



Rapport de projet Cassiopée

Sujet : Étude des statistiques mal spécifiées et
application à un processus autorégressif

LO Vincent
NECHNECH Idris
TALEB Merouane

Enseignants responsable :

FORTUNATI Stefano
PETETIN Yohan

Table des matières

1	Introduction	3
1.1	Statistique paramétrique	3
1.2	Statistique mal spécifiée	3
2	Calculs pour une t-distribution complexe	4
2.1	Estimation de la matrice de diffusion	4
2.2	Bornes de Cramer-Rao (CRB)	4
2.2.1	F_{Σ}	5
2.2.2	F_c	5
2.2.3	$F_{\lambda, \eta}$	5
2.2.4	T_2	6
2.2.5	U	6
2.2.6	CRB	6
2.3	Bornes de Cramer-Rao mal spécifiée (MCRB)	6
2.3.1	$A_{\bar{\Sigma}}$	7
2.3.2	$A_{\bar{\sigma}^2}$	7
2.3.3	A_c	7
2.3.4	$B_{\bar{\Sigma}}$	7
2.3.5	$B_{\bar{\sigma}^2}$	7
2.3.6	B_c	7
2.3.7	U	8
2.3.8	$MCRB$	8
2.4	Simulations	8
3	Calculs pour une AR(1)	9
3.1	Cas Gaussien	9
3.1.1	Cramér-Rao Bound (CRB)	9
3.1.2	Estimateur ML	10
3.2	Cas Student	10
3.2.1	Estimation de la CRB Student	11
3.2.2	Estimation de la MCRB Student	11
3.2.3	Estimateur ML par méthode de Newton	11
3.3	Simulations	12
3.3.1	Résultats des calculs précédents	12
3.3.2	Comparaisons en utilisant la MSE	13
4	Conclusion	15

1 Introduction

Dans de nombreux problèmes d'estimation, les données observées sont modélisées à l'aide de lois paramétriques dont la structure dépend d'un vecteur de paramètres inconnus. L'objectif est alors de proposer des estimateurs efficaces de ces paramètres, et d'évaluer leur précision à l'aide d'outils théoriques tels que la borne de Cramér-Rao.

Cependant, dans des situations réelles, les hypothèses faites sur la distribution des données peuvent être erronées. Il devient alors nécessaire d'élargir le cadre classique pour tenir compte de ces erreurs de spécification. Ce rapport explore ces deux approches complémentaires.

Nous introduisons tout d'abord les fondements de la statistique paramétrique classique et de la statistique mal spécifiée. Nous présentons ensuite des cas d'étude concrets: estimation de paramètres sous loi de Student multivariée, calculs des bornes dans le cadre paramétrique et mal spécifié; et analyse d'un modèle autorégressif AR(1) dans les contextes gaussien et non-gaussien.

1.1 Statistique paramétrique

La statistique paramétrique repose sur l'hypothèse selon laquelle les données observées suivent une loi de probabilité appartenant à une famille paramétrique bien définie. Dans ce cadre, les paramètres du modèle sont estimés via des méthodes telles que le maximum de vraisemblance ou les moindres carrés. Ces estimateurs sont alors étudiés en termes de biais, variance, et efficacité, et leurs performances peuvent être comparées à des bornes théoriques telles que la borne de Cramér-Rao (CRB), qui donne une limite inférieure à la variance d'un estimateur sans biais.

Formellement, soit une densité paramétrique $p(x; \theta)$ avec $\theta \in \Theta \subset \mathbb{R}^d$, et une suite i.i.d. de données $(X_i)_{i=1}^n$. L'estimateur du maximum de vraisemblance est donné par :

$$\hat{\theta}_{\text{ML}} = \arg \max_{\theta \in \Theta} \sum_{i=1}^n \log p(X_i; \theta)$$

La borne de Cramér-Rao s'écrit alors :

$$\text{Var}(\hat{\theta}) \succeq \text{CRB} = 1/n \cdot I(\theta)^{-1} \quad \text{avec} \quad I(\theta) = \mathbb{E} [\nabla_{\theta} \log p(X; \theta) \nabla_{\theta} \log p(X; \theta)^{\top}]$$

1.2 Statistique mal spécifiée

Dans un cadre idéal, les données suivent exactement la loi de probabilité supposée par le modèle. En pratique, cette hypothèse est souvent trop forte. Par exemple, on peut supposer que les bruits sont gaussiens alors qu'en réalité ils suivent une t-distribution. Cette erreur de spécification, bien que parfois inévitable, impacte la performance des estimateurs et la validité des bornes classiques comme la CRB. La statistique mal spécifiée s'intéresse justement à ces situations: elle permet d'adapter les outils d'estimation et d'évaluation, en introduisant des bornes plus réalistes, telles que la borne de Cramér-Rao mal spécifiée (MCRB), qui tient compte de l'écart entre le modèle supposé et la distribution réelle des données.

Formellement, si $p^*(x)$ désigne la vraie loi génératrice des données et $p(x; \theta)$ un modèle erroné, alors le vecteur score n'est en général plus centré :

$$\mathbb{E} p^* [\nabla \theta \log p(X; \theta)] \neq 0$$

On définit :

$$A(\theta) = \mathbb{E} p^* [\nabla^2 \log p(X; \theta)], \quad B(\theta) = \mathbb{E} p^* [\nabla \theta \log p(X; \theta) \nabla_{\theta} \log p(X; \theta)^{\top}]$$

La borne de Cramér-Rao mal spécifiée (MCRB) est alors donnée par :

$$\text{Var}(\hat{\theta}) \succeq \text{MCRB} = A(\theta)^{-1} B(\theta) A(\theta)^{-1}$$

On a toujours:

$$\text{MCRB} \succeq \text{CRB}$$

2 Calculs pour une t-distribution complexe

Considérons la suite de v.a. $(X_m)_{m \in [1, M]}$ à valeur dans \mathbb{C}^N et sont i.i.d. de loi t-distribuée ie. $\forall m, X_m \sim p(x_m; \Sigma, \lambda, \eta)$ avec

$$p(x_m; \Sigma, \lambda, \eta) = \frac{1}{\pi^N |\Sigma|} \frac{\Gamma(N + \lambda)}{\Gamma(\lambda)} \left(\frac{\lambda}{\eta}\right)^\lambda \left(\frac{\lambda}{\eta} + x_m^H \Sigma^{-1} x_m\right)^{-N-\lambda} \quad (1)$$

où $\lambda > 2$ est le paramètre de shape, η le paramètre de scale et Σ est la matrice de diffusion définie avec la matrice de covariance \mathbf{M} des X_m comme $\Sigma = N / \text{Tr}(\mathbf{M})$. On impose par ailleurs que $\text{Tr}(\Sigma) = N$. On suppose par ailleurs que (i) tous les X_m sont de moyenne nulle ($\gamma = \mathbb{E}[X_m] = 0$) et que (ii) la matrice de diffusion Σ est réelle et de rang plein.

