Cassiopée Project Report

Subject : Study of Misspecified Statistics and Application to an Autoregressive Process

NECHNECH Idris
LO Vincent
TALEB Merouane

Supervising Professors :

FORTUNATI Stefano
PETETIN Yohan

Table of Contents

1 Introduction

In many estimation problems, observed data are modeled using parametric distributions whose structure depends on an unknown parameter vector. The objective is then to propose efficient estimators for these parameters, and to evaluate their precision using theoretical tools such as the Cramér-Rao bound. However, in real-world situations, assumptions made about data distribution can be erroneous. It then becomes necessary to broaden the classical framework to account for these specification errors. This report explores these two complementary approaches. We first introduce the foundations of classical parametric statistics and misspecified statistics. We then present concrete case studies: parameter estimation under a multivariate t-distribution, calculations of bounds in both the parametric and misspecified frameworks; and analysis of an AR(1) autoregressive model in Gaussian and non-Gaussian contexts.

1.1 Parametric Statistics

Parametric statistics relies on the assumption that observed data follow a probability distribution belonging to a well-defined parametric family. In this framework, model parameters are estimated using methods such as maximum likelihood or least squares. These estimators are then studied in terms of bias, variance, and efficiency, and their performance can be compared to theoretical bounds such as the Cramér-Rao bound (CRB), which provides a lower limit to the variance of an unbiased estimator.
Formally, let $p(x;\theta)$ be a parametric density with $\theta \in \Theta \subset \mathbb{R}^d$, and an i.i.d. sequence of data $(X_i)_{i=1}^n$. The maximum likelihood estimator is given by:

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

The Cramér-Rao bound is then written as:

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

1.2 Misspecified Statistics

In an ideal framework, data exactly follow the probability distribution assumed by the model. In practice, this assumption is often too strong. For example, one might assume that the noise is Gaussian when it actually follows a t-distribution. This specification error, though sometimes unavoidable, impacts the performance of estimators and the validity of classical bounds like the CRB. Misspecified statistics precisely address these situations: it allows adapting estimation and evaluation tools by introducing more realistic bounds, such as the Misspecified Cramér-Rao Bound (MCRB), which accounts for the discrepancy between the assumed model and the true data distribution. Formally, if $p^*(x)$ denotes the true data generating distribution and $p(x;\theta)$ an erroneous model, then the score vector is generally no longer centered:

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

We define :

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

The Misspecified Cramér-Rao Bound (MCRB) is then given by:

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

We always have:

$$MCRB \succeq CRB$$

2 Calculation for a complex t-distribution

Let us consider the sequence of random variables $(X_m)_{m\in[1,M]}$ taking values in $\mathbb{C}^N$ which are i.i.d. following a t-distribution, i.e. $\forall m,\ X_m \sim p(x_m;\Sigma,\lambda,\eta)$ with

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

where $\lambda > 2$ is the shape parameter, η is the scale parameter Σ is the scatter matrix defined with the covariance matrix $\mathbf{M}$ of X_m as $\Sigma = N/Tr(\mathbf{M})$. We also impose that $Tr(\Sigma) = N$.We further assume that (i) all X_m are zero-mean ($\gamma = \mathbb{E}[X_m] = 0$) and that (ii) the scatter matrix Σ is real and full rank. In the following we will denote

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

2.1 Estimation of the Scatter Matrix

We now seek to estimate the scatter matrix. Let's calculate the maximum likelihood estimator of this matrix by writing the log-likelihood:

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

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

We finally obtain:

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

2.2 Cramér-Rao Bounds (CRB)

We seek to estimate the parameter $\theta = \begin{bmatrix} vecs(\Sigma)^T & \lambda & \eta \end{bmatrix}^T$ for t-distributed data. The Fisher information matrix is written as follows:

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

Thus, by denoting $l = log(p(x;\theta))$:

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

2.2.1 F_Σ

Let's start with theΣ block. We will use the following formulas :

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

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

$$\frac{\partial\Sigma^{-1}}{\partial\Sigma} = -\Sigma^{-1}\otimes\Sigma^{-1} \tag{8}$$

thus

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

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

After vectorization, calculation, and taking the expectation, we find that:

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

2.2.2 F_c

Referring to equation (9) :

