Cauchetaux
Maxence

2024 - 2025
École des Ponts

Intervalles de confiance.

Exercice 3.A.3.

Dans le modèle gaussien $\mathcal{P} = \{\mathcal{N}(\mu, \sigma^2) \mid \mu \in \mathbb{R}, \sigma^2 > 0\}$, avec X_i iid $\sim \mathcal{N}(\mu, \sigma^2)$, on sait par le cours que :

$$\boxed{Z_{n-1} = \frac{n-1}{\sigma^2} S_n^2 \sim \chi^2(n-1)} \quad \text{où} \quad S_n^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \overline{X}_n)^2.$$

En notant $\chi^2_{n-1, 1-\alpha}$ le quantile d'ordre $1-\alpha$ de la loi $\chi^2(n-1)$, on a, pour $\alpha \in]0, \frac{1}{2}[$,

$$1-\alpha = \mathbb{P}_{\mu,\sigma^2}\left(Z_{n-1} \in \left[\chi^2_{n-1, \frac{\alpha}{2}}, \chi^2_{n-1, 1-\frac{\alpha}{2}}\right]\right)$$

Or : $\left\{Z_{n-1} \in \left[\chi^2_{n-1, \frac{\alpha}{2}}, \chi^2_{n-1, 1-\frac{\alpha}{2}}\right]\right\}$

$$= \left\{\frac{n-1}{\sigma^2} S_n^2 \in \left[\chi^2_{n-1, \frac{\alpha}{2}}, \chi^2_{n-1, 1-\frac{\alpha}{2}}\right]\right\}$$

$$= \left\{\sigma^2 \in \left[\frac{(n-1) S_n^2}{\chi^2_{n-1, 1-\frac{\alpha}{2}}}, \frac{(n-1) S_n^2}{\chi^2_{n-1, \frac{\alpha}{2}}}\right]\right\}$$

Bilan :
$$\mathbb{P}_{\mu,\sigma^2}\left(\sigma^2 \in \left[\frac{(n-1) S_n^2}{\chi^2_{n-1, 1-\frac{\alpha}{2}}}, \frac{(n-1) S_n^2}{\chi^2_{n-1, \frac{\alpha}{2}}}\right]\right) = 1-\alpha$$

Ainsi $\boxed{\left[\frac{(n-1) S_n^2}{\chi^2_{n-1, 1-\frac{\alpha}{2}}}, \frac{(n-1) S_n^2}{\chi^2_{n-1, \frac{\alpha}{2}}}\right]}$ (dont les bornes sont des statistiques), est un intervalle de confiance de niveau $1-\alpha$ pour σ^2.

Exercice 3.A.4.

1/ • $X_1 \leq 1$ ps, donc $X_1 \in L^1$

• De plus :

$$\mathbb{E}_\theta(X_1) = \int_{\mathbb{R}} 2x \left[\mathbb{1}_{\{0 < x \leq \theta\}} \frac{x}{\theta} + \mathbb{1}_{\{\theta < x \leq 1\}} \frac{1-x}{1-\theta}\right] dx$$ (transfert)

$$= \frac{2}{\theta}\int_0^\theta x^2 dx + \frac{2}{1-\theta}\int_\theta^1 x(1-x)\, dx$$

$$= \frac{2}{\theta}\frac{\theta^3}{3} + \frac{2}{1-\theta}\left(\frac{1}{6} - \frac{\theta^2}{2} + \frac{\theta^3}{3}\right)$$

$$= \frac{2}{3}\theta^2 + \frac{2}{1-\theta}\left(\frac{1}{6} - \frac{\theta^2}{2} + \frac{\theta^3}{3}\right)$$

$$= \frac{1}{3(1-\theta)}\left[2\theta^2(1-\theta) + 1 - 3\theta^2 + 2\theta^3\right]$$

$$= \frac{1}{3(1-\theta)}\left[2\theta^2 - \cancel{2\theta^3} + 1 - 3\theta^2 + \cancel{2\theta^3}\right]$$

$$= \frac{1}{3(1-\theta)}(1-\theta^2)$$

$$\boxed{\mathbb{E}_\theta(X_1) = \frac{1+\theta}{3}}$$

- $3\overline{X}_n - 1$ n'est pas un estimateur de θ. En effet, il n'est pas à valeurs dans $]0,1[$ ps.