Dans la suite on notera

$$\begin{aligned} l((x_m)_m; \Sigma, \lambda, \eta) &= \log(p(x_m; \Sigma, \lambda, \eta)) = -N \log(\pi) - \log(|\Sigma|) + \log(\Gamma(N + \lambda)) - \log(\Gamma(\lambda)) \\ &\quad + \lambda \log\left(\frac{\lambda}{\eta}\right) - (N + \lambda) \log\left(\frac{\lambda}{\eta} + x_m^H \Sigma^{-1} x_m\right) \end{aligned} \quad (2)$$

2.1 Estimation de la matrice de diffusion

Nous cherchons maintenant à estimer la matrice de diffusion. Calculons l'estimateur de vraisemblance de cette matrice en écrivant la log-vraisemblance :

$$\begin{aligned} L(\Sigma, \lambda, \eta, (x_m)_m) &= \log\left(\prod_{m=1}^M p(x_m; \Sigma, \lambda, \eta)\right) \\ &= -MN \log(\pi) + M \log(|\Sigma^{-1}|) + M \log\left(\frac{\Gamma(N + \lambda)}{\Gamma(\lambda)} \left(\frac{\lambda}{\eta}\right)^\lambda\right) \\ &\quad - (N + \lambda) \sum_{m=1}^M \log\left(\frac{\lambda}{\eta} + x_m^H \Sigma^{-1} x_m\right) \end{aligned} \quad (3)$$

$$\frac{\partial L}{\partial \Sigma^{-1}} = M \Sigma - (N + \lambda) \sum_{m=1}^M \frac{x_m x_m^H}{\frac{\lambda}{\eta} + x_m^H \Sigma^{-1} x_m} = 0 \quad (4)$$

On obtient finalement

$$\hat{\Sigma}_{ML} = \frac{N + \lambda}{M} \sum_{m=1}^M \frac{x_m x_m^H}{\frac{\lambda}{\eta} + x_m^H \Sigma^{-1} x_m} \quad s.t. \quad \text{Tr}(\Sigma) = N \quad (5)$$

2.2 Bornes de Cramer-Rao (CRB)

Nous cherchons à estimer le paramètre $\theta = [\text{vecs}(\Sigma)^T \quad \lambda \quad \eta]^T$ pour les données t-distribuées. La matrice d'information de Fisher s'écrit de la manière suivante :

$$F_\theta = T_2^T \begin{bmatrix} F_\Sigma & F_c \\ F_c^T & F_{\lambda, \eta} \end{bmatrix} T_2 = \mathbb{E}[(\nabla_\theta \log(p(x; \theta)))^T (\nabla_\theta \log(p(x; \theta)))]$$

On a donc que, en notant $l = \log(p(x; \theta))$:

$$F_\Sigma = \mathbb{E}\left[-\frac{\partial^2 l}{\partial \Sigma^2}\right], \quad F_c = \begin{bmatrix} \mathbb{E}\left[-\frac{\partial^2 l}{\partial \lambda \partial \Sigma}\right] & \mathbb{E}\left[-\frac{\partial^2 l}{\partial \eta \partial \Sigma}\right] \end{bmatrix} \quad \text{et} \quad F_{\lambda, \eta} = \begin{bmatrix} \mathbb{E}\left[-\frac{\partial^2 l}{\partial \lambda^2}\right] & \mathbb{E}\left[-\frac{\partial^2 l}{\partial \lambda \partial \eta}\right] \\ \mathbb{E}\left[-\frac{\partial^2 l}{\partial \eta \partial \lambda}\right] & \mathbb{E}\left[-\frac{\partial^2 l}{\partial \eta^2}\right] \end{bmatrix}$$

2.2.1 F_Σ

Commençons par le bloc en Σ . On va utiliser les formules suivantes :

$$\frac{\partial(\log|\Sigma|)}{\partial\Sigma} = \Sigma^{-1} \quad (6)$$

$$\frac{\partial(x_m^H \Sigma^{-1} x_m)}{\partial\Sigma} = -\Sigma^{-1} x_m x_m^H \Sigma^{-1} \quad (7)$$

$$\frac{\partial\Sigma^{-1}}{\partial\Sigma} = -\Sigma^{-1} \otimes \Sigma^{-1} \quad (8)$$

donc

$$\frac{\partial l}{\partial\Sigma} = -\Sigma^{-1} - (N + \lambda) \frac{\Sigma^{-1} x_m x_m^H \Sigma^{-1}}{\frac{\lambda}{\eta} + x_m^H \Sigma^{-1} x_m} \quad (9)$$

$$\begin{aligned} \frac{\partial^2 l}{\partial\Sigma^2} &= \Sigma^{-1} \otimes \Sigma^{-1} - \frac{N + \lambda}{(\frac{\lambda}{\eta} + x_m^H \Sigma^{-1} x_m)^2} (\Sigma^{-1} x_m x_m^H \Sigma^{-1}) (\Sigma^{-1} x_m x_m^H \Sigma^{-1})^T \\ &\quad - \frac{N + \lambda}{\frac{\lambda}{\eta} + x_m^H \Sigma^{-1} x_m} \cdot \frac{\partial(\Sigma^{-1} x_m x_m^H \Sigma^{-1})}{\partial\Sigma} \end{aligned} \quad (10)$$

Après vectorisation puis calcul et passage à l'espérance, on trouve que :

$$F_\Sigma = -\frac{1}{N + \lambda + 1} \text{vec}(\Sigma^{-1}) \text{vec}(\Sigma^{-1})^T + \frac{N + \lambda}{N + \lambda + 1} (\Sigma^{-1} \otimes \Sigma^{-1}) \quad (11)$$

2.2.2 F_c

En reprenant l'équation () :

$$\frac{\partial^2 l}{\partial\lambda \partial \text{vec}(\Sigma)} = \frac{\partial}{\partial\lambda} \left(\frac{N + \lambda}{(\frac{\lambda}{\eta} + x_m^H \Sigma^{-1} x_m)^2} \right) \text{vec}(\Sigma^{-1} x_m x_m^H \Sigma^{-1}) \quad (12)$$

$$= \frac{1}{\frac{\lambda}{\eta} + x_m^H \Sigma^{-1} x_m} \cdot \left(1 - \frac{N + \lambda}{\frac{\lambda}{\eta} + x_m^H \Sigma^{-1} x_m} \right) \cdot \left(\frac{1}{\eta} \right) \text{vec}(\Sigma^{-1} x_m x_m^H \Sigma^{-1}) \quad (13)$$

$$\frac{\partial^2 l}{\partial\eta \partial \text{vec}(\Sigma)} = \frac{N + \lambda}{(\frac{\lambda}{\eta} + x_m^H \Sigma^{-1} x_m)^2} \cdot \left(\frac{\lambda}{\eta^2} \right) \text{vec}(\Sigma^{-1} x_m x_m^H \Sigma^{-1}) \quad (14)$$

