Petite Classe 9 - MAP 311 - Groupes 1 et 12

Josselin Garnier

26 juin 2017 - Salle PC 37

Exercice 1. Sur un échantillon représentatif de 1000 personnes, on rapporte les avis favorables pour un homme politique. En novembre, il y avait 38% d'avis favorables, et 36% en décembre. Un éditorialiste dans son journal prend très au sérieux cette chute de 2 points. Le but de cet exercice est de confirmer ou d'infirmer la position du journaliste.

On note p la proportion d'avis favorables en novembre, et q cette proportion en décembre, et on se propose de test $H_0 : p - q = 0$ contre $H_1 : p - q \neq 0$ au niveau 5%. On note $\widehat{p}_n$ la proportion d'avis favorables dans un échantillon de n personnes en novembre (et de même $\widehat{q}_n$ pour décembre).

1. Démontrer que $\sqrt{n}(\widehat{p}_n - p, \widehat{q}_n - q)$ converge en loi vers un vecteur gaussien dont on précisera la matrice de covariance. En déduire que
$$\sqrt{n}\big(\widehat{p}_n - \widehat{q}_n - (p - q)\big) \xrightarrow[n\to\infty]{\text{loi}} \mathcal{N}(0, p(1-p) + q(1-q)).$$
2. Analyser le comportement de
$$T_n = \frac{\sqrt{n}(\widehat{p}_n - \widehat{q}_n)}{\sqrt{\widehat{p}_n(1-\widehat{p}_n) + \widehat{q}_n(1-\widehat{q}_n)}}$$
sous H_0 et sous H_1.
3. Construire un test de niveau asymptotique α de H_0 contre H_1 en se servant de T_n. Faire l'application numérique.

Solution.

1. D'après le théorème central limite, on sait que $\sqrt{n}(\widehat{p}_n - p) \to \mathcal{N}(0, p(1-p))$ et $\sqrt{n}(\widehat{q}_n - q) \to \mathcal{N}(0, q(1-q))$. Comme $\widehat{p}_n$ et $\widehat{q}_n$ sont indépendants, ces deux convergences ont lieu conjointement. Ainsi,
$$\sqrt{n}(\widehat{p}_n - p, \widehat{q}_n - q) \xrightarrow[n\to\infty]{\text{loi}} \mathcal{N}(0, \Sigma), \qquad \text{avec} \quad \Sigma = \begin{pmatrix} p(1-p) & 0 \\ 0 & q(1-q) \end{pmatrix}$$
Par continuité de l'application $(x, y) \mapsto x - y$, on en déduit que
$$\sqrt{n}\big(\widehat{p}_n - \widehat{q}_n - (p - q)\big) \xrightarrow[n\to\infty]{\text{loi}} \mathcal{N}(0, p(1-p) + q(1-q)).$$
2. D'après la loi forte des grands nombres, $\widehat{p}_n(1-\widehat{p}_n) + \widehat{q}_n(1-\widehat{q}_n)$ converge presque sûrement vers $p(1-p) + q(1-q)$. D'après le lemme de Slutsky, sous H_0,
$$T_n = \frac{\sqrt{n}(\widehat{p}_n - \widehat{q}_n)}{\sqrt{\widehat{p}_n(1-\widehat{p}_n) + \widehat{q}_n(1-\widehat{q}_n)}} \xrightarrow[n\to\infty]{\text{loi}} \mathcal{N}(0, 1).$$
Sous H_1, on a $\widehat{p}_n - \widehat{q}_n \to p - q \neq 0$ p.s., donc $|T_n| \to +\infty$ p.s..

3. On choisit un test de région critique $\{|T_n| > c\}$.

   On choisit c tel que $\mathbb{P}(|\mathcal{N}(0,1)| > c) = 5\%$ (ainsi $c \simeq 1{,}96$), de sorte que sous H_0, $\mathbb{P}(|T_n| > c) \to 5\%$. On rejette donc H_0 avec un risque asymptotique de première espèce 5% si $|T_n| > 1{,}96$.

