<u>**Corrigé du TD no 1**</u> Modèles linéaires et généralisations

EXERCICE 1

Question (1)**.** *Soient $X_1, \dots, X_n$ i.i.d. de densité $\theta x^{-\theta-1}\mathbb{1}(x \geq 1)$, où $\theta > 0$. On suppose que $\theta > 1$, estimer θ par la méthode des moments.*

On calcule l'espérance de X_1 :

$$\mu_1(\theta) = E_\theta(X_1) = \int_1^\infty \frac{\theta x}{x^{\theta+1}}\, dx = \left.\frac{\theta x^{1-\theta}}{(1-\theta)}\right|_1^\infty = \frac{\theta}{\theta-1}.$$

La fonction $\mu_1(.)$ établit une bijection de $\Theta =]1, +\infty[$ dans $]1, +\infty[$, d'inverse $\mu_1^{-1} = \mu_1$. Comme $\bar{X}$ appartient p.s. à $]1, +\infty[$, l'estimateur par la méthode des moments est bien défini p.s. sur $\Theta =]1, +\infty[$ et est l'unique θ tel que $\mu_1(\theta) = \bar{X}$, c'est-à-dire

$$\hat{\theta}_n^{MM} = \frac{\bar{X}}{\bar{X}-1}.$$

Question (2)**.** *On suppose maintenant que $\Theta = \{\theta > 0\}$. Estimer θ par la méthode des moments généralisée et du maximum de vraisemblance.*

Montrons d'abord que la méthode des moments ne peut pas être appliquée dans ce cas.
On calcule les moments d'ordre r de X_1 pour un $r > 0$:

$$\mu_r(\theta) = E_\theta(X_1^r) = \int_1^\infty \frac{\theta x^r}{x^{\theta+1}}\, dx = \begin{cases} \left.\frac{\theta x^{r-\theta}}{(r-\theta)}\right|_1^\infty = \frac{\theta}{\theta-r}, & \text{si } \theta > r, \\ +\infty, & \text{si } \theta \leq r. \end{cases}$$

On voit qu'il n'existe pas de $r > 0$ tel que la fonction $\theta \in \Theta =]0, +\infty[\mapsto \mu_r(\theta)$ soit injective. Cela veut dire que l'estimateur par la méthode des moments ne peut pas être défini dans ce cas. (On ne pourra jamais trouver un ensemble D contenant Θ pour lequel $\theta \in D \mapsto \mu_r(\theta)$ soit injective.) En revanche, on peut définir l'estimateur par la méthode des moments généralisée. On pose $g(x) = 1/x$. Dans ce cas

$$\mu_g(\theta) = E_\theta(g(X_1)) = \int_1^\infty \frac{\theta}{x^{2+\theta}}\, dx = \frac{\theta}{1+\theta}.$$

La fonction $\mu_g(.)$ établit une bijection de $\Theta =]0, +\infty[$ dans $]0, 1[$, d'inverse $\mu_g^{-1}(y) = y/(1-y)$. Comme $\overline{1/X} = n^{-1}\sum_{i=1}^n g(X_i)$ appartient p.s. à $]0, 1[$, l'estimateur par la méthode des moments est bien défini p.s. sur $\Theta =]0, +\infty[$ et est l'unique θ tel que $\mu_g(\theta) = \overline{1/X}$, c'est-à-dire

$$\hat{\theta}_n^{MM} = \frac{\overline{1/X}}{1-\overline{1/X}}.$$

Calculons à présent l'estimateur du maximum de vraisemblance sur $\Theta =]0, +\infty[$, on a

$$L((x_1, \dots, x_n), \theta) = \theta^n \mathbb{1}(x_{(1)} \geq 1) \prod_{i=1}^n \frac{1}{x_i^{\theta+1}}.$$

Comme $X_{(1)} \geq 1$ p.s., l'égalité ci-dessus implique

$$l_n(\theta) = -\log\theta + \frac{(\theta+1)}{n}\sum_{i=1}^n \log X_i$$

La fonction de log-vraisemblance $l_n(.)$ est dérivable sur $\Theta =]0, +\infty[$ et on a $l_n'(\theta) = -\theta^{-1} + n^{-1}\sum_{i=1}^n \log X_i$, de sorte que

$$l_n'(\theta) \geq 0 \Longleftrightarrow \theta \geq \left(\frac{1}{n}\sum_{i=1}^n \log X_i\right)^{-1}.$$

Comme l'inverse de $n^{-1}\sum_{i=1}^n \log X_i$ est p.s. dans $\Theta =]0, +\infty[$, la fonction de log-vraisemblance admet un unique maximum sur $\Theta =]0, +\infty[$ et

$$\hat{\theta}_n^{MV} = \left(\frac{1}{n}\sum_{i=1}^n \log(X_i)\right)^{-1}.$$

On remarque que l'EMV aurait pu être obtenu en appliquant la méthode des moments avec $g(x) = \log(x)$.