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

$$= \frac{1}{\frac{\lambda}{\eta} + x_m^H\Sigma^{-1}x_m}\ .(1 - \frac{N+\lambda}{\frac{\lambda}{\eta} + x_m^H\Sigma^{-1}x_m})\ .(\frac{1}{\eta})\ vec(\Sigma^{-1}x_m x_m^H\Sigma^{-1}) \tag{13}$$

$$\frac{\partial^2 l}{\partial\eta\partial vec(\Sigma)} = \frac{N+\lambda}{(\frac{\lambda}{\eta} + x_m^H\Sigma^{-1}x_m)^2}\ .(\frac{\lambda}{\eta^2})\ vec(\Sigma^{-1}x_m x_m^H\Sigma^{-1}) \tag{14}$$

After taking the expectation, we obtain

$$F_c = -vec(\Sigma^{-1})\begin{bmatrix}\frac{1}{(N+\lambda)(N+\lambda+1)} & \frac{\lambda}{\eta(N+\lambda+1)}\end{bmatrix}$$

2.2.3 $F_{\lambda,\eta}$

Using the digamma function $\psi : z \mapsto \psi(z) = \frac{\Gamma'(z)}{\Gamma(z)}$, its trigamma derivative$\frac{d}{dz}\psi : z \mapsto \sum_{k=0}^{+\infty}\frac{1}{(z+k)^2}$ and denoting $Q := x_m^H\Sigma^{-1}x_m$, we have

$$\frac{\partial l}{\partial\lambda} = \psi(N+\lambda) - \psi(\lambda) + log(\frac{\lambda}{\eta}) + 1 - log(\frac{\lambda}{\eta} + Q) - \frac{N+\lambda}{\frac{\lambda}{\eta}+Q}\ .\frac{1}{\eta} \tag{15}$$

$$\frac{\partial^2 l}{\partial\lambda^2} = \sum_{k=0}^{N-1}\frac{1}{(\lambda+k)^2} + \frac{1}{\lambda} - \frac{1}{\frac{\lambda}{\eta}+Q}\ .\frac{1}{\eta} - (\frac{1}{\frac{\lambda}{\eta}+Q}\ .\frac{1}{\eta} - \frac{N+\lambda}{(\frac{\lambda}{\eta}+Q)^2}\ .\frac{1}{\eta^2}) \tag{16}$$

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

After simplification and taking the expectation, we obtain:

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

Similarly for η :

$$\frac{\partial l}{\partial \eta} = -\frac{\lambda}{\eta} + \frac{N+\lambda}{\frac{\lambda}{\eta}+Q}.\frac{\lambda}{\eta^2} \tag{18}$$

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

We obtain

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

And to finish with the cross-derivative term

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

Taking the expectation gives:

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

2.2.4 T_2

Let us set $\ell = \frac{N(N+1)}{2}$.
We have:

$$T_2 = \begin{pmatrix} D_N & 0 \\ 0 & I_2 \end{pmatrix} \tag{21}$$

with D_N the unique matrix of size $N^2 \times \ell$ satisfying $D_N \operatorname{vecs}(A) = \operatorname{vec}(A)$, for any symmetric matrix A

2.2.5 U

Let $U \in \mathbb{R}^{(l+2)\times(l+1)}$ a matrix whose columns form an orthonormal basis for the null space of $\nabla f(\theta)$, that is, $\nabla f(\theta) U = 0$ and $U^\top U = I$.
The matrix U pcan be calculated numerically, for example using singular value decomposition (SVD), by extracting the $l+1$ orthonormal eigenvectors associated with the zero eigenvalue of $\nabla f(\theta)$, with

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

where $\mathbf{1}_I$is an l-dimensional column vector defined by:

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

2.2.6 CRB

Finally, we have:

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

2.3 Misspecified Cramér-Rao Bound (MCRB)

Here, we estimate the same θ, in the misspecified framework where we assume the data are Gaussian. Here :

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

The information matrices are:

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

And :

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

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

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

We have:

$$A_{\Sigma} = -\Sigma^{-1} \otimes \Sigma^{-1} \tag{46}$$

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

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

2.3.3 A_c

We have:

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

where $A_i = \frac{\partial \Sigma}{\partial \theta_i}$ is a 0-1 symmetric matrix.
We deduce:

