Full paper

Inference with the Median of a Prior

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

[1]School of Intelligent Systems (IPM) and Amirkabir University of Technology (Dept. of Stat.), Tehran, Iran; E-mail: `adel@aut.ac.ir`
[2]LSS (CNRS-Supélec-Univ. Paris 11), Supélec, Plateau de Moulon, 91192 Gif-sur-Yvette, France
E-mail: `mohammadpour@lss.supelec.fr`, `djafari@lss.supelec.fr`

Abstract: We consider the problem of inference on one of the two parameters of a probability distribution when we have some prior information on a nuisance parameter. When a prior probability distribution on this nuisance parameter is given, the marginal distribution is the classical tool to account for it. If the prior distribution is not given, but we have partial knowledge such as a fixed number of moments, we can use the maximum entropy principle to assign a prior law and thus go back to the previous case. In this work, we consider the case where we only know the median of the prior and propose a new tool for this case. This new inference tool looks like a marginal distribution. It is obtained by first remarking that the marginal distribution can be considered as the mean value of the original distribution with respect to the prior probability law of the nuisance parameter, and then, by using the median in place of the mean.

Keywords: Nuisance parameter, maximum entropy, marginalization, incomplete knowledge.

MSC 2000 codes: 62F30

1   Introduction

We consider the problem of inference on a parameter of interest θ of a probability distribution when we have some prior information on a nuisance parameter ν from a finite number of samples of this probability distribution. Assume that we know the expressions of either the cumulative distribution function (cdf) $F_{\boldsymbol{X}|\mathcal{V},\theta}(\boldsymbol{x}|\nu,\theta)$ or its corresponding probability density function (pdf) $f_{\boldsymbol{X}|\mathcal{V},\theta}(\boldsymbol{x}|\nu,\theta)$, where $\boldsymbol{X} = (X_1,\cdots,X_n)'$ and $\boldsymbol{x} = (x_1,\cdots,x_n)'$. $\mathcal{V}$ is a random parameter on which we have an *a priori* information and θ is a fixed unknown parameter. This prior information can either be of the form of a prior cdf $F_{\mathcal{V}}(\nu)$ (or a pdf $f_{\mathcal{V}}(\nu)$) or, for example, only the knowledge of a finite number of its moments. In the first case, the marginal cdf

$$
\begin{aligned}
F_{\boldsymbol{X}|\theta}(\boldsymbol{x}|\theta) &= \int_{-\infty}^{+\infty} F_{\boldsymbol{X}|\mathcal{V},\theta}(\boldsymbol{x}|\nu,\theta) f_{\mathcal{V}}(\nu)\,\mathrm{d}\nu \\
&= E_{\mathcal{V}}\big(F_{\boldsymbol{X}|\mathcal{V},\theta}(\boldsymbol{x}|\mathcal{V},\theta)\big),
\end{aligned} \tag{1}
$$

is the classical tool for doing any inference on θ. For example the Maximum Likelihood (ML) estimate, $\widehat{\theta}_{ML}$ of θ is defined as

$$
\widehat{\theta}_{ML} = \arg\max_{\theta} \big\{ f_{\boldsymbol{X}|\theta}(\boldsymbol{x}|\theta) \big\},
$$

where $f_{\boldsymbol{X}|\theta}(\boldsymbol{x}|\theta)$ is the pdf corresponding to the cdf $F_{\boldsymbol{X}|\theta}(\boldsymbol{x}|\theta)$.

In the second case the Maximum Entropy (ME) principle ([4, 5]), can be used for assigning the probability law $f_{\mathcal{V}}(\nu)$ and thus go back to the previous case, e.g. [1] page 90.