   Application numérique. On trouve $t_n \simeq 0{,}93$. On ne peut pas rejeter H_0 au niveau (asymptotique) 5%, infirmant ainsi la position du journaliste.

Exercice 2. Soit $X = (X_1, \dots, X_n)$ un échantillon de taille n de variables aléatoires réelles indépendantes de même loi exponentielle de paramètre $1/\theta > 0$ avec θ inconnu. Soient $\theta_0 > 0$ et $\alpha \in]0,1[$. On pose $S_n = X_1 + \cdots + X_n$.

On souhaite tester $H_0 = \{\theta = \theta_0\}$ contre $H_1 = \{\theta > \theta_0\}$ au niveau α.

1. Rappeler que S_n suit une loi Γ.

2. Démontrer que sous H_0,
$$Z_n = \frac{2S_n}{\theta_0}$$
   suit une loi du χ^2 à un nombre de de degrés de liberté d_n qu'on précisera.

   On rappelle qu'une loi du χ^2 à d degrés de liberté est la loi de la somme des carrés de d gaussiennes centrées réduites indépendantes, et qu'une densité de cette loi est
$$\frac{1}{2^{\frac{d}{2}}\Gamma(d/2)} t^{\frac{d}{2}-1} e^{-\frac{t}{2}} \mathbf{1}_{t \geq 0}.$$

   On rappelle que la densité de la loi $\Gamma(a, \lambda)$ est $\frac{1}{\Gamma(a)} \lambda^a x^{a-1} e^{-\lambda x} \mathbf{1}_{x>0}$.

3. En déduire une région critique de la forme
$$W_n = \{S_n \geq c\}$$
   avec c une constante qu'on exprimera en fonction de θ_0 et $z_{1-\alpha}$, le quantile d'ordre $1-\alpha$ de la loi du χ^2 à d_n degrés de liberté.

4. On suppose que θ est le temps moyen d'attente du RER B à la gare de Lozère. Une association d'usagers et la RATP souhaitent tester si le RER B respecte un temps d'attente moyen réglementaire d'au plus $\theta_0 = 15\,\text{min}$. Teste-t-on
$$H_0 = \{\theta \leq \theta_0\} \text{ contre } H_1 = \{\theta > \theta_0\} \qquad \text{ou} \qquad H_0 = \{\theta \geq \theta_0\} \text{ contre } H_1 = \{\theta < \theta_0\}?$$

Solution.

1. S_n est une somme de n variables aléatoires indépendantes de loi exponentielle de paramètre λ, donc S_n suit la loi $\Gamma(n, \lambda)$, d'où le résultat.

2. On utilise la méthode de la fonction muette. Pour une fonction continue bornée, sous H_0,
$$\mathbb{E}(f(Z_n)) = \int_0^\infty f\left(\frac{2x}{\theta_0}\right) \frac{1}{\Gamma(n)} \theta_0^{-n} x^{n-1} e^{-x/\theta_0} dx.$$
Le changement de variable $u = 2x/\theta_0$ donne
$$\mathbb{E}(f(Z_n)) = \int_0^\infty f(u) \frac{1}{2^n \Gamma(n)} u^{n-1} e^{-u/2} du.$$
Ainsi, Z_n suit une loi du χ^2 à $d_n = 2n$ degrés de liberté.

3. Ainsi,
$$\mathbb{P}_{\theta_0}\left(\frac{2S_n}{\theta_0} \geq z_{1-\alpha}\right) = \mathbb{P}_{\theta_0}(Z_n \geq z_{1-\alpha}) = \alpha.$$
La région critique est donc de la forme $W_n = \{S_n \geq \frac{z_{1-\alpha}\theta_0}{2}\}$.

4. Du point du vue de la RATP, il faut éviter d'annoncer que le RER B présente des retards alors que ce n'est pas le cas (et ce n'est pas très grave d'annoncer qu'il n'y a pas de retard alors qu'il y en a...). Il faut donc minimiser l'erreur d'annoncer $\theta > \theta_0$ alors que $\theta \leq \theta_0$.