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

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

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

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

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

2.3.6 B_c

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

Thus :

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

2.3.7 U

Let $U \in \mathbb{R}^{(l+2)\times(l+1)}$ be a matrix whose columns form an orthonormal basis for the kernel of $\nabla f(\theta)$,that is,

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

The matrix U can be calculated numerically, for example using singular value decomposition (SVD), by extracting the $l+1$ orthonormal eigenvectors associated with the zero eigenvalue of $\nabla f(\theta)$,with

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

2.3.8 $MCRB$

Finally, we have:

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

2.4 Simulations

We observe that MCRB > CRB and that:
- For small values of λ (highly non-Gaussian data), the estimation losses due to model mismatch are very large.
- When $\lambda \to +\infty$, meaning the data become Gaussian, the MCRB and the CRB tend toward the same limit. This is expected because for Gaussian data, the MCRB and CRB are equal.

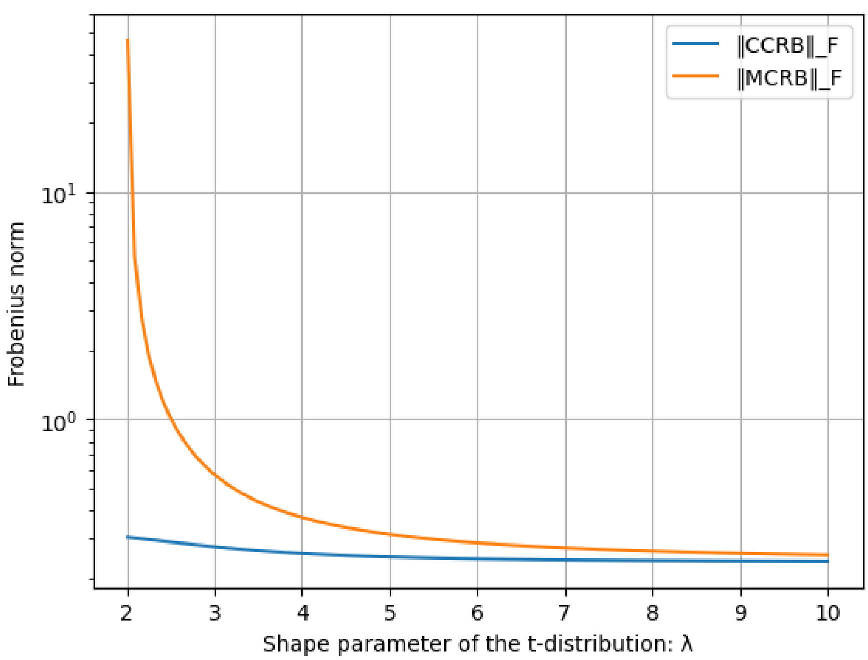

Figure 1: Comparison of MCRB and CRB

3 Calculation for an AR(1)

Let $N \in \mathbb{N}$, we define the random vectors $(X_i)_{i\in[0,N]}$, such that $\forall i, X_i \in \mathbb{R}$ and

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

with $\theta \in \mathbb{R}$ et ε_n a zero-mean noise such that the $(\epsilon_i)_{i\in[0,N]}$ are i.i.d.
We further assume that $X_0 \sim \mathcal{N}(0,1)$. In the following, we consider only $\theta \in]0,1[$ to avoid divergences in our process.

Moreover, we can recursively write each of our X_n only in terms of noise and the first term:

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

3.1 Gaussian Case

We assume that $\forall i, \epsilon_n \sim \mathcal{N}(0, \sigma^2)$. For all $i \in [0, N-1]$, we have $X_{i+1}|X_i \sim \mathcal{N}(\theta X_i, \sigma^2)$
By the Markov property:

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

We have:

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

And since $\forall i,\ X_{i+1} \mid X_i \sim \mathcal{N}(\theta X_i, \sigma^2)$, we have :

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

By summing:

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

Thus:

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

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

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

3.1.1 Cramér-Rao Bound (CRB)

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

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

With:

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

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

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

3.1.2 Maximum Likelihood Estimator (MLE)

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

Let's expand:

$$\sum_{n=0}^{N-1} X_n X_{n+1} - \theta_{\mathrm{ML}} \sum_{n=0}^{N-1} X_n^2 = 0 \tag{48}$$