In this paper we consider the case where we only know the median of the nuisance parameter $\mathcal{V}$. If we had a complementary knowledge about the finite support of pdf of $\mathcal{V}$, then we could again use the ME principle to assign a prior and go back to the previous case, e.g. [3]. But if we are given the median of $\mathcal{V}$ and if the support is not finite, then in our knowledge, there is not any solution for this case. The main object of this paper is to propose a solution for it. For this aim, in place of $F_{\boldsymbol{X}|\theta}(\boldsymbol{x}|\theta)$ in (1), we propose a new inference tool $\widetilde{F}_{\boldsymbol{X}|\theta}(\boldsymbol{x}|\theta)$ which can be used to infer on θ (we will show that $\widetilde{F}_{\boldsymbol{X}|\theta}(\boldsymbol{x}|\theta)$ is a cdf under a few conditions). For example we can define

$$
\widetilde{\theta} = \arg\max_{\theta} \big\{ \widetilde{f}_{\boldsymbol{X}|\theta}(\boldsymbol{x}|\theta) \big\},
$$

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

This new tool is deduced from the interpretation of $F_{\boldsymbol{X}|\theta}(\boldsymbol{x}|\theta)$ as the mean value of the random variable $T = T(\mathcal{V};\boldsymbol{x}) = F_{\boldsymbol{X}|\mathcal{V},\theta}(\boldsymbol{x}|\mathcal{V},\theta)$ as given by (1). Now, if in place of the mean value, we take the median, we obtain this new inference tool $\widetilde{F}_{\boldsymbol{X}|\theta}(\boldsymbol{x}|\theta)$ which is defined as

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

and can be used in the same way to infer on θ.

As far as the authors know, there is no work on this subject except recently presented conference papers by the authors, [9, 8, 7]. In the first article we introduced an alternative inference tool to total probability formula, which is called a new inference tool in this paper. We calculated directly this new inference tool (such as Example A in Section 2) and a numerical method suggested for its approximation. In the second one, we used this new tool for parameter estimation. Finally, in the last one, we reviewed the content of two previous papers and mentioned its use for the estimation of a parameter with incomplete knowledge on a nuisance parameter in the one dimensional case. In this paper we give more details and more results with proofs using weaker conditions, with a new overlook on the problem. We also extend the idea to the multivariate case. In the following, first we give more precise definition of $\widetilde{F}_{\boldsymbol{X}|\theta}(\boldsymbol{x}|\theta)$. Then we present some of its properties. For example, we show that under some conditions, $\widetilde{F}_{\boldsymbol{X}|\theta}(\boldsymbol{x}|\theta)$ has all the properties of a cdf, its calculation is very easy and depends only on the median of prior distribution. Then, we give a few examples and finally, we compare the relative performances of these two tools for the inference on θ. Extensions and conclusion are given in the last two sections.

2   A New Inference Tool

Hereafter in this section to simplify the notations we omit the parameter θ, and we assume that the random variables X_i, $i = 1, \cdots, n$ and random parameter $\mathcal{V}$ are continuous and real. We also use *increasing* and *decreasing* instead of *non-decreasing* and *non-increasing* respectively.

Definition 1 *Let $\boldsymbol{X} = (X_1, \cdots, X_n)'$ have a cdf $F_{\boldsymbol{X}|\mathcal{V}}(\boldsymbol{x}|\nu)$ depending on a random parameter $\mathcal{V}$ with pdf $f_{\mathcal{V}}(\nu)$, and let the random variable $T = T(\mathcal{V}; \boldsymbol{x}) = F_{\boldsymbol{X}|\mathcal{V}}(\boldsymbol{x}|\mathcal{V})$ have a unique median for each fixed $\boldsymbol{x}$. The new inference tool, $\widetilde{F}_{\boldsymbol{X}}(\boldsymbol{x})$, is defined as the median of T:*

$$F_{F_{\boldsymbol{X}|\mathcal{V}}(\boldsymbol{x}|\mathcal{V})}(\widetilde{F}_{\boldsymbol{X}}(\boldsymbol{x})) = \frac{1}{2}, \quad or \quad P(F_{\boldsymbol{X}|\mathcal{V}}(\boldsymbol{x}|\mathcal{V}) \le \widetilde{F}_{\boldsymbol{X}}(\boldsymbol{x})) = \frac{1}{2}. \tag{2}$$