Question (3)**.** *Le modèle statistique en question est-il régulier ? Calculer l'information de Fisher $I(\theta)$.*

Pour prouver que le modèle est régulier, il faut vérifier les hypothèses (D), (H1)-(H4) (voir cours). L'hypothèse (D) est évidemment satisfaite avec la mesure de Lebesgue.

L'hypothèse (H1) équivaut à vérifier que pour différentes valeurs de θ, les densités f_θ ont le même support. Dans notre cas il est évident que le support de f_θ est $[1, \infty[$ et il est indépendant de θ.

La fonction $\theta \mapsto l(x, \theta) = \log f_\theta(x) = \log\theta - (1+\theta)\log x$ définie sur l'intervalle ouvert $\theta > 0$ est infiniment dérivable, donc (H2) est également satisfaite.

Pour vérifier (H3), on remarque que pour tout $\theta^* > 0$ et pour tout $\theta \in]\theta^*/2, 2\theta^*[:= U_{\theta^*}$, on a

$$\begin{aligned}
|l'(x,\theta)| &= |\theta^{-1} - \log x| \leq \frac{2}{\theta^*} + \log x, \quad \forall x > 1 \\
|(l'(x,\theta))^2| &\leq \left(\frac{2}{\theta^*} + \log x\right)^2 \leq \frac{8}{(\theta^*)^2} + 2\log^2(x), \quad \forall x > 1 \\
|l''(x,\theta)| &= |-\theta^{-2}| \leq \frac{4}{(\theta^*)^2}, \quad \forall x > 1 \\
\sup_{\theta \in U_{\theta^*}} f_\theta(x) &\leq \frac{2\theta^*}{x^{\theta^*/2+1}}, \quad \forall x > 1.
\end{aligned}$$

L'hypothèse (H3) découle immédiatement de ces inégalités, en posant $\Lambda(x) = \frac{2}{\theta^*} + \log x + \frac{8}{(\theta^*)^2} + 2\log^2(x)$.

Calculons à présent l'information de Fisher, ce qui nous permettra de vérifier (H4). Par définition,

$$I(\theta) = E_\theta\left[\left(\frac{d\log f_\theta}{d\theta}(X)\right)^2\right] = E_\theta\left[\left(\frac{1}{\theta} - \log X\right)^2\right].$$

On pose $Y_i = \log(X_i)$; par intégration par parties, on trouve

$$\mathbf{E}(Y_i) = \theta\int_1^\infty \frac{\log(x)}{x^{\theta+1}}\,dx = -\frac{\log(x)}{x^\theta}\Big|_1^\infty + \int_1^\infty \frac{1}{x^{\theta+1}}\,dx = -\frac{1}{\theta x^\theta}\Big|_1^\infty = \frac{1}{\theta},$$

et

$$\mathbf{E}(Y_i^2) = \theta\int_1^\infty \frac{\log^2 x}{x^{\theta+1}}\,dx = \int_1^\infty \frac{2\log x}{x^{\theta+1}}\,dx = \frac{2\mathbf{E}(Y_1)}{\theta} = \frac{2}{\theta^2}.$$

Par suite $I(\theta) = \mathbf{Var}(Y_1) = \theta^{-2} > 0$, ce qui nous permet de conclure que le modèle statistique qu'on considère est régulier.

Question (4)**.** *Sans utiliser le théorème général, étudier la loi limite de l'estimateur du maximum de vraisemblance $\hat\theta_n^{EMV}$. Est ce que la variance asymptotique de $\sqrt{n}(\hat\theta_n^{EMV} - \theta)$ est égale à $I(\theta)^{-1}$? Pourquoi ?*

Pour trouver la loi limite de θ_n^{MV}, on utilise la loi forte des grands nombres et le théorème central limite pour les variables $Y_i = \log(X_i)$ i.i.d. : d'une part, on remarque que $Y_i \geq 0$ p.s. car $X_1 \geq 1$ p.s., de sorte que $\mathbf{E}(|Y_1|) = \mathbf{E}(Y_1) = 1/\theta < \infty$ et en utilisant la loi forte des grands nombres on obtient

$$\hat\theta_n^{MV} = \frac{1}{\bar Y} \xrightarrow[n\to\infty]{p.s.} \frac{1}{\mathbf{E}(Y_1)} = \theta. \tag{1}$$

Pour déterminer la loi limite, on utilise le théorème central limite (car $\mathbf{E}(Y_i^2) = 2/\theta^2 < \infty$),

$$\sqrt{n}\,(\bar Y - \theta^{-1}) \xrightarrow[n\to\infty]{D} N(0, \theta^{-2}). \tag{2}$$

En utilisant les convergences (1), (2), la représentation

$$\sqrt{n}\,(\hat\theta_n^{MV} - \theta) = \sqrt{n}\,(\bar Y^{-1} - \theta) = -\frac{\theta\sqrt{n}}{\bar Y}\left(\bar Y - \theta^{-1}\right)$$

ainsi que le théorème de Slutsky, on obtient

$$\sqrt{n}\,(\hat\theta_n^{MV} - \theta) \xrightarrow[n\to\infty]{D} -N(0, \theta^2) = N(0, \theta^2).$$

La variance asymptotique de $\sqrt{n}(\hat\theta_n^{EMV} - \theta)$ est donc égale à $I(\theta)^{-1}$. Ceci est cohérent avec le résultat général du cours sur la distribution asymptotique de l'EMV, valable dans tout modéle régulier lorsque l'EMV est consistant (donc par exemple, tout à fait applicable dans le contexte présent).