And we isolate θ_{ML} :

$$\theta_{\mathrm{ML}} = \frac{\sum_{n=0}^{N-1} X_n X_{n+1}}{\sum_{n=0}^{N-1} X_n^2} \tag{49}$$

3.2 t-distribution case

We assume that the noises $(\epsilon_i)_i$ follow a centered t-distribution with ν degrees of freedom and a scale parameter σ, i.e. $\epsilon_i \sim t_\nu(0, \sigma)$.In this case, we have $X_{i+1} \mid X_i \sim t_\nu(\theta X_i, \sigma)$

The conditional density then becomes:

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

The likelihood function is therefore, by a calculation similar to the Gaussian case:

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

And :

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

We now calculate the gradients:

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

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

3.2.1 t-distribution CRB Estimation

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

It is not possible to obtain an explicit form for the Fisher information matrix, so this quantity is evaluated using a Monte-Carlo method.

3.2.2 t-distribution MCRB Estimation

For the misspecified case, with θ and σ fixed, we estimate the following quantities:

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

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

We have:

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

Since the ϵ_k are independent of each other, and as $X_0 \sim \mathcal{N}(0,1)$ is independent of the ε_k, we have:

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

Now we have $\mathbb{E}[X_0^2] = 1$ and since $\epsilon \sim t_\nu(0,\sigma)$ thus $\mathrm{Var}(\epsilon) = \frac{\nu\sigma^2}{\nu-2}$

Thus $\mathbb{E}[X_n^2] = \theta^{2n} + \frac{\nu\sigma^2}{\nu-2}\cdot\frac{1-\theta^{2n}}{1-\theta^2}$ si $|\theta|<1$. Thus, we obtain:

$$\begin{aligned} A(\theta) &= -\frac{1}{\sigma^2}\sum_{n=0}^{N-1}\mathbb{E}[X_n^2] && (60)\\ &= -\frac{1}{\sigma^2}\sum_{n=0}^{N-1}\left(\theta^{2n} + \frac{\nu\sigma^2}{\nu-2}\cdot\frac{1-\theta^{2n}}{1-\theta^2}\right) && (61)\\ &= -\frac{1}{\sigma^2}\left[\frac{1-\theta^{2N}}{1-\theta^2} + \frac{\sigma^2}{\nu-2}\cdot\frac{1}{1-\theta^2}\left(N - \frac{1-\theta^{2N}}{1-\theta^2}\right)\right. && (62) \end{aligned}$$

We estimate $B(\theta)$ using the Monte Carlo method, then we calculate $\mathrm{MCRB}(\theta) = \frac{B(\theta)}{A(\theta)^2}$

3.2.3 MLE using Newton's Method

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

We cannot calculate this value numerically, so we will estimate θ_{ML} using Newton's method:

$$\theta^{(k+1)} = \theta^{(k)} - \frac{\frac{d}{d\theta}\log L(\theta^{(k)})}{\frac{d^2}{d\theta^2}\log L(\theta^{(k)})} \quad \textit{with} \quad \theta^{(0)} = \frac{\sum_{i=0}^{N-1} X_iX_{i+1}}{\sum_{i=0}^{N-1} X_i^2} \tag{64}$$

the value we would have obtained via the maximum likelihood estimator for Gaussian data.

3.3 Simulations

3.3.1 Results from Previous Calculations

Figure 2 compares the true Cramér-Rao Bound (CRB), calculated by correctly assuming t-distributed noise, with the Misspecified Cramér-Rao Bound (MCRB), obtained when the noise is incorrectly assumed to be Gaussian, even though it is truly t-distributed.

The two curves converge to a similar value when $\nu \to +\infty$, which is consistent: a t-distribution with many degrees of freedom approaches a normal distribution. Therefore, in this case, assuming Gaussian noise is a good approximation.

For low values of ν,the MCRB is much larger than the CRB. This demonstrates that the Gaussian approximation is poor. Furthermore, the MCRB exhibits significant instability for values of ν close to 2, with visible peaks. This reflects the sensitivity of the Gaussian model to the violation of its assumptions when the noise is not truly normal.

Furthermore, a logarithmic scale is used to highlight the order-of-magnitude differences between the CRB and the MCRB.

