Université de Pau et des Pays de l'Adour | **Semestre printemps 2023–2024**
L3 Math | Introduction à la Statistique

Feuille TD 3 : Intervalle de confiance et test statistique Correction

---

Exercice 3.1 Durée vie d'une ampoule

1. Déterminer l'estimateur du maximum de vraisemblance $\hat{\theta}^{MV}$ du paramètre θ, au vu d'observations $x_1, \cdots, x_n$ obtenues par un échantillon $X_1, \ldots, X_n$ régi par un une loi exponentielle $\mathcal{E}(\theta)$, $\theta > 0$, i.e.

$$f_\theta(x) = \theta \mathrm{e}^{-\theta x}, \text{ pour } x > 0.$$

La log-vraisemblance de ce modèle est

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

Donc en dérivant par rapport à θ, on trouve que

$$\partial_\theta \log L_n(x_1, \ldots, x_n; \theta) = \frac{n}{\theta} - \sum_{i=1}^{n} x_i.$$

Donc par une étude de fonction, on trouve que

$$\hat{\theta}_{MV} = \frac{n}{\sum_{i=1}^{n} X_i}.$$

2. **Application :** la durée de vie (en heures) d'un certain type d'ampoule est modélisée par une loi exponentielle de paramètre θ^* inconnu. On a testé la durée de vie d'un échantillon de 30 de ces ampoules et on a obtenu les résultats suivants :

5153.89	667.51	249.14	289.22	1529.90	1398.49	1611.51	1016.90
1284.25	1426.83	487.94	527.57	1681.54	1414.59	880.41	1847.25
60.64	2240.23	986.03	981.95	794.00	157.77	147.75	505.42
1168.15	287.22	33.26	2377.38	732.37	923.60		

La moyenne de cet échantillon est $\overline{x} = 1095.424$. Donner une estimation de θ^* au vu des résultats obtenus.

3. Construire un intervalle de confiance asymptotique de niveau $1 - \alpha$ pour θ.

Exercice 3.2 Paramètres d'un intervalle de confiance

La quantité de liquide dans un flacon est supposée suivre une loi $\mathcal{N}(\mu, \sigma^2)$, où $\sigma = 1.5$mL, et μ inconnue. On effectue une mesure sur 40 flacons, et on obtient une moyenne observée de 357 ml.

1. Donner un IC à 95% de μ. On appelle *marge d'erreur* d'un intervalle de confiance I la quantité

$$d = (\sup I - \inf I)/2.$$

On avait obtenu que

$$\mathrm{IC} = \left[\bar{X}_n \pm \frac{\sigma}{\sqrt{n}} q_{1-\alpha/2} \right].$$

2. Dans la question précédente, ce sont le nombre d'observations n et le niveau α qui ont déterminé la marge d. Sans changer n, trouver un niveau permettant d'avoir une marge $d = 0.375$.
La marge vaut

$$d = \frac{\bar{X}_n + \frac{\sigma}{\sqrt{n}} q_{1-\alpha/2} - \bar{X}_n + \frac{\sigma}{\sqrt{n}} q_{1-\alpha/2}}{2} = \frac{\sigma}{\sqrt{n}} q_{1-\alpha/2}.$$

Donc si $d = 0.375$, $\sigma = 1.5$ et $n = 40$ alors

$$q_{1-\alpha/2} = \frac{d\sqrt{n}}{\sigma} \approx 1.58.$$

On cherche sur la table de le quantile d'ordre $\beta = 1 - \alpha/2$ valant 1.58, et on voit que $\beta = 0.9429$. Donc $\alpha = 2(1-\beta) \approx 11.4\%$.
3. Maintenant, on souhaite avoir un IC de niveau de confiance $\alpha = 0.01$ et une marge d'erreur faible avec $d = 0.1$mL. Déterminer le nombre d'observations ?
Maintenant, c'est n que l'on souhaite faire varier donc

$$\sqrt{n} = \frac{q_{1-\alpha/2}\sigma}{d}.$$

En remplaçant, $d = 0.1$, $\sigma = 1.5$ et $q_{1-\alpha/2} \approx 2.58$, on a $n \approx 1496$.