Exercice 2

On observe X_1 de loi $U[0,1]$ sous H_0, ou $U[2,3]$ sous H_1. Proposer un test de l'hypothèse H_0 contre l'alternative H_1, et calculer ses erreurs de première et deuxième espèce.

Vu que les intervalles $[0,1]$ et $[2,3]$ sont disjoints, le test le plus naturel consiste à accepter H_0 si la valeur observée X_1 appartient à $[0,1]$ et accepter H_1 dans le cas contraire.

L'erreur de première espèce est alors la probabilité de rejeter H_0 alors qu'elle était vraie :

$$\mathbf{P}_{U[0,1]}(X_1 \notin [0,1]) = 0.$$

De la même faÁon, on vérifie que l'erreur de second espèce est également nulle. C'est donc un test idéal, car la probabilité de commettre une erreur est nulle.

Exercice 3

On suppose que l'on observe $X_1, \ldots, X_n$ i.i.d. de loi $\mathcal{N}(\mu, 1)$. On veut tester $H_0 : \mu = 0$ contre $H_1 : \mu = m$ o ? m est un nombre réel négatif fixé.

Question (1)**.** *Donner la forme du test de Neyman-Pearson de niveau $\alpha \in (0,1)$ pour ce problème.*

Pour définir le test de Neyman-Pearson (N-P), on détermine d'abord la fonction de vraisemblance :

$$L(\mu, X_1, \ldots, X_n) = (2\pi)^{-n/2} \prod_{i=1}^{n} e^{-(X_i-\mu)^2/2}$$

et le rapport des vraisemblances :

$$\begin{aligned}\frac{L(\mu_1, X_1, \ldots, X_n)}{L(\mu_0, X_1, \ldots, X_n)} &= \prod_{i=1}^{n} \exp\big(-[(X_i-\mu_1)^2 - (X_i-\mu_0)^2]/2\big) \\ &= \exp\big(n(\mu_1-\mu_0)\bar{X} + \frac{1}{2}n(\mu_0^2-\mu_1^2)\big) = e^{-\frac{1}{2}nm^2+mn\bar{X}},\end{aligned}$$

o ? la dernière égalité est due au fait que dans notre cas $\mu_0 = 0$ et $\mu_1 = m$. Cette expression du rapport de vraisemblance implique que le test N-P est défini par la région critique

$$R = \{e^{-\frac{1}{2}nm^2+nm\bar{X}} > c_\alpha\}$$

o ? c_α vérifie $\mathbf{P}_0(e^{-\frac{1}{2}nm^2+nm\bar{X}} \geq c_\alpha) = \alpha$. Si l'on pose $c'_\alpha = (2n^{-1}\log c_\alpha + m^2)/2m$, alors la région R se simplifie :

$$R = \{\bar{X} < c'_\alpha\}.$$

Il est évident que sous H_0, la variable aléatoire $\bar{X}$ suit la loi $\mathcal{N}(0, 1/n) = n^{-1/2}\mathcal{N}(0,1)$. Donc

$$\alpha = \mathbf{P}_0(R) = \mathbf{P}(n^{-1/2}N(0,1) < c'_\alpha) = \Phi(\sqrt{n}\, c'_\alpha),$$

o ? Φ désigne la fonction de répartition de $\mathcal{N}(0,1)$. On en déduit que

$$\sqrt{n}\, c'_\alpha = \Phi^{-1}(\alpha) = q_\alpha^N, \qquad \text{ou encore} \qquad c'_\alpha = n^{-1/2}\Phi^{-1}(\alpha) = n^{-1/2} q_\alpha^N,$$

o ? q_α^N est le quantile d'ordre α de la loi normale $\mathcal{N}(0,1)$.

