An Alternative Criterion to Likelihood for Parameter Estimation Accounting for Prior Information on Nuisance Parameter

Adel Mohammadpour[1] and Ali Mohammad-Djafari[2]

[1] Permanent address: Department of Statistics, Faculty of Mathematics & Computer Science, Amirkabir University of Technology, 424 Hafez Ave., 15914 Tehran, and School of intelligent Systems, IPM, Iran, adel@aut.ac.ir

[2] [1] Laboratoire des Signaux et Systèmes, Unité mixte de recherche 8506 (CNRS-Supélec-UPS) Supélec, Plateau de Moulon, 91192 Gif-sur-Yvette, France {mohammadpour,djafari}@lss.supelec.fr

In this paper we propose an alternative tool to marginal likelihood for parameter estimation when we want to account for some prior information on a nuisance parameter. This new criterion is obtained using the median in place of the mean when using a prior distribution on the nuisance parameter. We first give the precise definition of this new criterion and its properties and then present a few examples to show their differences.

1 Introduction

Assume that we are given an observation x with cumulative distribution function (cdf) $F_{X|\mathcal{V},\theta}(x|\nu,\theta)$ (or probability density function (pdf) $f_{X|\mathcal{V},\theta}(x|\nu,\theta)$) with two unknown parameters ν and θ. We assume that ν is a nuisance parameter on which we have an *a priori* information translated by a prior distribution $F_{\mathcal{V}}(\nu)$ (or a pdf $f_{\mathcal{V}}(\nu)$) and we want to infer on θ.

If ν was given, i.e. $\nu = \nu_0$, then the classical Maximum Likelihood (ML) estimate of θ is defined as the optimizer of the likelihood function

$$l(\theta) = f_{X|\nu,\theta}(x|\nu_0,\theta).$$

The question is now how to account for the prior $F_{\mathcal{V}}(\nu)$. Again the classical solution is to integrate out ν to obtain the marginal pdf

$$f_{X|\theta}(x|\theta) = \int f_{X|\mathcal{V},\theta}(x|\nu,\theta) f_{\mathcal{V}}(\nu)\,\mathrm{d}\nu$$

and then estimate θ by optimizing the likelihood function

$$l(\theta) = f_{X|\theta}(x|\theta).$$

In this work, we propose a new inference tool $\widetilde{F}_{X|\theta}(x|\theta)$ which can be used to do the same inference on θ. This new inference tool is deduced from the interpretation of $F_{X|\theta}(x|\theta)$ as the mean value $F_{X|\mathcal{V},\theta}(x|\nu,\theta)$ using the pdf of $f_{\mathcal{V}}(\nu)$. Now, if in place of the mean value we take the median we obtain this new inference tool $\widetilde{F}_{X|\theta}(x|\theta)$ which is defined as

$$\widetilde{F}_{X|\theta}(x|\theta) : \mathrm{P}\left(F_{X|\mathcal{V},\theta}(x|\mathcal{V},\theta) \leq \widetilde{F}_{X|\theta}(x|\theta)\right) = 1/2$$

and can be used in the same way to estimate θ by optimizing

$$\widetilde{l}(\theta) = \widetilde{f}_{X|\theta}(x|\theta),$$

where $\widetilde{f}_{X|\theta}(x|\theta)$ is the pdf corresponding to the cdf $\widetilde{F}_{X|\theta}(x|\theta)$.

As far as the authors know, this alternative tool is newly presented [1, 2] and applied for hypothesis testing. In this paper we consider its use for parameter estimation.

In the following, first we give more precise definition of $\widetilde{F}_{X|\theta}(x|\theta)$. Then we present some of its properties, for example, we show that under some conditions $\widetilde{F}_{X|\theta}(x|\theta)$ has all the properties of a cdf, its calculation is very simple and is robust relative to the prior distribution. Then, we give a few examples and finally, we compare the relative performances of these two tools for estimating of θ.

2 A New Inference Tool

Hereafter in this section to simplify the notations we omit the parameter θ.

