



HAL
open science

Fast and asymptotically efficient estimation for t and $\log(t)$ distributions

Alexandre Brouste, Youssef Esstafa, Cécile Malique

► **To cite this version:**

Alexandre Brouste, Youssef Esstafa, Cécile Malique. Fast and asymptotically efficient estimation for t and $\log(t)$ distributions. 2024. hal-04612255

HAL Id: hal-04612255

<https://hal.science/hal-04612255>

Preprint submitted on 14 Jun 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Fast and asymptotically efficient estimation for t and $\log(t)$ distributions

Alexandre Brouste^{1*}, Youssef Esstafa¹ and Cécile Malique¹

^{1*}Laboratoire Manceau de Mathématiques, Le Mans Université, France.

*Corresponding author(s). E-mail(s): alexandre.brouste@univ-lemans.fr;
Contributing authors: youssef.esstafa@univ-lemans.fr;
cecile.malique.etu@univ-lemans.fr;

Abstract

Fast and asymptotically efficient one-step and two-steps estimation procedures for the parameters of the location-scale t and $\log(t)$ distributions are proposed. They are based on two possible initial guess estimators: the first one is the maximum likelihood estimator on a subsample and the second one is a combination of the empirical median for the location, a slowly converging Hill estimator for the degree of freedom and a maximum likelihood type estimator for the scale. Then, one step or two steps of the Fisher scoring gradient descent method are done in order to correct the initial estimation and reach asymptotical efficiency. The performances of the estimation procedures are evaluated on samples of finite size in terms of mean square error and computation time. An application in economics is also proposed.

Keywords: extreme value distribution, one-step procedure, Hill estimator, efficiency

1 Introduction

In various fields, including economics or finance and environmental science, extremely large observations cannot be ignored and has to be considered. Modeling the occurrence of extreme values can be done with heavy-tailed probability distributions (see [7] and the references therein). For instance, the Student distribution and the log Student distribution are used to model extreme phenomena due to their heavy tails and flexibility.

The Student distribution (or t distribution) is a generalization of the Cauchy distribution [2]. The estimation of the parameters of the t distribution has historically been one of the more intractable cases with the presence of the degree-of-freedom parameter [12].

For the t or $\log(t)$ distributions, the parameters are generally estimated by the maximum likelihood estimator which is asymptotically efficient. For the R software, the `fitdistr` function in the `fitdistrplus` package or the `fitdistr` function in the `MASS` package use numerical optimization method to compute the maximum likelihood estimator (MLE). The starting point used corresponds to the estimators for the location and scale in a Cauchy distribution (respectively the median and half of the interquartile range) and the degree of freedom is fixed to 10. This non-consistent initial guess estimator can leads to computation errors. Moreover, for large dataset and real-time applications, the MLE can be time consuming.

Fast and asymptotically efficient estimation one-step or two-steps procedures for the parameters of the location-scale t and $\log(t)$ distributions are considered in this paper. They are based on two possible initial guess estimators. The first one is the maximum likelihood estimator computed on a subsample. The second one is a combination of the empirical median for the location, the slowly converging Hill estimator for the degree of freedom and a maximum likelihood type estimator for the scale. Then, one-step or two steps of the Fisher scoring gradient descent method are done to correct the initial estimation and reach efficiency.

The aforementioned statistical procedures are directly inspired by the Le Cam one-step estimation procedure [13]. The asymptotical efficiency of the one-step estimator has been initially proved for a \sqrt{n} -consistent initial guess estimator and a uniformly continuous Fisher information matrix. It had been extended to $n^{\frac{\delta}{2}}$ -consistent initial guess estimator and Lipschitz continuous Fisher information matrix and also to multi-step estimation (see for instance [11]). For the R software, these methods were implemented for samples of independent and identically distributed univariate random variables in the `OneStep` package [3] and for several multivariate distribution in the `MLEce` package (see also [9, 10] for the Gamma and Weibull distributions). It is worth mentioning that the method has also been extended recently to independent but not identically distributed random variables [14], counting processes [4, 6], Markov processes [11], *etc.*

The notations and the classical estimation properties in this setting are detailed in Section 2. Main results on the one-step and two-steps estimation procedures are described in Section 3. In Section 4, the performances of the estimation procedures are evaluated on samples of finite size in terms of mean square error and computation time. An application in economics is also proposed in Section 5.

2 Notations

2.1 The t distribution

Let $\vartheta = (\mu, \sigma, \nu) \in \Theta \subset \mathbb{R} \times \mathbb{R}_*^+ \times \mathbb{R}_*^+$. The scale-location Student (or t) distribution is considered and characterized by its probability density function

$$g(x, \vartheta) = f\left(\frac{x - \mu}{\sigma}, \nu\right) \quad (1)$$

where

$$f(x, \nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\nu\pi}} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}.$$

Here ν is the degree of freedom of the Student distribution. This distribution is denoted $t(\mu, \sigma, \nu)$ in the following. The Cauchy distribution is obtained for $\nu = 1$. Let us also recall that for $X_1 \sim t(\mu, \sigma, \nu)$, the first moment only exists for $\nu > 1$ with $\mathbf{E}_\vartheta(X_1) = \mu$ and the variance only exists for $\nu > 2$ with $\mathbf{Var}_\vartheta(X_1) = \sigma^2 \frac{\nu}{\nu-2}$.

Let $\ell(\vartheta, x) = \log g(x, \vartheta)$. Direct computations, which are postponed in Appendix 1, gives the score function

$$\dot{\ell}(\vartheta, x) = \begin{pmatrix} \frac{\nu+1}{\sigma} \cdot \frac{y/\nu}{1+y^2/\nu} \\ -\frac{1}{\sigma} + \frac{\nu+1}{\sigma} \cdot \frac{y^2/\nu}{1+y^2/\nu} \\ \frac{1}{2}\psi^{(1)}(\frac{\nu+1}{2}) - \frac{1}{2}\psi^{(1)}(\frac{\nu}{2}) - \frac{1}{2\nu} - \frac{1}{2}\log(1 + \frac{y^2}{\nu}) + \frac{(\nu+1)}{2} \cdot \frac{y^2/\nu^2}{(1+y^2/\nu)} \end{pmatrix}$$

with $y = \frac{x-\mu}{\sigma}$ and $\psi^{(n)}$ is the polygamma functions (see [1, section 6.4.1, page 260]) defined by $\psi^{(n)}(\alpha) = \frac{\partial^n}{\partial \alpha^n} \log \Gamma(\alpha)$. The Fisher information matrix can also be obtained and reads

$$\begin{aligned} \mathcal{I}(\vartheta) &= -\mathbf{E}_\vartheta \left(\ddot{\ell}(\vartheta, X_1) \right) \\ &= \begin{pmatrix} \frac{\nu+1}{\sigma^2(\nu+3)} & 0 & 0 \\ 0 & \frac{2\nu}{\sigma^2(\nu+3)} & -\frac{2}{\sigma(\nu+1)(\nu+3)} \\ 0 & -\frac{2}{\sigma(\nu+1)(\nu+3)} & -\frac{1}{4}\psi^{(2)}\left(\frac{\nu+1}{2}\right) + \frac{1}{4}\psi^{(2)}\left(\frac{\nu}{2}\right) - \frac{(\nu+5)}{2\nu(\nu+1)(\nu+3)} \end{pmatrix}. \quad (2) \end{aligned}$$