   Du point de vue des usagers, il faut éviter d'annoncer qu'il n'y a pas de retard alors qu'il y en a (et ce n'est pas très grave d'annoncer qu'il y a des retards alors que ce n'est pas le cas).

   Ainsi, la RATP va préférer le test $H_0 = \{\theta \leq \theta_0\}$ contre $H_1 = \{\theta > \theta_0\}$ alors que l'assocation d'usagers va préférer le test $H_0 = \{\theta \geq \theta_0\}$ contre $H_1 = \{\theta < \theta_0\}$.

Exercice 3. La hauteur maximale en mètres de la crue annuelle d'un fleuve est modélisée par une variable aléatoire de Rayleigh de paramètre a qui a pour densité :
$$f(x) = \frac{x}{a} e^{-\frac{x^2}{2a}} \mathbf{1}_{[0,+\infty[}(x)$$
où a est un paramètre inconnu.

1. Déterminer l'estimateur du maximum de vraisemblance $\widehat{a}_n$ de a.

2. Si X suit une loi de Rayleigh de paramètre a, démontrer que X^2 suit une loi exponentielle de paramètre $1/(2a)$.

3. L'estimateur $\widehat{a}_n$ est-il sans biais ?

4. Trouver un intervalle de confiance asymptotique pour a au niveau 95%.

   On rappelle que la variance d'une loi exponentielle de paramètre λ est $1/\lambda^2$.

5. Une compagnie d'assurance estime qu'une catastrophe correspond à une crue de plus de 6m. Trouver un intervalle de confiance asymptotique pour la probabilité p qu'une catastrophe se produise durant une année au niveau 95%.

Solution. Soit $(X_i)_{i\geq 1}$ une suite de variables aléatoires indépendantes de loi de Rayleigh.

1. Une densité de $(X_1, \ldots, X_n)$ au point $(x_1, \ldots, x_n)$ est

$$p(x_1, \ldots, x_n; a) = \frac{\prod_{i=1}^n x_i}{a^n} e^{-\frac{\sum_{i=1}^n x_i^2}{2a}}.$$

Pour chercher quelle valeur de a maximise cette quantité, on calcule sa dérivée par rapport à a, qui vaut

$$\begin{aligned} \partial_a p(x_1, \ldots, x_n; a) &= -n\frac{\prod_{i=1}^n x_i}{a^{n+1}} e^{-\frac{\sum_{i=1}^n x_i^2}{2a}} + \frac{\prod_{i=1}^n x_i}{a^n} \cdot \frac{\sum_{i=1}^n x_i^2}{2a^2} \cdot e^{-\frac{\sum_{i=1}^n x_i^2}{2a}} \\ &= \frac{\prod_{i=1}^n x_i}{a^{n+2}} e^{-\frac{\sum_{i=1}^n x_i^2}{2a}} \left(-an + \frac{\sum_{i=1}^n x_i^2}{2} \right). \end{aligned}$$

Ainsi,

$$\widehat{a}_n = \frac{\sum_{i=1}^n X_i^2}{2n}.$$

2. On utilise la méthode de la fonction muette. Soit $F : \mathbb{R} \to \mathbb{R}$ une fonction continue bornée. Alors, avec le changement de variable $u = x^2$ (et donc $x = \sqrt{u}$ et $dx = \frac{du}{2\sqrt{u}}$):

$$\mathbb{E}(F(X)) = \int_0^\infty F(x^2) \frac{x}{a} e^{-\frac{x^2}{2a}} dx = \int_0^\infty F(u) \frac{1}{2a} e^{-\frac{u}{2a}} du,$$

d'où le résultat.

3. Comme X_1^2 suit une loi exponentielle de paramètre $1/(2a)$, on a $\mathbb{E}(X_1^2) = 2a$, et par linéarité de l'espérance $\widehat{a}_n$ est sans biais.