Definition 1. *Let X have a cdf depending on random parameter $\mathcal{V}$ with pdf $f_{\mathcal{V}}(\nu)$. The marginal cdf of X based on median, $\widetilde{F}_X(x)$, is defined as the median of $F_{X|\mathcal{V}}(x|\nu)$ over $f_{\mathcal{V}}(\nu)$.*

To simplify calculations of $\widetilde{F}_X(x)$, we use definition of median in statistics. That is we calculate $\widetilde{F}_X(x)$ by solving the following equation

$$F_{F_{X|\mathcal{V}}(x|\mathcal{V})}(\widetilde{F}_X(x)) = \frac{1}{2}, \quad \text{or equivalently} \quad P(F_{X|\mathcal{V}}(x|\mathcal{V}) \leq \widetilde{F}_X(x)) = \frac{1}{2}. \quad (1)$$

Theorem 1. *Let $\widetilde{F}_X(x)$ be the function defined in (1).*

1. *$\widetilde{F}_X(x)$ is a non-decreasing function.*
2. *If $F_{X|\mathcal{V}}(x|\nu)$ and $F_{\mathcal{V}}(\nu)$ are continuous cdfs and the random variable $T = F_{X|\mathcal{V}}(x|\mathcal{V})$ has an increasing cdf (for all fixed x) then $\widetilde{F}_X(x)$ is a continuous function.*

3. $0 \le \widetilde{F}_X(x) \le 1$.

Proof: 1. Let $x_1 < x_2$. For $i = 1, 2$, take

$$k_i = \widetilde{F}_{X_i}(x_i) \quad \text{and} \quad Y_i = F_{X_i|\mathcal{V}}(x_i|\nu).$$

Then using 1 we have

$$P(Y_1 \le k_1) = P(Y_2 \le k_2) = \frac{1}{2}.$$

We also have

$$Y_1 \le Y_2.$$

Therefore,

$$P(Y_1 \le k_1) = P(Y_2 \le k_2) \le P(Y_1 \le k_2),$$

i.e. $k_1 \le k_2$ or equivalently $\widetilde{F}_X(x)$ is non-decreasing.

2. If $\widetilde{F}_X(x)$ is a non-decreasing function, then

$$\widetilde{F}_X(x_-) = \lim_{t \uparrow x} \widetilde{F}_X(t) \;\text{ and }\; \tilde{F}_X(x_+) = \lim_{t \downarrow x} \widetilde{F}_X(t)$$

exist and are *finite* (e.g. [3]).

Further, $F_{X|\mathcal{V}}(x|\nu)$ is continuous with respect to x, and so

$$P(F_{X|\nu}(x_-|\nu) \le \widetilde{F}_X(x_-)) = P(F_{X|\mathcal{V}}(x|\nu) \le \widetilde{F}_X(x_-)),$$

$$P(F_{X|\nu}(x_+|\nu) \le \widetilde{F}_X(x_+)) = P(F_{X|\mathcal{V}}(x|\nu) \le \widetilde{F}_X(x_+)).$$

And by (1) we have

$$\begin{aligned} P(F_{X|\mathcal{V}}(x|\nu) \le \widetilde{F}_X(x_-)) &= P(F_{X|\mathcal{V}}(x|\nu) \le \widetilde{F}_X(x)) \\ &= P(F_{X|\mathcal{V}}(x|\nu) \le \widetilde{F}_X(x_+)). \end{aligned} \tag{2}$$

If $Y = F_{X|\mathcal{V}}(x|\nu)$ has an increasing distribution function, then

$$\widetilde{F}_X(x_-) = \widetilde{F}_X(x) = \widetilde{F}_X(x_+)$$

and by (2) $\widetilde{F}_X(x)$ is *continuous.*

3. On the other hand $\widetilde{F}_X(x)$ is the median of Y, $0 \le Y \le 1$, and so

$$0 \le \widetilde{F}_X(x) \le 1.$$

Remark 1 *By part 1 of Theorem 1,* $\lim_{x \uparrow +\infty} \widetilde{F}_X(x)$ *and* $\lim_{x \downarrow -\infty} \widetilde{F}_X(x)$ *exist. Therefore* $\widetilde{F}_X(x)$ *is a continuous cdf if conditions of Theorem 1 hold and* $\lim_{x \downarrow -\infty} \widetilde{F}_X(x) = 0$, $\lim_{x \uparrow \infty} \widetilde{F}_X(x) = 1$.

