3. Théorèmes asymptotiques. Tests, Théorème de Neyman-Pearson

Objectifs : Savoir appliquer la loi des grands nombres, le théorème central limite, la méthode delta et le Lemme de Slutsky. Savoir appliquer le théorème de Neyman-Pearson pour des test d'hypothèses simples. Les exercices 3.1 à 3.2 sont à faire pendant le TD, les 3.3 à 3.6 sont à chercher de votre côté.

Exercice 3.1 (Théorèmes asymptotiques)**.** Supposons que $X_1, \ldots, X_n$ sont i.i.d., centrées, de variance finie $\sigma^2 > 0$ inconnue. On note classiquement $\bar{X}$ la moyenne empirique et $\hat{\sigma}^2$ l'estimateur non biaisé de la variance.

1. Écrire $\hat{\sigma}^2$ en fonction des X_i. *Solution. On a déjà vu en cours que l'estimateur non biaisé de la variance s'écrit*

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 .$$

2. Étudier la convergence presque sûre de $\bar{X}$ et de $\hat{\sigma}^2$. *Solution. La loi forte des grands nombres implique $\bar{X} \xrightarrow[n\to\infty]{p.s.} 0$. De plus, on a aussi (calcul classique):*

$$\hat{\sigma}^2 = \frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2 - (\overline{X})^2\right),$$

et en appliquant la loi forte des grands nombres à chacun des termes, on a sans grande suprise

$$\hat{\sigma}^2 \xrightarrow[n\to\infty]{p.s.} 1 \times (\mathbb{E}[X_1^2] - \mathbb{E}[X_1]^2) = \mathrm{Var}(X_1) = \sigma^2 .$$

3. Montrer que

$$\frac{\sqrt{n}\bar{X}}{\sqrt{\hat{\sigma}^2}} \xrightarrow[n\to\infty]{(d)} \mathcal{N}(0,1) .$$

. *Solution. Le théorème central limite donne*

$$\frac{\sqrt{n}\bar{X}}{\sigma} \xrightarrow[n\to\infty]{(d)} \mathcal{N}(0,1) .$$

On veut remplacer σ par son estimée dans cette convergence, c'est possible grâce au Lemme de Slutsky : $\sqrt{\hat{\sigma}^2} \xrightarrow[n\to\infty]{p.s.} \sigma$ d'après 3.1.1. Cette convergence a donc également lieu en loi. D'après le Lemme de Slutsky,

$$\frac{\sqrt{n}\bar{X}}{\sqrt{\hat{\sigma}^2}} = \underbrace{\frac{\sigma}{\sqrt{\hat{\sigma}^2}}}_{\xrightarrow[n\to\infty]{p.s.} 1} \cdot \underbrace{\frac{\sqrt{n}\bar{X}}{\sigma}}_{\xrightarrow[n\to\infty]{(d)} \mathcal{N}(0,1)} \xrightarrow[n\to\infty]{(d)} \mathcal{N}(0,1) .$$

.

4. On suppose ici que $\sigma^2 = 1$ et qu'il est connu. Donner la limite en loi de

$$\sqrt{n}(e^{\bar{X}/2} - 1) .$$

. *Solution. C'est une application directe de la méthode Delta. Ici $g = \exp$ et $g'(0) = \frac{1}{2}e^{0/2} = 1/2 \neq 0$, d'où $\sqrt{n}(e^{\bar{X}/2} - 1) \xrightarrow[n\to\infty]{(d)} g'(0)\mathcal{N}(0,\sigma^2) = \mathcal{N}(0,\sigma^2 g'(0)^2) = \mathcal{N}(0,1/4)$.*

Exercice 3.2 (Test optimal pour la loi exponentielle). Soient $X_1, \ldots, X_n$ des variables aléatoires i.i.d. de loi exponentielle de paramètre[2] $\lambda > 0$. On souhaite tester les hypothèses simples $\mathcal{H}_0 : \lambda = \lambda_0$ contre $\mathcal{H}_1 : \lambda = \lambda_1$ avec $\lambda_1 > \lambda_0$.

On rappelle la propriété suivante : la somme de n variables aléatoires indépendantes de loi exponentielle de paramètre $\lambda = 1$ suit une loi Gamma $\Gamma(n, 1)$, dont on notera $q_\beta^{\Gamma,n}$ le quantile d'ordre β.