Question (2)**.** *Calculer la puissance $m \in \mathbb{R}_- \mapsto \pi_n(m)$ de ce test et tracer son graphe. Etudier la convergence simple de π_n lorsque n tend vers $+\infty$. Peut–on parler de convergence uniforme sur $\mathbb{R}_-$?*

La puissance de ce test est définie par

$$\pi_n(m) = \mathbf{P}_m(\bar{X} < c'_\alpha).$$

Lorsque $\bar{X} \sim \mathcal{N}(m, n^{-1})$, on a $\sqrt{n}(\bar{X} - m) \sim \mathcal{N}(0,1)$. Donc

$$\begin{aligned}\pi_n(m) &= \mathbf{P}\big(\mathcal{N}(0,1) < \sqrt{n}(c'_\alpha - m)\big) = \Phi(\sqrt{n}c'_\alpha - \sqrt{n}\, m) \\ &= \Phi\big(q_\alpha^N - \sqrt{n}\, m\big).\end{aligned}$$

On a évidemment $\pi_n(0) = \alpha$ et $\lim_{-\infty} \pi_n = \lim_{+\infty} \Phi = 1$. De plus, comme Φ est croissante, π_n est décroissante. On peut aussi calculer la pente de π_n en 0 : $\pi'_n(0) = -\sqrt{n}\phi(q_\alpha^N)$, o ? ϕ est la densité d'une loi normale centrée réduite.

Il est facile de voir que $(\pi_n)_n$ converge simplement vers la fonction $\mathbb{1}_{]-\infty,0[} + \alpha\delta_0$. Comme cette fonction limite n'est pas continue alors que les π_n le sont, on en déduit que la convergence n'est pas uniforme (théorème de Dini).

Question (3)**.** *On considère l'alternative $H_1 : \mu = -Cn^{-\gamma}$, avec $C > 0$ et $\gamma \in \mathbb{R}$. Étudier le comportement de la puissance du test en fonction de γ lorsque $n \to \infty$.*

Posons $\tilde{\pi}_n(\gamma) = \pi_n(-Cn^{-\gamma})$. En substituant m par $-Cn^{-\gamma}$ dans l'expression de la puissance de la question précédente, on obtient :

$$\tilde{\pi}_n(\gamma) = \Phi\big(q_\alpha^N + Cn^{\frac{1}{2}-\gamma}\big).$$

Si $\gamma > 1/2$, alors $n^{\frac{1}{2}-\gamma} \to 0$ et donc $\tilde{\pi}_n(\gamma) \to \Phi(q_\alpha^N) = \alpha$. Si $\gamma < 1/2$, alors $n^{\frac{1}{2}-\gamma} \to +\infty$ et donc $\tilde{\pi}_n(\gamma) \to \Phi(q_\alpha^N + \infty) = 1$. Si $\gamma = 1/2$, on a $\tilde{\pi}_n(\gamma) = \Phi\big(q_\alpha^N + C\big)$. En conclusion,

$$\lim_{n\to\infty} \tilde{\pi}_n(\gamma) = \begin{cases} \alpha, & \text{si } \ \gamma > 1/2, \\ \Phi\big(q_\alpha^N + C\big), & \text{si } \ \gamma = 1/2, \\ 1, & \text{si } \ \gamma < 1/2. \end{cases}$$

On voit que pour $\gamma \geq 1/2$ ce test n'est pas consistant ($\pi \not\to 1$). Ceci implique qu'on ne peut pas distinguer l'alternative de l'hypothèse si elles sont "trop proches".

Exercice 4

Un marchand de graines a l'habitude de fournir à un ingénieur agronome un mélange de 6 types de graines qui entrent toutes dans le mélange avec les mêmes proportions. Quelque soit le type, les graines ont sensiblement le même poids la même taille. Elles diffèrent d'un type à l'autre par leurs formes. On désigne ces 6 types par A,...,F.

Un jour, l'ingénieur lui demande de préparer un mélange spécial dans lequel il y aura $1/4$ *de graines A,* $1/2$ *de B, les autres graines restent en proportion inchangées.*

Le jour de livraison, un employé un peu trop zélé a déplacé les sacs. Le grainetier, fort perplexe, croit que le premier sac est bon, mais il n'en est pas certain. Comment faire ?

Évidemment, on pourrait trier les graines de ce sac, mais ceci serait fastidieux. On propose ici un test pour apporter une solution à ce problème : il s'agit, à partir d'un échantillon de graines du sac, de décider entre deux hypothèses : " ce sac est bon" et "ce sac est ordinaire".

Question (1)**.** *A partir des observations de la forme "la i-ième graine tirée est de type A ou non", formuler un modèle statistique et poser le problème de test d'hypothèses.*

Soient Y_i une variable aléatoire qui vaut 1 si le i-ème graine de l'échantillon choisi est de type A, et 0 dans le cas contraire. Il est bien évident que Y suit une loi de Bernoulli qu'on notera $Be(p)$, pour un paramètre d'intérêt $p \in]0, 1[$. L'hypothèse H_0 est alors "le sac est bon", qui se traduit par H_0 :$p = 1/4$, l'alternative étant "le sac est ordinaire" $H_1 : p = 1/6$.

