MAP433 Statistique

PC 6: Introduction aux tests statistiques

2 octobre 2015

1 Tests d'égalité et d'inégalité

Soient $(X_1, ..., X_n)$ des variables gaussiennes indépendantes respectivement de loi $\mathcal{N}(\mu, \sigma^2)$.

1. On suppose σ^2 connu. Soient $\mu_0 \neq \mu_1 \in \mathbb{R}$. On considère les hypothèses

   (H-0) $\mu = \mu_0$.

   (H-1) $\mu = \mu_1$.

   i Déterminer le test δ_α de Neyman-Pearson de niveau $\alpha \in (0, 1)$.

   ii Dans le cas $\mu_0 < \mu_1$, en donner une forme simplifiée consistant à comparer la statistique
   $$\bar{X}_n = \frac{1}{n} \sum_{k=1}^{n} X_k$$
   à un seuil s (appelé aussi *valeur critique*) à déterminer.

   iii Calculer sa puissance. Quel est son comportement quand α et n varient ?

2. On conserve la même forme de test qu'à la question précédente mais pour les hypothèses

   (H'-0) $\mu \leq \mu_0$.

   (H'-1) $\mu > \mu_0$.

   i Déterminer le niveau α' et la puissance du test δ_α pour ces nouvelles hypothèses.

   ii Soit $\delta'_{\alpha'}$ un test de niveau α' pour (H'-0) contre (H'-1). En utilisant le fait que ce test peut servir aussi pour tester (H-0) contre (H-1) avec $\mu_1 > \mu_0$ arbitraire, comparer la puissance de δ'_α à celle de δ_α.

   iii Conclure.

3. On considère à présent les hypothèses

   (H"-0) $\mu = \mu_0$.

   (H"-1) $\mu \neq \mu_0$.

   i Construire un test δ''_α de niveau α pour tester ces hypothèses, basé sur la même statistique de test que précédemment.

   ii Calculer la puissance du test δ''_α et le comparer au test δ_α. Ce test est-il UPP ?

4. Comment adapter ces différents tests au cas où σ^2 est inconnu ?

Application interactive : `http://webpopix.org:8080/X/TestMean`

Corrigé :
Modèle statistique : $\mathbf{X} = (X_1, \cdots, X_n) \in \mathbb{R}^n$; muni de sa tribu borélienne. Famille de lois $\{\mathbb{P}_\mu, \mu \in \mathbb{R}\}$; modèle dominé par la mesure de Lebesgue sur $\mathbb{R}^n$, et la densité de $\mathbb{P}_\mu$ est

$$p_\mu(x_{1:n}) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left\{\frac{-1}{2\sigma^2}\sum_{k=1}^n (x_k-\mu)^2\right\}$$