C'est pourquoi on pose $\boxed{Z_n = \min(1, 3\overline{X}_n - 1)}$ qui est bien un estimateur.

Par loi forte et continuité de $\min(1,\cdot)$, on obtient que $\boxed{Z_n \xrightarrow[n\infty]{ps} \min(1,\theta) = \theta}$

ie (Z_n) est un estimateur fortement consistant de θ.

2/ X_1 est L^2 car bornée et on a

$$\mathbb{V}_\theta(X_1) = \mathbb{E}(X_1^2) - \mathbb{E}(X_1)^2$$

$$= \mathbb{E}(X_1^2) - \left(\frac{1+\theta}{3}\right)^2$$

Et $\mathbb{E}_\theta(X_1^2) = \int_{\mathbb{R}} 2x^2\left(\mathbb{1}_{\{0<x\leq\theta\}}\frac{x}{\theta} + \mathbb{1}_{\{\theta<x<1\}}\frac{1-x}{1-\theta}\right)dx$ (transfert)

$$= \frac{2}{\theta}\int_0^\theta x^3\,dx + \frac{2}{1-\theta}\int_\theta^1 x^2(1-x)\,dx$$

$$= \frac{2}{\theta}\times\frac{\theta^4}{4} + \frac{2}{1-\theta}\left(\frac{1}{3} - \frac{1}{4} - \frac{1}{3}\theta^3 + \frac{1}{4}\theta^4\right)$$

$$= \frac{\theta^3}{2} + \frac{2}{1-\theta}\left(\frac{1}{12} - \frac{1}{3}\theta^3 + \frac{1}{4}\theta^4\right)$$

$$= \frac{\theta^3(1-\theta) + 4\left(\frac{1}{12} - \frac{1}{3}\theta^3 + \frac{1}{4}\theta^4\right)}{2(1-\theta)}$$

$$= \frac{\theta^3 - \cancel{\theta^4} + \frac{1}{3} - \frac{4\theta^3}{3} + \cancel{\theta^4}}{2(1-\theta)}$$

$$\boxed{\mathbb{E}_\theta(X_1^2) = \frac{1-\theta^3}{6(1-\theta)}}$$

Puis : $\mathbb{V}_\theta(X_1) = \frac{1-\theta^3}{6(1-\theta)} - \frac{1+2\theta+\theta^2}{9}$

$$= \frac{1}{9(1-\theta)}\left[\frac{9}{6}(1-\theta^3) - (1+2\theta+\theta^2)(1-\theta)\right]$$

$$= \frac{1}{9(1-\theta)}\left(\frac{1}{2} - \theta + \theta^2 - \frac{1}{2}\theta^3\right)$$

$$= \frac{1}{18(1-\theta)}\left(1 - 2\theta + 2\theta^2 - \theta^3\right)$$

$$\boxed{\mathbb{V}_\theta(X_1) = \frac{1-\theta+\theta^2}{18}.}$$

- On écrit :

$$\begin{aligned}\sqrt{n}(Z_n-\theta) &= \sqrt{n}\left(\min(1, 3\overline{X}_n-1)-\theta\right)\\ &= \min\left(\sqrt{n}(1-\theta), \sqrt{n}(3\overline{X}_n-1-\theta)\right)\\ &= \min\Big(\underbrace{\sqrt{n}(1-\theta)}_{=X_n}, \underbrace{3\sqrt{n}\left(\overline{X}_n-\tfrac{1+\theta}{3}\right)}_{=Y_n}\Big)\\ &= \min(X_n, Y_n)\\ &= \underbrace{X_n \mathbb{1}\{X_n \leq Y_n\}}_{=\tilde{X}_n} + \underbrace{Y_n \mathbb{1}\{Y_n \leq X_n\}}_{=\tilde{Y}_n}\\ &= \boxed{\tilde{X}_n + \tilde{Y}_n}\end{aligned}$$

bien

- Or : $Y_n = 3\sqrt{n}(\overline{X}_n - \mathbb{E}_\theta(X_1)) \xrightarrow[n\to\infty]{\text{loi}} \mathcal{N}\left(0, \frac{1}{2}(1-\theta+\theta^2)\right)$ par théorème central limite.

et $\mathbb{1}\{Y_n \leq X_n\} \xrightarrow[n\to\infty]{ps} 1$ (car $\overline{X}_n - \frac{1+\theta}{3} \xrightarrow[n\to\infty]{ps} 0$) et donc a fortiori en $\mathbb{P}$.