Question (2)**.** *Quelle est, sous chacune des ces hypothèses, la loi du nombre X de graines de type A dans un échantillon de taille n ? Donner les valeurs de* $\mathbf{E}(X)$ *et* $\mathbf{Var}(X)$ *correspondant à chaque cas.*

La variable X est simplement la somme des Y_i, donc elle suit la loi binomiale :

$$X = \sum_{i=1}^{n} Y_i \sim B(n, p).$$

Sous H_0 cela donne $X \sim B(n, 1/4)$, alors que sous H_1, $X \sim B(n, 1/6)$. On a déjà calculé l'espérance et la variance de la loi binomiale :

$$\begin{array}{lll} \text{sous } H_0 & \mathbf{E}(X) = n\mathbf{E}(Y_1) = np = n/4, & \mathbf{Var}(X) = n\mathbf{Var}(Y_1) = np(1-p) = 3n/16, \\ \text{sous } H_1 & \mathbf{E}(X) = n\mathbf{E}(Y_1) = np = n/6, & \mathbf{Var}(X) = n\mathbf{Var}(Y_1) = np(1-p) = 5n/36. \end{array}$$

Question (3). *Donner la forme du test Neyman–Pearson de niveau α pour tester l'hypothèse "ce sac est bon" contre l'alternative "c'est un sac ordinaire".*

Pour tout p, la vraisemblance de ce modèle de Bernoulli s'écrit

$$L(Y_1 \dots Y_n, p) = \prod_{i=1}^{n} \left(p^{Y_i}(1-p)^{1-Y_i}\right) = p^X(1-p)^{n-X},$$

de sorte que le rapport de vraisemblance entre $p_1 = 1/6$ et $p_0 = 1/4$ est

$$\frac{L(Y_1 \dots Y_n, p_1)}{L(Y_1 \dots Y_n, p_0)} = \left(\frac{p_1}{p_0}\right)^X \left(\frac{1-p_1}{1-p_0}\right)^{n-X} = \left(\frac{4}{6}\right)^X \left(\frac{5}{6}\cdot\frac{4}{3}\right)^{n-X} = \left(\frac{3}{5}\right)^X \left(\frac{10}{9}\right)^n.$$

Par conséquent, la région critique du test de N-P est

$$R^* = \{L(Y_1 \dots Y_n, p_1) > c_\alpha L(Y_1 \dots Y_n, p_0)\} = \{X < c'_\alpha\}$$

o? $c'_\alpha = (n\log(10/9) - \log(c_\alpha))/\log(5/3)$. La valeur c'_α est telle que

$$\alpha = \mathbf{P}_{1/4}(R^*) = \mathbf{P}_{1/4}(X < c'_\alpha) = \mathbf{P}(Z < c'_\alpha) \text{ avec } Z \sim B(n, 1/4).$$

Question (4). *En utilisant une approximation de la loi de la variable aléatoire X par une loi normale (sous chacune des hypothèses), déterminer la taille minimale n_0 de l'échantillon et la région critique pour avoir un test de niveau de confiance α et de puissance plus grande que $1-\beta$ avec $0 < \alpha = \beta < 1$. Que vaut la région de rejet et n_0 pour $\alpha = \beta = 0.05$? Conclure.*

Comme les variables Y_i sont i.i.d. telles que $\mathbf{E}(|Y|^2) = \mathbf{E}(Y) < \infty$, on peut leur appliquer le théorème central limite : pour tout p,

$$\frac{X - np}{\sqrt{np(1-p)}} = \sqrt{n}\frac{\bar{Y} - p}{\sqrt{p(1-p)}} \xrightarrow[n\to\infty]{D} \mathcal{N}(0,1).$$

Après calcul, ceci nous donne pour $p_0 = 1/4$ et $p_1 = 1/6$ les convergences :

$$\text{sous } H_0 : \quad \frac{4X - n}{\sqrt{3n}} \xrightarrow{D} \mathcal{N}(0,1) \tag{3}$$

$$\text{sous } H_1 : \quad \frac{6X - n}{\sqrt{5n}} \xrightarrow{D} \mathcal{N}(0,1) \tag{4}$$

Choisissons à présent la constante c'_α de tel sorte que $\alpha = \mathbf{P}_{1/4}\big(X < c'_\alpha\big)$, en utilisant l'approximation (3) :

$$\begin{aligned}\alpha = \mathbf{P}_{1/4}\big(X < c'_\alpha\big) &= \mathbf{P}_{1/4}\left(\frac{4X-n}{\sqrt{3n}} < \frac{4c'_\alpha - n}{\sqrt{3n}}\right) \\ &\approx \mathbf{P}\left(\mathcal{N}(0,1) < \frac{4c'_\alpha - n}{\sqrt{3n}}\right) = \Phi\left(\frac{4c'_\alpha - n}{\sqrt{3n}}\right).\end{aligned}$$