To make our point clear we begin with the following simple example, called *Example A*. Let $F_{X|\mathcal{V}}(x|\nu) = 1 - e^{-\nu x}$, $x > 0$, be the cdf of an exponential random variable with scale parameter $\nu > 0$. We assume that the prior pdf of $\mathcal{V}$ is known and also is exponential with parameter 1, i.e. $f_{\mathcal{V}}(\nu) = e^{-\nu}$, $\nu > 0$. We define the random variable $T = F_{X|\mathcal{V}}(x|\mathcal{V}) = 1 - e^{-\mathcal{V}x}$, for any fixed value $x > 0$. The random variable $0 \le T \le 1$ has the following cdf

$$F_T(t) \;\; = \;\; P(1 - e^{-\mathcal{V}x} \le t) \;\; = \;\; 1 - (1-t)^{\frac{1}{x}}, \;\; 0 \le t \le 1.$$

Therefore, pdf of T is $f_T(t) = \frac{1}{x}(1-t)^{(\frac{1}{x}-1)}, \quad 0 \leq t \leq 1$. Now, we can calculate the mean of the random variable T as follow

$$E(T) = \int_0^1 t \, \frac{1}{x}(1-t)^{(\frac{1}{x}-1)} \, dt = 1 - \frac{1}{x+1}.$$

Let $Med(T)$ be the median of the random variable T, then it can be calculated by

$$F_T(Med(T)) = \frac{1}{2} \implies Med(T) = 1 - e^{-x\ln(2)}.$$

Mean value of the random variable T is a cdf with respect to (wrt) x. This fact is always true; because $E(T)$ is the marginal cdf of random variable X, i.e. $F_X(x)$. The marginal cdf is well known, well defined and can also be calculated directly by (1). On the other hand, in this example, it is obvious that $Med(T)$ is a cdf wrt x, which is called $\widetilde{F}_X(x)$ in Definition 1, see Figure 1. However, we have not a short cut for calculating $\widetilde{F}_X(x)$ such as $F_X(x)$ in (1).

In the following theorem and remark, first we show that under a few conditions, $\widetilde{F}_{\mathbf{X}}(\mathbf{x})$ has all the properties of a cdf. Then, in Theorem 2, we drive a simple expression for calculating $\widetilde{F}_{\mathbf{X}}(\mathbf{x})$ and show that, in many cases, the expression of $\widetilde{F}_{\mathbf{X}}(\mathbf{x})$ depends only on the median of the prior and can be calculated simply, see Remark 2. In Theorem 3 we state separability property of $\widetilde{F}_{\mathbf{X}}(\mathbf{x})$ versus exchangeability of $F_{\mathbf{X}}(\mathbf{x})$.

Theorem 1 *Let $\mathbf{X}$ have a cdf $F_{\mathbf{X}|\mathcal{V}}(\mathbf{x}|\nu)$ depending on a random parameter $\mathcal{V}$ with pdf $f_{\mathcal{V}}(\nu)$ and the real random variable $T = F_{\mathbf{X}|\mathcal{V}}(\mathbf{x}|\mathcal{V})$ have a unique median for each fixed $\mathbf{x}$. Then:*

1. *$\widetilde{F}_{\mathbf{X}}(\mathbf{x})$ is an increasing function in each of its arguments.*

2. *If $F_{\mathbf{X}|\mathcal{V}}(\mathbf{x}|\nu)$ and $F_{\mathcal{V}}(\nu)$ are continuous cdfs then $\widetilde{F}_{\mathbf{X}}(\mathbf{x})$ is a continuous function in each of its arguments.*

3. *$0 \leq \widetilde{F}_{\mathbf{X}}(\mathbf{x}) \leq 1$.*

Proof:
1. Let $\mathbf{y} = (y_1, \cdots, y_n)'$, $\mathbf{z} = (z_1, \cdots, z_n)'$, $y_j < z_j$ for fixed j and $y_i = z_i$ for $i \neq j$, $1 \leq i,j \leq n$ and take

$$k_{\mathbf{y}} = \widetilde{F}_{\mathbf{X}}(\mathbf{y}), \ k_{\mathbf{z}} = \widetilde{F}_{\mathbf{X}}(\mathbf{z}) \ \text{ and } \ Y = F_{\mathbf{X}|\mathcal{V}}(\mathbf{y}|\mathcal{V}), \ Z = F_{\mathbf{X}|\mathcal{V}}(\mathbf{z}|\mathcal{V}).$$

