OneStep : Le Cam's One-step Estimation Procedure

by Alexandre Brouste, Christophe Dutang and Darel Noutsa Mieniedou

Abstract The **OneStep** package proposes principally an eponymic function that numerically computes Le Cam's one-step estimator, which is asymptotically efficient and can be computed faster than the maximum likelihood estimator for large datasets. Monte Carlo simulations are carried out for several examples (discrete and continuous probability distributions) in order to exhibit the performance of Le Cam's one-step estimation procedure in terms of efficiency and computational cost on observation samples of finite size.

1 Introduction

In the statistical experiments generated by i.i.d. observation samples, the sequence of maximum likelihood estimators (MLE) is known to be asymptotically efficient under very general assumptions and consequently presents the fastest convergence rate and the lowest possible asymptotic variance.

Although the sequence of MLE is asymptotically efficient, it is generally not expressed in a closed form and requires time consuming numerical computations. On the other hand, the other generic estimation procedures which can sometimes be computed faster, as the method of moments, do not generally reach the optimal asymptotic variance.

In R, the package **fitdistrplus** (see Delignette-Muller and Dutang (2015)) is commonly used to infer the parameters of univariate probability distributions. For non-censored datasets, **fitdistrplus** allows four different estimation methods: maximum likelihood, moment matching, quantile matching, and maximum goodness-of-fit estimation. The unified approach is provided by `fitdist()` which returns an S3-object having usual generic functions (`plot`, `summary`, `logLik`, ...) as well as dedicated `fitdist`-functions (`gofstat`, `bootdist`, ...).

Other packages with similar purposes are **EstimationTools** and **DistributionFitR**, providing MLE for non-censored data. Many other packages also provide estimation procedure on the basis of one function per probability distribution; see, e.g., packages **univariateML**, **propagate**. A final package, which is worth mentioning, is **fitter**, which fits a set of probability distributions on a given dataset sequentially.

However, as soon as the Fisher information matrix is sufficiently regular with respect to the parameter to be estimated, Le Cam's one-step estimation procedures can be achieved (see Le Cam (1956)). They are based on an initial sequence of guess estimators and a single Newton step or a Fisher scoring step on the loglikelihood function. Namely, they write

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

for the Fisher scoring type procedure, where $(\vartheta_n^*, n \geq 1)$ is the initial sequence of guess estimators, $\ell(\vartheta)$ is the loglikelihood function, the dot is the notation for the gradient with respect to ϑ, and $\mathcal{I}(\vartheta) = -\mathbf{E}_\vartheta(\ddot{\ell}(\vartheta))$ is the Fisher information matrix and

$$\overline{\vartheta}_n = \vartheta_n^* + \widehat{\mathcal{I}}_n(\vartheta_n^*)^{-1} \cdot \frac{1}{n} \sum_{j=1}^{n} \dot{\ell}(\vartheta_n^*, X_j), \quad n \geq 1, \tag{2}$$

for the Newton type procedure, where $\widehat{\mathcal{I}}_n(\vartheta) = -\frac{1}{n} \sum_{j=1}^{n} \ddot{\ell}(\vartheta, X_j)$ is the opposite of the Hessian for the loglikelihood function.

The sequence of Le Cam's one-step estimators presents certain advantages over the sequence of MLE and over the initial sequence of estimators (method of moments, quantile matching method, etc.) in terms of computational cost and asymptotic variance. It is much less computationally expensive than the MLE, while it has the same rate and the same asymptotic variance. Since there is no full numerical optimization (but only one computation of the Newton step or the Fisher scoring step), the procedure is faster and appropriate for very large datasets. On the other hand, it is asymptotically optimal in terms of asymptotic variance, which is generally not the case for the initial sequence of guess estimators.

For several probability distributions (Gaussian, Exponential, Lognormal, Poisson, Geometric), a fast computable sequence of estimators is already asymptotically efficient, and no correction is

needed. When the Fisher information is given in a closed form (Gamma, Beta, χ^2, Weibull, Pareto II, Cauchy), Le Cam's one-step estimation procedure (1) is executed. In all the other cases, the estimation procedure (2) is used either with the score function and the Hessian in a closed form (Negative Binomial) or with the numerical approximation of the score function and the Hessian with the package **numDeriv** (see Gilbert and Varadhan (2019)).

Le Cam's one-step estimation procedures are implemented in the package **OneStep**, and the eponymic function **onestep** is described in this paper and used on different examples. Monte Carlo simulations are performed for several examples in order to exhibit the performance of Le Cam's one-step procedure on samples of finite size. They exhibit their asymptotic efficiency simultaneously with their better computational cost for several probability distributions.

2   The function onestep

Let $(X_1, X_2, \ldots, X_n)$ be a sample of i.i.d. random variables. The probability density function of X_1 is denoted f and depends on an unknown parameter $\vartheta \in \Theta \subset \mathbb{R}^p$ which is to be estimated.

Le Cam's one-step estimation procedure is based on an initial sequence of guess estimators and a Fisher scoring step or a single Newton step on the loglikelihood function. For the novice user, the function **onestep** automatically chooses the best procedure to be used. There is, consecutively, a single command for the user, which is

```
onestep(data,distr)
```

The function **onestep** presents several procedures internally depending on whether the initial sequence of guess estimators is in a closed-form or not and whether the score and the Fisher information matrix can be elicited in a closed-form.

Here is the list of the procedures taken into account in the **onestep** function:

1. Distributions for which the MLE is already explicit: `norm`, `exp`, `lnorm`, `invgauss` for continuous probability distributions, and `pois`, `geom` for discrete probability distributions. For this class, the explicit MLE is returned.

2. Distributions for which the initial sequence of guess estimators, the score and the Fisher information matrix have been elicited in a closed-form: `gamma`, `beta`, `chisq` with an initial sequence of moment estimators, `cauchy` with an initial sequence of quantile matching estimators, and `weibull` with an initial sequence of graphical plot estimators. For this class, Le Cam's one-step procedure (1) is applied.

3. Distributions for which the initial sequence of guess estimators, the score, and the Hessian have been elicited in a closed-form: `nbinom`. For this class, Le Cam's one-step procedure (2) is applied.

4. Distributions for which the initial sequence is numerically computed on a subsample, but the score and the Fisher information matrix have closed-form: `pareto`. For this class, Le Cam's one-step procedure (1) is executed.