Après passage à l'espérance, on obtient

$$F_c = -\text{vec}(\Sigma^{-1}) \left[\frac{1}{(N + \lambda)(N + \lambda + 1)} \quad \frac{\lambda}{\eta(N + \lambda + 1)} \right]$$

2.2.3 $F_{\lambda, \eta}$

En utilisant la fonction digamma $\psi : z \mapsto \psi(z) = \frac{\Gamma'(z)}{\Gamma(z)}$, sa dérivée trigamma $\frac{d}{dz}\psi : z \mapsto \sum_{k=0}^{+\infty} \frac{1}{(z+k)^2}$ et en notant $Q := x_m^H \Sigma^{-1} x_m$, on a

$$\frac{\partial l}{\partial\lambda} = \psi(N + \lambda) - \psi(\lambda) + \log\left(\frac{\lambda}{\eta}\right) + 1 - \log\left(\frac{\lambda}{\eta} + Q\right) - \frac{N + \lambda}{\frac{\lambda}{\eta} + Q} \cdot \frac{1}{\eta} \quad (15)$$

$$\frac{\partial^2 l}{\partial\lambda^2} = \sum_{k=0}^{N-1} \frac{1}{(\lambda + k)^2} + \frac{1}{\lambda} - \frac{1}{\frac{\lambda}{\eta} + Q} \cdot \frac{1}{\eta} - \left(\frac{1}{\frac{\lambda}{\eta} + Q} \cdot \frac{1}{\eta} - \frac{N + \lambda}{(\frac{\lambda}{\eta} + Q)^2} \cdot \frac{1}{\eta^2} \right) \quad (16)$$

$$= \sum_{k=0}^{N-1} \frac{1}{(\lambda + k)^2} + \frac{1}{\lambda} - \frac{2}{\frac{\lambda}{\eta} + Q} \cdot \frac{1}{\eta} + \frac{N + \lambda}{(\frac{\lambda}{\eta} + Q)^2} \cdot \frac{1}{\eta^2} \quad (17)$$

Après simplification et passage à l'espérance, on obtient :

$$[F_{\lambda, \eta}]_{1,1} = \sum_{k=0}^{N-1} \frac{1}{(\lambda + k)^2} + \frac{(\lambda + 1)(N + \lambda)}{\lambda(N + \lambda + 1)} - \frac{2N}{\lambda(N + \lambda)} - 1$$

De même pour η :

$$\frac{\partial l}{\partial \eta} = -\frac{\lambda}{\eta} + \frac{N + \lambda}{\frac{\lambda}{\eta} + Q} \cdot \frac{\lambda}{\eta^2} \quad (18)$$

$$\frac{\partial^2 l}{\partial \eta^2} = \frac{\lambda}{\eta^2} - \frac{2(N + \lambda)}{(\frac{\lambda}{\eta} + Q)^2} \cdot \frac{\lambda}{\eta^3} + \frac{N + \lambda}{\frac{\lambda}{\eta} + Q} \cdot \frac{2\lambda}{\eta^3} \quad (19)$$

On obtient

$$[F_{\lambda, \eta}]_{1,1} = \frac{N\lambda}{\eta^2(N + \lambda + 1)}$$

Et pour finir le terme des dérivées croisées

$$\frac{\partial^2 l}{\partial \eta \partial \lambda} = \frac{\partial^2 l}{\partial \lambda \partial \eta} = -\frac{1}{\eta} + \frac{1}{\frac{\lambda}{\eta} + Q} - \frac{N + \lambda}{(\frac{\lambda}{\eta} + Q)^2} \cdot \frac{-1}{\eta^2} \quad (20)$$

Le passage à l'espérance donne :

$$[F_{\lambda, \eta}]_{1,2} = [F_{\lambda, \eta}]_{2,1} = \frac{N}{\eta(N + 1)(N + \lambda + 1)}$$

2.2.4 T_2

Posons $\ell = \frac{N(N+1)}{2}$.

On a :

$$T_2 = \begin{pmatrix} D_N & 0 \\ 0 & I_2 \end{pmatrix} \quad (21)$$

avec D_N l'unique matrice de taille $N^2 \times \ell$ vérifiant $D_N \text{vecs}(A) = \text{vec}(A)$ pour toute matrice A symétrique.

2.2.5 U

Soit $U \in \mathbb{R}^{(l+2) \times (l+1)}$ une matrice dont les colonnes forment une base orthonormée du noyau de $\nabla f(\theta)$, c'est-à-dire que $\nabla f(\theta)U = 0$ et $U^\top U = I$.

La matrice U peut être calculée numériquement, par exemple à l'aide de la décomposition en valeurs singulières (SVD), en extrayant les $l + 1$ vecteurs propres orthonormés associés à la valeur propre nulle de $\nabla f(\theta)$, avec

$$\nabla f(\theta) = (\mathbf{1}_I^\top \quad 0 \quad 0)$$

où $\mathbf{1}_I$ est un vecteur colonne de dimension l défini par :

$$(\mathbf{1}_I)_i = \begin{cases} 1 & \text{si } i \in I \\ 0 & \text{sinon} \end{cases}$$

2.2.6 CRB

Finalement, on a :

$$CRB(\theta) = U(U^\top F_\theta U)^{-1} U^\top \quad (22)$$

2.3 Bornes de Cramer-Rao mal spécifiée (MCRB)

On estime ici le même θ , dans le cadre mal spécifié où on suppose les données gaussiennes. Ici :

$$f_X(x_m; \theta) f_X(x_m; \Sigma, \sigma^2) = \frac{1}{(\pi \sigma^2)^N |\Sigma|} \exp \left(-\frac{x_m^H \Sigma^{-1} x_m}{\sigma^2} \right) \quad (23)$$

Les matrices d'information sont:

$$A_{\theta_0} = T_1^\top \begin{pmatrix} A_{\bar{\Sigma}} & A_c \\ A_c^\top & A_{\bar{\sigma}^2} \end{pmatrix} T_1 \quad (24)$$