Then using (2) we have

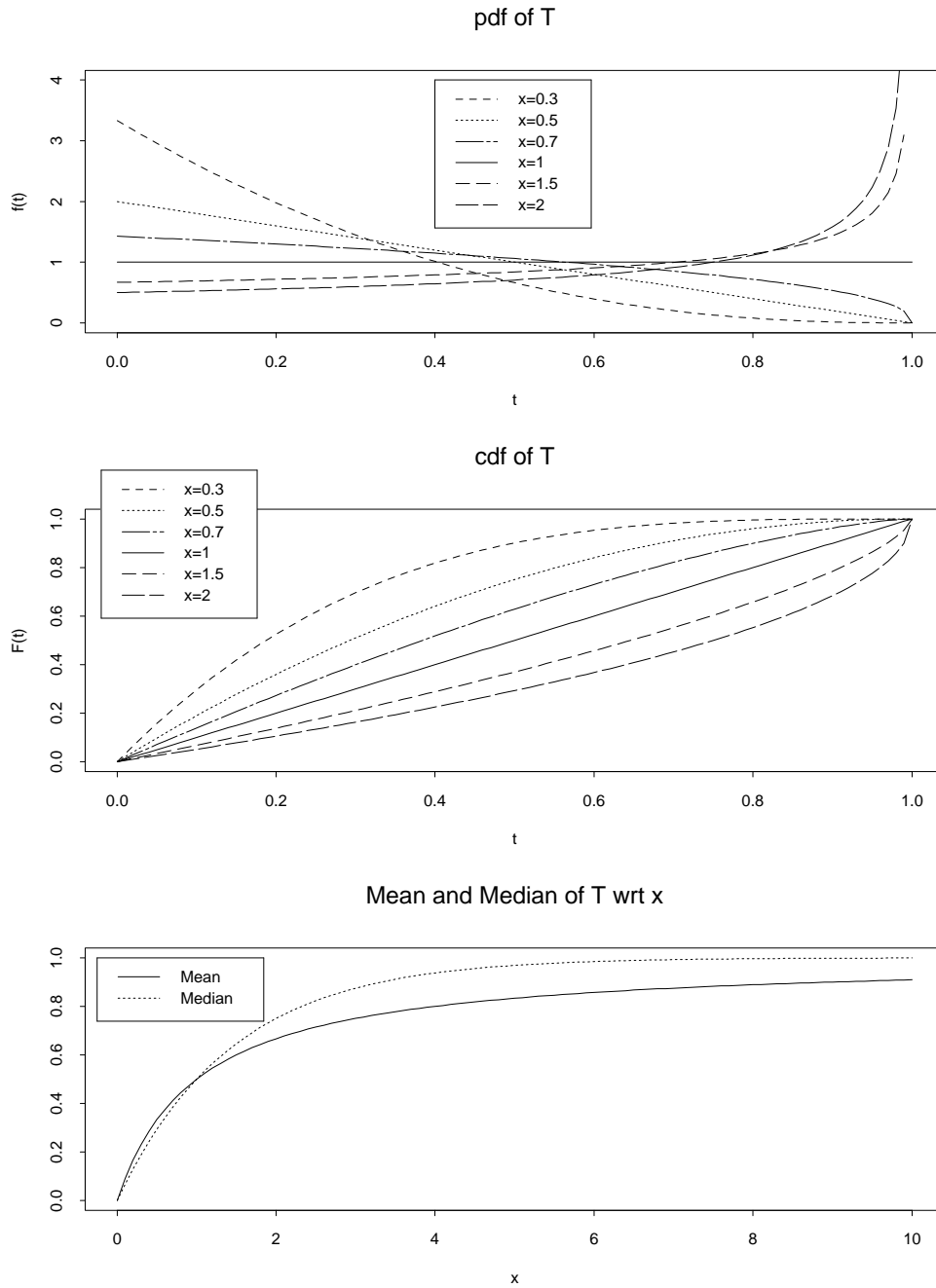$$P(Y \leq k_{\mathbf{y}}) = P(Z \leq k_{\mathbf{z}}) = \frac{1}{2}.$$