Theorem 2. *Let $F_{X|\mathcal{V}}(x|\nu)$ and $F_{\mathcal{V}}(\nu)$ be continuous cdfs. If $L(\nu) = F_{X|\mathcal{V}}(x|\nu)$ is a monotone function with respect to ν, then $\widetilde{F}_X(x) = L(F_{\mathcal{V}}^{-1}(\frac{1}{2}))$.*

Proof By (1) we have,

$$
\begin{aligned}
& P(L(\mathcal{V}) \leq \widetilde{F}_X(x)) = \frac{1}{2} \\
\Leftrightarrow & \begin{cases} P(\mathcal{V} \leq L^{-1}(\widetilde{F}_X(x))) = \frac{1}{2} & \text{if } L \text{ is an increasing function} \\ P(\mathcal{V} \geq L^{-1}(\widetilde{F}_X(x))) = \frac{1}{2} & \text{if } L \text{ is a decreasing function} \end{cases} \\
\Leftrightarrow & \begin{cases} F_{\mathcal{V}}(L^{-1}(\widetilde{F}_X(x))) = \frac{1}{2} & \text{if } L \text{ is an increasing function} \\ 1 - F_{\mathcal{V}}(L^{-1}(\widetilde{F}_X(x))) = \frac{1}{2} & \text{if } L \text{ is a decreasing function} \end{cases} \\
\Leftrightarrow & F_{\mathcal{V}}(L^{-1}(\widetilde{F}_X(x))) = \frac{1}{2} \\
\Leftrightarrow & \widetilde{F}_X(x) = L(F_{\mathcal{V}}^{-1}(\frac{1}{2})),
\end{aligned}
$$

where the last inequality follows from non-decreasing property of the cdf of $\mathcal{V}$.

Remark 2 *Let conditions of Theorem 2 be hold. If $\widetilde{F}_X(x)$ is a cdf then $\widetilde{F}_X(x)$ belong to the family of distribution $F_{X|\mathcal{V}}(x|\nu)$, because $\widetilde{F}_X(x) = F_{X|\mathcal{V}}(x|F_{\mathcal{V}}^{-1}(\frac{1}{2}))$.*

Remark 3 *$\widetilde{F}_X(x)$ is depend on the median of prior distribution, $F_{\mathcal{V}}^{-1}(\frac{1}{2})$, (but for calculating the marginal distribution of X we need the prior). Therefore $\widetilde{F}_X(x)$ is robust relative to prior distributions with the same medians.*

In the following two theorems we show that some important families of cdfs have a monotone distribution function with respect to their parameters and so, calculating of $\widetilde{F}_X(x)$ is very easy by using Theorem 2.

Theorem 3. *Let $L(\nu) = F_{X|\mathcal{V}}(x|\nu)$ and $F_{X|\mathcal{V}}(x|\nu)$ be an increasing function with respect to x. If ν is a location (scale) parameter then $L(\nu)$ is decreasing (monotone) with respect to ν.*

Theorem 4. *Let $X|\nu$ be distributed according to an exponential family with pdf*

$$f_{X|\mathcal{V}}(x|\nu) = h(x) \exp\left(\nu T(x) - A(\nu)\right),$$

where T and A are real functions. Then $L(\nu) = F_{X|\mathcal{V}}(x|\nu)$ is a monotone function with respect ν.

3 Examples

In this section we give a few simple examples to illustrate properties of the new inference tool. First we consider the exponential distribution with the following distribution function

$$F_{X|\mathcal{V},\theta}(x|\nu,\theta) = 1 - \exp\left(-\nu(x-\theta)\right),\ x > \theta,\ \theta > 0,\ f_{\mathcal{V}}(\nu) = \exp\left(-\nu\right),\ \nu > 0,$$

where $1/\nu$ is the scale parameter. We can use Theorem 3 for calculating $\widetilde{F}_{X|\theta}(x|\theta)$. By noting that, the median of $\mathcal{V}$ is $\ln 2$, we have

$$\widetilde{F}_{X|\theta}(x|\theta) = 1 - \exp\left(-\ln 2(x-\theta)\right),\ x > \theta,\ \theta > 0.$$

On the other hand we can calculate $F_{X|\theta}(x|\theta)$ by integrating over $f_{\mathcal{V}}(\nu)$ i.e.

$$F_{X|\theta}(x|\theta) = 1 - \frac{1}{(x-\theta)+1},\ \ x > \theta,\ \theta > 0.$$

Note that in this problem MLE based on $\widetilde{l}(\theta) = \widetilde{f}_{X|\theta}(x|\theta)$ and $l(\theta) = f_{X|\theta}(x|\theta)$ are the same.