5. For all other distributions, if the density function is well defined, the numerical computation of the Newton step in Le Cam's one-step procedure (2) is proposed with an initial sequence of guess estimators, which is the sequence of maximum likelihood estimators computed on a subsample.

As is described, all explicit distributions of the `mmedist` function from **fitdistrplus** have been corrected in the **onestep** function, except `unif` and `logis`. The example `unif` is a famous example of the singular behavior of the MLE. The correction of `logis` is of very small gain from the method of moments estimator to the MLE. For these two distributions, the method of moments is returned.

However, the package also offers several new explicit computations as `cauchy`, `chisq` and `weibull` and applies to distributions coming from the **actuar** package (see Dutang et al. (2008)) such as `invgauss` and `pareto`.

The function `onestep` allows the user to propose its own initial guess estimation in the one-step procedure by specifying the parameter `init` in the command

```
onestep(data,distr,init)
```

The user can consequently use different initial guess estimators for the aforementioned classical distributions (moment estimators, estimators based on the characteristic function, quantile matching estimators, Bayesian estimators, mode-type estimators, graphical methods, *etc.*). Several examples can be found in the documentation of the `onestep` function.

Monte Carlo simulations are done for several examples (discrete and continuous probability distributions) in order to exhibit the performance of Le Cam's one-step estimation procedure in terms of efficiency and computational cost on observation samples of finite size.

For the assessment of the efficiency, the proximity of the renormalized statistical errors $(Y_1, Y_2, \ldots, Y_M)$ to the centered Gaussian asymptotic distribution (for a coordinate in the multivariate setting) is evaluated with the Cramer-Von Mises statistic

$$T = M\omega_M^2 = M \int_{\mathbb{R}} \left(F_M(y) - F_*(y)\right)^2 dF_*(y) = \frac{1}{12M} + \sum_{i=1}^{M} \left(F_*(Y_{(i)}) - \frac{2i-1}{2M}\right)^2,$$

where the order statistics are denoted $Y_{(1)} \leq Y_{(2)} \leq \cdots \leq Y_{(M)}$, $F_*(\cdot)$ is the theoretical Gaussian asymptotic cumulative distribution function and

$$F_M(y) = \frac{1}{M} \sum_{i=1}^{M} \mathbb{1}_{\{Y_i \leq y\}}$$

is the empirical cumulative distribution function. The asymptotic distribution of T is tabulated, for instance, in the goftest package. Note that a value of the statistic below 0.7434 corresponds to accept the null (and the equivalence to theoretical Gaussian asymptotic distribution) with an error of type I equal to 1%.

Timing performance (given in seconds) is done with the proc.time function ("elapsed" time) on a laptop with an Intel Core i7 2.7 GHz processor with 8GB RAM. The onestep function is compared to the MLE computed with mledist of the **fitdistrplus** package and to the sequence of initial guess estimators (for instance, the moment estimator is computed with mmedist for the gamma, beta, and nbinom distributions).

The functions benchonestep and benchonestep.replicate were used to compare the performance of estimators (see documentation of the **OneStep** package).

Closed-form sequence of initial guess estimators

For the majority of the "closed formula" cases, the initial sequence of guess estimators of the unknown parameter ϑ is the sequence of moment estimators.

Let $(X_1, X_2, \ldots, X_n)$ be a sample of i.i.d. random variables. Let us denote the theoretical moments $m_k(\vartheta) = \mathbf{E}_\vartheta(X_1^k)$ and the empirical moments $\tilde{m}_k = \frac{1}{n} \sum_{j=1}^{n} X_j^k$. The sequence of moment estimators (ME) is generally defined as the solution of the system of equations

$$m_k(\vartheta_n^*) = \tilde{m}_k, \quad k = 1, \ldots, p.$$

Indeed, under very mild conditions, the sequence of moment estimators $(\vartheta_n^*, n \geq 1)$ is asymptotically normal, and therefore, $\sqrt{n}$-consistent (see Ibragimov and Has'minskii (1981)). Namely,

$$\sqrt{n}\left(\vartheta_n^* - \vartheta\right) \longrightarrow \mathcal{N}\left(0, J^{-1}(\vartheta) A(\vartheta) J^{-T}(\vartheta)\right) \tag{3}$$

in law as $n \to \infty$, where

$$J(\vartheta) = \left(\frac{\partial}{\partial \vartheta_j} m_i(\vartheta)\right)_{1 \leq i \leq p, 1 \leq j \leq p}$$

and

$$A(\vartheta) = \mathbf{E}_\vartheta\left((X_1^i - m_i(\vartheta))(X_1^j - m_j(\vartheta))\right)_{1 \leq i \leq p, 1 \leq j \leq p}.$$

Here, the notation M^T means the transpose matrix of M. Several examples (Gamma, Beta, and Negative Binomial) are given later on in this section.

However, other $\sqrt{n}$-consistent sequences of initial guess estimators can be used. For instance, an initial sequence of quantile matching estimators is employed for the Cauchy distribution.

It had been shown in Le Cam (1956) that for a $\sqrt{n}$–consistent initial sequence of guess estimators and an uniformly continuous Fisher information matrix, the sequence of Le Cam's one-step estimators defined in (1) or (2) is consistent, asymptotically normal and efficient (in the Fisher sense) with

$$\sqrt{n}\left(\widehat{\vartheta}_n - \vartheta\right) \to \mathcal{N}\left(0, \mathcal{I}(\vartheta)^{-1}\right).$$

In other words, for an initial sequence which is asymptotically rate but not variance efficient, the new sequence is asymptotically rate and variance efficient.