L'approximation (3) nous suggère donc de choisir c'_α telle que $\alpha = \Phi\big((4c'_\alpha - n)/\sqrt{3n}\big)$ (en toute rigueur, il faudrait noter cette constante c''_α car elle diffère légèrement de c'_α ; on la note c'_α pour simplifier). Par suite, en notant $q^N_\alpha = \Phi^{-1}(\alpha)$ le quantile d'ordre α de la loi normale centrée réduite, on obtient $\frac{4c'_\alpha - n}{\sqrt{3n}} = q^N_\alpha$ et donc

$$c'_\alpha = \frac{q^N_\alpha\sqrt{3n} + n}{4}.$$

Par ailleurs, ce choix de c'_α étant effectué, on cherche à choisir n_0 assez grand, tel que pour tout $n \geq n_0$, le test de région $\{X < c'_\alpha\}$ est une puissance d'au moins $1-\beta$. Ceci équivaut à $\forall n \geq n_0$,

$$1-\alpha = 1-\beta \leq \mathbf{P}_{1/6}(X < c'_\alpha) = \mathbf{P}_{1/6}\left(\frac{6X-n}{\sqrt{5n}} > \frac{6c'_\alpha - n}{\sqrt{5n}}\right).$$

A l'approximation (4) près, et en utilisant

$$\begin{aligned} 1-\alpha \leq \Phi\left(\frac{6c'_\alpha - n}{\sqrt{5n}}\right) &\iff -q^N_\alpha \leq \frac{6c'_\alpha - n}{\sqrt{5n}} \\ &\iff \frac{-\sqrt{5n}q^N_\alpha + n}{6} \leq c'_\alpha \\ &\iff \frac{-\sqrt{5n}q^N_\alpha + n}{6} \leq \frac{q^N_\alpha\sqrt{3n}+n}{4} \\ &\iff 2n \geq (-q^N_\alpha)(4\sqrt{5n} + 6\sqrt{3n}) \\ &\iff n \geq (q^N_\alpha)^2(2\sqrt{5}+3\sqrt{3})^2, \end{aligned}$$

on déduit que n_0 est le premier entier supérieur ou égal à $(q^N_\alpha)^2(2\sqrt{5}+3\sqrt{3})^2$.
Pour $\alpha = 0.05$, on a d'après l'énoncé $q^N_\alpha \simeq -1.65$, une calculatrice donne $R^* = \{X < 52.34\}$ et $n_0 = 255$.
Finalement, en utilisant le test de N-P (et l'approximation d'une loi binomiale par une loi normale), pour garantir un niveau inférieure à 0.05 et pour garantir une puissance plus grande que 0.95 , le grainetier doit tirer 255 graines et doit décider que le sac est ordinaire si et seulement si le nombre de graines de type A tirées est plus petit que 53.

Exercice 5

Soit $X_1, \ldots, X_n$ des v.a. i.i.d., dont la loi admet la densité $f(x-\theta)$, o ? $f(x) = 2(1-x)\mathbb{1}_{[0,1]}(x)$. On veut tester $H_0 : \theta \geq 1$ contre $H_1 : \theta < 1$. Pour ceci, on introduit les régions critiques

$$R_c = \{X_{(1)} < c\} \qquad et \qquad \tilde{R}_c = \{X_{(n)} < c\}.$$

Le but de cet exercice est de comparer les tests basés sur R_c et $\tilde{R}_c$.

Question (1). *Calculer la fonction puissance π_c associée à R_c et vérifier que cette fonction est monotone.*

Par définition,

$$\begin{aligned} \pi_c(\theta) = \mathbf{P}_\theta(R_c) &= \mathbf{P}_\theta(X_{(1)} < c) = 1 - \mathbf{P}_\theta(X_{(1)} \geq c) \\ &= 1 - \prod_{i=1}^n \mathbf{P}_\theta(X_i \geq c) = 1 - (1-F_\theta(c))^n, \end{aligned}$$

o ? $F_\theta(x)$ désigne la fonction de répartition ayant la densité $f(x-\theta)$, c'est-à-dire

$$F_\theta(x) = \int_{-\infty}^x 2(1-y+\theta)\mathbb{1}_{[\theta,\theta+1]}(y)\,dy = \begin{cases} 0, & x < \theta, \\ -(1-y+\theta)^2\Big|_\theta^x, & x \in [\theta, \theta+1], \\ 1, & x > \theta+1. \end{cases}$$

$$= \left(1-(1-x+\theta)^2\right)\mathbb{1}_{[\theta,\theta+1]}(x) + \mathbb{1}_{]\theta+1,\infty[}(x).$$

On en conclut que
$\pi_c(\theta) = 1 - [\mathbb{1}_{]-\infty,\theta[}(c) + (1+\theta-c)^{2n}\mathbb{1}_{[\theta,\theta+1]}(c)] = \big(1-(1-c+\theta)^{2n}\big)\mathbb{1}_{[c-1,c]}(\theta) + \mathbb{1}_{]-\infty,c-1[}(\theta).$
Il est évident que c'est une fonction constante sur $]-\infty, c-1[\cup]c,+\infty[$ et décroissante sur $[c-1,c]$.