1. On a l'égalité (en supposant $\mu_2 > \mu_1$; adapter le raisonnement dans le cas contraire)
$$\left\{x_{1:n} : \ln \frac{p_{\mu_1}(x_{1:n})}{p_{\mu_0}(x_{1:n})} > c_\alpha\right\} = \left\{x_{1:n} : \bar{x}_n > \frac{1}{n}\frac{\sigma^2 \ln c_\alpha}{\mu_1 - \mu_0} + \frac{\mu_0+\mu_1}{2}\right\}$$
dont on déduit que NP est un test non randomisé de zone de rejet de la forme $\{\bar{X}_n > s_\alpha\}$. Le test NP de niveau α est obtenu en prenant
$$s_\alpha = \mu_0 + \frac{\sigma}{\sqrt{n}} z_{1-\alpha},$$
où z_q est le quantile d'ordre q d'une loi $\mathcal{N}(0,1)$. La puissance du test est donnée par
$$\pi(\mu_1) = \mathbb{P}\left(\xi > z_{1-\alpha} - \frac{\sqrt{n}}{\sigma}(\mu_1 - \mu_0)\right)$$
où sous $\mathbb{P}$, $\xi \sim \mathcal{N}(0,1)$.
2. • Pour la règle de décision δ_α définie à la question précédente, le niveau du test de H_0' contre H_1' est
$$\sup_{\mu \le \mu_0} \mathbb{P}_\mu(\bar{X}_n > s_\alpha) = \mathbb{P}(\xi > z_{1-\alpha}) = \alpha.$$
La fonction puissance est pour tout $\mu > \mu_0$,
$$\mu \mapsto \mathbb{P}_\mu\left(\bar{X}_n > s_\alpha\right) = 1 - \Phi\left(z_{1-\alpha} - \frac{\sqrt{n}}{\sigma}(\mu - \mu_0)\right)$$
où Φ est la cdf d'une loi $\mathcal{N}(0,1)$.
• Pour une autre règle de décision δ_α' de fonction puissance π' : (i) comme $H_0 = H_0'$, le niveau de la règle δ_α' pour tester H_0 contre H_1 est α ; (ii) mais δ_α est UPP$_\alpha$ pour H_0 contre H_1 donc les fonctions puissance vérifient $\pi(\mu_1) \geq \pi'(\mu_1)$; (iii) cette relation est vraie pour tout $\mu_1 > mu_0$.
• Si on cherche de H_0' contre H_1' dont la zone de rejet est de la forme $\{\bar{X}_n > s_\alpha'\}$ au niveau $\alpha\%$, on trouve $s_\alpha' = s_\alpha$. Donc les deux règles de décision coïncident. La question précédente entraine que cette règle est UPP$_\alpha$ pour tester H_0' contre H_1'.
3. On propose un test non randomisé de zone de rejet de la forme
$$\{x_{1:n} : |\bar{x}_n - \mu_0| > s''\}$$
Pour qu'il soit de niveau α, on prend $s'' = \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2}$. La puissance de ce test est donnée, pour tout $\mu \neq \mu_0$,
$$\mu \mapsto 1 - \Phi\left(z_{1-\alpha/2} + \frac{\sqrt{n}}{\sigma}(\mu_0 - \mu)\right) + \Phi\left(-z_{1-\alpha/2} + \frac{\sqrt{n}}{\sigma}(\mu_0 - \mu)\right)$$
Ce test n'est pas UPP$_\alpha$ pour tester H_0'' contre H_1''.
4. La construction des tests (via la définition de leur zone de rehetà va utiliser
$$\frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{\sqrt{S_n}} \sim t_{n-1} \qquad \text{avec } S_n = \frac{1}{n-1}\sum_{k=1}^n (X_k - \bar{X}_n)^2.$$
où t_q désigne une loi de Student à q degrés de liberté. Ce sont donc les quantiles de cette loi et non pas ceux de la loi $\mathcal{N}(0,1)$ qui vont intervenir dans la définition des seuils critiques.

2 Test pour une certification bio

Pour avoir la certification "bio", un fabriquant de produits "bios" doit garantir pour chaque lot un pourcentage d'OGM inférieur à 1%. Il prélève donc $n = 25$ produits par lot et teste si le pourcentage d'OGM est inférieur à 1%. On note X_i le logarithme du pourcentage d'OGM du paquet numéro i.
Modèle : On suppose que les X_i sont indépendants et suivent une loi gaussienne $\mathcal{N}(\theta, 1)$.

1. Pour le fabriquant, le pourcentage d'OGM est inférieur à 1% sauf preuve du contraire. Il veut tester l'hypothèse $\mathbf{H}_0 : \theta \leq 0$ contre $\mathbf{H}_1 : \theta > 0$ et il souhaite que pour $\theta \leq 0$ le test se trompe avec une probabilité inférieure à 5%. Calculez un seuil $t_{25,5}$ tel que
$$\sup_{\theta\leq 0} \mathrm{P}_\theta(\bar{X}_{25} > t_{25,5}) = 5\%.$$
On pourra utiliser que $\mathrm{P}(Z > 1.645) \approx 5\%$, pour $Z \sim \mathcal{N}(0, 1)$.
2. Une association "anti-OGM" veut s'assurer qu'il n'y a effectivement pas plus de 1% d'OGM dans les produits labélisés "bio". En particulier, elle s'inquiète de savoir si le test parvient à éliminer les produits pour lesquels le pourcentage d'OGM dépasse de 50% le maximum autorisé. Quelle est la probabilité que le test ne rejette pas $\mathbf{H}_0$ lorsque le pourcentage d'OGM est de 1.5% ?
3. Scandalisée par le résultat précédent, l'association milite pour que le test du fabriquant prouve effectivement que le pourcentage d'OGM est inférieur à 1%. Pour elle, le pourcentage d'OGM est supérieur à 1% sauf preuve du contraire, donc $\mathbf{H}_0$ est $\theta > 0$ et $\mathbf{H}_1$ est $\theta \leq 0$. Proposez un test de $\mathbf{H}_0$ contre $\mathbf{H}_1$ tel que la probabilité que le test rejette à tort $\mathbf{H}_0$ soit inférieure à 5%.
4. Les agences de régulation accepte le test statistique proposé par le fabricant ($\mathbf{H}_0 : \theta \leq 0$ contre $\mathbf{H}_1 : \theta > 0$) mais exige qu'un dépassement de 10% du maximum autorisé soit détecté dans 80% des cas. Le niveau du test restant fixé à 5% que doit faire le fabricant pour se soumettre à cette législation ?.