Gamma The first example is the joint estimation of the shape parameter α and scale parameter β in the statistical experiment generated by a sample $(X_1, X_2, \ldots, X_n)$ of i.i.d. Gamma random variables

whose probability density function is given by

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x), \quad x > 0.$$

Let us denote $\vartheta = (\alpha, \beta)$. In this statistical experiment, the sequence of maximum likelihood estimators $(\hat{\vartheta}_n, n \geq 1)$ of ϑ is not in a closed-form. The sequence of MLE satisfies

$$\sqrt{n}\left(\hat{\vartheta}_n - \vartheta\right) \to \mathcal{N}\left(0, \mathcal{I}(\vartheta)^{-1}\right),$$

where

$$\mathcal{I}(\vartheta) = \begin{pmatrix} \psi^{(2)}(\alpha) & -\frac{1}{\beta} \\ -\frac{1}{\beta} & \frac{\alpha}{\beta} \end{pmatrix}.$$

Here, $\psi^{(n)}$ is the polygamma functions (see (Abramowitz and Stegun, 1972, section 6.4.1, page 260)) defined by $\psi^{(n)}(\alpha) = \frac{\partial^n}{\partial \alpha^n} \log \Gamma(\alpha)$. Consider that

$$m_k(\vartheta) = \frac{\alpha(\alpha+1)\dots(\alpha+k-1)}{\beta^n}.$$

Consecultively, the sequence of moment estimators $(\vartheta_n^* = (\alpha_n^*, \beta_n^*), n \geq 1)$ given by

$$\alpha_n^* = \frac{\tilde{m}_1^2}{\tilde{m}_2 - \tilde{m}_1^2} \quad \text{and} \quad \beta_n^* = \frac{\tilde{m}_1}{\tilde{m}_2 - \tilde{m}_1^2}$$

is asymptotically normal (see (3)) with

$$J = \begin{pmatrix} \frac{1}{\beta} & -\frac{\alpha}{\beta^2} \\ \frac{1+2\alpha}{\beta^2} & -\frac{2\alpha(\alpha+1)}{\beta^3} \end{pmatrix} \quad \text{and} \quad A(\vartheta) = \begin{pmatrix} \frac{\alpha}{\beta^2} & \frac{2\alpha(\alpha+1)}{\beta^3} \\ \frac{2\alpha(\alpha+1)}{\beta^3} & \frac{\alpha(4\alpha+6)(\alpha+1)}{\beta^4} \end{pmatrix},$$

and consequently does not reach asymptotical efficiency.

We can see in Figure 1 (and in Table 1 with the Cramer-Von Mises statistics) that the sequence of Le Cam's one-step estimators (LCE) reaches efficiency and naturally overperforms the initial sequence of ME in terms of asymptotic variance. Moreover, this sequence is faster to be computed on $M = 10000$ Monte Carlo simulations than the sequence of MLE, as shown in Table 1, displaying total computation times over M.

	MLE	ME	LCE
Computation time (s)	1559.74	29.72	37.46
CvM statistic alpha	0.63	1.04	0.23
CvM statistic beta	0.47	0.86	0.19

Table 1: Computation time and CvM statistics

As it was mentioned in the introduction, the major advantage of the sequence of one-step estimators is that it is computed faster than the maximum likelihood estimator for large datasets. We illustrate this fact in Table 2, where the average computation times (over 10 Monte Carlo simulations) are done for different sample sizes $n = 10^r, r = 3, \dots, 9$ for both MLE and LCE. LCE is between 20 times and 55 times faster than MLE, especially for large sample sizes when there is memory overload.

	10^3	10^4	10^5	10^6	10^7	10^8	10^9
MLE	0.0121	0.1193	1.0056	10.4735	99.4203	2339.0437	31600.8313
LCE	0.0007	0.0053	0.0252	0.2361	2.3152	40.0440	540.7393

Table 2: Average computation time (s)

Beta The second example is the joint estimation of the first shape parameter α and the second shape parameter β in the statistical experiment generated by a sample $(X_1, X_2, \dots, X_n)$ of i.i.d. Beta random variables whose probability density function is given by

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \quad x \in [0,1].$$
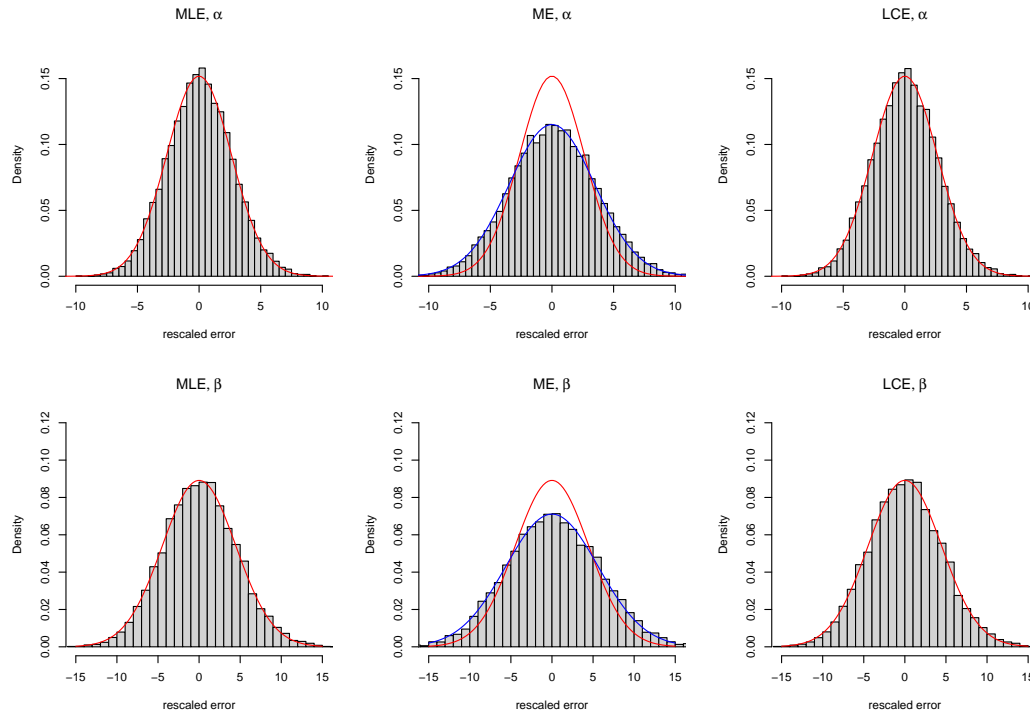
Figure 1: Histograms for the $M = 10000$ Monte Carlo simulations of the rescaled statistical error of the MLE, ME, and LCE for the Gamma distribution with $(\alpha, \beta) = (2, 3)$ and $n = 10000$. Superimposed red and blue lines are the theoretical centered Gaussian asymptotic distributions of the MLE and the ME, respectively.