It is worth noting that for $\nu = 1$, the Fisher information matrix reduces (see for instance [3]) to

$$\mathcal{I}_1(\mu, \sigma) = \begin{pmatrix} \frac{1}{2\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^2} \end{pmatrix}$$

and, for $\nu \rightarrow \infty$, to

$$\mathcal{I}_\infty(\mu, \sigma) = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{2}{\sigma^2} \end{pmatrix}$$

which is the Fisher information matrix for the joint estimation of the mean and standard-deviation of a Gaussian random variable.

2.2 The $\log(t)$ distribution

Let $\vartheta = (\mu, \sigma, \nu) \in \Theta \subset \mathbb{R} \times \mathbb{R}_*^+ \times \mathbb{R}_*^+$. The log-Student (or $\log(t)$) distribution is also considered and characterized by its probability density function

$$h(x, \vartheta) = \frac{1}{x} g(\log(x), \vartheta)$$

where g is defined in (1).

For this distribution, the score function is given by $\dot{\ell}(\vartheta, \log(x))$ and the Fisher information matrix is similar to (2). This distribution is denoted by $\log(t)(\mu, \sigma, \nu)$ in the following.

2.3 Maximum likelihood estimator

Let $X^{(n)} = (X_1, X_2, \dots, X_n)$ be a sample of independent and identically distributed location-scale t distribution (or $\log(t)$ distribution) random variables. The maximum likelihood estimator $\hat{\vartheta}_n$ (when it exists) satisfies

$$\dot{\ell}_n(\hat{\vartheta}_n) = 0_{\mathbb{R}^3} \quad (3)$$

where

$$\ell_n(\vartheta) = \begin{cases} \sum_{i=1}^n \ell(\vartheta, X_i) & \text{if } X_1 \sim t(\mu, \sigma, \nu), \\ \sum_{i=1}^n \ell(\vartheta, \log(X_i)) - \log(X_i) & \text{if } X_1 \sim \log(t)(\mu, \sigma, \nu). \end{cases}$$

It can be shown that, in both cases, the MLE is consistent, asymptotically normal with

$$\sqrt{n}(\hat{\vartheta}_n - \vartheta) \implies \mathcal{N}(0, \mathcal{I}(\vartheta)^{-1}) \quad \text{as } n \rightarrow \infty,$$

and asymptotically efficient in these statistical experiments. Since no explicit form can be exhibited, the computation of the MLE needs numerical methods and can be time consuming for large samples (see Section 4). Moreover, the optimization method generally uses a starting point which is practically not chosen consistently.

3 Main results

One-step and two-steps estimation procedures for the parameter $\vartheta = (\mu, \sigma, \nu)$ in the case of t and $\log(t)$ distributions are proposed in this section and are shown to be fast and asymptotically efficient. Since heavy-tail modeling is considered, we only consider distributions with $\nu < 4$.

3.1 One step and two-steps procedures

Let $X^{(n)} = (X_1, X_2, \dots, X_n)$ be a sample of i.i.d. random variables of t or $\log(t)$ distribution.

Since the Fisher information matrix is explicit, a fast and asymptotically efficient (Fisher scoring) multi-step procedure can be applied to estimate the 3-dimensional parameter ϑ . Namely, if we denote ϑ_n^* an initial guess estimator (which is not rate efficient but fast to be computed), the one-step estimator $\bar{\vartheta}_n^{(1)}$ is defined by

$$\bar{\vartheta}_n^{(1)} = \vartheta_n^* + \mathcal{I}(\vartheta_n^*)^{-1} \cdot \frac{1}{n} \sum_{j=1}^n \dot{\ell}(\vartheta_n^*, X_j), \quad n \geq 1, \quad (4)$$

and the two-steps estimator $\bar{\vartheta}_n^{(2)}$ is defined by

$$\bar{\vartheta}_n^{(2)} = \bar{\vartheta}_n^{(1)} + \mathcal{I}(\bar{\vartheta}_n^{(1)})^{-1} \cdot \frac{1}{n} \sum_{j=1}^n \dot{\ell}(\bar{\vartheta}_n^{(1)}, X_j), \quad n \geq 1, \quad (5)$$

for the t distribution. It also reads

$$\bar{\vartheta}_n^{(1)} = \vartheta_n^* + \mathcal{I}(\vartheta_n^*)^{-1} \cdot \frac{1}{n} \sum_{j=1}^n \dot{\ell}(\vartheta_n^*, \log(X_j)), \quad n \geq 1, \quad (6)$$

$$\bar{\vartheta}_n^{(2)} = \bar{\vartheta}_n^{(1)} + \mathcal{I}(\bar{\vartheta}_n^{(1)})^{-1} \cdot \frac{1}{n} \sum_{j=1}^n \dot{\ell}(\bar{\vartheta}_n^{(1)}, \log(X_j)), \quad n \geq 1, \quad (7)$$

for the $\log(t)$ distribution.

Theorem 1. *Let us consider an initial guess estimator ϑ_n^* that satisfy*

$$n^{\delta/2} (\vartheta_n^* - \vartheta) \quad \text{is tight in } \mathbb{R}^3 \quad \text{as } n \rightarrow \infty, \quad \frac{1}{4} < \delta \leq 1.$$

For $\frac{1}{2} < \delta \leq 1$, the one-step estimator $\bar{\vartheta}_n^{(1)}$ is asymptotically normal with

$$\sqrt{n} \left(\bar{\vartheta}_n^{(1)} - \vartheta \right) \Longrightarrow \mathcal{N} \left(0, \mathcal{I}(\vartheta)^{-1} \right) \quad \text{as } n \rightarrow \infty.$$

For $\frac{1}{4} < \delta \leq 1$, the two-steps estimator $\bar{\vartheta}_n^{(2)}$ is asymptotically normal with

$$\sqrt{n} \left(\bar{\vartheta}_n^{(2)} - \vartheta \right) \Longrightarrow \mathcal{N} \left(0, \mathcal{I}(\vartheta)^{-1} \right) \quad \text{as } n \rightarrow \infty.$$

Proof. The proof relies on the Lipschitz regularity of the explicit Fisher information matrix and is postponed in Appendix A. \square

The choice of an initial estimation is a key point of the whole procedure. For the t or $\log(t)$ distributions, we face the following challenge: when the parameter ν is high, moment estimators can be defined but the 3-3 component of the Fisher information matrix tends to zero, meaning that the estimation cannot be precise in practice. When ν is low, no moment exists for the random variable and no moment estimator can be defined. Later on, two initial guess estimators are proposed:

1. The (non-explicit) MLE on a subsample improved by a one-step procedure.
2. An ingenious combination of the empirical median for the location parameter μ the slowly converging Hill estimator for the degree-of-freedom ν and a maximum likelihood type estimator for the scale parameter σ , improved by a two-step procedure.