4. D'après le théorème central limite,

$$\sqrt{n} \left(\frac{\sum_{i=1}^n X_i^2}{n} - 2a \right) \underset{n\to\infty}{\longrightarrow} \mathcal{N}(0, \mathrm{Var}(X_1^2)) = \mathcal{N}(0, 4a^2)$$

en loi. On en déduit en utilisant le lemme de Slutsky que

$$\frac{\sqrt{n}}{\widehat{a}_n} (\widehat{a}_n - a) \underset{n\to\infty}{\longrightarrow} \mathcal{N}(0, 1).$$

En notant $z_{1-\alpha/2} \simeq 1{,}96$ le quantile de niveau $1 - \alpha/2$ de la loi gaussienne centrée réduite, on en déduit que

$$\widehat{I}_n = \left[\widehat{a}_n \left(1 - \frac{z_{1-\alpha/2}}{\sqrt{n}} \right), \widehat{a}_n \left(1 + \frac{z_{1-\alpha/2}}{\sqrt{n}} \right) \right]$$

est un intervalle de confiance asymptotique pour a au niveau 95%.

5. La probabilité d'une catastrophe est

$$p = \int_6^\infty \frac{x}{a} e^{-\frac{x^2}{2a}} dx = \left[e^{-\frac{x^2}{2a}} \right]_6^\infty = e^{-18/a}.$$

Ainsi, un intervalle de confiance asymptotique pour p au niveau 95% est :

$$\widehat{J}_n = \left[\exp\left(-\frac{18}{\widehat{a}_n \left(1 - \frac{z_{1-\alpha/2}}{\sqrt{n}}\right)} \right), \exp\left(-\frac{18}{\widehat{a}_n \left(1 + \frac{z_{1-\alpha/2}}{\sqrt{n}}\right)} \right) \right].$$

Exercice 4. On se place dans le modèle gaussien à espérance et variance égales $\boldsymbol{\mathcal{P}} = \{\mathcal{N}(\theta, \theta), \theta > 0\}$.

1. Déterminer l'Estimateur du Maximum de Vraisemblance $\hat{\theta}_n$ de θ. Cet estimateur est-il convergent ?

2. Montrer que $\hat{\theta}_n$ est asymptotiquement normal et identifier sa variance asymptotique $\sigma^2(\theta)$.

3. On note $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ et $\overline{V}_n = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)^2$ la moyenne et la variance empiriques et pour $\lambda \in [0,1]$, on pose $T_n^\lambda = \lambda \overline{X}_n + (1-\lambda)\overline{V}_n$. Montrer que T_n^λ est un estimateur convergent et asymptotiquement normal de θ dont on calculera la variance asymptotique $\sigma^2(\lambda, \theta)$. Vérifier que $\min_{\lambda \in [0,1]} \sigma^2(\lambda, \theta) = \sigma^2(\theta)$. Dans la classe des estimateurs $(T_n^\lambda)_{\lambda \in [0,1]}$, existe-il un estimateur aussi performant que $\hat{\theta}_n$ pour tout $\theta > 0$ en termes de variance asymptotique ?

Pour $\theta_0 > 0$ fixé, on souhaite tester l'hypothèse nulle $H_0 = [0, \theta_0]$ contre l'hypothèse alternative $H_1 =]\theta_0, +\infty[$.

4. Construire un test asymptotique de niveau α à l'aide de $\hat{\theta}_n$.
   A.N.: $\theta_0 = 3{,}8$, $n = 100$, $\overline{x}_n = 4{,}18$, $\overline{v}_n = 3{,}84$, et $\alpha = 0{,}05$.

5. Construire un test asymptotique de niveau α à l'aide de T_n^λ.
   A.N.: $\lambda \in \{0, 0{,}5, 1, \lambda_0\}$ où λ_0 minimise $\lambda \in [0,1] \mapsto \sigma^2(\lambda, \theta_0)$.

Solution.

