EXAMEN STATISTIQUE - 1MF2E

Lundi 22 mars 2021, 8h00-9h45

Partiel sans document (Une feuille A4 recto-verso autorisée)

Exercice 1 : Estimation (10 points)

On considère n variables aléatoires indépendantes $X_1, ..., X_n$ suivant la même loi normale $\mathcal{N}(0, \sigma^2)$ où $\sigma^2 > 0$ est un paramètre inconnu. L'objectif de cet exercice est d'étudier différentes méthodes d'estimation du paramètre $\theta = \sigma^2$.

1. (2pts) Déterminer la vraisemblance de $X_1, ..., X_n$ et montrer qu'elle admet un unique maximum global pour une valeur de σ^2 que l'on déterminera. En déduire l'estimateur du maximum de vraisemblance du paramètre σ^2 noté $\widehat{\sigma}^2_{\text{MV}}$ en fonction des variables X_i.

2. (2pts) Déterminer la borne de Cramér-Rao pour un estimateur non biaisé du paramètre σ^2 à partir de l'observation des variables $X_1, ..., X_n$. L'estimateur $\widehat{\sigma}^2_{\text{MV}}$ est-il l'estimateur efficace de σ^2 ?
Indication : on rappelle que si $X \sim \mathcal{N}(0, \sigma^2)$, on a $E[X^4] = 3\sigma^4$.

3. Les données $x_1, ..., x_n$ sont soumises à une non-linéarité quadratique et on ne peut qu'observer les données transformées $y_i = x_i^2$. On désire estimer le paramètre θ à partir des données $y_1, ..., y_n$.

   - (2pts) Montrer que la densité de la variable aléatoire $Y_i = X_i^2$ est une loi gamma $\Gamma\left(\frac{1}{2}, \frac{1}{2\sigma^2}\right)$.
   - (1pt) Déterminer l'estimateur du maximum de vraisemblance du paramètre σ^2 noté $\widetilde{\sigma}^2_{\text{MV}}$ construit à partir des variables Y_i.
   - (1pt) Déterminer la borne de Cramér-Rao pour un estimateur non biaisé du paramètre σ^2 à partir de l'observation des variables $Y_1, ..., Y_n$.
   - (1pt) Peut-on espérer obtenir de meilleurs estimateurs de σ^2 en utilisant les données $y_1, ..., y_n$ plutôt que les données $x_1, ..., x_n$?

4. (3pts) On suppose maintenant qu'on dispose d'une information a priori sur le paramètre σ^2 résumée dans la loi a priori $p(\sigma^2)$ définie comme une loi inverse gamma de paramètres $\theta = \sigma_0^2$ et $\nu = 2$ (la moyenne de cette loi a priori est égale à σ_0^2), c'est-à-dire

$$p(\sigma^2) = \frac{1}{\sigma^6} \exp\left(-\frac{\sigma_0^2}{\sigma^2}\right) \mathcal{I}_{\mathbb{R}^+}(\sigma^2)$$

   où $\mathcal{I}_{\mathbb{R}^+}$ est la fonction indicatrice sur $\mathbb{R}^+$ ($\mathcal{I}_{\mathbb{R}^+}(x) = 1$ si $x > 0$ et $\mathcal{I}_{\mathbb{R}^+}(x) = 0$ si $x \leq 0$). Montrer que la loi a posteriori de $\sigma^2 | y_1, ... y_n$ est une loi inverse gamma dont on déterminera les paramètres. En déduire que l'estimateur de la moyenne a posteriori du paramètre σ^2 construit à partir des données $y_1, ..., y_n$ s'écrit

$$\widehat{\sigma}^2_{\text{MMSE}} = \beta(n) \left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) + [1 - \beta(n)]\, \sigma_0^2$$

   où $\beta(n)$ est une fonction de n qu'on déterminera. Expliquer le comportement de $\beta(n)$ lorsque $n \to \infty$ et $n \to 0$.