Let us denote $\vartheta = (\alpha, \beta)$. In this statistical experiment, the sequence of maximum likelihood estimators $(\widehat{\vartheta}_n, n \geq 1)$ of ϑ is not in a closed-form. The sequence of MLE satisfies

$$\sqrt{n}\left(\widehat{\vartheta}_n - \vartheta\right) \to \mathcal{N}\left(0, \mathcal{I}(\vartheta)^{-1}\right),$$

where

$$\mathcal{I}(\vartheta) = \begin{pmatrix} \psi^{(2)}(\alpha) - \psi^{(2)}(\alpha + \beta) & -\psi^{(2)}(\alpha + \beta) \\ -\psi^{(2)}(\alpha + \beta) & \psi^{(2)}(\beta) - \psi^{(2)}(\alpha + \beta) \end{pmatrix}.$$

It is worth mentioning that for the Beta distribution

$$m_k(\vartheta) = \frac{\Gamma(\alpha + k)\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\alpha + \beta + k)}$$

that allows one to build closed-form moment estimators. Namely, denoting $\tilde{w}_n = \frac{\tilde{m}_1(1 - \tilde{m}_1)}{\tilde{m}_2 - \tilde{m}_1^2} - 1$, we obtain

$$\alpha_n^* = \tilde{m}_1 \tilde{w}_n \quad \text{and} \quad \beta_n^* = (1 - \tilde{m}_1)\tilde{w}_n.$$

Asymptotic variance in (3) of the sequence of moment estimators $(\vartheta_n^* = (\alpha_n^*, \beta_n^*), n \geq 1)$ can also be computed with

$$J = \begin{pmatrix} \frac{\beta}{(\alpha+\beta)^2} & -\frac{\alpha}{(\alpha+\beta)^2} \\ \frac{2\alpha^2\beta + 2\alpha\beta^2 + 2\alpha\beta + \beta^2 + \beta}{(\alpha+\beta)^2(\alpha+\beta+1)^2} & -\frac{\alpha(\alpha+1)(2\alpha+2\beta+1)}{(\alpha+\beta)^2(\alpha+\beta+1)^2} \end{pmatrix}$$

and

$$A(\vartheta) = \begin{pmatrix} \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} & \frac{2\alpha\beta(\alpha+1)}{(\alpha+\beta)^2(\alpha+\beta+1)(\alpha+\beta+2)} \\ \frac{2\alpha\beta(\alpha+1)}{(\alpha+\beta)^2(\alpha+\beta+1)(\alpha+\beta+2)} & \frac{\alpha(\alpha+1)(2\alpha^3 + 6\alpha^2\beta + 4\alpha\beta^2 + 14\alpha\beta + 4\alpha^2 + 4\alpha + 6\beta^2 + 6\beta)}{(\alpha+\beta)^2(\alpha+\beta+1)(\alpha+\beta+2)(\alpha+\beta+3)} \end{pmatrix},$$

and consequently the sequence does not reach asymptotical efficiency.

Here again, the sequence of Le Cam's one-step estimators naturally overperforms the initial sequence of ME in terms of asymptotic variance (see Figure 2 and the next Table for CvM statistics).

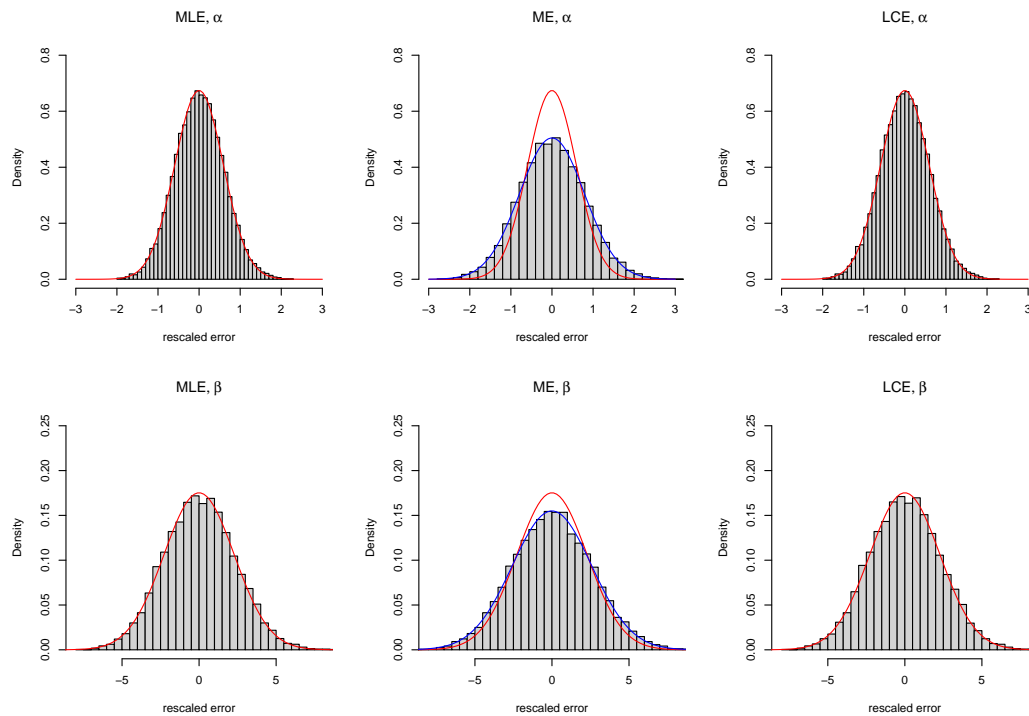It is computed faster than the sequence of MLE, as shown in Table 3.

Figure 2: Histograms for the $M = 10000$ Monte Carlo simulations of the rescaled statistical error of the MLE, ME, and LCE for the Beta distribution with $(\alpha, \beta) = (0.5, 1.5)$ and $n = 10000$. Superimposed red and blue lines are the theoretical centered Gaussian asymptotic distributions of the MLE and the ME, respectively.