Corrigé :

1. On trouve $t_{25,5} = 1.645/\sqrt{n} \approx 0.33$.
2. $$\mathbb{P}_{\ln(3/2)}\left(\bar{X}_n \leq t_{25,5}\right) = \mathbb{P}\left(\mathcal{N}(0,1) \leq 1.645 - 5\ln(3/2)\right) \approx 0.35$$
3. On propose le test de zone de rejet
$$\{x_{1:n} : \bar{x}_n \leq -t_{25,5}\}.$$
4. on cherche un seuil s et une taille d'échantillon n tels que
$$\sup_{\theta\leq 0} \mathbb{P}_\theta\left(\bar{X}_n > s\right) = 0.05$$
$$\inf_{\theta\geq \ln(11/10)} \mathbb{P}_\theta\left(\bar{X}_n > s\right) \geq 0.8.$$

3 Des finanicers sans scrupules

Roger a lu dans le journal que 20% des professionnels de la finance pensent qu'il faut enfreindre la loi pour réussir :

`http://www.slate.fr/story/101785/wall-street-enfeindre-loi-reussir`

Il pense que ce pourcentage est sous-estimé et se propose d'enquêter pour vérifier son hypothèse. Il a dans son carnet d'adresses seulement deux noms de financiers qu'il peut interroger et à qui il va poser la question suivante : *"Pensez-vous qu'il faut enfreindre la loi pour réussir ?"*.

1. En supposant que la conclusion de Roger ne dépende que des réponses à cette question, quelles règles de décision peut-il mettre en place, et quelles sont alors les risques pour Roger de conclure à tort ?
2. Roger souhaite que la probabilité de conclure à tort que ce pourcentage est supérieur à 20% soit exactement de 5%. Il réalise alors qu'il n'a pas besoin d'interroger qui que ce soit pour que cette contrainte soit satisfaite : il lui suffit de mettre dans sa poche 19 jetons noir et un rouge. Pourquoi ? Quel est le défaut de cette méthode ?
3. Roger choisit maintenant une carte dans un jeu de 32 cartes et note soigneusement le résultat avant de faire son enquête auprès de ses deux connaissances. Pourquoi a-t-il adopté cette nouvelle stratégie ? Quelles sont alors les risques pour Roger de conclure à tort ?
4. Roger sent bien qu'une telle décision ne peut être prise sur la base d'un tirage aléatoire... Que devrait-il faire alors ?

Corrigé :
Soit p la proportion de professionnels de la finance pensant qu'il faut enfreindre la loi pour réussir. On cherche à tester $H_0 : p = 0.2$ contre $H_1 : p > 0.2$.