1. Déterminer un test uniformément plus puissant (UPP) de niveau $\alpha \in]0, 1[$ pour ce problème. Montrer que c'est un test dont la région de rejet peut s'écrire

$$R = \left\{ \sum_{i=1}^{n} X_i < \frac{u(\alpha)}{\lambda_0} \right\},$$

où $u(\alpha)$ est une constante dépendant de α dont on ne demande pas l'expression dans cette question.

Solution. *Le rapport de vraisemblance est donné par :*

$$\frac{L(z; \lambda_1)}{L(z; \lambda_0)} = \left(\frac{\lambda_1}{\lambda_0}\right)^n \exp\left(-(\lambda_1 - \lambda_0) \sum_{i=1}^{n} X_i \right).$$

D'après le théorème de Neyman-Pearson (cas continu), un test UPP de niveau $\alpha \in]0, 1[$ a une région de rejet de la forme

$$R = \left\{ \frac{L(z; \lambda_1)}{L(z; \lambda_0)} > t \right\}$$

avec t à calibrer pour être de niveau α. Ici, le rapport de vraisemblance est une fonction décroissante de $\lambda_0 \sum_{i=1}^n X_i$. On peut donc réexprimer la région de rejet optimale de la forme :

$$R = \left\{ \sum_{i=1}^{n} X_i < \frac{u(\alpha)}{\lambda_0} \right\},$$

avec $u(\alpha)$ tel que $\mathbb{P}_0(\sum_{i=1}^n X_i < u/\lambda_0) = \alpha$.

2. Montrer que sous $\mathcal{H}_0$, $\lambda_0 X_1 \sim \mathrm{Exp}(1)$. En déduire la valeur de $u(\alpha)$. *Solution.* *Via la fonction de répartition. Ensuite, sous $\mathcal{H}_0$, $\lambda_0 \sum X_i \sim \Gamma(n, 1)$, donc $u = q_\alpha^{\Gamma,n}$.*

3. On notera F_n la fonction de répartition de la loi $\Gamma(n, 1)$. Exprimer la puissance $\pi(\lambda_1)$ du test précédent pour toute valeur de $\lambda_1 > \lambda_0$, en fonction de F_n. Commenter sa monotonie en λ_1.

Solution. *Rappelons que la puissance du test est la probabilité de rejeter $\mathcal{H}_0$ sous l'hypothèse alternative $\mathcal{H}_1$. Elle vaut :*

$$\pi(\lambda_1) = \mathbb{P}_1\left(\sum_{i=1}^{n} X_i < \frac{q_\alpha^{\Gamma,n}}{\lambda_0} \right) = \mathbb{P}_1\left(\lambda_1 \sum_{i=1}^{n} X_i < \frac{\lambda_1}{\lambda_0} \cdot q_\alpha^{\Gamma,n} \right) = F_n\left(\frac{\lambda_1}{\lambda_0} \cdot q_\alpha^{\Gamma,n} \right),$$

en utilisant que par un raisonnement similaire à la question 2, sous $\mathcal{H}_1$, $\lambda_1 \sum X_i \sim \Gamma(n, 1)$. Notons que cette puissance est croissante en λ_1, ce qui est cohérent avec le fait que la puissance augmente lorsque λ_1 s'éloigne de λ_0 (il est alors plus facile de les distinguer).

4. Montrer par ailleurs que le test précédent est uniformément plus puissant (UPP) pour tester $\mathcal{H}_0 : \lambda = \lambda_0$ contre $\mathcal{H}_1 : \lambda > \lambda_0$.

[2] on rappelle que c'est loi à densité $x \mapsto \mathbb{1}_{x \geq 0} \lambda e^{-\lambda x}$.

Solution. *Le test déterminé ci-dessus ne dépend pas de la valeur de λ_1, tant que $\lambda_1 > \lambda_0$. Il est donc optimal pour tout $\lambda_1 > \lambda_0$, ce qui en fait un test UPP pour tester $\mathcal{H}_0 : \lambda = \lambda_0$ contre $\mathcal{H}_1 : \lambda > \lambda_0$.*

Exercice 3.3 (Asymptotique du maximum de vraisemblance)**.** Reprenons le modèle où les $X_1, \ldots, X_n$ sont i.i.d. et dont la loi est à densité sur $\mathbb{R}$ donnée par

$$x \mapsto \mathbb{1}_{x \geq 1} \frac{a-1}{x^a}$$

avec $a > 1$ un paramètre. On a établi que l'estimateur du maximum de vraisemblance de a est donné par

$$\hat{a}_{MV} = 1 + \left(\frac{1}{n} \sum_{i=1}^{n} \log X_i \right)^{-1}$$

dans l'exercice 2.3.