1. La vraisemblance est :

$$p_n(\boldsymbol{x}, \theta) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\theta}} \exp\Big(- \frac{(x_i - \theta)^2}{2\theta} \Big) = \big(2\pi\theta\big)^{-n/2} \exp\Big(- \frac{\sum_{i=1}^n x_i^2}{2\theta} - \frac{n\theta}{2} + \sum_{i=1}^n x_i \Big)$$

   La log-vraisemblance est :

$$l_n(\boldsymbol{x}, \theta) = -\frac{n}{2}\ln\theta - \frac{\sum_{i=1}^n x_i^2}{2\theta} - \frac{n\theta}{2} - \frac{n}{2}\ln(2\pi) + \sum_{i=1}^n x_i$$

   et sa dérivée est :

$$\partial_\theta l_n(\boldsymbol{x}, \theta) = -\frac{n}{2\theta} + \frac{\sum_{i=1}^n x_i^2}{2\theta^2} - \frac{n}{2}$$

Sur $]0,+\infty[$, la dérivée de la log vraisemblance s'annule uniquement en

$$\hat{\theta}(\boldsymbol{x}) = \frac{\frac{2}{n}\sum_{i=1}^n x_i^2}{1+\sqrt{1+\frac{4}{n}\sum_{i=1}^n x_i^2}}$$

Elle est positive sur $]0,\hat{\theta}(\boldsymbol{x})[$ et négative sur $]\hat{\theta}(\boldsymbol{x}),+\infty[$. Donc elle atteint son unique maximum en $\hat{\theta}(\boldsymbol{x})$ ce qui montre que l'EMV de θ est

$$\hat{\theta}_n = \frac{\frac{2}{n}\sum_{i=1}^n X_i^2}{1+\sqrt{1+\frac{4}{n}\sum_{i=1}^n X_i^2}}$$

D'après la loi forte des grands nombres, sous $\mathbb{P}_\theta$, les X_i sont i.i.d. de loi $\mathcal{N}(\theta,\theta)$ et donc

$$\frac{1}{n}\sum_{i=1}^n X_i^2 \overset{n\to+\infty}{\longrightarrow} \mathbb{E}_\theta(X_1^2) = \theta+\theta^2$$

p.s.. Ceci montre que l'EMV est convergent :

$$\hat{\theta}_n \overset{n\to+\infty}{\longrightarrow} \frac{2(\theta+\theta^2)}{1+\sqrt{1+4(\theta+\theta^2)}} = \frac{2(\theta+\theta^2)}{1+(1+2\theta)} = \theta$$

p.s..

2. On invoque le théorème de la limite centrale :

$$\sqrt{n}\Big(\frac{1}{n}\sum_{i=1}^n X_i^2 - (\theta+\theta^2)\Big) \overset{n\to+\infty}{\longrightarrow} \mathcal{N}\big(0,\mathrm{Var}_\theta(X_1^2)\big)$$

en loi, avec

$$\mathrm{Var}_\theta(X_1^2) = \mathbb{E}_\theta(X_1^4) - \mathbb{E}_\theta(X_1^2)^2 = (\theta^4+6\theta^3+3\theta^2)-(\theta^4+2\theta^3+\theta^2) = 4\theta^3+2\theta^2 = 2\theta^2(1+2\theta)$$

[Pour calculer $\mathbb{E}[X^n]$ avec $X\sim\mathcal{N}(m,\sigma^2)$, développer $\mathbb{E}[(m+\sigma Z)^n]$ avec $Z\sim\mathcal{N}(0,1)$]. On a $\hat{\theta}_n = g(S_n)$ avec $S_n=\sum_{i=1}^n X_i^2/n$ et $g(S)=2S/(1+\sqrt{1+4S})$. Avec la méthode Delta on trouve

$$\sqrt{n}\big(\hat{\theta}_n-\theta\big) \overset{n\to+\infty}{\longrightarrow} \mathcal{N}\big(0, g'(S^\theta)^2\mathrm{Var}_\theta(X_1^2)\big)$$