	MLE	ME	LCE
Computation time (s)	2324.99	43.26	52.02
CvM statistic alpha	0.22	0.06	0.05
CvM statistic beta	0.34	0.11	0.20

Table 3: Computation time and CvM statistics

Cauchy The third example is the joint estimation of the location parameter $m \in \mathbb{R}$ and the scale parameter $d > 0$ in the statistical experiment generated by a sample $(X_1, X_2, \ldots, X_n)$ of i.i.d. Cauchy random variables whose probability density function is given by

$$f(x) = \frac{d}{\pi \left(d^2 + (x - m)^2 \right)}, \quad x \in \mathbb{R}. \tag{4}$$

Let us denote $\vartheta = (m, d)$. In this statistical experiment, the sequence of maximum likelihood estimators $(\widehat{\vartheta}_n, n \geq 1)$ of ϑ is not in a closed-form. The sequence of MLE satisfies

$$\sqrt{n} \left(\widehat{\vartheta}_n - \vartheta \right) \to \mathcal{N} \left(0, \mathcal{I}(\vartheta)^{-1} \right),$$

where

$$\mathcal{I}(\vartheta) = \begin{pmatrix} \frac{1}{2d^2} & 0 \\ 0 & \frac{1}{2d^2} \end{pmatrix}.$$

It is worth mentioning that the Cauchy distribution has no first moment, and consecutively its unknown parameter ϑ cannot be estimated via the classical method of moments. But a quantile matching method allows one to define

$$m_n^* = Q_n \left(\frac{1}{2} \right) \quad \text{and} \quad d_n^* = \frac{1}{2} \left(Q_n \left(\frac{3}{4} \right) - Q_n \left(\frac{1}{4} \right) \right), \tag{5}$$

where the sample quantile is usually computed for $p \in (0, 1)$ as

$$Q_n(p) = X_{(\lfloor np \rfloor)},$$

with the order statistics denoted $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$. This sequence of estimators ($\vartheta_n^* = (m_n^*, d_n^*), n \geq 1$) can be shown to be consistent and asymptotically normal, namely

$$\sqrt{n} \left(\vartheta_n^* - \vartheta \right) \longrightarrow \mathcal{N} \left(0, \Gamma_1 \right),$$

where

$$\Gamma_1 = \begin{pmatrix} \frac{1}{4f(m)^2} & 0 \\ 0 & \frac{1}{16f(q_1)^2} \end{pmatrix}. \tag{6}$$

Here q_1 is the theoretical first quartile. Consequently, this estimator does not reach the efficient asymptotic variance.

As mentioned previously, the sequence of Le Cam's one-step estimators overperforms the initial sequence of quantile matching estimators (QME) in terms of asymptotic variance (see Figure 3 and the next Table for CvM statistics).
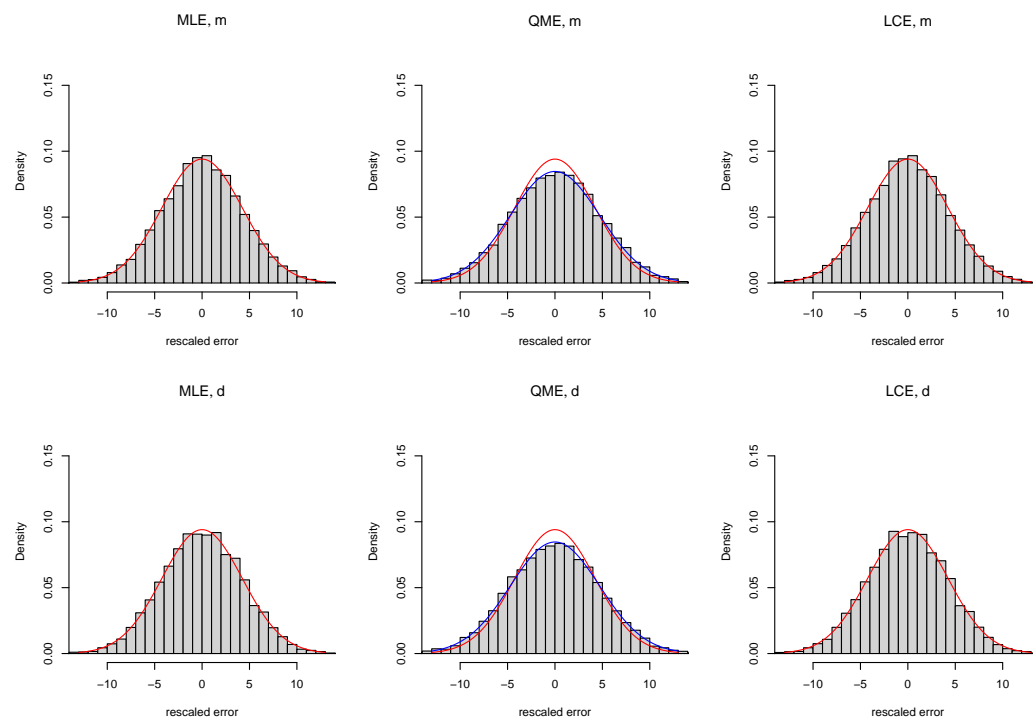

Figure 3: Histograms for the $M = 10000$ Monte Carlo simulations of the rescaled statistical error of the MLE, QME, and LCE for the Cauchy distribution with $(m, d) = (2, 3)$ and $n = 10000$. Superimposed red and blue lines are the theoretical centered Gaussian asymptotic distributions of the MLE and the QME, respectively.

It is computed faster than the sequence of MLE, as shown in Table 4.

	MLE	QME	LCE
Computation time (s)	225.32	7.40	27.59
CvM statistic m	0.08	0.42	0.08
CvM statistic d	0.04	0.10	0.04

Table 4: Computation time and CvM statistics

Pólya (negative binomial) The fourth example is the joint estimation of the size parameter r and the mean parameter μ in the statistical experiment generated by a sample $(X_1, X_2, \ldots, X_n)$ of i.i.d.

negative binomial random variables whose (discrete) probability density function is given by

$$f(x) = \frac{\Gamma(r+x)}{\Gamma(r)x!} \left(\frac{r}{\mu+r}\right)^r \left(\frac{\mu}{\mu+r}\right)^x, \quad x \in \mathbb{N}.$$

Let us denote $\vartheta = (r, \mu)$. In this statistical experiment, the sequence of maximum likelihood estimators $(\widehat{\vartheta}_n, n \geq 1)$ of ϑ is not in a closed-form and the Fisher information matrix neither.