Et :

$$A_{\theta_0} = T_1^\top \begin{pmatrix} A_{\bar{\Sigma}} & A_c \\ A_c^\top & A_{\bar{\sigma}^2} \end{pmatrix} T_1 \quad (25)$$

$$B_{\theta_0} = T_1^\top \begin{pmatrix} B_{\bar{\Sigma}} & B_c \\ B_c^\top & B_{\bar{\sigma}^2} \end{pmatrix} T_1 \quad (26)$$

2.3.1 $A_{\bar{\Sigma}}$

On a :

$$A_{\bar{\Sigma}} = -\Sigma^{-1} \otimes \Sigma^{-1} \quad (46)$$

2.3.2 $A_{\bar{\sigma}^2}$

$$A_{\bar{\sigma}^2} = \mathbb{E}_p \left[\frac{\partial^2}{\partial^2 \sigma^2} \ln f_X(x_m; \theta_0) \right] = -\frac{N}{\sigma^4} \quad (27)$$

2.3.3 A_c

On a :

$$[A_c]_{i,1} = \mathbb{E}_p \left[\frac{\partial^2}{\partial \sigma^2 \partial \text{vec}(\Sigma)_i} \ln f_X(x_m; \theta_0) \right] = -\frac{1}{\sigma^2} \text{tr}(\Sigma^{-1} A_i) = -\frac{1}{\sigma^2} \text{vec}(\Sigma^{-1})^\top \text{vec}(A_i)$$

où $A_i = \frac{\partial \Sigma}{\partial \theta_i}$ est une matrice symétrique 0-1.

On en déduit :

$$A_c = -\frac{1}{\sigma^2} \text{vec}(\Sigma^{-1})^\top \quad (28)$$

2.3.4 $B_{\bar{\Sigma}}$

$$B_{\bar{\Sigma}} = \frac{1}{\lambda - 2} \text{vec}(\Sigma^{-1}) \text{vec}(\Sigma^{-1})^\top + \frac{\lambda - 1}{\lambda - 2} \Sigma^{-1} \otimes \Sigma^{-1} \quad (29)$$

2.3.5 $B_{\bar{\sigma}^2}$

$$B_{\bar{\sigma}^2} = \mathbb{E}_p \left[\left(\frac{\partial}{\partial \sigma^2} \ln f_X(x_m; \theta_0) \right)^2 \right] = \frac{N(N + \lambda - 1)}{(\lambda - 2)\sigma^4} \quad (30)$$

2.3.6 B_c

$$(B_c)_{i,1} = \mathbb{E} \left[\frac{\partial \ln f_X(x_m; \theta_0)}{\partial \sigma^2} \cdot \frac{\partial \ln f_X(x_m; \theta_0)}{\partial (\text{vec} \Sigma)_i} \right] = \frac{N + (\lambda - 1)}{\sigma^2(\lambda - 2)} \text{tr}(\Sigma^{-1} A_i)$$

Donc :

$$B_c = \frac{N + (\lambda - 1)}{\sigma^2(\lambda - 2)} \text{vec}(\Sigma^{-1}) \quad (31)$$

2.3.7 U

Soit $U \in \mathbb{R}^{(l+2) \times (l+1)}$ une matrice dont les colonnes forment une base orthonormée du noyau de $\nabla f(\theta)$, c'est-à-dire que

$$\nabla f(\theta)U = 0 \quad \text{et} \quad U^\top U = I.$$

La matrice U peut être calculée numériquement, par exemple à l'aide de la décomposition en valeurs singulières (SVD), en extrayant les $l+1$ vecteurs propres orthonormés associés à la valeur propre nulle de $\nabla f(\theta)$, avec

$$\nabla f(\theta) = \begin{pmatrix} 1_I^\top & 0 \end{pmatrix}$$

2.3.8 $MCRB$

Finalement, on a:

$$MCRB(\theta_0) = M^{-1}U \left(U^\top A_{\theta_0} U \right)^{-1} U^\top B_{\theta_0} U \left(U^\top A_{\theta_0} U \right)^{-1} U^\top, \quad \theta_0 \in \Theta; \quad (32)$$

2.4 Simulations

On constate qu'on a effectivement $MCRB > CRB$ et que:

- Pour des petites valeurs de λ (données fortement non gaussiennes), les pertes d'estimation dues à une inadéquation du modèle sont très grandes.
- Quand $\lambda \rightarrow +\infty$, c'est-à-dire quand les données deviennent gaussiennes, la MCRB et la CRB tendent vers la même limite, c'est normal car pour des données gaussiennes la MCRB et la CRB sont égales.

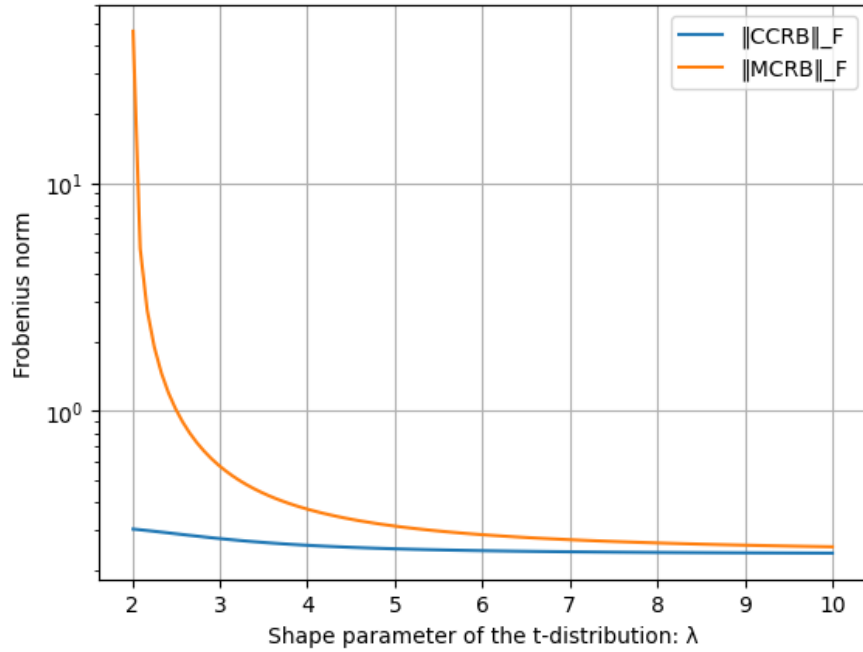


Figure 1: Comparaison de la MCRB et de la CRB

3 Calculs pour une AR(1)

Soit $N \in \mathbb{N}$, on définit les vecteurs aléatoires $(X_i)_{i \in [0, N]}$, tels que $\forall i, X_i \in \mathbb{R}$ et

$$\forall n \in \{0, \dots, N-1\}, \quad X_{n+1} = \theta X_n + \varepsilon_n \quad (33)$$

avec $\theta \in \mathbb{R}$ et ε_n un bruit centré tel que les $(\varepsilon_i)_{i \in [0, N]}$ soient i.i.d.