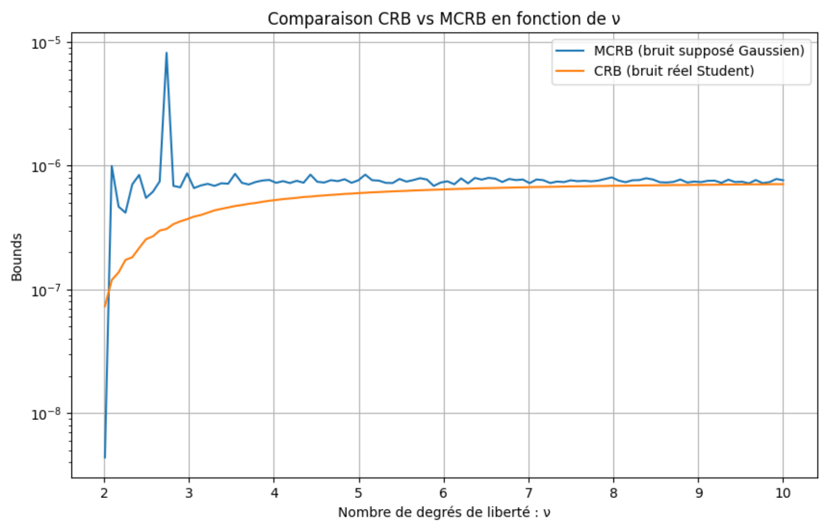

Figure 2: CRB and MCRB as a Function of Degrees of Freedom (N = 100, $\theta = 1$, $\sigma = 1$)

We're now studying this phenomenon by varying the number of data points (N) for the autoregressive process (Figure 3). The larger N gets, the more both bounds shift downwards. This reflects the intuition that more data allows for more precise estimation. However, the gap between MCRB and CRB remains significant, especially for small values of ν, even when N is large.

We observe few changes in the general shape of the curves for different values of N. However, it's notable that the convergence between the CRB and the MCRB occurs "earlier" with respect to the degrees of freedom ν of the t-distribution. This phenomenon is explained by the Law of Large Numbers: the empirical averages used in the MCRB expressions (based on residuals assumed to be Gaussian) converge towards their expected values under the true distribution (t-distribution). When N is large, extreme values are compensated for by the volume of data, and the estimator becomes more robust to misspecification.

Due to high complexity, we only went as far as $N = 5000$ And also reduced the number of samples for calculating the bounds using the Monte-Carlo method (1000 instead of 10,000). A trial run up to $N = 10000$ and with 10,000 Monte Carlo samples, led to an execution time of over 9 hours.

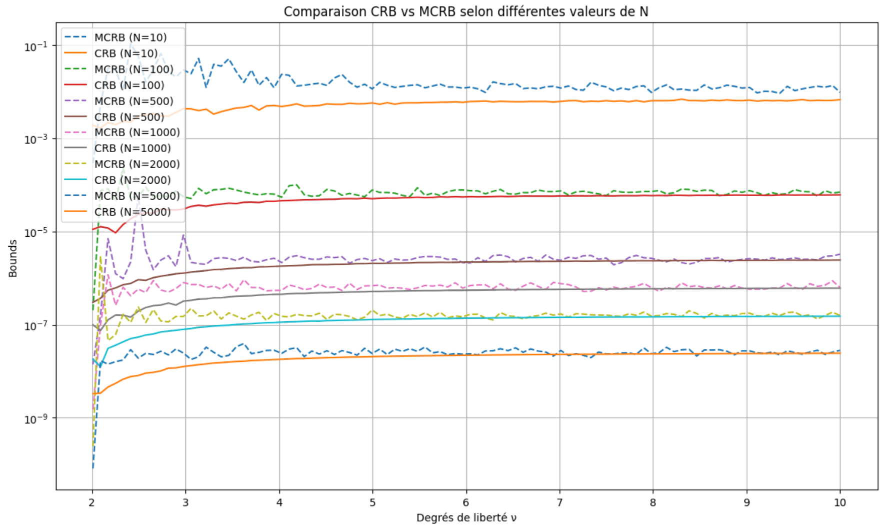

Figure 3: CRB and MCRB as a Function of Degrees of Freedom, Varying N from 10 to 5000

We observe that θ is better estimated using Newton's method in the misspecified case rather than in the well-specified case.