For this distribution,

$$m_1(\vartheta) = \mu \quad \text{and} \quad m_2(\vartheta) = \mu^2 + \mu + \frac{\mu^2}{r}$$

that gives closed-form sequence of moment estimators $(\vartheta_n^* = (r_n^*, \mu_n^*), n \geq 1)$ given by

$$r_n^* = \frac{\tilde{m}_1^2}{(\tilde{m}_2 - \tilde{m}_1^2) - \tilde{m}_1} \quad \text{and} \quad \mu_n^* = \tilde{m}_1.$$

In this discrete case, the sequence of Le Cam's one-step estimators (2) defined with the Hessian still overperforms the initial sequence of ME in terms of asymptotic variance (see Figure 4 and the next Table for CvM statistics). Let us mention that the CvM statistics are computed with the estimated variances for MLE and LCE.
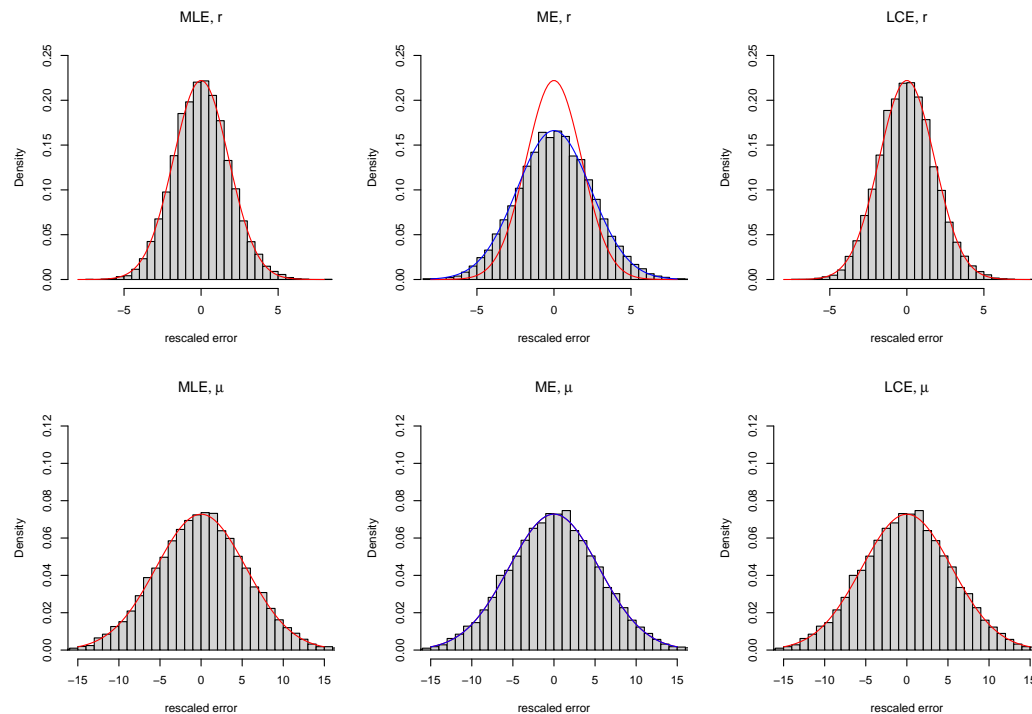

Figure 4: Histograms for the $M = 10000$ Monte Carlo simulations of the rescaled statistical error of the MLE, ME, and LCE for the Polya distribution with $(r, \mu) = (1, 5)$ and $n = 10000$. Superimposed red line is the empirical Gaussian asymptotic distributions of the MLE. Superimposed blue line is the theoretical centered Gaussian asymptotic distributions of the ME.

It also overperforms the sequence of MLE in terms of computation time, as shown in Table 5.

	MLE	ME	LCE
Computation time (s)	2038.33	39.59	104.75
CvM statistic r	0.18	0.71	0.19
CvM statistic mu	0.05	0.05	0.05

Table 5: Computation time and CvM statistics

Weibull The fifth example is the joint estimation of the shape parameter $\tau > 0$ and the rate parameter $\beta > 0$ in the statistical experiment generated by a sample $(X_1, X_2, \ldots, X_n)$ of i.i.d. Weibull random

variables whose probability density function is given by

$$f(x) = \beta\tau(\beta x)^{\tau-1} \exp\left(-(\beta x)^\tau\right), \quad x \in \mathbb{R}_*^+.$$

Let us denote $\vartheta = (\tau, \beta)$. In this example, neither the ME nor the MLE is in closed-form. However, the score and the Fisher information matrix can be explicitly computed. For instance,

$$\mathcal{I}(\vartheta) = \begin{pmatrix} \frac{1}{\tau^2}\left(\psi^{(2)}(1) + [\psi^{(1)}(2)]^2\right) & \frac{1}{\beta}\psi^{(1)}(2) \\ \frac{1}{\beta}\psi^{(1)}(2) & \frac{\tau^2}{\beta^2} \end{pmatrix}.$$

Moreover, it is possible to define a closed-form initial sequence of graphical plot estimators (based on the explicit form of the cumulative distribution function). Let us denote $x_i = \log(-\log(1 - i/(n+1)))$ and $Z_i = \log\left(X_{(i)}\right)$ for $i = 1, \dots, n$. The sequence of ordinary least square (OLS) based estimators $(\vartheta_n^* = (\tau_n^*, \beta_n^*), n \geq 1)$ is defined by

$$\tau_n^* = \left(\frac{\sum\limits_{i=1}^n (x_i - \bar{x}_n)Z_i}{\sum\limits_{i=1}^n (x_i - \bar{x}_n)^2}\right)^{-1} \quad \text{and} \quad \beta_n^* = \exp\left(\bar{Z}_n - (\tau_n^*)^{-1}\bar{x}_n\right),$$

where $\bar{Z}_n$ stands for the average of $Z_1, \dots, Z_n$.

The sequence of Le Cam's one-step estimators overperforms the initial sequence of OLS-based estimators in terms of asymptotic variance (see Figure 5 and the next Table). Since OLS presents a small bias for samples of finite size, its CvM statistics are bigger in the next Table.
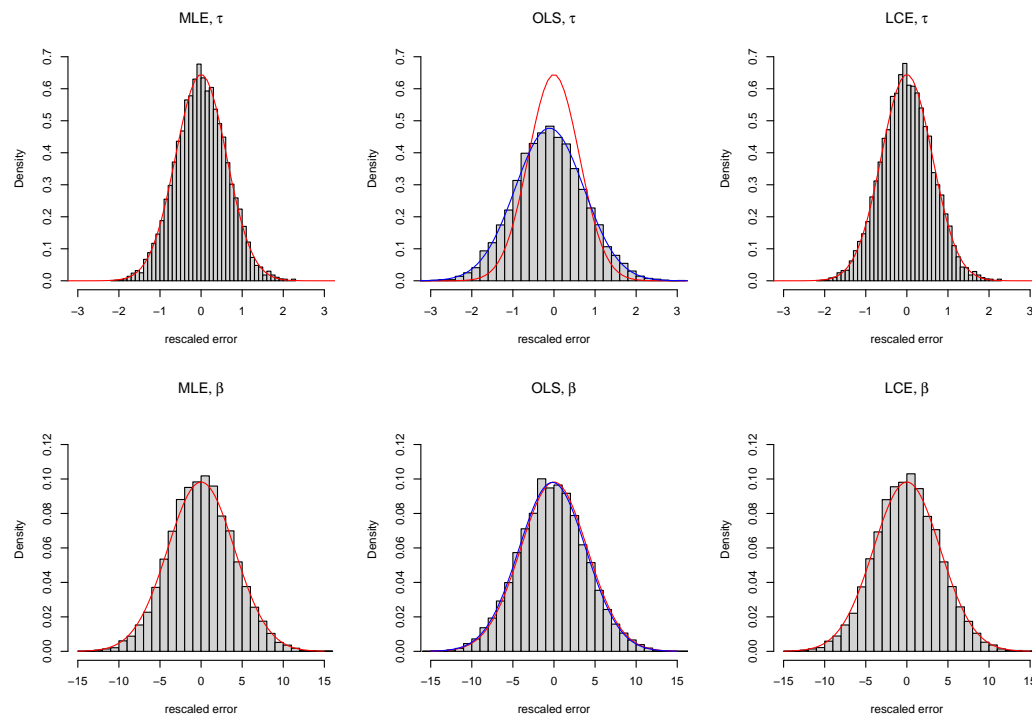

