    (d) Etudier son comportement asymptotique.
    (e) Construire un intervalle de confiance bilatéral symétrique asymptotique pour α au niveau de confiance 95%.

Exercice 24.
On considère la famille de densités à deux paramètres inconnus $\alpha \in \mathbb{R}$ et $\beta > 0$:

$$f_{\alpha,\beta}(x) = \frac{1}{\beta}\exp\left(-\frac{x-\alpha}{\beta}\right) I(x \geq \alpha).$$

1. La densité $f_{\alpha,\beta}$ appartient-elle à la famille exponentielle?

2. Soit X de densité $f_{\alpha,\beta}$. Déterminer les moments d'ordre k de X pour $k = 1, 2, 3, 4$ et calculer $\mathrm{Var}(X)$.
   Indication: si Y est une variable aléatoire de loi exponentielle de paramètre égal à 1, alors $\mathbb{E}[Y^3] = 6$ et $\mathbb{E}[Y^4] = 24$.

3. Soit $(X_1, \dots, X_n)$ un échantillon i.i.d. de densité $f_{\alpha,\beta}$. Déterminer l'estimateur de la méthode des moments $\left(\widetilde{\alpha}_n, \widetilde{\beta}_n\right)$ de (α, β).

4. Etablir la consistance et le comportement asymptotique de $\left(\widetilde{\alpha}_n, \widetilde{\beta}_n\right)$.

5. Déterminer l'estimateur du maximum de vraisemblance $\left(\widehat{\alpha}_n, \widehat{\beta}_n\right)$ de (α, β).

6. Etablir la consistance et le comportement asymptotique de $\widehat{\alpha}_n$.

7. Etablir la consistance et le comportement asymptotique de $\widehat{\beta}_n$.

8. Le modèle est-il régulier?

Exercice 25.

Rappel: Notons $\widehat{F}_n(.)$ est la fonction de répartition empirique de F déterminée à partir d'un n-échantillon i.i.d. distribué selon F. Définissons le quantile théorique d'ordre $p \in]0, 1[$ par

$$F^{-1}(p) = \inf\{t \in \mathbb{R} : F(t) \geq p\}$$

et le quantile empirique d'ordre $p \in]0, 1[$ par

$$\widehat{F}_n^{-1}(p) = \inf\{t \in \mathbb{R} : \widehat{F}_n(t) \geq p\}\,.$$

On admettra que la convergence suivante a lieu presque-sûrement lorsque $n \to \infty$:

$$\sup_{0<p<1} \left| \widehat{F}_n^{-1}(p) - F^{-1}(p) \right| \to 0\,.$$

De plus, si la fonction de répartition F est continue et dérivable en $F^{-1}(p)$ pour $0 < p < 1$, de dérivée $f(F^{-1}(p)) > 0$, alors la convergence suivante a lieu lorsque $n \to \infty$:

$$\sqrt{n}\left(\widehat{F}_n^{-1}(p) - F^{-1}(p)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{p(1-p)}{f(F^{-1}(p))^2}\right)\,.$$

Soit $(X_1, ..., X_n)$ un échantillon de variables aléatoires i.i.d. de variable parente X dont la distribution admet la densité suivante par rapport à la mesure de Lebesgue sur $\mathbb{R}$:

$$f_\mu(x) = \frac{1}{2}\exp\left(-|x-\mu|\right) \quad \text{avec } \mu \in \mathbb{R}.$$

1. Vérifier qu'il s'agit bien d'une densité de probabilité.

2. Le modèle appartient-il à la famille exponentielle?

3. Déterminer l'espérance, la variance et la médiane de X.

4. Déterminer l'information de Fisher du modèle probabiliste (à une observation).

5. Déterminer $\widetilde{\mu}_n$ l'estimateur de μ obtenu par la méthode des moments. Est-il biaisé? Etablir son comportement asymptotique.

6. Déterminer $\widehat{\mu}_n$ l'estimateur de μ obtenu par la méthode du maximum de vraisemblance. Etablir son comportement asymptotique.