3.2 Initial guess estimator

In this section, two procedures are proposed: a one-step estimator with the MLE on a subsample as initial guess estimator and a two-steps estimator with an initial unbiased Hill estimator.

3.2.1 MLE on a subsample

The MLE is asymptotically normal with rate \sqrt{n} ; consequently the MLE on a subsample (subMLE) of size $\lfloor n^\delta \rfloor$, denoted $\vartheta_n^* = \widehat{\vartheta}_{\lfloor n^\delta \rfloor}$, satisfy

$$\sqrt{n^\delta} (\vartheta_n^* - \vartheta) \implies \mathcal{N}(0, \mathcal{I}(\vartheta)^{-1}) \quad \text{as } n \rightarrow \infty,$$

and is consequently $n^{\delta/2}$ -consistent (see also [3] for other examples).

Choosing $1/2 < \delta \leq 1$, the subMLE can be used as an initial guess estimator for the one-step estimator $\widehat{\vartheta}_n^{(1)}$ in (4) or (6). This choice works well in practice but we face the same problem of the choice of non-consistent starting point in the numerical optimization scheme.

The choice of the subMLE with $\frac{1}{4} < \delta \leq \frac{1}{2}$ as an initial guess in the two-step procedure has been excluded from our study leading to numerical errors in the optimization procedure for very small samples.

3.2.2 Slowly converging Hill based estimator

Motivation:

Let $Y^{(n)} = (Y_1, Y_2, \dots, Y_n)$ be a sample of i.i.d. standard $t(0, \sigma, \nu)$ random variables. The cumulative distribution function is given by

$$F_{\sigma, \nu}(y) = P(Y_1 \leq y) = \frac{1}{2} + \frac{y}{\sigma} \Gamma\left(\frac{\nu+1}{2}\right) \frac{{}_2F_1\left(\frac{1}{2}, \frac{\nu+1}{2}, \frac{3}{2}, -\frac{y^2}{\nu\sigma^2}\right)}{\sqrt{\pi\nu}\Gamma\left(\frac{\nu}{2}\right)} \quad (8)$$

where ${}_2F_1$ is the hypergeometric function (see [1, section 15.1.1, page 556]). Let us denote $\overline{F}(u) = 1 - F_{\sigma, \nu}(y)$ the survival function of the distribution and $\gamma = \frac{1}{\nu}$.

Let $Y_{1,n} \leq Y_{2,n} \leq \dots \leq Y_{n,n}$ be the order statistic. For this distribution, the (upper) tail is characterized by

$$\lim_{t \rightarrow \infty} \frac{\overline{F}(yt)}{\overline{F}(t)} = y^{-1/\gamma}, \quad y > 0,$$

and the second order variation by

$$\lim_{t \rightarrow \infty} \frac{1}{A(1/\bar{F}(t))} \left(\frac{\bar{F}(yt)}{\bar{F}(t)} - y^{-1/\gamma} \right) = y^{-1/\gamma} \frac{y^{\rho/\gamma} - 1}{\gamma\rho}$$

where $\rho = -2/\nu$, $A(t) = \gamma b t^\rho$ and $b = \frac{\nu(\nu+1)}{\nu+2} \left(\frac{\Gamma(\frac{\nu+1}{2}) \nu^{(\nu-1)/2}}{\sqrt{\pi\nu} \Gamma(\frac{\nu}{2})} \right)^\rho$. The proof is postponed in Appendix B. In this setting, the Hill estimator

$$\bar{\gamma}_n = \frac{1}{k_n} \sum_{i=1}^{k_n} \log \frac{Y_{n-i+1,n}}{Y_{n-k_n,n}}, \quad (9)$$

is generally biased, namely for $k_n \rightarrow \infty$, $k_n/n \rightarrow 0$ and $\sqrt{k_n} A(n/k_n) \rightarrow \lambda \in \mathbb{R}$,

$$\sqrt{k_n} (\bar{\gamma}_n - \gamma) \Rightarrow \mathcal{N} \left(\frac{\lambda}{1-\rho}, \gamma^2 \right) \quad \text{as } n \rightarrow \infty. \quad (10)$$

For the Student distribution, taking $k_n = n^\delta$, $1/4 < \delta \leq 1$, we have the following cases:

- If $\delta < \frac{4}{4+\nu}$ then $\lambda = 0$,
- If $\delta = \frac{4}{4+\nu}$, then $\lambda = \gamma b < \infty$,
- If $\delta > \frac{4}{4+\nu}$, then $\lambda \rightarrow \infty$.

Therefore for $\nu < 4$ and a choice of $\frac{1}{4} < \delta < 1/2$, independently of the value of σ , the Hill estimator is unbiased and can be used as an initial guess estimator in our two-steps procedure.

Construction of the estimator of $\vartheta = (\mu, \sigma, \nu)$:

Let us consider the sample $X^{(n)} = (X_1, X_2, \dots, X_n)$ composed of i.i.d. random variables of $t(\mu, \sigma, \nu)$ distribution. For the joint estimation, we proceed as follows:

1. Estimate the parameter μ with the empirical median $\tilde{\mu}_n$.
2. Apply the Hill estimator (9) on the recentered sample composed of $Z_i = X_i - \tilde{\mu}_n$ which are approximatively $t(0, \sigma, \nu)$ distributed. Here σ acts as a nuisance parameter.
3. Estimate, finally, the scale σ by a MLE-type estimator $\tilde{\sigma}_n$ with the second component of the score function, namely

$$\frac{\partial}{\partial \sigma} \ell_n(\tilde{\mu}_n, \tilde{\sigma}_n, \tilde{\nu}_n) = 0 \quad (11)$$

where $\tilde{\mu}_n$ and $\tilde{\nu}_n$ are previously computed.