Exercice 2 : Test Statistique (10 points)

On considère n variables aléatoires $Y_1, ..., Y_n$ indépendantes de densités

$$f(y_i;\theta) = \frac{1}{\sqrt{2\pi\theta y_i}} \exp\left(-\frac{y_i}{2\theta}\right) \mathbb{I}_{\mathbb{R}^+}(y_i)$$

avec $\theta > 0$ et où $\mathbb{I}_{\mathbb{R}^+}$ indique la fonction indicatrice sur $\mathbb{R}^+$. On veut à l'aide d'observations $y_1, ..., y_n$ de ces variables aléatoires tester les deux hypothèses suivantes

$$H_0 : \theta = \theta_0 \text{ contre } H_1 : \theta = \theta_1 \qquad \text{avec } \theta_1 > \theta_0.$$

1. (2pts) À l'aide du théorème de Neyman-Pearson, calculer la statistique T_n du test le plus puissant et indiquer la région critique de ce test (comme d'habitude, on retiendra pour T_n la fonction seule des observations).
2. (2pts) Montrer que $\frac{Y_i}{\theta}$ possède une loi du χ_1^2. En déduire que $\frac{T_n}{\theta}$ suit une loi du χ_n^2.
3. (2pts) En déduire une expression des risques de première et seconde espèce α et β du test en fonction des paramètres θ_0 et θ_1 et de la fonction de répartition d'une loi du χ_n^2 notée F_n.
4. (2pts) Déterminer les courbes COR associées à ce test et expliquer comment les performances de ce test dépendent des valeurs de θ_0 et θ_1.
5. (2pts) On désire à partir d'un ensemble d'observations $(y_1, ..., y_n)$ déterminer s'il est raisonnable de supposer que ces observation sont issues de la densité $f(y_i; \theta)$. Expliquer le principe du test de Kolmogorov pour résoudre ce problème: on précisera avec soin 1) comment calculer la statistique de ce test notée D_n, 2) la région critique du test, 3) l'expression du seuil du test en fonction du risque de première espèce α et de la loi asymptotique de $\sqrt{n}D_n$ et 4) la manière de prendre en compte le fait que θ soit inconnu.

LOIS DE PROBABILITÉ CONTINUES

m : moyenne $\quad \sigma^2$: variance $\quad$ **F. C. :** fonction caractéristique

LOI	Densité de probabilité	m	σ^2	F. C.
Uniforme	$f(x) = \frac{1}{b-a}$ $x \in]a, b[$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{itb} - e^{ita}}{it(b-a)}$
Gamma $\mathcal{G}(\nu, \theta)$	$f(x) = \frac{\theta^\nu}{\Gamma(\nu)} e^{-\theta x} x^{\nu-1}$ $\theta > 0,\ \nu > 0$ $x \geq 0$ avec $\Gamma(n+1) = n!\ \forall n \in \mathbb{N}$	$\frac{\nu}{\theta}$	$\frac{\nu}{\theta^2}$	$\frac{1}{\left(1 - i\frac{t}{\theta}\right)^\nu}$
Inverse gamma $\mathcal{IG}(\nu, \theta)$	$f(x) = \frac{\theta^\nu}{\Gamma(\nu)} e^{-\frac{\theta}{x}} \frac{1}{x^{\nu+1}}$ $\theta > 0,\ \nu > 0$ $x \geq 0$ avec $\Gamma(n+1) = n!\ \forall n \in \mathbb{N}$	$\frac{\theta}{\nu-1}$ si $\nu > 1$	$\frac{\theta^2}{(\nu-1)^2(\nu-2)}$ si $\nu > 2$	(*)
Première loi de Laplace	$f(x) = \frac{1}{2} e^{-\lvert x \rvert}, \quad x \in \mathbb{R}$	0	2	$\frac{1}{1+t^2}$
Normale univariée $\mathcal{N}(m, \sigma^2)$	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}, \quad x \in \mathbb{R}$	m	σ^2	$e^{imt - \frac{\sigma^2 t^2}{2}}$
Normale multivariée $\mathcal{N}_p(\boldsymbol{m}, \boldsymbol{\Sigma})$	$f(x) = K e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{m})^T \Sigma^{-1} (\boldsymbol{x}-\boldsymbol{m})}$ $K = \frac{1}{\sqrt{(2\pi)^p \det(\Sigma)}}$ $x \in \mathbb{R}^p$	$\boldsymbol{m}$	$\boldsymbol{\Sigma}$	$e^{i\boldsymbol{u}^T \boldsymbol{m} - \frac{1}{2}\boldsymbol{u}^T \boldsymbol{\Sigma} \boldsymbol{u}}$
Khi2 χ^2_ν $\Gamma\left(\frac{1}{2}, \frac{\nu}{2}\right)$	$f(x) = k e^{-\frac{x}{2}} x^{\frac{\nu}{2}-1}$ $k = \frac{1}{2^{\frac{\nu}{2}} \Gamma\left(\frac{\nu}{2}\right)}$ $\nu \in \mathbb{N}^*,\ x \geq 0$	ν	2ν	$\frac{1}{(1-2it)^{\frac{\nu}{2}}}$
Cauchy $c_{\lambda,\alpha}$	$f(x) = \frac{1}{\pi\lambda\left(1 + \left(\frac{x-\alpha}{\lambda}\right)^2\right)}$ $\lambda > 0,\ \alpha \in \mathbb{R}$	(-)	(-)	$e^{i\alpha t - \lambda \lvert t \rvert}$
Beta $B(a, b)$	$f(x) = k x^{a-1} (1-x)^{b-1}$ $k = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$ $a > 0,\ b > 0$ $x \in]0, 1[$ avec $\Gamma(n+1) = n!\ \forall n \in \mathbb{N}$	$\frac{a}{a+b}$	$\frac{ab}{(a+b)^2(a+b+1)}$	(*)