1. Soit X le nombre de "oui" reçu en interrogeant ses deux amis : $X \sim \mathcal{B}(2, p)$. Deux règles de décision possibles

$$\begin{cases} \text{rejette } H_0 & \text{si } X \geq 1 \\ \text{accepte } H_0 & \text{si } X = 0 \end{cases} \qquad \begin{cases} \text{rejette } H_0 & \text{si } X = 2 \\ \text{accepte } H_0 & \text{si } X \leq 1 \end{cases}$$

   Pour la première règle de décision, le niveau est 0.36 et la fonction puissance $p \mapsto p^2 + 2p(1-p)$. Pour la seconde règle de décision, le niveau est 0.04 et la fonction puissance $p \mapsto p^2$.
2. $1/20 = 0.05$. On peut construire un test de niveau 5% sans utiliser les données, mais en tirant un jeton et en rejettant H_0 si l'on tire le jeton rouge.
3. Soit le test randomisé

$$\begin{cases} \text{rejette } H_0 & \text{si } X = 2 \\ \text{on rejette } H_0 \text{ avec probabilité } q \text{ et on rejette sinon} & \text{si } X = 1 \\ \text{accepte } H_0 & \text{si } X = 0 \end{cases}$$

   En prenant $q = 1/32$, on obtient un niveau de test égal à 5% et une fonction puissance $p \mapsto p^2 + p(1-p)/16$.
4. Accepter le fait que son plan d'expérience le contraint à ne mettre en place que des tests (raisonnables) de niveaux donnés. Augmenter la taille de l'échantillon permet de mieux contrôler ce niveau.

4 Piscine

Un constructeur de piscine veut comparer 2 produits à dissoudre dans l'eau qui permettent de tuer les bactéries présentes. Les deux produits garantissent que 95 % des bactéries seront éliminées. Par contre il se peut que le pH soit plus basique et donc plus nocif pour les utilisateurs. Il mène une expérience dans son magasin où 2 piscines remplies sont exposées. Il met le produit A dans la piscine 1 et le produit B dans la piscine 2. Il a un pH-mètre vendu dans le commerce et il se doute que la mesure n'est pas fiable. Il prend donc 10 mesures dans chaque piscine.
Dans la piscine 1, il trouve $x_1, ..., x_{10}$ qui valent respectivement

$$7.33\ ;\ 6.17\ ;\ 7.46\ ;\ 8.13\ ;\ 6.68\ ;\ 6.76\ ;\ 7.97\ ;\ 6.76\ ;\ 6.81\ ;\ 8.40$$

La moyenne empirique vaut $\bar{x} = 7.247$ et la variance empirique (celle divisée par "$n-1$") vérifie $\sqrt{\widehat{\sigma_x^2}} = 0.73$.
Dans la piscine 2, il trouve $y_1, ..., y_{10}$ qui valent respectivement

$$10.40\ ;\ 7.27\ ;\ 8.99\ ;\ 7.28\ ;\ 9.18\ ;\ 9.10\ ;\ 7.96\ ;\ 7.71\ ;\ 9.59\ ;\ 9.61$$

Le moyennes et variance empiriques valent $\bar{y} = 8.709$ et $\sqrt{\widehat{\sigma_y^2}} = 1.08$.
Le constructeur n'a aucun parti pris entre les 2 produits et veut juste conseiller au mieux ses clients. Il voudrait pouvoir leur dire clairement "préférez A", "préférez B" ou "faites ce que vous voulez, les deux produits sont indiscernables vu mes mesures".
On sait que $pH = 7$ est neutre pour la peau tandis que $pH = 9$ est basique et donc nocif pour la peau.

1. Supposons que les observations sont les réalisations de variables aléatoires indépendantes gaussiennes de même variance σ^2 et de moyenne m_1 quand la mesure est prise dans la piscine 1, respectivement m_2 quand la moyenne est prise dans la piscine 2, m_1 et m_2. Préciser le modèle statistique sous la forme d'un seul vecteur gaussien dont on donnera la moyenne et la matrice de covariance.
2. Supposons que σ^2 est connue, égale à 1.
   i Donner l'expression la plus simple possible du test du rapport de vraisemblance de $\mathbf{H}_0$: "$m_1 = m_2 = 7$" contre $\mathbf{H}_1$: "$m_1 = 7, \quad m_2 = 9$" au niveau α.
   ii Expliquer pourquoi ce test est uniformément le plus puissant parmi tous les tests de niveau α de ces mêmes hypthèses.
   iii Calculer la p-valeur du test et conclure.
3. Toujours dans le cas où σ^2 est connu, construire un test de $\mathbf{H}_0$: "$m_1 = m_2$" contre $\mathbf{H}_1$: "$m_1 < m_2$" au niveau α.
4. Revenons au cas général (plus réaliste vu l'énoncé) où σ^2 est inconnue.
   i Donner les lois de $Z = \bar{Y} - \bar{X}$ et de $W = 9\,\widehat{\sigma_X^2} + 9\,\widehat{\sigma_Y^2}$. Expliquer pourquoi W est indépendante de Z.
   ii Toujours dans le cas où σ^2 est inconnue, construire un test de $\mathbf{H}_0$: "$m_1 = m_2$" contre $\mathbf{H}_1$: "$m_1 < m_2$" au niveau α. Calculer la p-valeur du test et conclure par l'une des 3 phrases "préférez A", "préférez B" ou "faites ce que vous voulez, les deux produits sont indiscernables vu les mesures".