Theorem 2. *The aforementioned estimation procedure leads to a $n^{\frac{\delta}{2}}$ -consistent initial guess estimator with $\frac{1}{4} < \delta < \frac{1}{2}$.*

Proof. Let $X^{(n)} = (X_1, X_2, \dots, X_n)$ be a sample of i.i.d. random variables of $t(\mu, \sigma, \nu)$ distribution. The empirical median $\tilde{\mu}_n$ is a \sqrt{n} -consistent and asymptotically normal estimator of μ .

The parameter σ acts as a nuisance parameter in the estimation of ν with the aforementioned Hill type procedure. By plug-in, the proposed estimator $\tilde{\nu}_n$ is $\sqrt{k_n}$ -consistent with asymptotic zero mean for $\frac{1}{4} < \delta < \frac{1}{2}$ (the asymptotic variance depends on σ). Namely,

$$\begin{aligned}\tilde{\gamma}_n &= \frac{1}{k_n} \sum_{i=1}^{k_n} \log \frac{Z_{n-i+1,n}}{Z_{n-k_n,n}} = \frac{1}{k_n} \sum_{i=1}^{k_n} \log \frac{Y_{n-i+1,n} + (\tilde{\mu}_n - \mu)}{Y_{n-k_n,n} + (\tilde{\mu}_n - \mu)} \\ &= \bar{\gamma}_n + (\tilde{\mu}_n - \mu) \frac{1}{k_n} \sum_{i=1}^{k_n} \left(\frac{1}{Y_{n-i+1,n}} - \frac{1}{Y_{n-k_n,n}} \right) + o(\tilde{\mu}_n - \mu)\end{aligned}$$

where $\bar{\gamma}_n$ is defined in (9). Consequently, one obtains

$$\begin{aligned}\sqrt{k_n}(\tilde{\gamma}_n - \gamma) &= \sqrt{k_n}(\bar{\gamma}_n - \gamma) + \sqrt{k_n}(\tilde{\mu}_n - \mu)A + o(\sqrt{k_n}(\tilde{\mu}_n - \mu)) \\ &= \sqrt{k_n}(\bar{\gamma}_n - \gamma) + \sqrt{\frac{k_n}{n}} \cdot \sqrt{n}(\tilde{\mu}_n - \mu) \cdot A + o\left(\sqrt{\frac{k_n}{n}} \cdot \sqrt{n}(\tilde{\mu}_n - \mu)\right)\end{aligned}\quad (12)$$

where

$$\begin{aligned}A &= \frac{1}{k_n} \sum_{i=1}^{k_n} \left(\frac{1}{Y_{n-i+1,n}} - \frac{1}{Y_{n-k_n,n}} \right) \\ &= \frac{1}{Y_{n-k_n,n}} \cdot \left(\frac{1}{k_n} \sum_{i=1}^{k_n} \frac{Y_{n-k_n,n}}{Y_{n-i+1,n}} - 1 \right).\end{aligned}$$

By Rényi representation [7, Lemma 3.2.3 p.71 and Exercise 3.2 p.125], we have $\frac{1}{k_n} \sum_{i=1}^{k_n} \frac{Y_{n-k_n,n}}{Y_{n-i+1,n}} \rightarrow \frac{1}{1+\gamma}$ in probability as $n \rightarrow \infty$ and by [7, Lemma 3.2.1 p.69] we have $\frac{1}{Y_{n-k_n,n}} \rightarrow 0$ in probability as $n \rightarrow \infty$ that gives the convergence in probability of A to zero. Since $k_n/n \rightarrow 0$ and $\tilde{\mu}_n$ is \sqrt{n} -consistent, we have the convergence in probability of the second term in the r.h.s. of (12) to zero as $n \rightarrow \infty$.

Finally, plugging the two estimates $(\tilde{\mu}_n, \tilde{\nu}_n)$ in the score function to estimate σ leads to a $\sqrt{k_n}$ -consistent estimator that ends the proof. Namely, the estimator is defined by $\frac{\partial}{\partial \sigma} \ell_n(\tilde{\vartheta}_n) = 0$, $\tilde{\mu}_n$ and $\tilde{\nu}_n$ being defined previously.

$$\frac{\partial}{\partial \sigma} \ell_n(\tilde{\vartheta}_n) = \frac{\partial}{\partial \sigma} \ell_n((\tilde{\mu}_n, \sigma, \tilde{\nu}_n)) + (\tilde{\sigma}_n - \sigma) \frac{\partial^2}{\partial \sigma^2} \ell_n(\tilde{\mu}_n, \tilde{\sigma}_n, \tilde{\nu}_n)$$

with $\sigma < \tilde{\sigma}_n < \tilde{\sigma}_n$ say. On the other hand,

$$\begin{aligned}\frac{\partial}{\partial \sigma} \ell_n(\tilde{\mu}_n, \sigma, \tilde{\nu}_n) &= \frac{\partial}{\partial \sigma} \ell_n(\mu, \sigma, \nu) + (\tilde{\mu}_n - \mu) \frac{\partial^2}{\partial \mu \partial \sigma} \ell_n(\tilde{\mu}_n, \sigma, \tilde{\nu}_n) + \\ &\quad + (\tilde{\nu}_n - \nu) \frac{\partial^2}{\partial \nu \partial \sigma} \ell_n(\tilde{\mu}_n, \sigma, \tilde{\nu}_n)\end{aligned}$$

with $\mu < \check{\mu}_n < \tilde{\mu}_n$ and $\nu < \check{\nu}_n < \tilde{\nu}_n$ say. We get

$$\begin{aligned} \sqrt{k_n}(\tilde{\sigma}_n - \sigma) = & - \left(\frac{1}{n} \frac{\partial^2}{\partial \sigma^2} \ell_n((\tilde{\mu}_n, \check{\sigma}_n, \tilde{\nu}_n)) \right)^{-1} \left(\sqrt{\frac{k_n}{n}} \sqrt{n}(\tilde{\mu}_n - \mu) \frac{1}{n} \frac{\partial^2}{\partial \mu \partial \sigma} \ell_n(\tilde{\mu}_n, \sigma, \check{\nu}_n) + \right. \\ & \left. + \sqrt{k_n}(\tilde{\nu}_n - \nu) \frac{1}{n} \frac{\partial^2}{\partial \nu \partial \sigma} \ell_n(\tilde{\mu}_n, \sigma, \check{\nu}_n) + \sqrt{\frac{k_n}{n}} \frac{1}{\sqrt{n}} \frac{\partial}{\partial \sigma} \ell_n(\mu, \sigma, \nu) \right). \end{aligned} \quad (13)$$