Par Slutsky, on obtient donc $(Y_n, \mathbb{1}\{Y_n \leq X_n\}) \xrightarrow[n\to\infty]{\text{loi}} (\mathcal{N}(0, \frac{1}{2}(1-\theta+\theta^2)), 1)$

et par continuité de $\varphi : (x,y) \mapsto xy$, on a $\boxed{\tilde{Y}_n \xrightarrow[n\to\infty]{\text{loi}} \mathcal{N}\left(0, \frac{1}{2}(1-\theta+\theta^2)\right)}$

- $\tilde{X}_n = X_n \mathbb{1}\{X_n \leq Y_n\} \xrightarrow[n\to\infty]{ps} 0$ (car $\overline{X}_n - \frac{1+\theta}{3} \xrightarrow[n\to\infty]{ps} 0$) et donc a fortiori en $\mathbb{P}$.

Ainsi, de nouveau par Slutsky, $(\tilde{X}_n, \tilde{Y}_n) \xrightarrow[n\to\infty]{\text{loi}} (0, \mathcal{N}(0, \frac{1}{2}(1-\theta+\theta^2)))$ et par continuité de $\tilde{\varphi} : (x,y) \mapsto x+y$, on obtient finalement

$$\boxed{\sqrt{n}(Z_n-\theta) = \tilde{X}_n + \tilde{Y}_n \xrightarrow[n\to\infty]{\text{loi}} \mathcal{N}\left(0, \frac{1}{2}(1-\theta+\theta^2)\right)}$$

[(Z_n) asymptotiquement normal de variance asymptotique $\frac{1}{2}(1-\theta+\theta^2)$]

4/ Par corollaire de l'inégalité de Hoeffding, comme les X_i sont à valeurs dans $[0,1]$, on a :

$$\forall r \geq 0, \forall n \geq 1 \quad \boxed{\mathbb{P}_\theta\left(|\overline{X}_n - \mathbb{E}_\theta(X_1)| \geq \frac{r}{\sqrt{n}}\right) \leq 2e^{-2r^2}}.$$

On prend $r = r^H(\alpha)$ tq $2e^{-2r^2} = \alpha$ ie $\boxed{r^H(\alpha) = \sqrt{-\frac{1}{2}\ln\left(\frac{\alpha}{2}\right)}}$ et on obtient donc

$$\mathbb{P}_\theta\left(|\overline{X}_n - \mathbb{E}_\theta(X_1)| \geq \frac{r^H(\alpha)}{\sqrt{n}}\right) \leq \alpha.$$

Puis : $1-\alpha = \mathbb{P}_\theta\left(|\overline{X}_n - \mathbb{E}_\theta(X_1)| \leq \frac{r^H(\alpha)}{\sqrt{n}}\right)$

Or :

$$\begin{aligned}\left\{|\overline{X}_n - \mathbb{E}_\theta(X_1)| \leq \frac{r^H(\alpha)}{\sqrt{n}}\right\} &= \left\{\frac{1+\theta}{3} - \overline{X}_n \in \left[\frac{-r^H(\alpha)}{\sqrt{n}}, \frac{r^H(\alpha)}{\sqrt{n}}\right]\right\}\\ &= \left\{\frac{1+\theta}{3} \in \left[\overline{X}_n - \frac{r^H(\alpha)}{\sqrt{n}}, \overline{X}_n + \frac{r^H(\alpha)}{\sqrt{n}}\right]\right\}\\ &= \left\{\theta \in \left[3\left(\overline{X}_n - \frac{r^H(\alpha)}{\sqrt{n}}\right) - 1, 3\left(\overline{X}_n + \frac{r^H(\alpha)}{\sqrt{n}}\right) - 1\right]\right\}.\end{aligned}$$

Bilan :
$$\mathbb{P}_\theta\left(\theta \in \left[3\left(\overline{X}_n - \frac{r^H(\alpha)}{\sqrt{n}}\right) - 1, 3\left(\overline{X}_n + \frac{r^H(\alpha)}{\sqrt{n}}\right) - 1\right]\right) \geq 1-\alpha$$
Donc l'intervalle $\boxed{\left[3\left(\overline{X}_n - \frac{r^H(\alpha)}{\sqrt{n}}\right) - 1, 3\left(\overline{X}_n + \frac{r^H(\alpha)}{\sqrt{n}}\right) - 1\right]}$ dont les bornes sont des statistiques, est un intervalle de confiance par excès de θ de niveau $1-\alpha$.