On suppose de plus que $X_0 \sim \mathcal{N}(0, 1)$. Dans la suite on ne considère uniquement $\theta \in]0, 1[$ pour éviter les divergences de notre processus.

Par ailleurs, récursivement, on peut écrire chacun de nos X_n uniquement en fonction des bruits :

$$\forall n, \quad X_n = \theta^n X_0 + \sum_{j=0}^{n-1} \theta^j \varepsilon_{n-1-j} \quad (34)$$

3.1 Cas Gaussien

On suppose que $\forall i, \varepsilon_n \sim \mathcal{N}(0, \sigma^2)$. Pour tout $i \in [0, N-1]$, on a $X_{i+1} | X_i \sim \mathcal{N}(\theta X_i, \sigma^2)$

Par propriété markovienne :

$$f(X_0, X_1, \dots, X_n | \theta) = f(X_0) \cdot \prod_{i=0}^{N-1} f(X_{i+1} | X_i, \theta) \quad (35)$$

On a :

$$X_0 \sim \mathcal{N}(0, 1) \quad \Rightarrow \quad \log f(X_0) = -\frac{1}{2} \log(2\pi) - \frac{1}{2} X_0^2 \quad (36)$$

Et comme $\forall i, X_{i+1} | X_i \sim \mathcal{N}(\theta X_i, \sigma^2)$, on a que :

$$\log f(X_{i+1} | X_i, \theta) = -\frac{1}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} (X_{i+1} - \theta X_i)^2 \quad (37)$$

En sommant :

$$\log L(X_0, \dots, X_n | \theta) = \log f(X_0) + \sum_{i=0}^{N-1} \log f(X_{i+1} | X_i, \theta) \quad (38)$$

Donc :

$$\log L(\theta) = -\frac{1}{2} \log(2\pi) - \frac{1}{2} X_0^2 - \frac{N}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=0}^{N-1} (X_{i+1} - \theta X_i)^2 \quad (39)$$

$$\frac{d}{d\theta} \log L(\theta) = \frac{1}{\sigma^2} \sum_{i=0}^{N-1} X_i (X_{i+1} - \theta X_i) \quad (40)$$

$$\frac{d^2}{d\theta^2} \log L(\theta) = -\frac{1}{\sigma^2} \sum_{i=0}^{N-1} X_i^2 \quad (41)$$

3.1.1 Cramér-Rao Bound (CRB)

$$\mathcal{I}(\theta) = -\mathbb{E} \left[\frac{d^2}{d\theta^2} \log L(\theta) \right] = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} \mathbb{E}[X_n^2] \quad (42)$$

$$\text{CRB}(\theta) = \frac{1}{N \mathcal{I}(\theta)} = \frac{\sigma^2}{N \sum_{n=0}^{N-1} \mathbb{E}[X_n^2]} \quad (43)$$

Avec:

$$\sum_{n=0}^{N-1} \mathbb{E}[X_n^2] = \sum_{n=0}^{N-1} \left(\theta^{2n} + \sigma^2 \frac{1 - \theta^{2n}}{1 - \theta^2} \right) \quad (44)$$

$$= \sum_{n=0}^{N-1} \theta^{2n} + \frac{\sigma^2}{1 - \theta^2} \sum_{n=0}^{N-1} (1 - \theta^{2n}) \quad (45)$$

$$= \frac{1 - \theta^{2N}}{1 - \theta^2} + \frac{\sigma^2}{1 - \theta^2} \left(N - \frac{1 - \theta^{2N}}{1 - \theta^2} \right) \quad (46)$$

3.1.2 Estimateur ML

$$\left. \frac{d}{d\theta} \log L(\theta) \right|_{\theta=\theta_{\text{ML}}} = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} X_n (X_{n+1} - \theta_{\text{ML}} X_n) = 0 \quad (47)$$

On développe :

$$\sum_{n=0}^{N-1} X_n X_{n+1} - \theta_{\text{ML}} \sum_{n=0}^{N-1} X_n^2 = 0 \quad (48)$$

Et on isole θ_{ML} :

$$\theta_{\text{ML}} = \frac{\sum_{n=0}^{N-1} X_n X_{n+1}}{\sum_{n=0}^{N-1} X_n^2} \quad (49)$$

3.2 Cas Student

On suppose que les bruits $(\epsilon_i)_i$ suivent une loi de Student centrée avec ν degrés de liberté et un paramètre d'échelle σ , ie. $\epsilon_i \sim t_\nu(0, \sigma)$. Dans ce cas on a que $X_{i+1} | X_i \sim t_\nu(\theta X_i, \sigma)$

La densité conditionnelle vaut alors:

$$\log f(X_{i+1} | X_i, \theta) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi} \Gamma(\frac{\nu}{2}) \sigma} \left(1 + \frac{(X_{i+1} - \theta X_i)^2}{\nu\sigma^2} \right)^{-\frac{\nu+1}{2}} \quad (50)$$

La vraisemblance vaut donc, par un calcul similaire à celui effectué dans le cas Gaussien :

$$\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{X_0^2}{2}\right) \cdot \prod_{i=0}^{N-1} \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi} \Gamma(\frac{\nu}{2}) \sigma} \left(1 + \frac{(X_{i+1} - \theta X_i)^2}{\nu\sigma^2} \right)^{-\frac{\nu+1}{2}} \quad (51)$$

Et :

$$\log L(\theta) = -\frac{1}{2} \log(2\pi) - \frac{1}{2} X_0^2 + N \log \left(\frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi} \Gamma(\frac{\nu}{2}) \sigma} \right) - \frac{\nu+1}{2} \sum_{i=0}^{N-1} \log \left(1 + \frac{(X_{i+1} - \theta X_i)^2}{\nu\sigma^2} \right) \quad (52)$$

On calcule maintenant les gradients :

$$\frac{d}{d\theta} \log L(\theta) = (\nu+1) \sum_{i=0}^{N-1} \frac{X_i (X_{i+1} - \theta X_i)}{\nu\sigma^2 + (X_{i+1} - \theta X_i)^2} \quad (53)$$

$$\frac{d^2}{d\theta^2} \log L(\theta) = (\nu+1) \sum_{i=0}^{N-1} X_i^2 \cdot \frac{(X_{i+1} - \theta X_i)^2 - \nu\sigma^2}{(\nu\sigma^2 + (X_{i+1} - \theta X_i)^2)^2} \quad (54)$$

3.2.1 Estimation de la CRB Student

$$\mathcal{I}(\theta) = -\mathbb{E} \left[\frac{d^2}{d\theta^2} \log L(\theta) \right] = (\nu + 1) \cdot \mathbb{E} \left[\left(\sum_{i=0}^{N-1} X_i^2 \cdot \frac{\epsilon_i^2 - \nu\sigma^2}{[\nu\sigma^2 + \epsilon_i^2]^2} \right) \right] \quad (55)$$

Il n'est pas possible d'obtenir une forme explicite de la matrice d'information de Fisher, donc on évalue cette quantité via une méthode de Monte-Carlo.