Figure 1: Top: pdf of random variable $T = T(\mathcal{V}; x) = F_{X|\mathcal{V}}(x|\mathcal{V}) = 1 - e^{-\mathcal{V}x}$, Middle: cdf of random variable T, and Bottom: mean and median of random variable T in Example A.

We also have $Y \leq Z$, because $F_{\mathbf{X}|\mathcal{V}}$ is an increasing function in each of its arguments. Therefore,

$$P(Y \leq k_{\mathbf{y}}) = P(Z \leq k_{\mathbf{z}}) \leq P(Y \leq k_{\mathbf{z}}),$$

$k_{\mathbf{y}}$ is the unique median of Y and so $k_{\mathbf{y}} \leq k_{\mathbf{z}}$ or equivalently $\widetilde{F}_{\mathbf{X}}(\mathbf{x})$ is increasing in its j-th argument.

2. Let $\mathbf{x}_. = (x_1, \cdots, x_{j-1}, x_., x_{j+1}, \cdots, x_n)'$ and $\mathbf{t} = (x_1, \cdots, x_{j-1}, t, x_{j+1}, \cdots, x_n)'$. By part 1, $\widetilde{F}_{\mathbf{X}}(\mathbf{x})$ is an increasing function in each of its arguments. Therefore,

$$\widetilde{F}_{\mathbf{X}}(\mathbf{x}_-) = \lim_{t \uparrow x_j} \widetilde{F}_{\mathbf{X}}(\mathbf{t}) \quad \text{and} \quad \widetilde{F}_{\mathbf{X}}(\mathbf{x}_+) = \lim_{t \downarrow x_j} \widetilde{F}_{\mathbf{X}}(\mathbf{t}),$$

exist and are finite, e.g. [11].

Further, $F_{\mathbf{X}|\mathcal{V}}(\mathbf{x}|\nu)$ is continuous wrt x_j, and so

$$P(F_{\mathbf{X}|\mathcal{V}}(\mathbf{x}_-|\mathcal{V}) \leq \widetilde{F}_{\mathbf{X}}(\mathbf{x}_-)) = P(F_{\mathbf{X}|\mathcal{V}}(\mathbf{x}|\mathcal{V}) \leq \widetilde{F}_{\mathbf{X}}(\mathbf{x}_-)),$$

$$P(F_{\mathbf{X}|\mathcal{V}}(\mathbf{x}_+|\mathcal{V}) \leq \widetilde{F}_{\mathbf{X}}(\mathbf{x}_+)) = P(F_{\mathbf{X}|\mathcal{V}}(\mathbf{x}|\mathcal{V}) \leq \widetilde{F}_{\mathbf{X}}(\mathbf{x}_+)),$$

and by (2) we have

$$
\begin{aligned}
P(F_{\mathbf{X}|\mathcal{V}}(\mathbf{x}|\mathcal{V}) \leq \widetilde{F}_{\mathbf{X}}(\mathbf{x}_-)) &= P(F_{\mathbf{X}|\mathcal{V}}(\mathbf{x}|\mathcal{V}) \leq \widetilde{F}_{\mathbf{X}}(\mathbf{x})) \\
&= P(F_{\mathbf{X}|\mathcal{V}}(\mathbf{x}|\mathcal{V}) \leq \widetilde{F}_{\mathbf{X}}(\mathbf{x}_+)).
\end{aligned}
\tag{3}
$$

But $\widetilde{F}_{\mathbf{X}}(\mathbf{x})$ is the unique median of $F_{\mathbf{X}|\mathcal{V}}(\mathbf{x}|\mathcal{V})$, therefore by (3),

$$\widetilde{F}_{\mathbf{X}}(\mathbf{x}_-) = \widetilde{F}_{\mathbf{X}}(\mathbf{x}) = \widetilde{F}_{\mathbf{X}}(\mathbf{x}_+),$$

and thus $\widetilde{F}_{\mathbf{X}}(\mathbf{x})$ is continuous.