The second example is the normal distribution with pdf

$$f_{X|\mathcal{V},\theta}(x|\nu,\theta) = \frac{1}{\sqrt{2\pi\theta}} \exp\left(-\frac{1}{2}(\frac{x-\nu}{\sqrt{\theta}})^2\right),\ \theta > 0,$$

where $\mathcal{V}$ has a standard normal distribution. In this case the median of $\mathcal{V}$ is zero and so (by using Theorem 3 or 4),

$$\widetilde{f}_{X|\theta}(x|\theta) = \frac{1}{\sqrt{2\pi\theta}} \exp\left(\frac{-x^2}{2\theta}\right),\ \theta > 0,$$

and also we can calculate

$$f_{X|\theta}(x|\theta) = \frac{1}{\sqrt{2\pi(\theta+1)}} \exp\left(\frac{-x^2}{2(\theta+1)}\right),\ \theta > 0.$$

The MLE based on $\widetilde{l}(\theta)$ is X^2 and base on $l(\theta)$ is $\max(X^2-1,0)$.

Figure 1 shows the estimated mean absolute error, $E_{X|\theta}(|\ \mathrm{MLE} - \theta\ |)$, of these estimators, computed using 200000 samples generated from $X \mid \mathcal{V},\theta \sim N(\mathcal{V},\theta)$ and $\mathcal{V} \sim N(0,1)$ for $\theta = [0.01 : 0.1 : 10]$.

4 Conclusion

We introduced an alternative inference tool $\widetilde{l}(\theta)$ to the marginal likelihood $l(\theta)$ for using prior information by defining a marginal function $\widetilde{F}_{X|\theta}(x|\theta)$ which is based on median in place of $F_{X|\theta}(x|\theta)$ which is the expected value of $\widetilde{F}_{X|\theta}(x|\theta)$ with respect to $f_{\mathcal{V}}(\nu)$. We proved that, based on a few conditions, $\widetilde{F}_{X|\theta}(x|\theta)$ is a cdf. Indeed, the computation of $\widetilde{l}(\theta)$ is easier than $l(\theta)$ for two important classes of distributions which are exponential and location-scale family of distributions. $\widetilde{l}(\theta)$ depends only on the median of $f_{\mathcal{V}}(\nu)$ and thus, it is robust with respect to prior distributions with the same median.

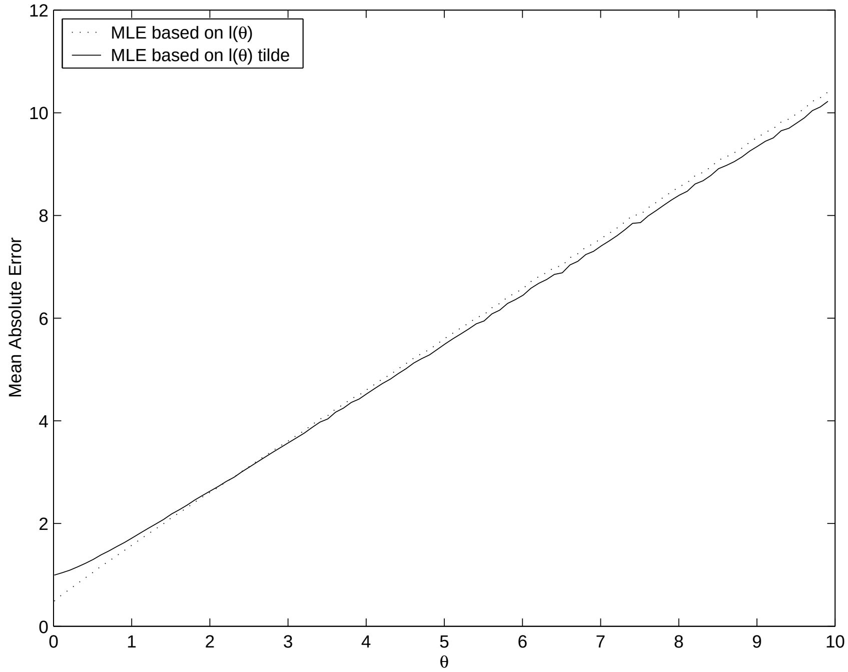

Fig. 1. Mean absolute errors for MLE based on $\widetilde{l}(\theta)$ and $l(\theta)$ i.e. X^2 and $\max(X^2 - 1, 0)$, respectively.

References

1. Mohammadpour A. (2003). Fuzzy Parameter and Its Application in Hypothesis Testing, Technical Report, School of Intelligent Systems, IPM, Tehran, Iran.
2. Mohammadpour A. and Mohammad-Djafari A. (2004). An alternative inference tool to total probability formula and its applications 2003, August, to appear in Proceeding of MAXENT23.
3. Rohatgi V.K. (1976). *An Introduction to Probability Theory and Mathematical Statistics*. John Wiley and Sons, New York.