3.2.2 Estimation de la MCRB Student

Pour le cas mal spécifié, on estime à θ et σ fixés les quantités:

$$A(\theta) = \mathbb{E} \left[\left(\frac{d^2}{d\theta^2} \log L(\theta) \right) \right] = -\frac{1}{\sigma^2} \sum_{i=0}^{N-1} \mathbb{E}[X_i^2] \quad (56)$$

$$B(\theta) = \mathbb{E} \left[\left(\frac{d}{d\theta} \log L(\theta) \right)^2 \right] = \frac{1}{\sigma^4} \mathbb{E} \left[\left(\sum_{i=0}^{N-1} X_i (X_{i+1} - \theta X_i) \right)^2 \right] = \frac{1}{\sigma^4} \mathbb{E} \left[\left(\sum_{i=0}^{N-1} X_i \epsilon_i \right)^2 \right] \quad (57)$$

On a:

$$X_n = \theta^n X_0 + \sum_{j=0}^{n-1} \theta^j \epsilon_{n-1-j} \quad (58)$$

Comme les ϵ_k sont indépendants entre eux et comme $X_0 \sim \mathcal{N}(0, 1)$ est indépendant des ϵ_k , on a :

$$\mathbb{E}[X_n^2] = \theta^{2n} \mathbb{E}[X_0^2] + \sum_{j=0}^{n-1} \theta^{2j} \mathbb{E}[\epsilon_{n-1-j}^2] \quad (59)$$

Or on a que $\mathbb{E}[X_0^2] = 1$ et comme $\epsilon \sim t_\nu(0, \sigma)$ alors $\text{Var}(\epsilon) = \frac{\nu\sigma^2}{\nu-2}$

Donc $\mathbb{E}[X_n^2] = \theta^{2n} + \frac{\nu\sigma^2}{\nu-2} \cdot \frac{1-\theta^{2n}}{1-\theta^2}$ si $|\theta| < 1$. Ainsi, on obtient :

$$A(\theta) = -\frac{1}{\sigma^2} \sum_{n=0}^{N-1} \mathbb{E}[X_n^2] \quad (60)$$

$$= -\frac{1}{\sigma^2} \sum_{n=0}^{N-1} \left(\theta^{2n} + \frac{\nu\sigma^2}{\nu-2} \cdot \frac{1-\theta^{2n}}{1-\theta^2} \right) \quad (61)$$

$$= -\frac{1}{\sigma^2} \left[\frac{1-\theta^{2N}}{1-\theta^2} + \frac{\sigma^2}{\nu-2} \cdot \frac{1}{1-\theta^2} \left(N - \frac{1-\theta^{2N}}{1-\theta^2} \right) \right] \quad (62)$$

On estime $B(\theta)$ par méthode de Monte-Carlo puis on calcule $\text{MCRB}(\theta) = \frac{B(\theta)}{A(\theta)^2}$

3.2.3 Estimateur ML par méthode de Newton

$$\left. \frac{d}{d\theta} \log L(\theta) \right|_{\theta=\theta_{\text{ML}}} = (\nu + 1) \sum_{i=0}^{N-1} \frac{X_i (X_{i+1} - \theta_{\text{ML}} X_i)}{\nu\sigma^2 + (X_{i+1} - \theta_{\text{ML}} X_i)^2} = 0 \quad (63)$$

On ne peut pas calculer cette valeur numériquement et on va donc estimer θ_{ML} par méthode de Newton:

$$\theta^{(k+1)} = \theta^{(k)} - \frac{\frac{d}{d\theta} \log L(\theta^{(k)})}{\frac{d^2}{d\theta^2} \log L(\theta^{(k)})} \quad \text{avec} \quad \theta^{(0)} = \frac{\sum_{i=0}^{N-1} X_i X_{i+1}}{\sum_{i=0}^{N-1} X_i^2} \quad (64)$$

la valeur qu'on aurait obtenu via l'estimateur du maximum de vraisemblance pour des données Gaussiennes.

3.3 Simulations

3.3.1 Résultats des calculs précédents

La Figure 2 compare la borne de Cramér-Rao (CRB) réelle calculée en supposant correctement un bruit de Student à la borne de Cramér-Rao modifiée (MCRB), obtenue lorsque le bruit est supposé Gaussien, alors qu'il est en réalité de Student.

Les deux courbes convergent vers une valeur proche lorsque $\nu \rightarrow +\infty$, ce qui est cohérent : une loi de Student avec beaucoup de degrés de liberté tend vers une loi normale. Donc dans ce cas, supposer un bruit gaussien est une bonne approximation.

Pour des valeurs de ν faible, la MCRB est beaucoup plus grande que la CRB. Cela montre que l'approximation gaussienne est mauvais. De plus, la MCRB montre une grande instabilité pour des valeurs de ν proches de 2, avec des pics visibles. Cela reflète la sensibilité du modèle Gaussien à la violation de ses hypothèses lorsque le bruit n'est pas réellement normal.

Par ailleurs, on utilise une échelle logarithmique pour mettre en valeur les différences d'ordre de grandeur entre la CRB et la MCRB.

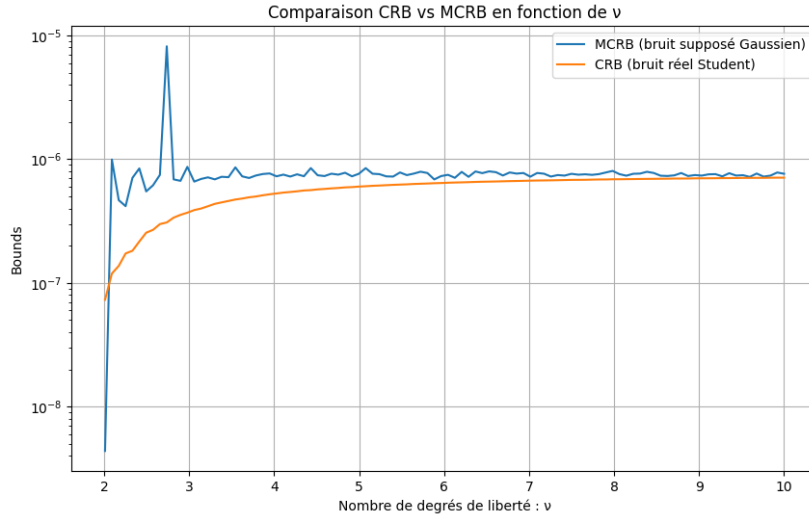


Figure 2: CRB et MCRB en fonction du degré de liberté ($N = 100$, $\theta = 1$, $\sigma = 1$)

On étudie maintenant le phénomène en variant le nombre de donnée pour le processus autorégressif. (Figure 3) Plus N est grand, plus les deux bornes se translate vers le bas. Cela reflète l'intuition que plus de données permettent une estimation plus précise. Cependant, l'écart entre MCRB et CRB reste significatif, en particulier pour les petites valeurs de ν , et ce même quand N est grand.