3. $\widetilde{F}_{\mathbf{X}}(\mathbf{x})$ is the median of random variable T, where $T = F_{\mathbf{X}|\mathcal{V}}(\mathbf{x}|\mathcal{V})$ and $0 \leq T \leq 1$, and so $0 \leq \widetilde{F}_{\mathbf{X}}(\mathbf{x}) \leq 1$. $\qquad \square$

Remark 1 *By part 1 of Theorem 1, $\lim_{x_j \uparrow +\infty} \widetilde{F}_{\mathbf{X}}(\mathbf{x})$ and $\lim_{x_j \downarrow -\infty} \widetilde{F}_{\mathbf{X}}(\mathbf{x})$ exist and are finite, [11]. Therefore $\widetilde{F}_{\mathbf{X}}(\mathbf{x})$ is a continuous cdf if conditions of Theorem 1 hold and*

1. *$\lim_{x_j \downarrow -\infty} \widetilde{F}_{\mathbf{X}}(\mathbf{x}) = 0$ for any particular j,*

2. *$\lim_{x_1 \uparrow +\infty, \cdots, x_n \uparrow +\infty} \widetilde{F}_{\mathbf{X}}(\mathbf{x}) = 1$,*

3. *$\Delta_{b_1 a_1} \cdots \Delta_{b_n a_n} \widetilde{F}_{\mathbf{X}}(\mathbf{x}) \geq 0$, where $a_i \leq b_i$, $i = 1, \cdots, n$, and*
   $\Delta_{b_j a_j} \widetilde{F}_{\mathbf{X}}(\mathbf{x}) = \widetilde{F}_{\mathbf{X}}((x_1, \cdots, x_{j-1}, b_j, x_{j+1}, \cdots, x_n)') - \widetilde{F}_{\mathbf{X}}((x_1, \cdots, x_{j-1}, a_j, x_{j+1}, \cdots, x_n)') \geq 0.$

In this case, we call $\widetilde{F}_{\boldsymbol{X}}(\boldsymbol{x})$ the marginal cdf of $\boldsymbol{X}$ based on median. When $\widetilde{F}_X(x)$ is a one dimensional cdf, the last condition follows from parts 1 and 3 of Theorem 1.

Theorem 2 *If $L(\nu) = F_{\boldsymbol{X}|\mathcal{V}}(\boldsymbol{x}|\nu)$ is a monotone function wrt ν and $\mathcal{V}$ has a unique median $F_{\mathcal{V}}^{-1}(\frac{1}{2})$, then $\widetilde{F}_{\boldsymbol{X}}(\boldsymbol{x}) = L(F_{\mathcal{V}}^{-1}(\frac{1}{2}))$.*

Proof: Let

$$L^-(u) = \begin{cases} \inf\{\nu; \ L(\nu) \geq u\} & \text{if } L \text{ is an increasing function} \\ \inf\{\nu; \ L(\nu) \leq u\} & \text{if } L \text{ is a decreasing function} \end{cases},$$

be the generalized inverse of L, e.g. [10] page 39. Noting that

$$\{(u,\nu): \ L^-(u) \leq \nu\} = \begin{cases} \{(u,\nu): \ u \leq L(\nu)\} & \text{if } L \text{ is an increasing function} \\ \{(u,\nu): \ u \geq L(\nu)\} & \text{if } L \text{ is a decreasing function} \end{cases},$$

and by (2) we have,

$$P(L(\mathcal{V}) \leq \widetilde{F}_{\boldsymbol{X}}(\boldsymbol{x})) = \frac{1}{2}$$