avec $g'(S)=1/\sqrt{1+4S}$ et $S^\theta=\theta+\theta^2$. En particulier $g'(S^\theta)=1/(1+2\theta)$ et donc

$$\sqrt{n}\big(\hat{\theta}_n-\theta\big) \overset{n\to+\infty}{\longrightarrow} \mathcal{N}\big(0,\sigma^2(\theta)\big)$$

en loi, avec

$$\sigma^2(\theta) = \frac{2\theta^2}{1+2\theta}$$

3. D'après la loi forte des grands nombres, $\overline{X}_n$ converge $\mathbb{P}_\theta$-p.s. vers $\mathbb{E}_\theta(X_1)=\theta$ et $\overline{V}_n = \frac{1}{n}\sum_{i=1}^n X_i^2 - \overline{X}_n^2$ converge $\mathbb{P}_\theta$-p.s. vers $\mathbb{E}_\theta(X_1^2)-\mathbb{E}_\theta(X_1)^2=\theta$. Donc T_n^λ est un estimateur convergent de θ. On invoque le théorème de la limite centrale vectoriel

$$\sqrt{n}\Big(\begin{pmatrix}\overline{X}_n\\ \frac{1}{n}\sum_{i=1}^n X_i^2\end{pmatrix} - \begin{pmatrix}\theta\\ \theta^2+\theta\end{pmatrix}\Big) \overset{n\to+\infty}{\longrightarrow} \mathcal{N}\Big(\begin{pmatrix}0\\0\end{pmatrix},\mathbf{C}_\theta\Big),$$

en loi, avec

$$\mathbf{C}_\theta = \begin{pmatrix} \mathbb{E}_\theta(X_1^2) - \mathbb{E}_\theta(X_1)^2 & \mathbb{E}_\theta(X_1^3) - \mathbb{E}_\theta(X_1)\mathbb{E}_\theta(X_1^2) \\ \mathbb{E}_\theta(X_1^3) - \mathbb{E}_\theta(X_1)\mathbb{E}_\theta(X_1^2) & \mathbb{E}_\theta(X_1^4) - \mathbb{E}_\theta(X_1^2)^2 \end{pmatrix} = \begin{pmatrix} \theta & 2\theta^2 \\ 2\theta^2 & 2\theta^2(1+2\theta) \end{pmatrix}$$

car $\mathbb{E}(X_1^3) = \theta^3 + 3\theta^2$. Or

$$T_n^\lambda = \lambda \overline{X}_n + (1-\lambda)\overline{V}_n = \lambda \overline{X}_n - (1-\lambda)\overline{X}_n^2 + (1-\lambda)\frac{1}{n}\sum_{i=1}^n X_i^2 = g\Big(\frac{1}{n}\sum_{i=1}^n X_i, \frac{1}{n}\sum_{i=1}^n X_i^2\Big)$$

avec

$$g(x,y) = \lambda x - (1-\lambda)x^2 + (1-\lambda)y$$

Donc

$$\sqrt{n}\big(T_n^\lambda - \theta\big) \overset{n\to+\infty}{\longrightarrow} \mathcal{N}\big(0, \sigma^2(\theta,\lambda)\big),$$

en loi, avec

$$\begin{aligned} \sigma^2(\theta,\lambda) &= \begin{pmatrix} g_x(\theta,\theta+\theta^2) \\ g_y(\theta,\theta+\theta^2) \end{pmatrix}^t \begin{pmatrix} \theta & 2\theta^2 \\ 2\theta^2 & 2\theta^2(1+2\theta) \end{pmatrix} \begin{pmatrix} g_x(\theta,\theta+\theta^2) \\ g_y(\theta,\theta+\theta^2) \end{pmatrix} \\ &= \begin{pmatrix} \lambda - 2(1-\lambda)\theta \\ 1-\lambda \end{pmatrix}^t \begin{pmatrix} \theta & 2\theta^2 \\ 2\theta^2 & 2\theta^2(1+2\theta) \end{pmatrix} \begin{pmatrix} \lambda - 2(1-\lambda)\theta \\ 1-\lambda \end{pmatrix} \\ &= \lambda^2\theta + 2\theta^2(1-\lambda)^2 \end{aligned}$$