Question (2)**.** *Quelle valeur critique $c = c_1$ faut-il choisir pour que le test associé à R_{c_1} soit exactement de niveau 5% ?*

L'ensemble des valeurs de θ correspondant à l'hypothèse H_0 est $\Theta_0 = [1,\infty[$. Il s'agit donc de trouver c tel que $\sup_{\theta\in\Theta_0}\mathbf{P}_\theta(R_c) \le 0.05$. Vu que la fonction puissance est décroissante, son supremum est atteint lorsque $\theta = 1$. On cherche donc la valeur $c = c_1$ vérifiant

$$0.95 = \mathbf{P}_1(X_{(1)} \ge c_1) = (1 - F_1(c_1))^n = \Big((1-c_1+\theta)^{2n}\mathbb{1}_{[\theta,\theta+1]}(c_1) + \mathbb{1}_{]-\infty,\theta[}(c_1)\Big)\Big|_{\theta=1}.$$

On en déduit facilement que

$$c_1 = 2 - (0.95)^{1/2n}.$$

Question (3)**.** *Calculer la fonction puissance $\tilde\pi_c$ associée à $\tilde R_c$ et trouver la valeur critique $c = c_2$ pour que le test associé à $\tilde R_{c_2}$ soit exactement de niveau* 5%*.*

Par définition,

$$\begin{aligned}
\tilde\pi_c(\theta) &= \mathbf{P}_\theta(\tilde R_c) = \mathbf{P}_\theta(X_{(n)} < c)\\
&= \prod_{i=1}^n \mathbf{P}_\theta(X_i < c) = [F_\theta(c)]^n = \begin{cases} 0, & c < \theta,\\ [1-(1+\theta-c)^2]^n, & c \in [\theta,\theta+1],\\ 1, & c > \theta+1.\end{cases}\\
&= [1-(1+\theta-c)^2]^n\mathbb{1}_{[c-1,c]}(\theta) + \mathbb{1}_{]-\infty,c-1[}(\theta).
\end{aligned}$$

Encore une fois, c'est une fonction décroissante, donc la valeur $c = c_2$ est définie par l'égalité

$$\mathbf{P}_1(\tilde R_{c_2}) = 0.05 \iff \tilde\pi_{c_2}(1) = 0.05.$$

On déduit que

$$\begin{aligned}
&[1-(2-c_2)^2]^n = 0.05\\
&(2-c_2)^2 = 1-(0.05)^{1/n}\\
&c_2 = 2 - \sqrt{1-(0.05)^{1/n}}.
\end{aligned}$$

Question (4)**.** *Comparer les fonctions puissance π_{c_1} et $\tilde\pi_{c_2}$ pour les tests de niveau 5%. Peut–on affirmer qu'un de ces tests est plus puissant que l'autre ?*

Pour $c_1 = 2-(0.95)^{1/2n}$, la fonction puissance vaut

$$\pi_{c_1}(\theta) = (1-[1-c_1+\theta]^{2n})\mathbb{1}_{[c_1-1,c_1]}(\theta) + \mathbb{1}_{]-\infty,c_1-1[}(\theta),$$

alors que pour $c_2 = 2-\sqrt{1-(0.05)^{1/n}}$ on a

$$\tilde\pi_{c_2}(\theta) = (1-[1-c_2+\theta]^2)^n\mathbb{1}_{[c_2-1,c_2]}(\theta) + \mathbb{1}_{]-\infty,c_2-1[}(\theta).$$

En dérivant la fonction $g(x) = (1-x)^n - (1-x^n)$, on peut vérifier qu'elle est décroissante sur $[0,1/2]$ et croissante sur $[1/2,1]$. Elle admet donc son maximum sur $[0,1]$ soit au point 0 soit au point 1. On en déduit que $g(x) < g(0) = g(1) = 0$ pour tout $x \in]0,1[$. En particulier, pour $x = (0.05)^{1/n}$, on obtient $1-(0.05)^{1/n} < (0.095)^{1/n}$ et donc $c_2 > c_1$.

Pour prouver que le test basé sur $\tilde{R}_c$ est plus puissant que celle basé sur R_c, il faut vérifier que

$$\tilde{\pi}_{c_2}(\theta) \geq \pi_{c_1}(\theta), \qquad \forall\, \theta \in \Theta_1 =]-\infty, 1[.$$

On a déjà prouvé que

$$1 < c_1 < c_2 < 2.$$

Cette inégalité implique que pour tout $\theta < c_2 - 1$, on a

$$\tilde{\pi}_{c_2}(\theta) = 1 \geq \pi_{c_1}(\theta).$$

En particulier $\tilde{\pi}_{c_2}(c_2-1) = 1 > \pi_{c_1}(c_2-1)$. Donc le premier test n'est pas plus puissant que le second.