Pour vous aider voici un tableau de valeurs qui donne t tel que $P(U > t) = a$

	a=0.05	a=0.025	a=0.01	a=0.005	a=0.0025	a=1.10^{-4}
U$\sim \mathcal{N}(0,1)$	1.645	1.960	2.326	2.576	2.807	3.719
U$\sim$T(9)	1.833	2.262	2.821	3.250	3.670	6.010
U$\sim$T(10)	1.812	2.228	2,763	3.169	3.581	5.694
U$\sim$T(18)	1.734	2.100	2.552	2.878	3.196	4.648
U$\sim$T(20)	1.724	2.085	2.528	2.845	3.153	4.538

où T(k) est la loi du Student à k degrés de liberté.

Corrigé :

1. On dispose de $2n$ observations ($n = 10$) ; $D_{2n} = (X_1, \cdots, X_n, Y_1, \cdots Y_n)$ à valeur dans $\mathbb{R}^{2n}$. On munit cet espace de la famille de lois $\{\mathbb{P}_m, m = (m_1, m_2) \in \mathbb{R}^2\}$ possédant une densité p_m par rapport à Lebesgue donnée par

$$\ln p_m(x_{1:n}, y_{1:n}) = C - \frac{1}{2\sigma^2}\left(\sum_{k=1}^{n}(x_k - m_1)^2 + \sum_{k=1}^{n}(y_k - m_2)^2\right)$$

ce qui est la densité d'un vecteur gaussien de moyenne et de matrice de covariance

$$\mathbb{E}[D_{2n}] = \begin{bmatrix} m_1 \\ \cdots \\ m_1 \\ m_2 \\ \cdots \\ m_2 \end{bmatrix} \qquad \mathrm{Var}(D_{2n}) = \sigma^2 I_{2n}.$$

2. (a) Notons $m = (7, 7)$ et $m' = (7, 9)$. Le ratio des lois s'exprime

$$\ln \frac{p_{m'}(x_{1:n}, y_{1:n})}{p_m(x_{1:n}, y_{1:n})} = -\frac{1}{2\sigma^2}\left(\sum_{k=1}^{n}(y_k - m_2')^2 - \sum_{k=1}^{n}(y_k - m_2)^2\right)$$

Par des calculs analogues à ce que l'on a fait dans l'exercice 1, on vérifie que le test du rapport de vraisemblance est un test non randomisé de zone de rejet de la forme $\{\bar{Y}_n > s\}$ où l'on choisit le seuil s en utilisant le fait que sous $\mathbb{P}_{(7,7)}$, $\bar{Y}_n \sim \mathcal{N}(7, \sigma^2/n)$. On trouve $s = 7 + \frac{1}{\sqrt{10}} z_{1-\alpha}$ où z_q est le quantile d'ordre q d'une loi $\mathcal{N}(0, 1)$.

   (b) Voir le cours sur NP.

   (c) Sur ce jeu de données, $\bar{Y}_n = 8.709$. On cherche $\mathbb{P}(T > 8.709)$ lorsque $T \sim \mathcal{N}(m_2, \sigma^2/n)$. Puisque $\sigma^2 = 1$, $n = 10$, $m_2 = 7$, on obtient $\mathbb{P}(\mathcal{N}(0, 1) > 1.709 * \sqrt{10}) \approx 0$. On rejette donc l'hypothèse nulle.