On observe peu de changements au niveau de l'allure des différents courbes pour chaque N . Toutefois on remarquera que la convergence entre la CRB et la MCRB se fait "plus tôt" par rapport au degré de liberté ν de la loi de Student. Ce phénomène s'explique par la loi des grands nombres : les moyennes empiriques utilisées dans les expressions de la MCRB (basées sur les résidus supposés gaussiens) convergent vers leurs espérances sous la vraie loi (Student). Quand N est grand, les valeurs extrêmes sont compensées par le volume de données, et l'estimateur devient plus robuste à la mauvaise spécification.

Pour des raisons de complexités élevées, on est allé uniquement jusqu'à $N = 5000$ et également réduire le nombre d'échantillon pour calculer les bornes par méthode de Monte-Carlo (1000 au lieu de 10000). Un essai allant jusqu'à $N = 10000$ et avec 10000 échantillons pour le Monte-Carlo nous a conduit à une exécution de plus de 9 heures.

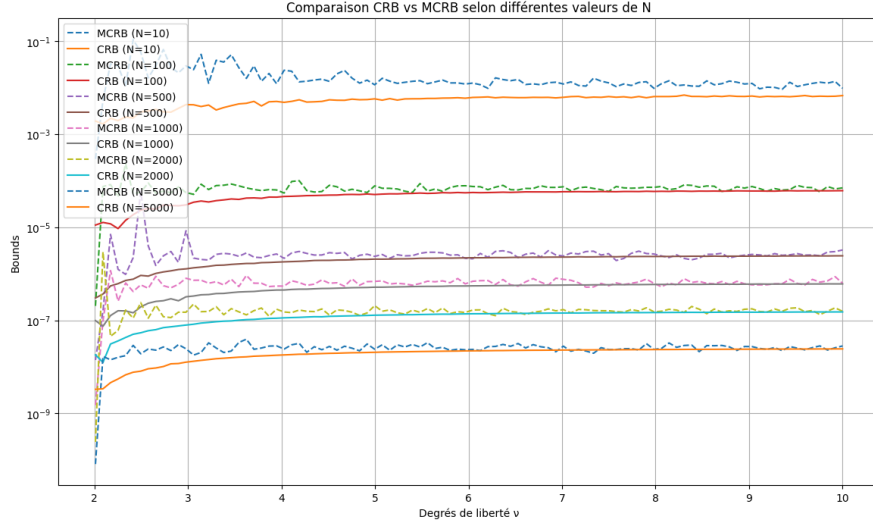


Figure 3: CRB et MCRB en fonction du degré de liberté et en faisant varier N de 10 à 5000

On observe que θ est mieux estimé via la méthode de Newton dans le cas mal spécifié plutôt que dans le cas bien spécifié.

Une estimation de θ dans le cas bien spécifié et mal spécifié nous donne ce résultat là, qui est peu représentatif car on a parfois un meilleur estimateur ML que Newton et inversement. (Figure 4)

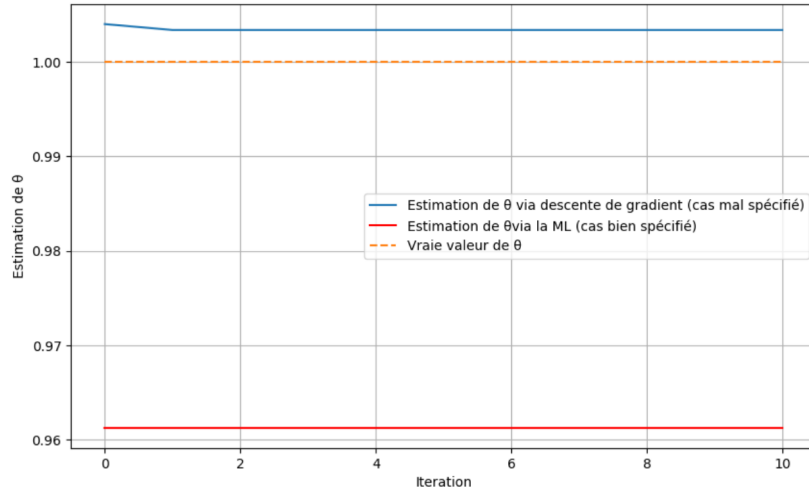


Figure 4: Estimation de θ ($N = 100$, $\sigma = 1$, $\nu = 3$)

3.3.2 Comparaisons en utilisant la MSE

Afin de comparer les différents estimateurs et évaluer les performances de la mal spécification et de vérifier quel estimateur est le meilleur, nous allons étudier les erreurs quadratiques moyennes (MSE) pour l'estimateur ML dans le cas Gaussien et l'estimateur issu de la méthode de Newton dans le cas Student. Pour cela on utilise la formule suivante, après avoir générer plusieurs estimations de θ qu'on notera, pour ML et Newton respectivement, $(\theta_{ML}^{(i)})_i$ et $(\theta_{NT}^{(i)})_i$:

$$MSE(\theta) = \frac{1}{N} \sum_{i=1}^N (\theta^{(i)} - \theta)^2 \quad \text{avec} \quad \forall i, \quad \theta^{(i)} = \theta_{ML}^{(i)} \text{ ou } \theta_{NT}^{(i)} \quad (65)$$

On compare donc la performance des deux estimateurs via leur erreur quadratique moyenne en fonction du nombre d'observations N qu'on fait varier entre 100 et 10000. (Figure 5)

Pour de petits échantillons ($N < 3000$), l'estimateur Newton présente une MSE significativement plus faible que le MLE classique. Lorsque N croît, les deux erreurs moyennes convergent vers 0, mais l'estimateur Newton reste légèrement plus performant en termes de MSE sur l'ensemble des tailles d'échantillons.