1. Etudier la distribution de $\log X_1$ en calculant sa fonction de répartition, et en déduire que $\hat{a}_{MV}$ est également consistant. *Solution.* *On trouve que $\log X_1 \sim \mathrm{Exp}(a-1)$.*
2. Calculer l'information de Fisher $I_1(a)$ de X_1. En admettant que le modèle est régulier, montrer que $\hat{a}_{MV}$ est asymptotiquement normal et donner ses caractéristiques. *Solution.* *Pour une seule variable, $\ell(a, x_1) = \log(a-1) - a \log x_1$ et $\frac{\partial}{\partial a} \ell(a, x_1) = \frac{n}{a-1} - \log x_1$, donc $I_1(a) = \mathrm{Var}_a(\frac{n}{a-1} - \log X_1) = \mathrm{Var}_a(\log X_1)$. Et comme d'après 1. $\log X_1 \sim \mathrm{Exp}(a-1)$, on a $I_1(a) = \frac{1}{(a-1)^2}$. La régularité étant admise, on applique directement le résultat du cours :*

$$\sqrt{n}(\hat{a}_{MV} - a) \sim \mathcal{N}(0, (a-1)^2)$$

Exercice 3.4 (Test de variance symétrique à deux échantillons)**.** *On admettra dans cet exercice que si Z_1, Z_2 sont deux variables i.i.d. gaussiennes standard, alors Z_1/Z_2 suit une loi de Cauchy de densité sur $\mathbb{R}$ donnée par $x \mapsto \frac{1}{\pi(1+x^2)}$.*

On travaille avec deux échantillons indépendants, X_1 et X_2. On suppose que $0 < \sigma_1^2 < \sigma_2^2$, et on souhaite tester

$$\mathcal{H}_0 : X_1 \sim \mathcal{N}(0, \sigma_1^2),\ X_2 \sim \mathcal{N}(0, \sigma_2^2)$$

contre

$$\mathcal{H}_1 : X_1 \sim \mathcal{N}(0, \sigma_2^2),\ X_2 \sim \mathcal{N}(0, \sigma_1^2)\,.$$

On souhaite travailler avec des tests symétriques ϕ, tels que $\phi(x_1, x_2) = 1 - \phi(x_2, x_1)$ presque partout.

1. Montrer que si ϕ est un test symétrique de niveau α et de puissance β, alors $\alpha = 1 - \beta$. *Solution.* *Observer que si (X_1, X_2) sont tirés sous $\mathcal{H}_0$, alors (X_2, X_1) sont tirés sous $\mathcal{H}_1$. On utilise alors cette symétrie pour écrire*

$$\alpha = \mathbb{E}_0[\phi(X_1, X_2)] = \mathbb{E}_1[\phi(X_2, X_1)] = 1 - \mathbb{E}_1[\phi(X_1, X_2)] = 1 - \beta\,.$$

2. Écrire le rapport de vraisemblance L, et construire un estimateur symétrique ϕ_0 basé sur L pour ce problème. *Solution.* *Le rapport de vraisemblance est donné par*

$$L = \frac{\frac{1}{\sqrt{2\pi}} \exp(-\frac{X_1^2}{2\sigma_2^2} - \frac{X_2^2}{2\sigma_1^2})}{\frac{1}{\sqrt{2\pi}} \exp(-\frac{X_1^2}{2\sigma_1^2} - \frac{X_2^2}{2\sigma_2^2})} = \exp\left((X_1^2 - X_2^2) \left(\frac{1}{2\sigma_1^2} - \frac{1}{2\sigma_2^2} \right) \right)$$