Consistency of the estimators $\tilde{\mu}_n$ and $\tilde{\nu}_n$ with the uniform continuity of the Fisher information matrix gives the convergence of the quantities $\frac{1}{n} \frac{\partial^2}{\partial \nu \partial \sigma} \ell_n$, $\frac{1}{n} \frac{\partial^2}{\partial \mu \partial \sigma} \ell_n$ and $\frac{1}{n} \frac{\partial^2}{\partial \sigma^2} \ell_n$ with the law of large numbers. Since $\frac{1}{\sqrt{n}} \frac{\partial}{\partial \sigma} \ell_n(\vartheta)$ is bounded in probability due to the central limit theorem, the quantity $\sqrt{n}(\tilde{\mu}_n - \mu)$ is tight and $k_n/n \rightarrow 0$ gives the convergence in probability of the first and third term in the r.h.s. in (13) to zero as $n \rightarrow \infty$. Finally, we deduce

$$\sqrt{k_n}(\tilde{\sigma}_n - \sigma) \implies \mathcal{N} \left(0, \gamma^2 \frac{\mathcal{I}(\vartheta)_{2,3}^2}{\mathcal{I}(\vartheta)_{3,3}^2} \right) \quad \text{as } n \rightarrow \infty. \quad (14)$$

□

Remark 1. *Our estimator reduces the optimization (or root solving) problem (3) of dimension 3 to the root solving problem (11) of dimension 1 for the sole scale parameter. In practice, we keep the starting point in this procedure to half of the interquartile range.*

Remark 2. *It is worth noting that the procedure works the same for the $\log(t)(\mu, \sigma, \nu)$ distribution, namely:*

1. *Estimate the parameter μ with the empirical median on the log values of the sample.*
2. *Applying the Hill estimator on the recentered sample composed of $Z_i = \log(X_i) - \tilde{\mu}_n$.*
3. *Finally, estimate the scale by a MLE-type estimator on the sole scale parameter.*

4 Simulations

The aforementioned methodologies:

- MLE: Maximum Likelihood Estimator
- subMLE: MLE computed on a subsample
- OS-subMLE: One-step procedure with initial subMLE guess estimator
- Hill: Slowly converging Hill based estimator introduced in Section 3.2.2
- TS-Hill: Two-steps procedure with initial Hill estimator

have been implemented with the R software for evaluating the performance of the estimators on samples of finite size in terms of mean square error and computation time. We present in this section the results for the t distribution but simulation results for the $\log(t)$ distribution are similar. In our simulations, the MLE are computed in two different ways:

1. with the optimization method used by the `fitdistr` function of the MASS package on the loglikelihood function. As mentioned above, the starting point used in the

- numerical procedure corresponds to the estimators for the location and scale in a Cauchy distribution, respectively the median and half of the interquartile range whereas the degree of freedom is fixed to 10. It is called MLE (max) in the following.
- with the `multiroot` function of the `rootSolve` package on our own score function mimicking Equation (3). The starting point is the true value in order to avoid numerical instability of this method. For this reason, this method cannot be used in practice but serves here for computation time comparison. It is called MLE (score) in the following.

The subMLE is only computed with the `fitdistr` function using the aforementioned starting point. It is the initial guess estimator in the OS-subMLE procedure (see Section 3.2.1). It is worth recalling that two-steps procedure with initial subMLE estimate has been excluded leading to numerical instability.

Finally, the estimator that combines the empirical median for the location, the slowly converging Hill estimator for the degree-of-freedom and the MLE-type estimator for the scale is called Hill in the following. As explained in Section 3.2.2, we are using the second component of the score to deduce the scale estimator. This is done with the `multiroot` function with a starting point which is the half of the interquartile range. This method is the initial guess estimator in the TS-Hill procedure.

For comparison in terms of mean square error and computation time, we generate $K = 4000$ Monte-Carlo simulations of samples of size $n = 10^4$ choosing $\mu = 2$, $\sigma = 0.5$ and $\nu = 3$. For the OS-SubMLE procedure, δ is fixed to 0.8. For the TS-Hill procedure, δ is fixed to 0.4.

Comparison in terms of mean square error:

The histograms of the renormalized error $\sqrt{n}(\tilde{\vartheta}_n - \vartheta)$ (for a generic estimation procedure $\tilde{\vartheta}_n$) have been illustrated on Figures 1, 2 and 3. The MLE (max), MLE (score), OS-subMLE and OS-Hill which are shown to be asymptotically efficient naturally overperform the initial guess estimators subMLE and Hill.

Comparison in terms of computation time:

The following table present the computation time of MLE (max), MLE (score), OS-subMLE and TS-Hill estimators. The OS-subMLE is 6 times faster than the MLE (max) and the TS-Hill procedure more than 130 times faster than the MLE (max). We recall that MLE (score) cannot be used in practice due to numerical instability but the TS-Hill procedure is still 2 times faster than the MLE (score).

	MLE (max)	MLE (score)	TS-Hill	OS-subMLE
time (s)	980.61	16.12	7.16	152.34

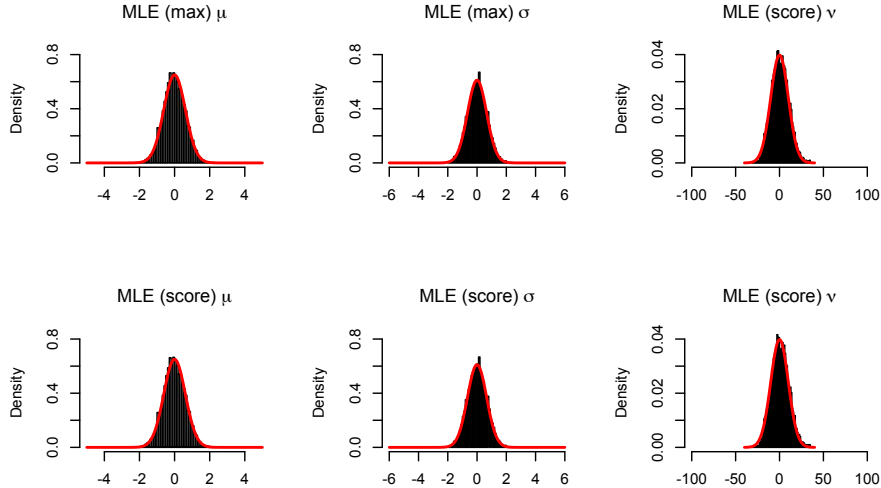


Fig. 1 Histograms of the renormalized error for the MLE (max) of μ , σ and ν . The superimposed red line is the asymptotical efficient distribution.

5 Applications

5.1 OneStep package

The previous estimation procedures (one-step with initial subMLE for $\frac{1}{2} < \delta \leq 1$ and two-steps estimator with Hill based estimator for $\frac{1}{4} < \delta < 1/2$) have been implemented in the `onestep` command of the `OneStep` package for the R software [15].