Le minimum de la variance asymptotique $\sigma^2(\theta,\lambda)$ en λ est atteint en λ_θ tel que $\partial_\lambda \sigma^2(\theta,\lambda) = 0$, c'est-à-dire $\lambda_\theta = 2\theta/(1+2\theta)$. On trouve alors

$$\min_{\lambda\in[0,1]} \sigma^2(\theta,\lambda) = \sigma^2(\theta,\lambda_\theta) = \frac{\theta}{1+2\theta} = \sigma^2(\theta)$$

Mais comme le λ optimal dépend de θ, on ne peut pas prendre parmi la famille $(T_n^\lambda)_{\lambda\in[0,1]}$, un estimateur aussi performant que $\hat\theta_n$ pour tout $\theta > 0$ en termes de variance asymptotique.

Remarque : On a

$$\hat\theta_n = \frac{2(\overline{V}_n + \overline{X}_n^2)}{1 + \sqrt{1 + 4(\overline{V}_n + \overline{X}_n^2)}}$$

qui n'est pas une combinaison linéaire de $\overline{V}_n$ et $\overline{X}_n^2$.

4. On construit un test de $H_0 = [0,\theta_0]$ contre H_0^c à l'aide de $\hat\theta_n$. La région critique est de la forme :

$$W_n = \{\hat\theta_n \geq c\}$$

et on choisit c de telle manière que $\lim_{n\to\infty} \sup_{\theta\in[0,\theta_0]} \mathbb{P}_\theta(\hat\theta_n \geq c) = \alpha$. On a

$$\mathbb{P}_\theta(\hat\theta_n \geq c) \simeq \mathbb{P}\Big(\theta + \frac{\sigma(\theta)}{\sqrt{n}} Z \geq c\Big) = \mathbb{P}\Big(Z \geq \frac{\sqrt{n}}{\sigma(\theta)}(c-\theta)\Big)$$

avec $Z \sim \mathcal{N}(0,1)$. Donc c est tel que

$$\mathbb{P}\Big(Z \geq \frac{\sqrt{n}}{\sigma(\theta_0)}(c - \theta_0)\Big) = \alpha$$

c'est-à-dire

$$c = \theta_0 + \frac{\sigma(\theta_0)}{\sqrt{n}}\Phi^{-1}(1-\alpha)$$

A.N. : Si $\theta_0 = 3{,}8$, $n = 100$, $\overline{x}_n = 4{,}18$ et $\overline{v}_n = 3{,}84$, alors on trouve $\hat{\theta}_n = 4{,}14$ et comme $\Phi^{-1}(0{,}95) = 1{,}65$, on a $c = 4{,}10$, donc on rejette l'hypothèse H_0.

5. On construit un test de $H_0 = [0, \theta_0]$ contre H_0^c à l'aide de T_n^λ. La région critique est de la forme :

$$W_{n,\lambda} = \{T_n^\lambda \geq c_\lambda\}$$

et on choisit c_λ de telle manière que $\lim_{n\to\infty} \sup_{\theta\in[0,\theta_0]} \mathbb{P}_\theta(T_n^\lambda \geq c_\lambda) = \alpha$. On a

$$\mathbb{P}_\theta(T_n^\lambda \geq c_\lambda) \simeq \mathbb{P}\Big(\theta + \frac{\sigma(\theta\lambda)}{\sqrt{n}}Z \geq c_\lambda\Big) = \mathbb{P}\Big(Z \geq \frac{\sqrt{n}}{\sigma(\theta,\lambda)}(c_\lambda - \theta)\Big)$$

avec $Z \sim \mathcal{N}(0,1)$. Donc c_λ est tel que

$$\mathbb{P}\Big(Z \geq \frac{\sqrt{n}}{\sigma(\theta_0,\lambda)}(c_\lambda - \theta_0)\Big) = \alpha$$

c'est-à-dire

$$c_\lambda = \theta_0 + \frac{\sigma(\theta_0,\lambda)}{\sqrt{n}}\Phi^{-1}(1-\alpha)$$