Exercice 3.3

On considère n variables aléatoires $X_1, \ldots, X_n$ i.i.d suivant la même loi de densité :

$$f(x, \sigma^2) = \frac{x}{\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right) \qquad \text{pour } x \geqslant 0$$

où $\sigma > 0$. Cette loi est appelée loi de Rayleigh notée $\mathcal{R}(\sigma^2)$. Dans cet exercice, on cherche à étudier des estimateurs de $\theta = \sigma^2$ obtenus par la méthode des moments.
1. Soit $X \sim \mathcal{R}(\sigma^2)$. Montrer, à l'aide d'une intégration par parties, que

$$\mathbb{E}(X) = \sqrt{\frac{\pi\sigma^2}{2}}.$$

On rappelle que $\displaystyle\int_0^{+\infty} e^{-t^2} dt = \frac{\sqrt{\pi}}{2}$.
Dans la suite, on admettra que $\mathbb{E}(X^2) = 2\sigma^2$, $\mathbb{E}(X^3) = 3\sqrt{\frac{\pi\sigma^6}{2}}$ et $\mathbb{E}(X^4) = 8\sigma^4$.
2. Déterminer un estimateur $\hat{\theta}^{MM}$ de θ par la méthode des moments en considérant le moment d'ordre 1.
Pour trouver un estimateur par la méthode des moments, on va résoudre l'équation

$$\bar{X}_n = \sqrt{\frac{\pi\hat{\theta}_{MM}}{2}}.$$

D'où en mettant au carré et en normalisant

$$\hat{\theta}_{MM} = \frac{2}{\pi}\overline{X}_n^2.$$

3. À l'aide de la normalité asymptotique de $\overline{X}_n$, déterminer un intervalle de confiance asymptotique de θ de niveau de confiance $1 - \alpha$.

On a d'après le Théorème Centrale Limite,

$$\sqrt{n}\frac{\bar{X}_n - \sqrt{\pi\theta/2}}{\sqrt{\theta(4-\pi)/2}} \overset{loi}{\to} Z \sim \mathcal{N}(0,1).$$

Par le Lemme de Slutsky, on peut remplacer le θ au dénominateur par $\hat{\theta}_{MM}$ en multipliant par

$$\frac{\sqrt{\theta(4-\pi)/2}}{\sqrt{\hat{\theta}_{MM}(4-\pi)/2}} \overset{p.s.}{\to} 1.$$

Donc on a

$$\sqrt{n}\frac{\bar{X}_n - \sqrt{\pi\theta/2}}{\sqrt{\hat{\theta}_{MM}(4-\pi)/2}} \overset{loi}{\to} Z \sim \mathcal{N}(0,1).$$

Soit $\alpha \in (0,1)$, on a par la définition de la convergence en loi, et la symétrie de la loi normale

$$Q_\theta\left(\sqrt{n}\frac{\bar{X}_n - \sqrt{\pi\theta/2}}{\sqrt{\hat{\theta}_{MM}(4-\pi)/2}} \in [-q_{1-\alpha/2}, q_{1-\alpha/2}]\right) \to 1-\alpha,$$

où $q_{1-\alpha/2}$ est le quantile d'ordre $1-\alpha/2$ de la loi normale centrée réduite. Il reste à présent à "inverser" l'intervalle pour que le θ soit tout seul,