The location-scale t distribution is available in the `extraDistr` package and named `lst`. After loading the `extraDistr` and `OneStep` packages, the two-steps estimator can be computed with the command `onestep(Z, "lst", control=list(delta=0.4))`

5.2 Economics

A dataset of 157 of the national consumer price index in Brazil are considered in [2]. The authors have considered a Cauchy distribution in this context. We propose to model the index with the t -distribution, adding the degree-of-freedom parameter in the modelling.

In order to compare the two models, we will compute the Akaike Information Criterion (AIC) defined

- for a general Student model $t(\mu, \sigma, \nu)$ and an estimator $\vartheta_n^* = (\mu_n^*, \sigma_n^*, \nu_n^*)$ by

$$AIC_t = 2 \times 3 - 2\ell_n(\mu_n^*, \sigma_n^*, \nu_n^*),$$

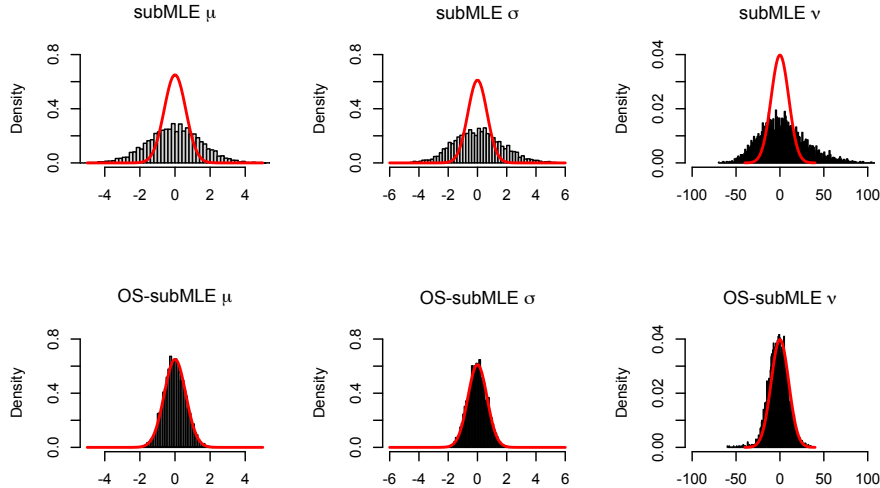


Fig. 2 Histograms of the renormalized error for the subMLE and OS-subMLE of μ , σ and ν . The superimposed red line is the asymptotical efficient distribution.

- for the Cauchy model $t(\mu, \sigma, 1)$ and an estimator $\vartheta_n^{**} = (\mu_n^{**}, \sigma_n^{**})$ by

$$AIC_C = 2 \times 2 - 2\ell_n(\mu_n^{**}, \sigma_n^{**}, 1).$$

We select the model which presents the lowest AIC.

A one-step estimator has been used for the Cauchy model $t(\mu, \sigma, 1)$ (see [3] for the definition) to compute the AIC. The estimation is 0.479 for the location parameter and 0.265 for the scale parameter. The AIC is equal to 282.5.

For the general Student model $t(\mu, \sigma, \nu)$, The two-steps estimator has been used to compute the AIC. The estimation is 0.510 for the location, 0.410 the scale and 2.68 for the degree-of-freedom. The AIC is 261.9 and this model is selected.

Acknowledgments: We would like to thank Gilles Stupfler for fruitful discussions on the topic. This research benefited from the support of the ANR project 'Efficient inference for large and high-frequency data' (ANR-21-CE40-0021), the chair 'Efficience et Sobriété Numériques', a joint initiative by Le Mans University and EREN Groupe under the aegis of the Institut Louis Bachelier and LabEx CHL ANR-11-LABX-020-01.

A Proof of Theorem 1

For an initial guess estimator ϑ_n^* which is $n^{\frac{\delta}{2}}$ -consistent with $\frac{1}{2} < \delta \leq 1$, the one-step estimator $\vartheta_n^{(1)}$ is shown to be \sqrt{n} -consistent with a centered normal limit with variance $\mathcal{I}(\vartheta)^{-1}$. The proof relies on the Lipschitz regularity of the explicit Fisher information matrix and is given in Section A.1

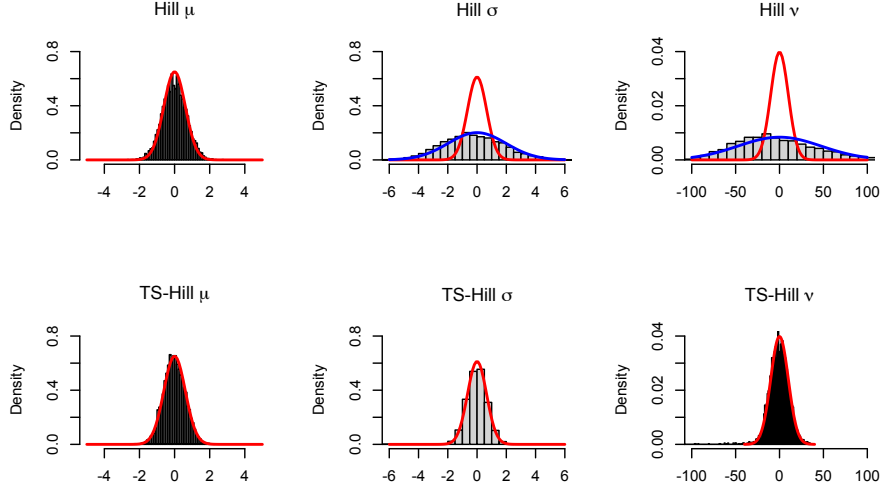


Fig. 3 Histograms of the renormalized error for the Hill and TS-Hill of μ , σ and ν . The superimposed red line is the asymptotical efficient distribution. The blue line corresponds to the theoretical asymptotic distribution of the Hill estimator of ν (see Equation (10)) and of the MLE-type estimator of σ (see Equation 14).

For the two-steps estimation procedure, with an initial guess estimator ϑ_n^* which is $n^{\frac{\delta}{2}}$ -consistent with $\frac{1}{4} < \delta \leq \frac{1}{2}$, we show firstly, that the one-step estimator $\vartheta_n^{(1)}$ is $n^{\frac{\gamma}{2}}$ -consistent with $\frac{1}{2} < \gamma = 2\delta \leq 1$ in Section A.2. Then, secondly, we use the previous result (Section A.1) to show that the two-steps estimator $\vartheta_n^{(2)}$ is \sqrt{n} -consistent with a centered normal limit with variance $\mathcal{I}(\vartheta)^{-1}$.