LOIS DE PROBABILITÉ DISCRÈTES

m : moyenne σ^2 : variance **F. C. :** fonction caractéristique

$p_k = P[X = k]$ $p_{1,\ldots,m} = P[X_1 = k_1, \ldots, X_m = k_m]$

LOI	**Probabilités**	**m**	σ^2	**F. C.**
Uniforme	$p_k = \frac{1}{n}$ $k \in \{1, \ldots, n\}$	$\frac{n+1}{2}$	$\frac{n^2-1}{12}$	$\dfrac{e^{it}(1-e^{itn})}{n(1-e^{it})}$
Bernoulli	$p_1 = P[X = 1] = p$ $p_0 = P[X = 0] = q$ $p \in [0,1]$ $q = 1-p$	p	pq	$pe^{it} + q$
Binomiale $B(n,p)$	$p_k = C_n^k p^k q^{n-k}$ $p \in [0,1]$ $q = 1-p$ $k \in \{0, 1, \ldots, n\}$	np	npq	$(pe^{it} + q)^n$
Binomiale négative	$p_k = C_{n+k-1}^{n-1} p^n q^k$ $p \in [0,1]$ $q = 1-p$ $k \in \mathbb{N}$	$n\dfrac{q}{p}$	$n\dfrac{q}{p^2}$	$\left(\dfrac{p}{1-qe^{it}}\right)^n$
Multinomiale	$p_{1,\ldots,m} = \frac{n!}{k_1!\ldots k_m!} p_1^{k_1} \ldots p_m^{k_m}$ $p_j \in [0,1]$ $q_j = 1-p_j$ $k_j \in \{0, 1, \ldots, n\}$ $\sum_{j=1}^m k_j = n$ $\sum_{j=1}^m p_j = 1$	np_j	Variance : np_jq_j Covariance : $-np_jp_k$	$\left(\sum_{j=1}^m p_j e^{it}\right)^n$
Poisson $P(\lambda)$	$p_k = e^{-\lambda}\dfrac{\lambda^k}{k!}$ $\lambda > 0$ $k \in \mathbb{N}$	λ	λ	$\exp\left[\lambda\left(e^{it}-1\right)\right]$
Géométrique	$p_k = pq^{k-1}$ $p \in [0,1]$ $q = 1-p$ $k \in \mathbb{N}^*$	$\dfrac{1}{p}$	$\dfrac{q}{p^2}$	$\dfrac{pe^{it}}{1-qe^{it}}$

Exercice 1

1) $L(x_1,\dots,x_n;\sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x_i^2}{2\sigma^2}\right)$

$= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2\right)$

donc $\ln L = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2$

$\frac{\partial \ln L}{\partial \sigma^2} \geq 0 \iff -\frac{n}{2}\frac{1}{\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n} x_i^2 \geq 0$

$\iff \sigma^2 \leq \frac{1}{n}\sum_{i=1}^{n} x_i^2 = S_n$

On a donc le tableau de variations suivant

σ^2	0		S_n		$+\infty$
$\frac{\partial \ln L}{\partial \sigma^2}$		$+$	0	$-$	
$\ln L$		$\nearrow$		$\searrow$	