Figure 5: Histograms for the $M = 10000$ Monte Carlo simulations of the rescaled statistical error of the MLE, OLS, and LCE for the Weibull distribution with $(\tau, \beta) = (0.8, 3)$ and $n = 10000$. Superimposed red line is the theoretical centered Gaussian asymptotic distributions of the MLE. Superimposed blue line is the empirical Gaussian asymptotic distributions of the OLS.

It also overperforms the sequence of MLE in terms of computation time for large datasets, as shown in Table 6.

	MLE	OLS	LCE
Computation time (s)	568.83	32.64	82.64
CvM statistic tau	0.13	17.26	0.06
CvM statistic beta	0.59	2.33	0.59

Table 6: Computation time and CvM statistics

Numerically computed initial sequence of guess estimators and numerical computations in the generic case

As soon as the ME is not in closed form or no other closed form estimators can be elicited, the use of a numerical ME as an initial sequence of guess estimator in Le Cam's one-step procedure is difficult to justify. Indeed, no gain will be given in terms of time computation and the use of the MLE is finally equivalent. We propose therefore to use in the **onestep** function an improvement of the classical Le Cam procedure.

Indeed, it can be shown (see Kamatani and Uchida (2015) or Kutoyants and Motrunich (2016)) that for a $n^{\delta/2}$–consistent initial sequence of guess estimators (with $\frac{1}{2} < \delta \leq 1$) and a Lipshitz Fisher information matrix, the sequence of Le Cam's one-step estimators is also consistent, asymptotically normal and efficient (in the Fisher sense). In this setting, for a initial sequence which is neither asymptotically rate nor variance efficient, the new sequence is asymptotically rate and variance efficient.

This result allows one to use the numerical computation of the MLE on a subsample (of size n^{δ}, for $\frac{1}{2} < \delta \leq 1$) as an initial sequence of guess estimators. Namely, for this initial sequence of guess estimators $(\vartheta_n^*, n \geq 1)$,

$$n^{\frac{\delta}{2}} (\vartheta_n^* - \vartheta) \to \mathcal{N} \left(0, \mathcal{I}(\vartheta)^{-1} \right).$$

Then the sequence of Le Cam's one-step estimators given by (2) is also consistent, asymptotically normal and efficient (in the Fisher sense) with

$$\sqrt{n} \left(\widehat{\vartheta}_n - \vartheta \right) \to \mathcal{N} \left(0, \mathcal{I}(\vartheta)^{-1} \right).$$

The choice of the exponent δ that measures the size of the subsample $\frac{1}{2} < \delta \leq 1$ is set to 0.9 by default and can be chosen by the user with the parameter control, for instance in the call

```
onestep(data, distr, control=list(delta=0.7))
```

The example of Pareto II distribution is interesting and shows the gain in terms of variance (with respect to the initial sequence of guess estimators) and in terms of computation time in comparison with the MLE.

This improved method is also used when the distribution does not belong to the closed-formula family. The initial sequence of maximum likelihood estimators is computed on a subsample with the mledist function of the **fitdistrplus** package. The score and the Hessian in the Newton step of the Le Cam procedure (2) are numerically computed with the functions grad and hessian of the **numDeriv** package (see Gilbert and Varadhan (2019)).

Pareto II The sixth example is the joint estimation of the shape parameter $\alpha > 0$ and the scale parameter $\sigma > 0$ in the statistical experiment generated by a sample $(X_1, X_2, \ldots, X_n)$ of i.i.d. Pareto II (Lomax) random variables whose probability density function is given by

$$f(x) = \frac{\alpha \sigma^{\alpha}}{(\sigma + x)^{\alpha+1}}, \quad x \in \mathbb{R}^+.$$

In this example neither the ME (when it exists for $\alpha > 2$) nor the MLE are in closed-form. But the score and the Fisher information matrix can be explicitly computed.

Let $\vartheta = (\alpha, \sigma)$. Then, considering the MLE computed on a subsample of size n^{δ}, $\frac{1}{2} < \delta \leq 1$, as an initial sequence of guess estimators $(\vartheta_n^*, n \geq 1)$, we get

$$n^{\frac{\delta}{2}} (\vartheta_n^* - \vartheta) \to \mathcal{N} \left(0, \mathcal{I}(\vartheta)^{-1} \right)$$

with

$$\mathcal{I}(\vartheta) = \begin{pmatrix} \frac{\alpha}{\sigma^2(\alpha+1)} & -\frac{1}{\sigma(\alpha+1)} \\ -\frac{1}{\sigma(\alpha+1)} & \frac{1}{\alpha^2} \end{pmatrix}.$$

The computation of the Fisher information matrix can be found in Brazauskas (2003).