An estimation of θ in both the well-specified and misspecified cases yielded these results (Figure 4). This outcome is not highly representative, as we sometimes observe the ML estimator performing better than Newton's method, and vice-versa.

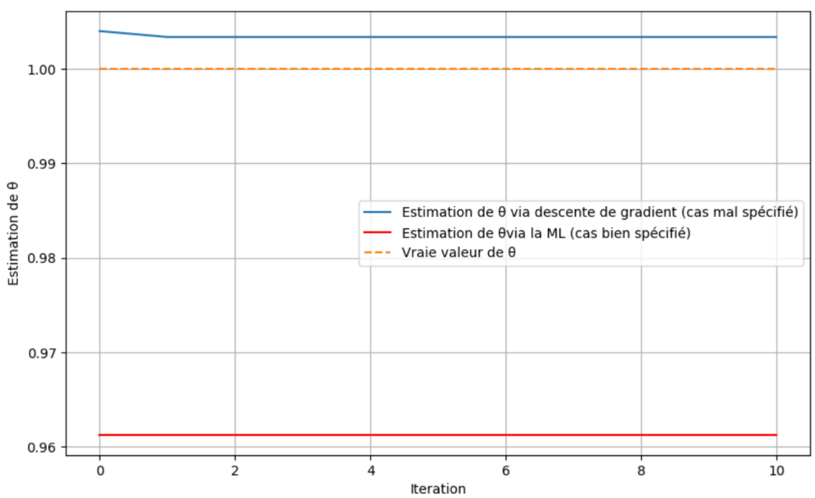

Figure 4: Estimation of θ (N $= 100$, $\sigma = 1$, $\nu = 3$)

3.3.2 Comparisons Using MSE

To compare the different estimators and evaluate the performance of misspecification, as well as to determine which estimator is best, we will study the Mean Squared Errors (MSE) for the ML estimator in the Gaussian case and the estimator derived from Newton's method in the t-distribution case. To do this, we use the following formula, after generating multiple estimations of θ which we will denote, for ML and Newton respectively, as $(\theta_{ML}^{(i)})_i$ and $(\theta_{NT}^{(i)})_i$:

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

We are comparing the performance of the two estimators by examining their Mean Squared Error (MSE) as a function of the number of observations (N), which we vary between 100 and 10,000. (Figure 5)

For small samples, ($N < 3000$), the Newton estimator shows a significantly lower MSE than the classical MLE. As N increases, both mean squared errors converge to 0, but the Newton estimator consistently remains slightly more performant in terms of MSE across all sample sizes.

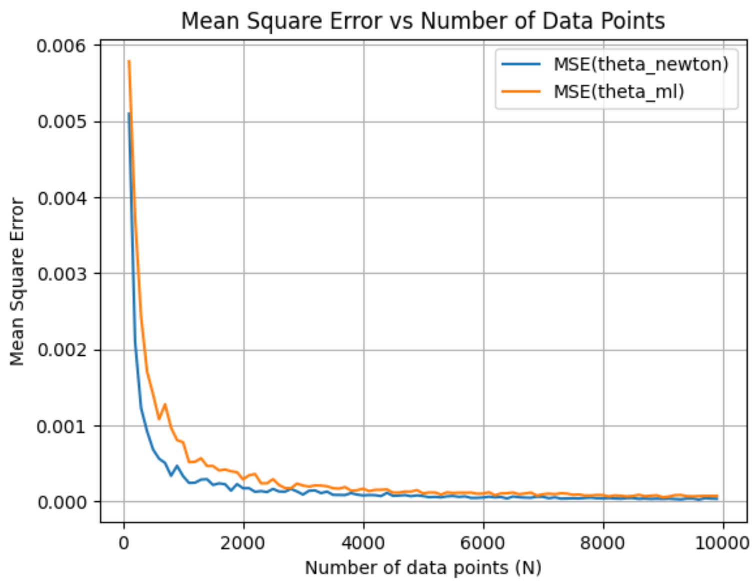

Figure 5: Mean Squared Error (MSE) for ML and Newton Estimators with N Ranging from 100 to 10000