La vraisemblance admet donc un maximum global unique pour $\sigma^2 = S_n$ d'où

$$\boxed{\hat{\sigma}^2_{MV} = \frac{1}{n}\sum_{i=1}^{n} X_i^2}$$

2) $\frac{\partial^2 \ln L}{\partial \sigma^2} = \frac{n}{2}\frac{1}{\sigma^4} + \frac{1}{2}\left(-\frac{2}{\sigma^6}\right)\sum_{i=1}^{n} x_i^2$

donc $E\left[\frac{\partial^2 \ln L}{\partial \sigma^2}\right] = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6} \times n \underbrace{E(X_i^2)}_{\sigma^2} = -\frac{n}{2\sigma^4}$

d'où $\boxed{BCR(\sigma^2) = \frac{2\sigma^4}{n}}$

$E[\hat{\sigma}^2_{MV}] = \frac{1}{n}\sum_{i=1}^{n} \underbrace{E[X_i^2]}_{\sigma^2} = \sigma^2$ donc $\hat{\sigma}^2_{MV}$ <u>est un estimateur non biaisé de σ^2</u>

$Var(\hat{\sigma}^2_{MV}) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i^2) = \frac{Var(X_1^2)}{n}$

avec $Var(X_1^2) = E(X_1^4) - E^2(X_1^2) = 3\sigma^4 - \sigma^4 = 2\sigma^4$

donc $\boxed{var(\hat{\sigma}^2_{MV}) = \frac{2\sigma^4}{n}}$

$\hat{\sigma}^2_{MV}$ non biaisé

$var(\hat{\sigma}^2_{MV}) = BCR(\sigma^2)$

donc $\boxed{\hat{\sigma}^2_{MV} \text{ est l'estimateur efficace de } \sigma^2}$

3) • Le changement de variables $y_i = x_i^2$ est bijectif par morceaux

$g_1 : \mathbb{R}^+ \to \mathbb{R}^+$
$x_i \mapsto y_i = x_i^2 \Leftrightarrow x_i = \sqrt{y_i}$ $\quad \left|\frac{dx_i}{dy_i}\right| = \frac{1}{2\sqrt{y_i}}$

$\pi_1(y_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{y_i}{2\sigma^2}\right) \frac{1}{2\sqrt{y_i}} \mathbb{1}_{\mathbb{R}^+}(y_i)$

$g_2 : \mathbb{R}^- \to \mathbb{R}^+$
$x_i \mapsto y_i = x_i^2 \Leftrightarrow x_i = -\sqrt{y_i}$ $\quad \left|\frac{dx_i}{dy_i}\right| = \frac{1}{2\sqrt{y_i}}$

$\pi_2(y_i) = \pi_1(y_i)$

La densité de Y_i est donc $\pi(y_i) = \pi_1(y_i) + \pi_2(y_i)$ soit

$$\boxed{\pi(y_i) = \frac{1}{\sqrt{2\pi\sigma^2 y_i}} \exp\left(-\frac{y_i}{2\sigma^2}\right) \mathbb{1}_{\mathbb{R}^+}(y_i)}$$

$\frac{y_i}{\sigma^2} = \frac{x_i^2}{\sigma^2} = \left(\frac{x_i}{\sigma}\right)^2 \sim \boxed{\chi^2_1}$ car $\frac{x_i}{\sigma} \sim \mathcal{N}(0,1)$

• $\pi(y_1, \dots, y_n ; \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \frac{1}{\prod_{i=1}^n \sqrt{y_i}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n y_i\right)$

donc $\ln \pi = K - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^n y_i$ qui a la même forme que $\ln L$ mais en remplaçant x_i^2 par y_i donc d'après ce qui précède

$$\boxed{\tilde{\sigma}^2_{MV} = \frac{1}{n}\sum_{i=1}^n Y_i}$$

• $\frac{\partial \ln \pi}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^n y_i$

$\frac{\partial^2 \ln \pi}{\partial \sigma^4} = \frac{n}{2\sigma^4} + \frac{1}{2}\left(-\frac{2}{\sigma^6}\right)\sum_{i=1}^n y_i$

donc $E\left[\frac{\partial^2 L_n \mathcal{L}}{\partial(\sigma^2)^2}\right] = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6} n E[Y_1]$ (3)

Puisque $Y_1 = X_1^2$, on a $E[Y_1] = \sigma^2$ d'où la borne suivante

$$\boxed{BCR_{Y_1,\dots,Y_n}(\sigma^2) = \frac{2\sigma^4}{n}}$$

- On a donc la même borne de Cramér-Rao, qu'on observe les variables $Y_1,\dots,Y_n$ ou $X_1,\dots,X_n$. Les estimateurs du maximum de vraisemblance basés sur ces deux suites sont les mêmes. En effet

$$\widehat{\sigma}^2_{MV} = \frac{1}{n}\sum_{i=1}^{n} x_i^2 = \frac{1}{n}\sum_{i=1}^{n} y_i = \widehat{\sigma}^2_{MV}$$