$$\begin{aligned}
\sqrt{n}\frac{\bar{X}_n - \sqrt{\pi\theta/2}}{\sqrt{\hat{\theta}_{MM}(4-\pi)/2}} \in [-q_{1-\alpha/2}, q_{1-\alpha/2}] &\Leftrightarrow \sqrt{\pi\theta/2} \in \left[\bar{X}_n \pm \frac{\sqrt{\hat{\theta}_{MM}(4-\pi)/2}}{\sqrt{n}} q_{1-\alpha/2}\right] \\
&\Leftrightarrow \sqrt{\pi\theta/2} \in \Big[\max\left(0, \hat{X}_n - \frac{\sqrt{\hat{\theta}_{MM}(4-\pi)/2}}{\sqrt{n}} q_{1-\alpha/2}\right) \\
&\quad \bar{X}_n + \frac{\sqrt{\hat{\theta}_{MM}(4-\pi)/2}}{\sqrt{n}} q_{1-\alpha/2}\Big], \text{ par positivité de la racine.} \\
&\Leftrightarrow \pi\theta/2 \in \Big[\max\left(0, \hat{X}_n - \frac{\sqrt{\hat{\theta}_{MM}(4-\pi)/2}}{\sqrt{n}} q_{1-\alpha/2}\right)^2 \\
&\quad \left(\bar{X}_n + \frac{\sqrt{\hat{\theta}_{MM}(4-\pi)/2}}{\sqrt{n}} q_{1-\alpha/2}\right)^2\Big], \text{ par crois. du carré sur } \mathbb{R}^+.
\end{aligned}$$

Pour finir, il reste à multiplier par $2/\pi$.

4. Montrer que

$$\mathbb{E}_\theta(\hat{\theta}^{MM}) = \frac{4+\pi(n-1)}{n\pi}\theta.$$

Que peut-on en déduire ?

On rappelle que $\hat{\theta}^{MM} = \frac{2}{\pi}\overline{X}_n^2$. Ainsi en développant le carré, on obtient une autre forme

$$\hat{\theta}^{MM} = \frac{2}{\pi n^2}\sum_{i=1}^n X_i^2 + \frac{2}{\pi n^2}\sum_{i\neq j} X_i X_j.$$

Pour $i \neq j$, l'indépendance entre X_i et X_j implique

$$\mathbb{E}_\theta[X_i X_j] = \mathbb{E}_\theta[X_i]\mathbb{E}_\theta[X_j] = \frac{\pi}{2}\theta.$$

Donc en se rappelant que le cardinale de $\{(i,j) \in \{1,\ldots,n\}^2 \mid i \neq j\}$ est $n(n-1)$, on a

$$\mathbb{E}_\theta(\hat{\theta}^{MM}) = \frac{4}{\pi n}\theta + \frac{2}{\pi n^2}\frac{\pi n(n-1)}{2}\theta = \frac{4+\pi(n-1)}{n\pi}\theta.$$

Cet estimateur est donc biaisé mais asymptotiquement sans biais.

Le calcul pour déterminer si $\hat{\theta}^{MM}$ est un estimateur efficace de θ est trop compliqué. On décide alors de déterminer un nouvel estimateur $\bar{\theta}^{MM}$ de θ par la méthode des moments mais en considérant le moment d'ordre 2.

5. Déterminer $\bar{\theta}^{MM}$.

On doit résoudre $U_2(n) = 2\bar{\theta}_{MM}$.

6. Déterminer un nouvel intervalle de confiance asymptotique de θ de niveau de confiance $1-\alpha$.

Semblable à la question 3.

7. Étudier le biais de $\bar{\theta}^{MM}$.

Sans biais.

8. $\bar{\theta}^{MM}$ est-il efficace ?

Pour tester l'efficacité, il faut vérifier si le risque de l'estimateur atteint la borne de Cramer-Rao. Il faut donc calculer le risque quadratique de l'estimateur ainsi que l'information de Fisher. Avec la décomposition biais/variance, on obtient que

$$\mathcal{R}(\bar{\theta}_{MM};\theta) = \theta^2/n.$$

Pour l'information de Fisher, comme les n observations sont indépendantes, on a

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

d'où

$$I_n(\theta) = n\mathrm{Var}_\theta(\partial_\theta \log L_1(X_1;\theta)) = n\mathrm{Var}_\theta(-1/\theta + X_1^2/2\theta^2) = n/\theta^2.$$

La borne de Cramer-Rao est l'inverse de cette quantité donc notre estimateur est bien efficace.