Pour prouver que le second test non plus n'est pas plus puissant que le premier, on remarque que, pour $n = 2$,

$$\begin{aligned}
\pi'_{c_1}(1) - \tilde{\pi}'_{c_2}(1) &= -2n(1-c_1+1)^{2n-1} + 2n\big(1-(2-c_2)^2\big)^{n-1}(2-c_2) \\
&= 4\big[\sqrt{0.05(1-\sqrt{0}.05)} - (0.95)^{3/4}\big] \\
&< 4\big[\sqrt{0.05} - 0.95\big] \\
&< 4(0.3 - 0.95) < 0.
\end{aligned}$$

Donc la fonction $h(\theta) = \pi_{c_1}(\theta) - \tilde{\pi}_{c_2}(\theta)$ est décroissante dans un voisinage de 1. En particulier, pour $\varepsilon > 0$ suffisamment petit, $h(1-\varepsilon) > h(1) = \pi_{c_1}(1) - \tilde{\pi}_{c_2}(1) = 0.05 - 0.05 = 0$. Donc

$$\tilde{\pi}_{c_2}(1-\varepsilon) < \pi_{c_1}(1-\varepsilon).$$

Ceci implique que le second test n'est pas plus puissant que le premier.

En conclusion, il n'y a pas un test plus puissant que l'autre. La Figure 1 représente les deux fonctions puissances dans le cas $n = 2$. Elle illustre bien qu'aucune courbe puissance domine uniformément l'autre.

Question (5)**.** *Etudier la convergence simple de π_c et $\tilde{\pi}_c$ sur $]-\infty,1[$ quand $n \to \infty$ et c reste fixé.*

Fixons $\theta < 1$ et rappelons que

$$\pi_c(\theta) = \big(1-(1-c+\theta)^{2n}\big)\mathbb{1}_{[c-1,c]}(\theta) + \mathbb{1}_{]-\infty,c-1[}(\theta).$$

Pour tout $\theta \in [c-1, c[$, on a $1-c+\theta \in [0,1[$. Donc $(1-c+\theta)^{2n} \xrightarrow[n\to\infty]{} 0$. Il en résulte que

$$\lim_{n\to\infty} \pi_c(\theta) = \mathbb{1}_{]-\infty,c[}(\theta).$$

De même

$$\tilde{\pi}(\theta) = [1-(1-c+\theta)^2]^n \mathbb{1}_{[c-1,c]}(\theta) + \mathbb{1}_{]-\infty,c-1[}(\theta) \xrightarrow[n\to\infty]{} \mathbb{1}_{]-\infty,c-1]}(\theta).$$

Question (6)**.** *Etudier la convergence simple de π_{c_1} et $\tilde{\pi}_{c_2}$ sur $]-\infty,1[$ quand $n \to \infty$, en remarquant que les quantités c_1 et c_2 (qui dépendent de n) vérifient $c_1 \to 1$ et $c_2 \to 2$. Les tests associés à R_{c_1} et $\tilde{R}_{c_2}$ sont ils consistants ?*

Figure 1. Pour $n = 2$, graphes des deux fonctions $\tilde{\pi}_{c_2}$ (trait pointillé) et π_{c_1} (trait plein). Haut : sur l'intervalle $[c_2 - 1, c_2]$. Bas : zoom près de $c_2 - 1$.

Fixons $\theta < 1$. Comme $c_1 = 2 - (0.95)^{1/2n} \to 1$, il existe un $\delta > 0$ tel que à partir d'un certain rang en n, on ait $\theta + \delta < c_1$. Ainsi, $(1 - c_1 + \theta)^{2n} \leq (1 - \delta)^{2n}$ tend vers 0 et

$$\begin{aligned} \pi_{c_1}(\theta) &= (1 - [1 - c_1 + \theta]^{2n}) \mathbb{1}_{[c_1 - 1, c_1]}(\theta) + \mathbb{1}_{]-\infty, c_1 - 1[}(\theta) \\ &\to \mathbb{1}_{]-\infty, 1[}(\theta) = 1. \end{aligned}$$

Pour $c_2 = 2 - \sqrt{1 - (0.05)^{1/n}} \to 2$ on a $c_2 - 1 > \theta$ à partir d'un certain rang et donc à partir de ce rang $\tilde{\pi}_{c_2}(\theta) = (1 - [1 - c_2 + \theta]^2)^n \mathbb{1}_{[c_2 - 1, c_2]}(\theta) + \mathbb{1}_{]-\infty, c_2 - 1[}(\theta)$ vaut 1. Ceci implique que $\tilde{\pi}_{c_2}(\theta)$ tend vers 1.

Finalement, les deux fonctions π_{c_1} et $\tilde{\pi}_{c_2}$ convergent simplement sur $\Theta_1 =]-\infty, 1[$ vers 1, donc les tests correspondant sont consistants.

la correction ne sera pas tjrs tapée (lol)...katia