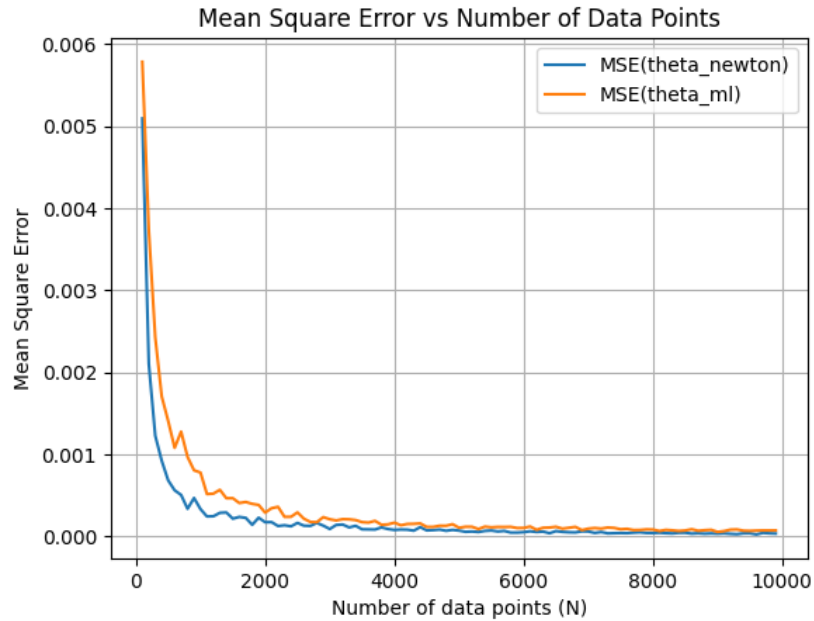


Figure 5: MSE pour les estimateurs ML et Newton pour N allant de 100 à 10000

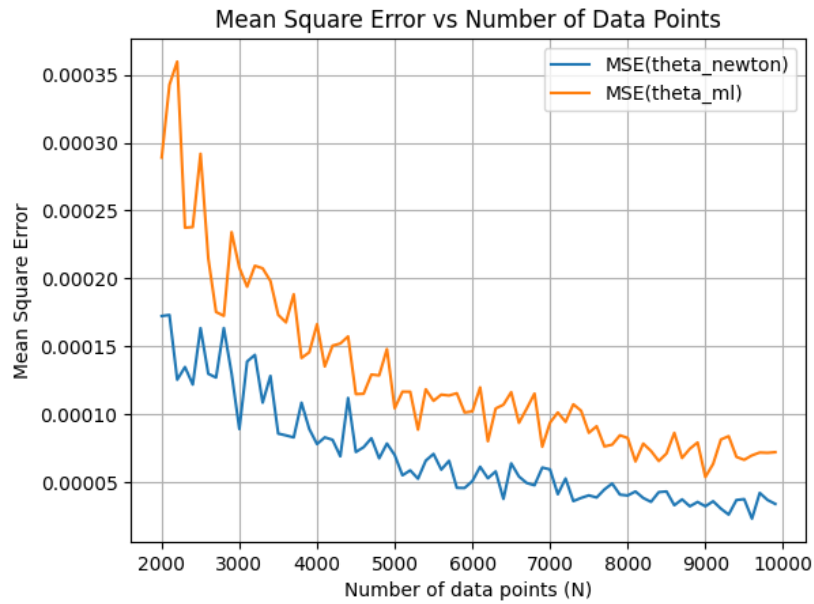


Figure 6: MSE pour les estimateurs ML et Newton pour N allant de 1000 à 10000

En complément pour observer davantage les meilleurs estimations par la méthode de Newton et la mal spécification (et donc meilleure modélisation des bruits), nous avons représenté la courbe des MSE

aux niveaux où les différences entre $\text{MSE}(\text{ML})$ et $\text{MSE}(\text{Newton})$ sont faibles (Figure 6). De même en traçant la différence en valeur absolu (Figure 7). Cela confirme donc le constat de la Figure 4.

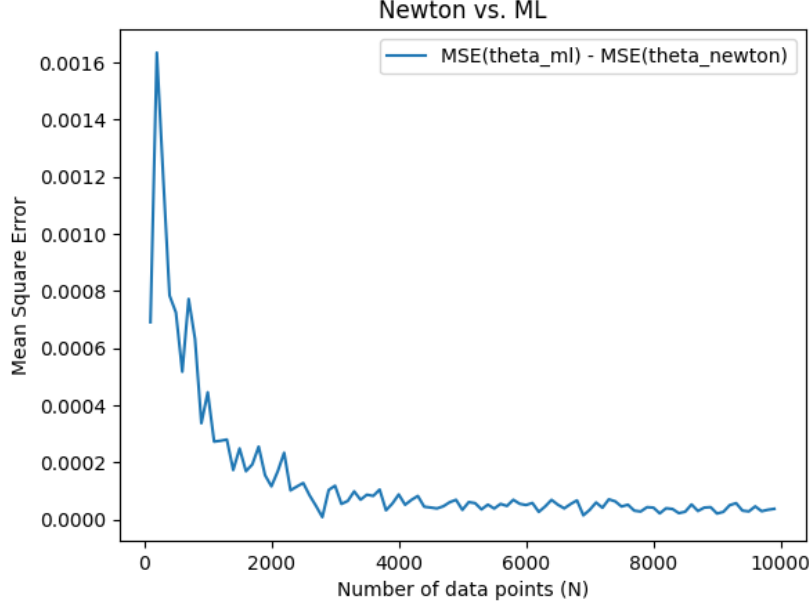


Figure 7: Différence $\text{MSE}(\text{ML}) - \text{MSE}(\text{Newton})$

4 Conclusion

Ce travail met en évidence l'importance de la statistique mal spécifiée dans les situations où les hypothèses du modèle, notamment la normalité des données, ne sont pas respectées. À travers l'étude de lois de Student et de processus $\text{AR}(1)$, nous avons montré que les outils classiques, comme la borne de Cramér-Rao (CRB), peuvent sous-estimer la variance réelle des estimateurs en cas de mauvaise spécification.

L'introduction de la borne de Cramér-Rao mal spécifiée (MCRB) permet d'obtenir une évaluation plus réaliste de la précision des estimateurs. Les simulations ont clairement montré que la MCRB devient significativement plus élevée que la CRB lorsque les données s'écartent de la gaussienneté, et que cet écart persiste même avec de grandes tailles d'échantillons.

De plus, l'utilisation d'estimateurs adaptés, comme la méthode de Newton dans le cas de bruits de Student, permet d'améliorer sensiblement les performances, notamment en termes d'erreur quadratique moyenne.

En somme, la statistique mal spécifiée offre un cadre rigoureux et plus robuste face aux imperfections des modèles, et s'impose comme un outil essentiel pour une estimation fiable en pratique.