3. On peut proposer un test non randomisé de zone de rejet $\{\bar{X}_n - \bar{Y}_n < s\}$ Puisque sous $\mathbb{P}_m$, $\bar{X}_n - \bar{Y}_n \sim \mathcal{N}(m_1 - m_2, \frac{2\sigma^2}{n})$, on choisit s pour que le niveau du test soit α. On trouve $s = z_\alpha \frac{\sqrt{2}\sigma}{\sqrt{n}}$.

4. (a) Sous $\mathbb{P}_m$ avec $m = (m_1, m_2)$,
   - $\bar{Y}_n - \bar{X}_n \sim \mathcal{N}(m_2 - m_1, 2\sigma^2/n)$ puisque les v.a. $\bar{X}_n$ et $\bar{Y}_n$ sont des gaussiennes indépendantes.
   - $9\,\widehat{\sigma_x^2}/\sigma^2 \sim \chi^2(9)$ et cette v.a. est une fonction de $\{X_1, \cdots, X_n\}$; et de même $9\,\widehat{\sigma_y^2}/\sigma^2 \sim \chi^2(9)$ et cette v.a. est une fonction de $\{Y_1, \cdots, Y_n\}$. Puisque $\{X_1, \cdots, X_n\}$ et $\{Y_1, \cdots, Y_n\}$ sont indépendantes, alors $\widehat{\sigma_x^2}$ et $\widehat{\sigma_y^2}$ sont des v.a. sont indépendantes. Donc W_n/σ^2 suit une loi $\chi^2(18)$.

– D'après Cochran et l'indep des v.a. $\{X_i, Y_j, 1 \leq i, j \leq n\}$: $\bar{X}_n$ est indépendant de $\widehat{\sigma_x^2}$ et de $\widehat{\sigma_y^2}$; de même, $\bar{Y}_n$ est indépendant de $\widehat{\sigma_y^2}$ et de $\widehat{\sigma_x^2}$. Dont on déduit (puisque $\bar{X}_n$ et $\bar{Y}_n$ sont indépendants) que W_n et Z_n sont indépendants.

(b) Sous $\mathbb{P}_{(m_1,m_2)}$,

$$\sqrt{n}(\bar{X}_n - m_1)/\sigma \sim \mathcal{N}(0,1) \qquad \sqrt{n}(\bar{Y}_n - m_2)/\sigma \sim \mathcal{N}(0,1)$$

et ces deux quantités sont indépendantes, donc

$$\frac{\sqrt{n}}{\sqrt{2}\sigma}(\bar{X}_n - \bar{Y}_n - m_1 + m_2) \sim \mathcal{N}(0,1).$$

D'autre part, $(9\widehat{\sigma_x^2} + 9\widehat{\sigma_y^2})/\sigma^2 \sim \chi^2(18)$.
Enfin, ces deux lois sont indépendantes d'après la question précédente. Donc sous $\mathbb{P}_{(m_1,m_1)}$ le ratio suivant suit une loi de Student à 18 degrés de liberté :

$$\frac{\sigma\sqrt{18}}{\sqrt{9\widehat{\sigma_x^2} + 9\widehat{\sigma_y^2}}}\frac{\sqrt{n}}{\sqrt{2}\sigma}(\bar{X}_n - \bar{Y}_n) = \frac{\sqrt{n}}{\sqrt{\widehat{\sigma_x^2} + \widehat{\sigma_y^2}}}(\bar{X}_n - \bar{Y}_n).$$

• On propose donc un test non randomisé de zone de rejet

$$\left\{\frac{\sqrt{n}}{\sqrt{\widehat{\sigma_x^2} + \widehat{\sigma_y^2}}}(\bar{X}_n - \bar{Y}_n) < t\right\}$$

où le seuil t est choisi comme le quantile d'ordre α d'une loi de Student à 18 degrés de liberté. On trouve $t = -1.734$ (en utilisant le tableau, puisque la loi de Student est une loi symétrique).