Donc on ne peut espérer obtenir de meilleurs estimateurs de σ^2 en utilisant $Y_1,\dots,Y_n$

(4) La loi a posteriori de $\sigma^2 \mid y_1,\dots,y_n$ s'écrit

$$p(\sigma^2 \mid y_1,\dots,y_n) \underset{\substack{\uparrow \\ \text{proportionnel à}}}{\propto} p(y_1,\dots,y_n \mid \sigma^2)\, p(\sigma^2)$$

$$\propto \frac{1}{(\sigma^2)^{\frac{n}{2}+3}} \exp\left[-\frac{1}{\sigma^2}\left(\sigma_0^2 + \frac{1}{2}\sum_{i=1}^{n} y_i\right)\right]$$

C'est une loi inverse gamma de paramètres $\theta = \sigma_0^2 + \frac{1}{2}\sum_{i=1}^{n} y_i$ et $\nu = \frac{n}{2}+2 = \frac{n+4}{2}$ donc

$$\boxed{\sigma^2 \mid y_1 \dots y_n \sim IG\left(\frac{n}{2}+2,\ \sigma_0^2 + \frac{1}{2}\sum_{i=1}^{n} y_i\right)}$$

L'estimateur MMSE de σ^2 est la moyenne de cette loi a posteriori qui d'après les tables s'écrit

$$\widehat{\sigma}^2_{MMSE} = E[\sigma^2 \mid y_1,\dots,y_n] = \frac{\theta}{\nu - 1}$$

soit

$$\widehat{\sigma}^2_{MMSE} = \frac{2}{n+2}\left[\sigma_0^2 + \frac{1}{2}\sum_{i=1}^{n} y_i\right]$$

ou encore

$$\boxed{\widehat{\sigma}^2_{MMSE} = \left(\frac{2}{n+2}\right)\sigma_0^2 + \left(\frac{n}{n+2}\right)\frac{1}{n}\sum_{i=1}^{n} y_i}$$

On a donc $\boxed{\beta(n) = \frac{n}{n+2}}$

$\beta(n) \xrightarrow[n \to +\infty]{} 1$ donc $\hat{\sigma}^2_{MMSE}$ se comporte comme $\tilde{\sigma}^2_{MV}$ quand $n \to +\infty$ (on fait confiance aux données)

$\beta(n) \xrightarrow[n \to 0]{} 0$ donc $\hat{\sigma}^2_{MMSE}$ se comporte comme σ_0^2 qui est la moyenne de la loi a priori (on fait confiance à l'information a priori)

Exercice 2

1. Le test de Neyman-Pearson s'écrit

Rejet de H_0 si $\frac{L(y_1, \ldots, y_n; \theta_1)}{L(y_1, \ldots, y_n; \theta_0)} > \text{seuil}$

soit

$$\frac{\frac{1}{(2\pi\theta_1)^{n/2}} \frac{1}{\pi\sqrt{y_i}} \exp\left(-\frac{1}{2\theta_1}\sum_{i=1}^n y_i\right)}{\frac{1}{(2\pi\theta_0)^{n/2}} \frac{1}{\pi\sqrt{y_i}} \exp\left(-\frac{1}{2\theta_0}\sum_{i=1}^n y_i\right)} > \text{seuil}$$

Un test équivalent est donc (on simplifie et on calcule le logarithme des deux membres de l'inégalité)

Rejet de H_0 si $\underbrace{\left(\frac{1}{\theta_0} - \frac{1}{\theta_1}\right)}_{> 0 \text{ car } \theta_1 > \theta_0} \sum_{i=1}^n y_i > \text{seuil}$

Soit $\boxed{\text{Rejet de } H_0 \text{ si } \sum_{i=1}^n y_i > S_\alpha}$

La statistique de test est donc $T_n = \sum_{i=1}^n y_i$ et la région critique du test est $\{(y_1, \ldots, y_n) / \sum_{i=1}^n y_i > S_\alpha\}$

5

2) on pose $Z_i = \frac{Y_i}{\theta}$ qui est un changement de variable bijectif de $\mathbb{R}^+$ dans $\mathbb{R}^+$ de Jacobien $\left|\frac{dy_i}{dz_i}\right| = \theta$ - On en déduit la densité de Z_i

$$q(z_i) = \frac{1}{\sqrt{2\pi z_i \theta^2}} \exp\left(-\frac{z_i}{2}\right)|\theta| \quad z_i > 0$$