A.1 Efficiency of the one-step procedure

Let $(\vartheta_n^*, n \geq 1)$ be $n^{\frac{\delta}{2}}$ -consistent initial sequence of guess estimators with $\frac{1}{2} < \delta \leq 1$. Recall that $(\hat{\vartheta}_n, n \geq 1)$ is the sequence of maximum likelihood estimator defined by (3). We show in the following the asymptotic equivalence of the one-step estimator and the maximum likelihood estimator, namely

$$\sqrt{n} (\hat{\vartheta}_n - \vartheta_n^*) \longrightarrow 0 \quad \text{in probability}$$

as $n \rightarrow \infty$.

The mean-value theorem gives,

$$\begin{aligned} \dot{\ell}_n(\vartheta_n^*) &= \dot{\ell}_n(\hat{\vartheta}_n) + \int_0^1 \ddot{\ell}_n(\hat{\vartheta}_n + v(\vartheta_n^* - \hat{\vartheta}_n)) dv \cdot (\vartheta_n^* - \hat{\vartheta}_n) \\ &= \int_0^1 \ddot{\ell}_n(\hat{\vartheta}_n + v(\vartheta_n^* - \hat{\vartheta}_n)) dv \cdot (\vartheta_n^* - \hat{\vartheta}_n). \end{aligned} \quad (15)$$

Based on (5), let us define a generic one-step procedure

$$\bar{\vartheta}_n = \vartheta_n^* + \mathcal{I}(\vartheta_n^*)^{-1} \cdot \frac{\dot{\ell}_n(\vartheta_n^*)}{n}, \quad n \geq 1.$$

We have

$$\left(\bar{\vartheta}_n - \widehat{\vartheta}_n\right) = \left(\vartheta_n^* - \widehat{\vartheta}_n\right) + \mathcal{I}(\vartheta_n^*)^{-1} \cdot \frac{\dot{\ell}_n(\vartheta_n^*)}{n}$$

and

$$\left(\bar{\vartheta}_n - \widehat{\vartheta}_n\right) = \left(I_3 + \mathcal{I}(\vartheta_n^*)^{-1} \int_0^1 \ddot{\ell}_n\left(\widehat{\vartheta}_n + v\left(\vartheta_n^* - \widehat{\vartheta}_n\right)\right) dv\right) \left(\vartheta_n^* - \widehat{\vartheta}_n\right)$$

where I_3 is the 3×3 identity matrix. It follows that

$$\begin{aligned} \sqrt{n} \left(\bar{\vartheta}_n - \widehat{\vartheta}_n\right) &= n^{\frac{1}{2}-\delta} \mathcal{I}(\vartheta_n^*)^{-1} \cdot \\ &\cdot n^{\frac{\delta}{2}} \left(\left[\mathcal{I}(\vartheta_n^*) + \frac{\ddot{\ell}_n(\widehat{\vartheta}_n)}{n} \right] + \frac{1}{n} \int_0^1 \left(\ddot{\ell}_n\left(\widehat{\vartheta}_n + v\left(\vartheta_n^* - \widehat{\vartheta}_n\right)\right) - \ddot{\ell}_n(\widehat{\vartheta}_n) \right) dv \right) \cdot \\ &\cdot n^{\frac{\delta}{2}} \left(\vartheta_n^* - \widehat{\vartheta}_n\right). \end{aligned}$$

A Lipschitz condition on the Fisher information matrix and the CLT give

$$\frac{1}{n} \left\| \ddot{\ell}_n\left(\widehat{\vartheta}\right) - \ddot{\ell}_n(\vartheta_n^*) \right\| \leq L \|\widehat{\vartheta}_n - \vartheta_n^*\|$$

and the boundedness of the middle term in the r.h.s..

A.2 The first step in the two step procedure

Let us now consider $(\vartheta_n^*, n \geq 1)$ an initial $n^{\frac{\delta}{2}}$ -consistent initial sequence of guess estimators with $\frac{1}{4} < \delta \leq 1/2$. We show that the one-step estimator $\bar{\vartheta}_n^{(1)}$ defined by

$$\bar{\vartheta}_n^{(1)} = \vartheta_n^* + \mathcal{I}(\vartheta_n^*)^{-1} \cdot \frac{\dot{\ell}_n(\vartheta_n^*)}{n}, \quad n \geq 1,$$

is $n^{\frac{\gamma}{2}}$ -consistent with $\frac{1}{2} < \gamma = 2\delta < 1$. Direct computations lead to

$$\begin{aligned} n^{\frac{\gamma}{2}} \left(\bar{\vartheta}_n^{(1)} - \vartheta\right) &= n^{\frac{\gamma}{2}} (\vartheta_n^* - \vartheta) + n^{\frac{\gamma}{2}} \mathcal{I}(\vartheta_n^*)^{-1} \cdot \frac{\dot{\ell}_n(\vartheta_n^*)}{n} \\ &= n^{\frac{\delta}{2}} \left(I_3 - \frac{1}{n} \int_0^1 \ddot{\ell}(\vartheta + v(\vartheta_n^* - \vartheta)) dv \right) n^{\frac{\delta}{2}} (\vartheta_n^* - \vartheta) + \\ &\quad + n^{\frac{\gamma}{2}-\frac{1}{2}} \mathcal{I}(\vartheta_n^*)^{-1} \cdot \frac{\dot{\ell}_n(\vartheta)}{\sqrt{n}}. \end{aligned}$$

The second term on the r.h.s. tends to zero in probability since $\frac{1}{\sqrt{n}}\dot{\ell}_n(\vartheta)$ is bounded in probability due to the central limit theorem and $n^{\frac{3}{2}-\frac{1}{2}}$ tends to zero as $n \rightarrow \infty$.

The first term of the r.h.s. is bounded in probability, with similar computations using the Lipschitz continuity of the Fisher information matrix and the fact that $n^{\frac{\delta}{2}}(\vartheta_n^* - \vartheta)$ is tight.