A.N. : Si $\lambda \in \{0, 0{,}5, 1, \lambda_0\}$ où λ_0 minimise $\lambda \in [0,1] \mapsto \sigma^2(\lambda, \theta_0)$, c'est-à-dire $\lambda_0 = 2\theta_0/(1+2\theta_0) = 0{,}88$. On a

$$c_0 = 4{,}69, \quad c_{0,5} = 4{,}27, \quad c_1 = 4{,}12, \quad c_{\lambda_0} = 4{,}10,$$

Comme $\hat{\theta}_n = 4{,}14$, on accepte H_0 dans les deux premiers cas, on rejette H_0 dans les deux derniers cas. Les deux derniers tests exploitent mieux les données que les deux premiers et arrivent à rejeter l'hypothèse nulle.

Exercice 5. On considère les décimales du nombre π et on souhaite savoir si elles sont uniformément réparties sur $\mathcal{X} = \{0, \ldots, 9\}$. On calcule π à 10^{-n} près, avec $n = 10000$, ce qui nous donne les n premières décimales de π. On trouve $N_0 = 968$, $N_1 = 1026$, $N_2 = 1021$, $N_3 = 975$, $N_4 = 1012$, $N_5 = 1046$, $N_6 = 1021$, $N_7 = 969$, $N_8 = 948$ et $N_9 = 1014$.
Tester l'hypothèse que les décimales de π sont uniformément réparties au niveau 5%.

Solution. On teste donc l'hypothèse $H_0 = \{(1/10, \ldots, 1/10)\}$. On a $\zeta_{10000}^{obs} \simeq 9{,}3$. Sous l'hypothèse H_0, ζ_{10000} suit une loi du χ^2 à 9 degrés de liberté. Si $Z \sim \chi_9^2$, on a $\mathbb{P}(Z \geq 16{,}9) = 0{,}05$. On ne peut donc pas rejeter l'hypothèse que les décimales de π sont uniformément réparties au niveau 5%.

Exercice 6. Soit $(X_i)_{i\geq 1}$ des variables aléatoires indépendantes et de même loi. On suppose que X_1 est de carré intégrable et que $\mathbb{E}(X_1) = 1$. Étudier la convergence en loi de $\sqrt{n}\frac{\sum_{i=1}^n (X_i - 1)}{\sum_{i=1}^n X_i}$.

Solution. Soit σ^2 la variance de X_1. On suppose que $\sigma^2 > 0$ (sinon $X_i = 1$ p.s.). On écrit :

$$\sqrt{n}\frac{\sum_{i=1}^n (X_i - 1)}{\sum_{i=1}^n X_i} = \frac{\sum_{i=1}^n (X_i - 1)}{\sqrt{n}} \frac{1}{\frac{\sum_{i=1}^n X_i}{n}}.$$

Posons $Y_n = \frac{\sum_{i=1}^n (X_i - 1)}{\sqrt{n}}$ et $Z_n = \frac{1}{\frac{\sum_{i=1}^n X_i}{n}}$. D'après le théorème central limite, Y_n converge en loi vers σN avec N une variable gaussienne centrée réduite et d'après la loi forte des grands nombres Z_n converge presque sûrement (donc en loi) vers 1. D'après le lemme de Slutsky, (Y_n, Z_n) converge en loi vers $(\sigma N, 1)$ et en composant par l'application continue $f(y, z) = y/z$, on en déduit que

$$\sqrt{n}\frac{\sum_{i=1}^n (X_i - 1)}{\sum_{i=1}^n X_i} \underset{n\to\infty}{\longrightarrow} \sigma N$$

en loi.