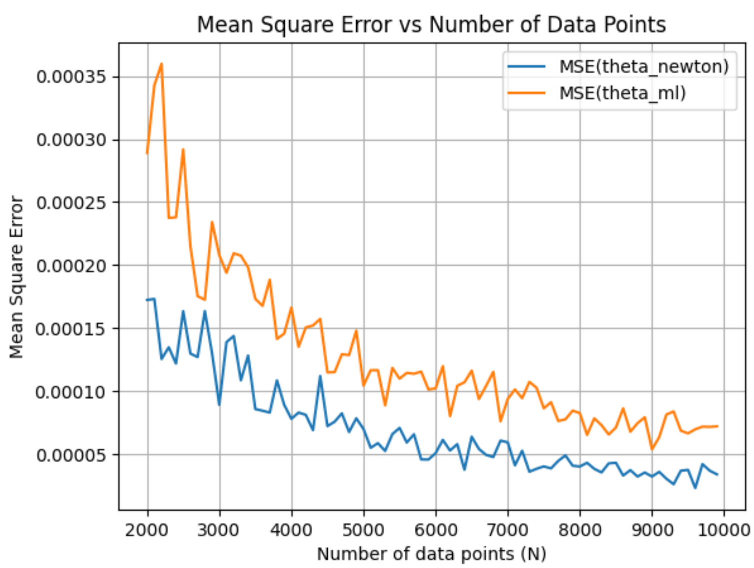

Figure 6: Mean Squared Error (MSE) for ML and Newton Estimators with N Ranging from 1000 to 10000

To further explore the superior performance of the Newton method and the impact of misspecification (and thus better noise modeling), we've plotted the MSE curves where the differences between MSE(ML) and MSE(Newton) are small (Figure 6). Similarly, we've plotted their absolute difference (Figure 7). This therefore confirms the observation from Figure 4.

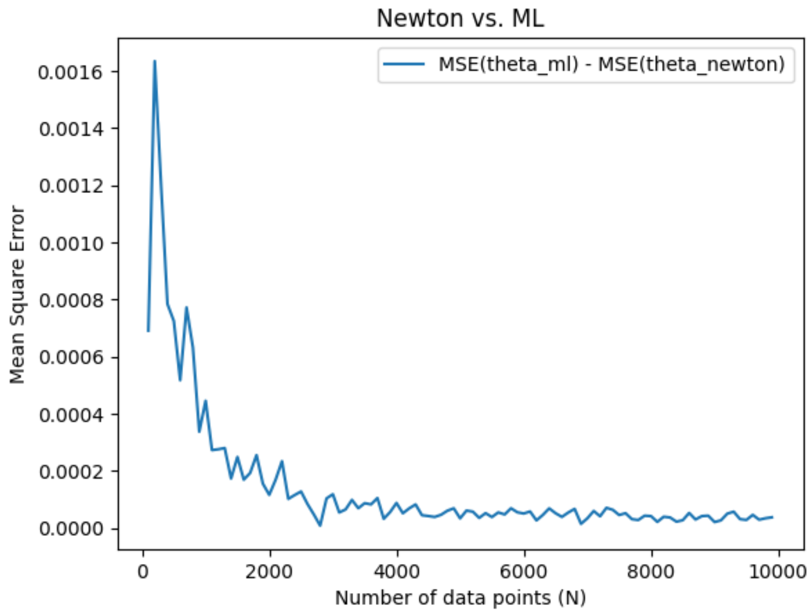

Figure 7: Difference: MSE(ML) - MSE(Newton)

4 Conclusion

This work highlights the importance of misspecified statistics in situations where the model assumptions, particularly the normality of the data, are not met. Through the study of t-distributions and AR(1) processes, we have shown that classical tools, such as the Cramér-Rao bound (CRB), can underestimate the actual variance of estimators in cases of misspecification.

The introduction of the Misspecified Cramér-Rao Bound (MCRB) provides a more realistic evaluation of estimator precision. Our simulations clearly demonstrated that the MCRB becomes significantly higher than the Cramér-Rao Bound (CRB) when data deviates from normality, and this discrepancy persists even with large sample sizes.

Furthermore, using adapted estimators, such as Newton's method for t-distributed noise, can significantly improve performance, particularly in terms of mean squared error.

In summary, misspecified statistics offer a rigorous and more robust framework for handling model imperfections. It stands out as an essential tool for reliable estimation in practice, especially when model assumptions, like data normality, aren't met.