Ainsi, un test basé sur $L > \tau$ est de la forme $X_1^2 - X_2^2 > C$, ce qui est symétrique presque sûrement si et seulement si $C = 0$. Notons qu'il est symétrique presque sûrement car $X_1^2 - X_2^2 = 0$ a une probabilité nulle. On construit donc le test symétrique

$$\phi_0(X_1, X_2) = \mathbb{1}_{X_1^2 - X_2^2 > 0}\,.$$

3. Montrer que le niveau $\alpha(\phi_0)$ et la puissance $\beta(\phi_0)$ de ϕ_0 vérifient :

$$\alpha(\phi_0) = 1 - \beta(\phi_0) = \frac{2}{\pi}\tan^{-1}\left(\frac{\sigma_1}{\sigma_2}\right).$$

Solution. *Calculons le niveau :*

$$\alpha(\phi_0) = \mathbb{P}_0(X_1^2 > X_2^2) = \mathbb{P}_0(\sigma_1^2 Z_1^2 > \sigma_2^2 Z_2^2) = \mathbb{P}_0(|Z_2/Z_1| < \sigma_1/\sigma_2) = \frac{2}{\pi}\tan^{-1}(\sigma_1/\sigma_2),$$

car $\tan^{-1}$ est la fonction de répartion d'une loi de Cauchy. On remarque que cette expression est toujours $\leq 1/2$, avec égalité si et seulement si $\sigma_1 = \sigma_2$, et tend vers 0 lorsque $\sigma_1 \ll \sigma_2$.

Exercice 3.5 (Plus vite que $\sqrt{n}$?)**.** Reprenons l'exercice 3.1, en supposant que $\sigma^2 = 1$ et qu'il est connu.

1. Quelle est la limite en loi de $\sqrt{n}(\cos(\bar{X}) - 1)$?
2. Trouver une suite $(a_n)_{n\geq 1}$ telle que $a_n(\cos(\bar{X}) - 1)$ converge en loi vers une v.a. Z qui n'est pas presque sûrement constante, dont on précisera la loi. *Solution.* *On revient à la preuve du théorème de la méthode Delta, mais on va jusqu'à l'ordre 2 ! On a $\cos(x) - 1 = -x^2/2 + x^2\varepsilon(x)$, avec ε continue sur $\mathbb{R}\setminus\{0\}$, et $\varepsilon(x) \xrightarrow[x\to 0]{} 0$. On peut alors prolonger ε par continuité en 0. Puisque $\bar{X} \to 0$ p.s., alors par continuité, $\varepsilon(\bar{X}) \to \varepsilon(0) = 0$ p.s., donc aussi en probabilité. Ainsi, pour tout n,*

$$\cos(\bar{X}) = 1 + 0 - \frac{(\bar{X})^2}{2}(1 + \varepsilon(\bar{X}))\,,$$

avec $1 + \varepsilon(\bar{X}) \xrightarrow[n\to\infty]{\mathbb{P}} 1$. On applique de nouveau le lemme de Slutsky pour conclure :

$$n(\cos(\bar{X}) - 1) = -(1 + \varepsilon(\bar{X})) \times \frac{(n\bar{X})^2}{2} \xrightarrow[n\to\infty]{(d)} -Z^2/2,$$

avec $Z \sim \mathcal{N}(0,1)$. Remarquons pour les plus sagaces qu'en appliquant une nouvelle fois la méthode Delta avec $\arccos$, on pourrait obtenir $n\bar{X} \to Z'$, ce qui contredirait le résultat de 3.1.2 (une vitesse en $n^{-1/2}$ y est établie, et non pas en n^{-1}). Mais, bien sûr, on ne peut pas faire cela (voyez-vous pourquoi ?). On pourra insister sur l'importance de la régularité dans les hypothèses de la méthode Delta.

Exercice 3.6. Reprenons l'exercice 3.2.

1. Sous quelle condition sur a ces variables sont-elles d'espérance finie ?
2. Sous la condition de la question 1, donner un estimateur $\hat{a}_{MM}$ de a via la méthode des moments. Prouver qu'il est consistant.
3. On suppose dans toute la suite que la condition de la question 1 est vérifiée. A l'aide du TCL et de la méthode Delta, montrer que $\hat{\alpha}_{MM}$ est asymptotiquement normal et donner ses caractéristiques.
4. Pour n grand, quel est l'estimateur à privilégier selon vous, entre $\hat{a}_{MM}$ et $\hat{a}_{MV}$?