5/. On a

θ	0	$1/2$	1
$\mathbb{V}_\theta(X_1)$	$\frac{1}{18}$ ↘	$\frac{1}{24}$	↗ $\frac{1}{18}$

(fonction polynomiale du 2nd degré)

Bilan : $\boxed{\sup_{\theta \in]0,1[} \mathbb{V}_\theta(X_1) = \frac{1}{18} < \frac{1}{4}}$

. Par Bienaymé-Chebychev, on a : $\forall r > 0,$
$$\mathbb{P}_\theta\left(|\overline{X}_n - \mathbb{E}_\theta(X_1)| \geq \frac{r}{\sqrt{n}}\right) \leq \frac{\mathbb{V}_\theta(\overline{X}_n)}{r^2/n} = \frac{\mathbb{V}_\theta(X_1)}{r^2} \leq \frac{1}{18r^2}.$$
On prend alors $r = \tilde{r}^{BC}(\alpha)$ tq $\frac{1}{18r^2} = \alpha$

ie $\boxed{\tilde{r}^{BC}(\alpha) = \frac{1}{\sqrt{18\alpha}} \left(< r^{BC}(\alpha) = \frac{1}{2\sqrt{\alpha}}\right)}$

De sorte que :
$$\mathbb{P}_\theta\left(|\overline{X}_n - \mathbb{E}_\theta(X_1)| \leq \frac{\tilde{r}^{BC}(\alpha)}{\sqrt{n}}\right) \geq 1-\alpha.$$
Or : $\left\{|\overline{X}_n - \mathbb{E}_\theta(X_1)| \leq \frac{\tilde{r}^{BC}(\alpha)}{\sqrt{n}}\right\} = \left\{\theta \in \left[3\left(\overline{X}_n - \frac{\tilde{r}^{BC}(\alpha)}{\sqrt{n}}\right) - 1, 3\left(\overline{X}_n - \frac{\tilde{r}^{BC}(\alpha)}{\sqrt{n}}\right) - 1\right]\right\}$

et donc
$$\mathbb{P}_\theta\left(\theta \in \left[3\left(\overline{X}_n - \frac{\tilde{r}^{BC}(\alpha)}{\sqrt{n}}\right) - 1, 3\left(\overline{X}_n - \frac{\tilde{r}^{BC}}{\sqrt{n}}\right) - 1\right]\right) \geq 1-\alpha.$$
Donc $\boxed{\left[3\left(\overline{X}_n - \frac{\tilde{r}^{BC}(\alpha)}{\sqrt{n}}\right) - 1, 3\left(\overline{X}_n - \frac{\tilde{r}^{BC}(\alpha)}{\sqrt{n}}\right) - 1\right)}$ est un intervalle de confiance par excès de θ, de niveau $1-\alpha$ et il est plus fin que I_n^{BC} car $\tilde{r}^{BC}(\alpha) < r^{BC}(\alpha)$.

6/ Cf graphe en bout de fichier.

Pour $\alpha < \alpha_0 \simeq 0,025$, l'intervalle donné par Q4/ est plus fin ;
Pour $\alpha > \alpha_0$, le plus fin est donné par Q5/.