• Avec les données de l'exercice, la statistique de test vaut

$$\frac{\sqrt{10}}{\sqrt{1.08^2 + 0.73^2}}(7.247 - 8.709) = -3.55$$

donc on rejette l'hypothèse nulle. La valeur moyenne du pH dans la piscine A est inférieure à celle dans la piscine B.

5 BONUS : Test à rapport de vraisemblance monotone

Dans cet exercice, on se donne un modèle statistique dominé paramétré par un sous-ensemble $\Theta \subset \mathbb{R}$. On note $L_n(\theta)$ la vraisemblance de l'échantillon en θ. On suppose qu'il existe une statistique $\widehat{S} = S(X_1, \ldots, X_n)$ et, pour tout $(\theta_0, \theta_1) \in \Theta^2$ tel que $\theta_0 < \theta_1$, une fonction g_{θ_0,θ_1} strictement croissante telles que $\frac{L_n(\theta_1)}{L_n(\theta_0)} = g_{\theta_0,\theta_1}(\widehat{S})$. On fixe $\theta_* \in \Theta$ et, pour tester $\mathbf{H}_0 \ : \ \theta \leq \theta_*$ contre $\mathbf{H}_1 \ : \ \theta > \theta_*$, on propose de rejeter $\mathbf{H}_0$ lorsque $r_n(\theta_*) \geq t_\alpha$ où

$$r_n(\theta_*) = \frac{\sup_{\theta_1 > \theta_*} L_n(\theta_1)}{\sup_{\theta_0 \leq \theta_*} L_n(\theta_0)}, \qquad t_\alpha = \inf\left\{t \in \mathbb{R}, \text{ tels que } \sup_{\theta_0 \leq \theta_*} \mathrm{P}_\theta\left(r_n(\theta_*) \geq t\right) \leq \alpha\right\}.$$

1. Montrer qu'il existe $s_\alpha \in \mathbb{R}$ tel que le test rejette $\mathbf{H}_0$ si $\widehat{S} \geq s_\alpha$.

On suppose maintenant qu'il existe s_α tel que $\mathrm{P}_{\theta_*}(\widehat{S} \geq s_\alpha) = \alpha$.

2. Soit $\theta_0 \leq \theta_*$. Montrer que $\mathrm{P}_{\theta_0}(\widehat{S} \geq s_\alpha) \leq \alpha$ et donc que le test est de niveau α.
   Indications : On pourra utiliser le test de Neyman-Pearson de $\mathbf{H}_0 : \theta = \theta_0$ contre $\mathbf{H}_1 : \theta = \theta_*$ de niveau $\alpha' = \mathrm{P}_{\theta_0}(\widehat{S} \geq s_\alpha)$ et comparer sa puissance au test trivial rejetant cette hypothèse lorsque $T = 1$ où T est une variable aléatoire de Bernoulli de paramètre α' indépendante des observations.
3. Soit maintenant un autre test de niveau α défini par sa région de rejet R et notons $x_{1:n} = (x_1, \ldots, x_n)$, $T : (x_{1:n}) \mapsto 1_{S(x_{1:n}) \geq s_\alpha}$ et $\phi : (x_{1:n}) \mapsto 1_{x_{1:n} \in R}$. Soit $\theta_1 > \theta_*$. Montrer que

$$\begin{aligned}
\mathrm{P}_{\theta_1}(\widehat{S} \geq s_\alpha) - \mathrm{P}_{\theta_1}(R) = g_{\theta_*,\theta_1}(s_\alpha) &\left(\mathrm{P}_{\theta_*}(\widehat{S} \geq s_\alpha) - \mathrm{P}_{\theta_*}(R)\right) \\
&+ \mathbb{E}_{\theta_*}\left[(T(X_{1:n}) - \phi(X_{1:n}))\left(\frac{L_n(\theta_1)}{L_n(\theta_*)} - g_{\theta_*,\theta_1}(s_\alpha)\right)\right] \\
&+ \mathbb{E}_{\theta_1}\left[(T(X_{1:n}) - \phi(X_{1:n}))1_{L_n(\theta_*)=0}\right]
\end{aligned}$$

4. Montrer que chacun des trois termes du membre de gauche est positif ou nul et conclure.