B Characterization of the survival function of the t distribution

Let us recall that, for $b - a \notin \mathbb{Z}$,

$$\begin{aligned} {}_2F_1(a, b, c, z) &= \frac{\Gamma(b-a)\Gamma(c)}{\Gamma(b)\Gamma(c-a)} \frac{1}{(-z)^a} {}_2F_1\left(a, a-c+1, a-b+1, \frac{1}{z}\right) + \\ &+ \frac{\Gamma(a-b)\Gamma(c)}{\Gamma(a)\Gamma(c-b)} \frac{1}{(-z)^b} {}_2F_1\left(b, b-c+1, b-a+1, \frac{1}{z}\right) \end{aligned} \quad (16)$$

and

$${}_2F_1(a, b, c, z) = \frac{\Gamma(c)}{\Gamma(a)\Gamma(b)} \sum_{n=0}^{\infty} \frac{\Gamma(a+n)\Gamma(b+n)}{\Gamma(c+n)} \frac{z^n}{n}. \quad (17)$$

With (16), we get

$$\begin{aligned} \bar{F}(y) &= -y^{-\nu} \frac{\Gamma((\nu+1)/2)\Gamma(-\nu/2)}{\Gamma(\nu/2)\Gamma(1-\nu/2)} \frac{\sigma^\nu(\sqrt{\nu})^\nu}{2\sqrt{\pi}} {}_2F_1((\nu+1)/2, \nu/2, 1+\nu/2, -\sigma^2\nu/y^2) \\ &= y^{-\nu} \frac{\Gamma((\nu+1)/2)}{\Gamma(1+\nu/2)} \frac{\sigma^\nu(\sqrt{\nu})^\nu}{2\sqrt{\pi}} {}_2F_1((\nu+1)/2, \nu/2, 1+\nu/2, -\sigma^2\nu/y^2). \end{aligned}$$

with $\Gamma(z)\Gamma(1-z) = \frac{\pi}{\sin(\pi z)}$. Then, one can use the first order expansion in (17) to obtain

$$\bar{F}(y) \simeq y^{-\nu} \alpha (1 + \beta y^{-2})$$

with

$$\alpha = \frac{\Gamma((\nu+1)/2)}{\Gamma(1+\nu/2)} \frac{\sigma^\nu(\sqrt{\nu})^\nu}{2\sqrt{\pi}} = \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)} \frac{\sigma^\nu(\sqrt{\nu})^{\nu-1}}{\sqrt{\pi\nu}} \quad \text{and} \quad \beta = -\frac{\nu}{2} \cdot \frac{\nu(\nu+1)}{(\nu+2)} \sigma^2$$

exploiting $\Gamma(z+1) = z\Gamma(z)$ when $\text{Re}(z) > 0$. Denoting $\gamma = \frac{1}{\nu}$, we get

$$\lim_{t \rightarrow \infty} \frac{\bar{F}(yt)}{\bar{F}(t)} = y^{-1/\gamma}, \quad y > 0.$$

Then, fixing $\rho = -\frac{2}{\nu}$, we deduce

$$\left(\frac{\bar{F}(yt)}{\bar{F}(t)} - y^{-1/\gamma}\right) \simeq \gamma\rho\beta t^{-2} \cdot y^{-1/\gamma} \frac{y^{\rho/\gamma} - 1}{\gamma\rho}.$$

Since $\bar{F}(t) \simeq \alpha t^{-\nu}$, we fix $A(u) = \gamma\rho\alpha^\rho\beta u^\rho$ to get

$$\lim_{t \rightarrow \infty} \frac{1}{A(1/\bar{F}(t))} \left(\frac{\bar{F}(yt)}{\bar{F}(t)} - y^{-1/\gamma}\right) = y^{-1/\gamma} \frac{y^{\rho/\gamma} - 1}{\gamma\rho}.$$

C Some results

Here are some results about the moments of the $t(0, 1, \nu)$ distribution. Namely

$$\mathbf{E}_\theta [(1 + Y_1^2/\nu)^{-m}] = \frac{(\nu/2 + m - 1) \cdots (\nu/2)}{(\nu + 1)/2 + m - 1 \cdots ((\nu + 1)/2)}.$$

The expectation of the score is null, for the second component, since

$$\mathbf{E}_\theta [Y_1^2/\nu \cdot (1 + Y_1^2/\nu)^{-1}] = \frac{1}{\nu + 1}.$$

For the computation of the Fisher information matrix, the following moment are also generally used:

$$\mathbf{E}_\theta [Y_1^2/\nu \cdot (1 + Y_1^2/\nu)^{-2}] = \frac{\nu}{(\nu + 3)(\nu + 1)}$$

and

$$\mathbf{E}_\theta [(Y_1^2/\nu)^2 \cdot (1 + Y_1^2/\nu)^{-2}] = \frac{3}{(\nu + 3)(\nu + 1)}.$$

References

- [1] M. Abramowitz and I. Stegun. Handbook of mathematical functions. National Bureau of Standards, 1972.
- [2] M. Amleh, A. Al-Natoor, B. Abughazaleh. An alternative method for estimating the parameters of log-Cauchy distribution, WSEAS Transactions on Mathematics, 22, 143–149, 2023.
- [3] A. Brouste, C. Dutang and D. Noutsia Mieniedou. OneStep - Le Cam's one-step estimation procedure, R Journal, 13(1), 366–377, 2021.
- [4] A. Brouste and C. Farinetto. Fast and asymptotically efficient estimation in the Hawkes processes, Japanese Journal of Statistics and Data Science, 6, 361–379, 2023.

- [5] F. Caeiro, M. Gomes and D. Pestana. Direct reduction of bias of the classical Hill estimator, *Revstat*, 3(2), 113–136, 2005.
- [6] A. Dabye, A. Gounoung and Y. Kutoyants. Method of moments estimators and multi-step MLE for Poisson processes. *Journal of Contemporary Mathematical Analysis*, 53(4), 187–196, 2018.
- [7] L. de Haan and A. Ferreira. *Extreme Value Theory*, Springer, 2006.
- [8] S. Girard, G. Stupfler and A. Usseglio-Carleve. On automatic bias reduction for extreme expectile estimation, *Statistics and Computing*, 32–64, 2022.
- [9] Y. Jang, J. Zhao, H. Kim, K. Yu, S. Kwon and S. Kim. New closed-form efficient estimator for multivariate gamma distribution, *Statistica Neerlandica*, 1–18, 2023
- [10] H. Kim and J. Zhao. New efficient estimators for the Weibull distribution, *Communication in Statistics - Theory and Methods*, 2023.
- [11] Y. Kutoyants and A. Motrunich. On multi-step MLE-process for Markov sequences, *Metrika*, 79, 705–724, 2016.
- [12] K. Lange, R. Little and J. Taylor. Robust Statistical Modeling Using the t Distribution, *Journal of the American Statistical Association*, 84, 408, 881–896, 1989.
- [13] L. Le Cam. On the asymptotic theory of estimation and testing hypothesis, *Proceedings of the 3rd Berkeley Symposium*, 1, 355–368, 1956.
- [14] T. Lumley. Fast Generalized Linear models by Database Sampling and one-step polishing, *Journal of Computational and Graphical Statistics*, 28(4), 1007–1010, 2019.
- [15] R Core Team. *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, <https://www.R-project.org>, 2023.