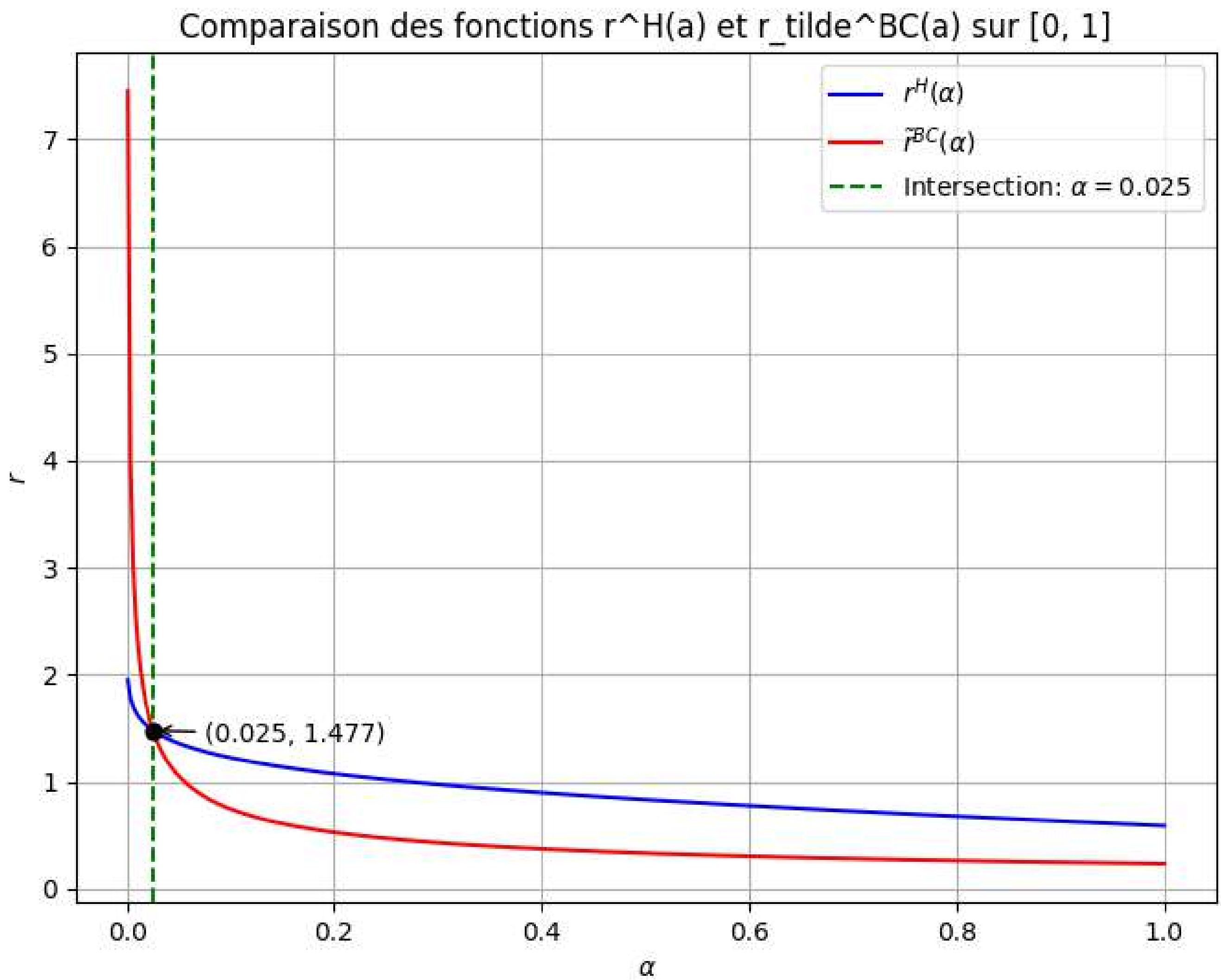
3/ L'estimateur Z_n de θ est consistant et asymptotiquement normal, de variance asymptotique égale à $V(\theta) = \frac{1}{2}(1-\theta+\theta^2)$

On pose $\boxed{\hat{V}_n = V(Z_n)}$. Il s'agit d'un estimateur de $V(\theta)$. C'est un estimateur fortement consistant (par loi forte et continuité de la fonction V)

Ainsi, par théorème 3.3.1, l'intervalle

$$\boxed{I_n = \left[Z_n - \Phi_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{V}_n}{n}} \,,\, Z_n + \Phi_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{V}_n}{n}} \right]} \quad (\alpha = 5\%)$$

est un intervalle de confiance asymptotique de niveau $1-\alpha = 95\%$ pour θ.



Comparaison des fonctions r^H(a) et r_tilde^BC(a) sur [0, 1]
$r^H(\alpha)$
$\tilde{r}^{BC}(\alpha)$
Intersection: $\alpha = 0.025$
(0.025, 1.477)
0
1
2
3
4
5
6
7
r
0.0
0.2
0.4
0.6
0.8
1.0
α