7. Recommencer avec la famille de densités à deux paramètres $\lambda > 0$ et $\alpha \in \mathbb{R}$ donnée pour $x \in \mathbb{R}$ par:
$$g_{\alpha,\lambda}(x) = \frac{\lambda}{2} \exp(-\lambda|x-\alpha|).$$

Exercice 26.
Soient n observations indépendantes $X_1, \ldots, X_n$. Pour $i = 1, \ldots, n$, on suppose que X_i suit la loi $\mathcal{N}(\mu\alpha_i, \sigma^2)$ où $(\alpha_i)_{i\geq 1}$ une suite connue de réels non nuls telle que $v(n) = \sum_{i=1}^{n} \alpha_i^2$ tende vers l'infini lorsque n tend vers l'infini. Le but est d'estimer les paramètres inconnus μ et σ^2.

1. (a) Déterminer l'estimateur des moindres carrés $\widetilde{\mu}_n$, c'est à dire la quantité qui minimise $r \to \sum_{i=1}^{n}(X_i - r\alpha_i)^2$ lorsque r varie dans $\mathbb{R}$.

   (b) Montrer qu'il existe une constante C telle que
   $$v(n)\mathbb{E}\left[(\widetilde{\mu}_n - \mu)^2\right] = C\,.$$
   Quelle est la vitesse de $\widetilde{\mu}_n$ lorsque $\alpha_i = 1$? Lorsque $\alpha_i = i$?

   (c) En remarquant que les variables ε_i définies pour $i = 1, \ldots, n$ par $\varepsilon_i = X_i - \mu\alpha_i$ forment une suite i.i.d. de loi $\mathcal{N}(0, \sigma^2)$, proposer un estimateur naturel $\widetilde{\sigma}_n^2$ de σ^2. Montrer que $\widetilde{\sigma}_n^2$ converge dans $\mathbb{L}_1$ vers σ^2.

2. (a) Montrer que ce modèle à n observations peut encore se déduire du modèle exponentiel canonique via la transformation
   $$\eta = Q(\mu, \sigma^2) = \left(\frac{1}{2\sigma^2}, \frac{\mu}{\sigma^2}\right)^t.$$
   Exhiber une statistique exhaustive pour (μ, σ^2) dans le modèle à n observations.

   (b) Calculer $\mathrm{Hess}_\eta B_n(\eta)$. En déduire la matrice d'information $I_n^*(\eta)$ associée au modèle canonique (à n observations). Calculer $I_n^*(Q(\mu, \sigma^2))$.

   (c) Quelle est la relation liant $I_n(\mu, \sigma^2)$, la matrice d'information du modèle initial, à $(DQ)(\mu, \sigma^2)$ et $I_n^*(Q(\mu, \sigma^2))$? En déduire $I_n(\mu, \sigma^2)$. Que vaut $I_n(\mu, \sigma^2)$ lorsque les coefficients α_i sont tous égaux à 1.

   (d) Déterminer $\widehat{\eta}_n$ l'estimateur du maximum de vraisemblance dans le modèle canonique.

   (e) En déduire les estimateurs du maximum de vraisemblance $\widehat{\mu}_n$ et $\widehat{\sigma}_n^2$ de μ et σ^2. Comparer ces estimateurs à $\widetilde{\mu}_n$ et $\widetilde{\sigma}_n^2$. En déduire que $\widehat{\mu}_n$ et $\widehat{\sigma}_n^2$ sont des estimateurs consistants de μ et σ^2.

(f) Quel théorème permet d'affirmer que $\widehat{\mu}_n$ et $\widehat{\sigma}_n^2$ sont indépendantes?

(g) Quelle est la loi de $\sqrt{v(n)}(\widehat{\mu}_n - \mu)$?

(h) Trouver la loi limite de $\sqrt{n}(\widehat{\sigma}_n^2 - \sigma^2)$.

(i) Peut-on dire que l'estimateur $(\widehat{\mu}_n, \widehat{\sigma}_n^2)^t$ est un estimateur asymptotiquement efficace de $(\mu, \sigma^2)^t$?