Pour calculer l'information de Fisher, on pouvait aussi dire que comme notre modèle est régulier, on a

$$I_1(\theta) = -\mathbb{E}_\theta[\partial_\theta^2 \log L_1(X_1;\theta)].$$

Exercice 3.4

Soit $X_1,\cdots,X_n$ un échantillon $i.i.d$ de loi parente $X \sim \mathcal{I}(\theta)$ dont la densité de probabilité est donnée par, pour tout $x \in \mathbb{R}_+^*$:

$$f_\theta(x) = \frac{1}{\theta x^2}e^{-\frac{1}{\theta x}}$$

où $\theta > 0$ est un paramètre inconnu.

1. Écrire le modèle statistique de l'échantillon. On admet que ce modèle est régulier.
2. Montrer, à l'aide d'un changement de variable adapté, que $\mathbb{E}(X) = +\infty$. Que peut-on en déduire pour la construction d'un estimateur de θ par la méthode des moments ?
3. Montrer que $\mathbb{E}(\frac{1}{X}) = \theta$ (*Indication : on pourra effectuer un changement de variable et une intégration par partie.*)
4. En déduire un estimateur $\hat{\theta}^{MM}$ de θ par la méthode des moments.
5. Montrer que $\mathrm{Var}\,(\frac{1}{X}) = \theta^2$.

6. Déterminer un intervalle de confiance asymptotique de θ d'un niveau de confiance $1-\alpha$ où $\alpha \in]0,1[$. On tâchera d'obtenir l'intervalle de confiance le plus précis possible.
7. Montrer que, pour tout $\theta > 0$, $I_n(\theta) = \dfrac{n}{\theta^2}$ où $I_n(\theta)$ est la quantité d'information de Fisher du modèle statistique de l'échantillon.
8. $\hat{\theta}^{MM}$ est-il un estimateur efficace de θ ?

Exercice 3.5 Vrai et Faux simultanément ?

On se place dans le modèle de n lancers de pièce indépendants $(\{0,1\}^n, (\mathbb{P}_\theta^{\otimes n})_{\theta\in(0,1)})$, où $\mathbb{P}_\theta$ suit une loi de Bernoulli de paramètre θ. On choisit la convention où Pile correspond à un succès. On souhaite tester l'hypothèse

$$(H_0) : "\theta = 0.43" \text{ contre } (H_1) : "\theta = 0.58".$$

1. Écrire une procédure de test de niveau $\alpha = 5\%$ avec l'estimateur $\bar{X}_n$.

Sous (H_0), on a d'après la LGN,

$$\bar{X}_n \overset{p.s.}{\to} \theta_0.$$

Ainsi, si (H_0) est vraie alors $\bar{X}_n$ est proche de $\theta_0 = 0.43$. Mais que veut dire proche ? On utilise le TCL pour répondre à cette question

$$\sqrt{n}\frac{\bar{X}_n - \theta_0}{\sqrt{\theta_0(1-\theta_0)}} \overset{loi}{\to} Z \sim \mathcal{N}(0,1).$$

Comme $\theta_0 = 0.43$, on peut remplacer

$$\sqrt{n}\frac{\bar{X}_n - 0.43}{\sqrt{0.43 \times 0.57}} \overset{loi}{\to} Z \sim \mathcal{N}(0,1).$$

Ainsi si (H_0) est vraie et que n est grand, on a

$$Q_{0.43}\left(\sqrt{n}\frac{\bar{X}_n - 0.43}{\sqrt{0.43 \times 0.57}} \in [\pm q_{1-\alpha/2}]\right) \approx 1-\alpha.$$

Ainsi pour α petit, si (H_0) est vraie alors il est crédible que d'observer l'évènement

$$\sqrt{n}\frac{\bar{X}_n - 0.43}{\sqrt{0.43 \times 0.57}} \in [\pm q_{1-\alpha/2}].$$

La procédure de test est donc rejeter si on n'observe pas cet événement, i.e.