Soit

$$\boxed{q(z_i) = \frac{1}{\sqrt{2\pi z_i}} e^{-z_i/2} \mathbb{1}_{\mathbb{R}^+}(z_i)}$$

On reconnait d'après les tables la densité d'une loi du χ^2_1 donc

$$\boxed{Z_i = \frac{Y_i}{\theta} \sim \chi^2_1}$$

$U_n = \frac{T_n}{\theta} = \sum_{i=1}^{n} \frac{Y_i}{\theta} = \sum_{i=1}^{n} Z_i$ qui admet la fonction caractéristique

$$\phi_{U_n}(t) = E\left[e^{itU_n}\right] \underset{\text{les variables } Z_i \text{ sont indépendants}}{=} \prod_{k=1}^{n} \underbrace{E\left[e^{itZ_k}\right]}_{\frac{1}{(1-2it)^{1/2}}} = \frac{1}{(1-2it)^{n/2}}$$

C'est la fonction caractéristique d'une loi du χ^2_n donc

$$\boxed{U_n \sim \chi^2_n}$$

3) $\alpha = P[\text{Rejeter } H_0 \mid H_0 \text{ vraie}] = P\left[\sum_{i=1}^{n} Y_i > s_\alpha \mid \theta_0 = \theta_0\right]$

$$= P\left[U_n = \frac{T_n}{\theta_0} > \frac{s_\alpha}{\theta_0} \mid \theta = \theta_0\right]$$

donc $\boxed{1 - F_n\left(\frac{s_\alpha}{\theta_0}\right) = \alpha}$ d'où $\boxed{s_\alpha = \theta_0 F_n^{-1}(1-\alpha)}$

$\beta = P[\text{Rejeter } H_1 \mid H_1 \text{ vraie}] = P\left[\sum_{i=1}^{n} Y_i \leq s_\alpha \mid \theta = \theta_1\right]$

$$= P\left[\frac{1}{\theta_1}\sum_{i=1}^{n} Y_i \leq \frac{s_\alpha}{\theta_1} \mid \theta = \theta_1\right]$$

$$\boxed{\beta = F_n\left(\frac{s_\alpha}{\theta_1}\right)}$$

4) Les courbes COR représentent $\Pi = 1-\beta$ en fonction de α.
Elles sont définies par

$$\Pi = 1 - F_n\left(\frac{s_\alpha}{\theta_1}\right) = \boxed{1 - F_n\left[\frac{\theta_0}{\theta_1} F_n^{-1}(1-\alpha)\right]}$$

Quand $\frac{\theta_1}{\theta_0} \nearrow$ (θ_1 s'éloigne de θ_0), $\frac{\theta_0}{\theta_1} \searrow$ donc $F_n\left[\frac{\theta_0}{\theta_1} F_n^{-1}(1-\alpha)\right] \searrow$

donc $\Pi \nearrow$

Le test est donc d'autant plus puissant que $\frac{\theta_1}{\theta_0}$ est grand, ce qui est naturel

5) La statistique du test de Kolmogorov se calcule comme suit

$$\boxed{D_n = \sup_{x \in \mathbb{R}} |\hat{F}(x) - F_0(x)| = \max_{i \in \{1,\dots,n\}} \{E_i^+, E_i^-\}}$$

avec $E_i^+ = \left|\frac{i}{n} - F_0(x_i^*)\right|$ et $E_i^- = \left|\frac{i-1}{n} - F_0(x_i^*)\right|$

où F_0 est la fonction de répartition de loi de densité $f(y;\theta)$

La région critique du test est $\boxed{\{(y_1,\dots,y_n) / D_n > s_\alpha\}}$

On a d'après le cours

$$P[\sqrt{n}\, D_n < y] \xrightarrow[n \to +\infty]{} K(y) = \sum_{k=-\infty}^{+\infty} (-1)^k \exp(-2k^2 y)$$

et $\boxed{s_\alpha \simeq \frac{1}{\sqrt{n}} K^{-1}(1-\alpha)}$

Quand θ est inconnu, on peut l'estimer à l'aide de la méthode du maximum de vraisemblance et tester si $(y_1,\dots,y_n)$ suit une loi de densité $f(y;\hat{\theta}_{MV})$ où $\hat{\theta}_{MV}$ est l'estimateur du maximum de vraisemblance de θ.