Here again, the sequence of Le Cam's one-step estimators naturally overperforms the initial sequence of MLE computed on a subsample in terms of asymptotic variance (see Figure 6 and the next Table for the CvM statistics).
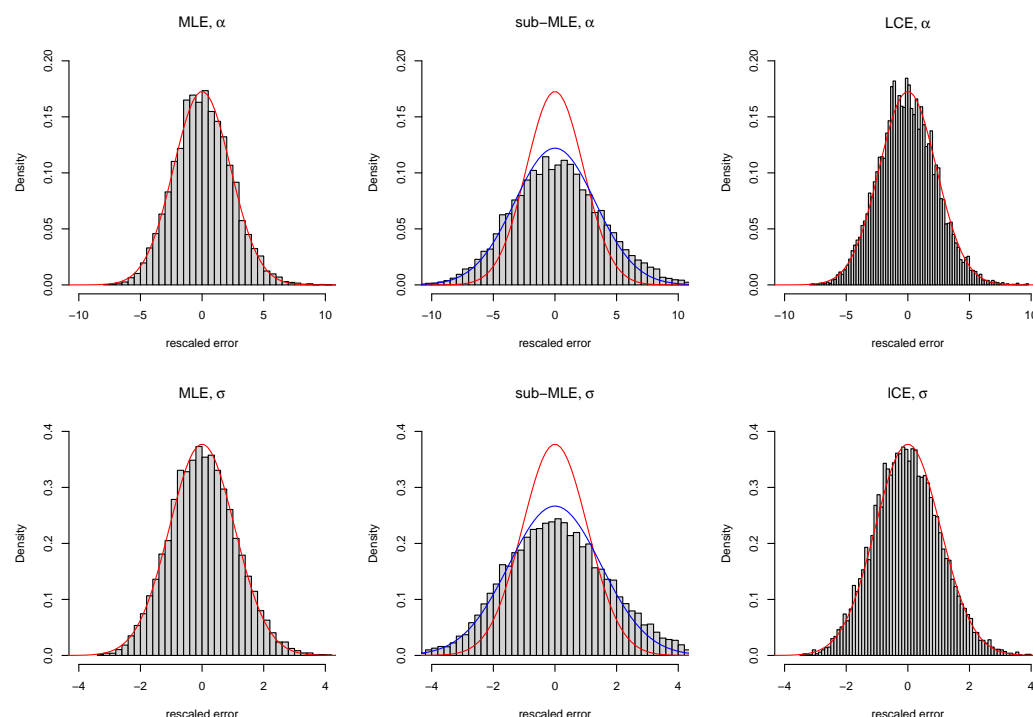

Figure 6: Histograms for the $M = 10000$ Monte Carlo simulations of the rescaled statistical error of the MLE, ME, and LCE for the Pareto II distribution with $(\alpha, \sigma) = (1.1, 0.3)$ and $n = 10000$. Superimposed red and blue lines are the theoretical centered Gaussian asymptotic distributions of the MLE and the MLE on a subsample, respectively.

It is also computed faster than the sequence of MLE, as shown in Table 7.

	MLE	MLEsub	LCE
Computation time (s)	1013.00	393.42	414.08
CvM statistic alpha	0.25	4.72	1.44
CvM statistic sigma	0.43	5.54	1.42

Table 7: Computation time and CvM statistics

Generic function For the generic example, we study the Weibull distribution again and force it to be numerically computed with the function parameter method="numeric" in order not to use the closed form processing.

We recall that, in the generic procedure, the score and the Hessian in the Newton step of the Le Cam procedure (2) are computed with the functions grad and hessian of the **numDeriv** package. By default, the initial sequence of maximum likelihood estimators is computed on a subsample if size n^δ with $\delta = 0.9$.

The numerically computed sequence of Le Cam's one-step estimators reaches asymptotic efficiency for a simulation of $M = 10000$ Monte-Carlo replications of samples of size $n = 10000$ as shown by the CvM statistics summarized in the next table. It is still computed faster than the sequence of MLE, see Table 8.

	MLE	LCE	GLCE
Computation time (s)	568.83	82.64	405.10
CvM statistic tau	0.13	0.06	0.25
CvM statistic beta	0.59	0.59	1.08

Table 8: Computation time and CvM statistics

paper.

Bibliography

M. Abramowitz and I. Stegun. Handbook of mathematical functions. *National Bureau of Standards*, 1972. [p383]

V. Brazauskas. Information matrix for Pareto (IV), Burr, and related distributions. *Communications in Statistics – Theory and Methods*, 32(2):315–325, 2003. URL https://doi.org/10.1081/sta-120018188. [p390]

M. L. Delignette-Muller and C. Dutang. fitdistrplus: An R package for fitting distributions. *Journal of Statistical Software*, 64(4):1–34, 2015. URL http://www.jstatsoft.org/v64/i04/. [p380]

C. Dutang, V. Goulet, and M. Pigeon. actuar: An R package for actuarial science. *Journal of Statistical Software*, 25(7):38, 2008. URL http://www.jstatsoft.org/v25/i07. [p381]

P. Gilbert and R. Varadhan. *numDeriv: Accurate Numerical Derivatives*, 2019. URL https://CRAN.R-project.org/package=numDeriv. R package version 2016.8-1.1. [p381, 389]

I. Ibragimov and R. Has'minskii. *Statistical Estimation - Asymptotic Theory*. Springer-Verlag, 1981. URL https://10.1007/978-1-4899-0027-2. [p382]

K. Kamatani and M. Uchida. Hybrid multi-step estimators for stochastic differential equations based on sampled data. *Stat Inference Stoch Process*, 18(2):177–204, 2015. URL https://doi.org/10.1007/s11203-014-9107-4. [p389]

Y. Kutoyants and A. Motrunich. On multi-step MLE-process for Markov sequences. *Metrika*, 79: 705–724, 2016. URL https://doi.org/10.1007/s00184-015-0574-4. [p389]

L. Le Cam. On the asymptotic theory of estimation and testing hypotheses. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, pages 129–156, Berkeley, Calif., 1956. University of California Press. URL https://projecteuclid.org/euclid.bsmsp/1200501652. [p380, 382]

Alexandre Brouste
Laboratoire Manceau de Mathématiques, Le Mans Université
Avenue Olivier Messiaen, 72085 LE MANS
France
ORCID: 0000-0001-6719-7432
Alexandre.brouste@univ-lemans.fr

Christophe Dutang
CEREMADE, CNRS, Université Paris-Dauphine, Université PSL
Place du Maréchal de Lattre de Tassigny, 75016 PARIS
France
ORCID: 0000-0001-6732-1501
christophe.dutang@dauphine.psl.eu

Darel Noutsa Mieniedou
Laboratoire Manceau de Mathématiques, Le Mans Université
Avenue Olivier Messiaen, 72085 LE MANS
France