$$\sqrt{n}\frac{\bar{X}_n - 0.43}{\sqrt{0.43 \times 0.57}} \notin [\pm q_{1-\alpha/2}],$$

et on ne rejette pas sinon. Une autre manière d'écrire la procédure : on définit la statistique de test

$$T(X_1,\ldots,X_n) = \mathbb{1}_{X_1,\ldots,X_n \in \mathcal{R}},$$

où $\mathcal{R}$ est la zone de rejet,

$$\mathcal{R} = \{x_1,\ldots,x_n\mathbb{R}^n \mid \bar{x}_n \notin [0.43 \pm q_{1-\alpha/2}\sqrt{0.43\times 0.57}/\sqrt{n}]\}.$$

2. Après $n = 100$ lancers, on observe 50 Piles. Que conclure ?

Ainsi en remplaçant tous les termes connus, $n = 100$, $\bar{X}_n = 50/100 = 0.5$ et $q_{1-\alpha/2} = 1.96$ pour $\alpha = 5\%$, on ne rejette pas l'hypothèse nulle.

3. Inverser les hypothèses (H_0) et (H_1), quelle est la nouvelle conclusion ?

On répond de la même manière que la question 1 en remplaçant 0.43 par 0.58. Dans ce cas là, on ne rejette pas non plus.

Exercice 3.6
Un négociant en vin s'intéresse à la contenance des bouteilles d'un producteur soupçonné par certains clients de frauder. Il souhaite s'assurer que cette contenance respecte bien en moyenne la limite légale de 75cL. À cet effet, il mesure le contenu de 10 bouteilles prises au hasard et obtient les valeurs suivantes (en cL) :

73.2	72.6	74.5	75	75.5	73.7	74.1	75.8	74.8	75

On suppose que la contenance des bouteilles (en cL) suit une loi gaussienne d'espérance θ inconnue, d'écart-type connu égal à 1.
1. Écrire le modèle statistique associé.
2. Le négociant décide de tester l'hypothèse

$$(H_0) : "\theta = 75" \text{ contre } (H_1) : "\theta < 75".$$

Quel point de vue le négociant adopte-t-il en choisissant ces hypothèses ? Justifier précisément la réponse.
3. Construire, à l'aide d'une règle de décision intuitive basée sur la moyenne empirique, un test de niveau 1% de (H_0) contre (H_1). Quelle est la conclusion de ce test ?
4. Le négociant veut pouvoir détecter, avec une probabilité élevée (99%), une contenance moyenne de 74.8cL tout en gardant un test de niveau 1%. Que doit-il faire ?

Exercice 3.7 Peut-on retarder sa mort ?
On prétend couramment que les mourants peuvent retarder leur décès jusqu'à certains événements importants. Pour tester cette théorie, Philips et King (1988, article paru dans *The Lancet*, prestigieux journal médical) ont collecté des données de décès aux environs d'une fête religieuse juive. Sur 1919 décès, 922 (resp. 997) ont eu lieu la semaine précédente (resp. suivante). Comment utiliser de telles données pour tester cette théorie grâce à un test asymptotique ?

Exercice 3.8 Test sur le support d'une loi uniforme
Soit $X_1, \ldots, X_n$ n variables aléatoires i.i.d. de loi de Uniforme sur $[0, \theta]$ avec $\theta \in \mathbb{R}^+$ inconnu. On rappelle que l'EMV de ce modèle est

$$\hat{\theta}_{MV} = \max(X_1, \ldots, X_n).$$

Nous avions déjà dit que ce modèle n'est pas régulier et donc le comportement asymptotique de cet estimateur n'est pas gaussien. Dans la suite, on cherche à tester l'hypothèse

$$(H_0) : "\theta = 1" \text{ contre } (H_1) : "\theta > 1".$$

1. On propose le test suivant : on rejette (H_0) lorsque $\hat{\theta}_{MV} > c$ (c constante donnée). Calculer la fonction de puissance.
2. Quelle valeur prendre pour c pour obtenir un niveau de 5% ?
3. Si $n = 20$ et que la valeur observée de $\hat{\theta}_{MV}$ est 0.96, que vaut la p-value ? Quelle conclusion tirer sur (H_0) ? Même question pour l'observation $\hat{\theta}_{MV} = 1.04$.