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

$$\Leftrightarrow \begin{cases} F_{\mathcal{V}}(L^-(\widetilde{F}_{\boldsymbol{X}}(\boldsymbol{x}))) = \frac{1}{2} & \text{if } L \text{ is an increasing function} \\ 1 - F_{\mathcal{V}}(L^-(\widetilde{F}_{\boldsymbol{X}}(\boldsymbol{x}))) = \frac{1}{2} & \text{if } L \text{ is a decreasing function} \end{cases}$$

$$\Leftrightarrow \ F_{\mathcal{V}}(L^-(\widetilde{F}_{\boldsymbol{X}}(\boldsymbol{x}))) = \frac{1}{2}$$

$$\Leftrightarrow \ L^-(\widetilde{F}_{\boldsymbol{X}}(\boldsymbol{x})) = F_{\mathcal{V}}^{-1}(\frac{1}{2}) \quad \text{by uniqueness of the median of } \mathcal{V}$$

$$\Leftrightarrow \ \widetilde{F}_{\boldsymbol{X}}(\boldsymbol{x}) = L(F_{\mathcal{V}}^{-1}(\frac{1}{2})),$$

where the last expression follows from

$$\{(u,\nu): \ L^-(u) = \nu\} \subseteq \{(u,\nu): \ L(\nu) = u\}.$$

$\square$

Remark 2 *If conditions of Theorem 2 hold, then $\widetilde{F}_{\boldsymbol{X}}(\boldsymbol{x})$ belongs to the family of distributions $F_{\boldsymbol{X}|\mathcal{V}}(\boldsymbol{x}|\nu)$. Because, $\widetilde{F}_{\boldsymbol{X}}(\boldsymbol{x}) = F_{\boldsymbol{X}|\mathcal{V}}(\boldsymbol{x}|F_{\mathcal{V}}^{-1}(\frac{1}{2}))$. Therefore $\widetilde{F}_{\boldsymbol{X}}(\boldsymbol{x})$ is a cdf and conditions in Remark 1 hold.*

Remark 3 $\widetilde{F}_{\boldsymbol{X}}(\boldsymbol{x})$ *depends only on the median of prior distribution,* $F_{\mathcal{V}}^{-1}(\frac{1}{2})$, *while the expression of* $F_{\boldsymbol{X}}(\boldsymbol{x})$ *needs the perfect knowledge of* $F_{\mathcal{V}}(\nu)$. *Therefore,* $\widetilde{F}_{\boldsymbol{X}}(\boldsymbol{x})$ *is robust relative to prior distributions with the same median.*

Remark 4 *If median of* T *is not unique then* $\widetilde{F}_{X}(x)$ *may not be a unique cdf wrt* x. *For example (called* Example B*), assume that* $\mathcal{V}$ *has the following cdf, in Example A, Figure 2-left:*

$$F_{\mathcal{V}}(\nu) = \begin{cases} 0 & \nu < 0 \\ \nu & 0 \le \nu < \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \le \nu < \frac{3}{4} \\ \nu - \frac{1}{4} & \frac{3}{4} \le \nu < \frac{5}{4} \\ 1 & \frac{5}{4} \le \nu \end{cases}.$$

Then, $T = T(\mathcal{V}; x) = F_{X|\mathcal{V}}(x|\mathcal{V}) = 1 - e^{-\mathcal{V}x}$ *has the following cdf*

$$F_{T}(t) = \begin{cases} 0 & t < 0 \\ \frac{-\ln(1-t)}{x} & 0 \le t < 1 - e^{-\frac{1}{2}x} \\ \frac{1}{2} & 1 - e^{-\frac{1}{2}x} \le t < 1 - e^{-\frac{3}{4}x} \\ \frac{-\ln(1-t)}{x} - \frac{1}{4} & 1 - e^{-\frac{3}{4}x} \le t < 1 - e^{-\frac{5}{4}x} \\ 1 & 1 - e^{-\frac{5}{4}x} \le t \end{cases}.$$

Therefore, the median of T *is an arbitrary point in the following interval: (see Figure 2-right)*

$$\left[1 - e^{-\frac{1}{2}x} \,,\; 1 - e^{-\frac{3}{4}x} \right).$$
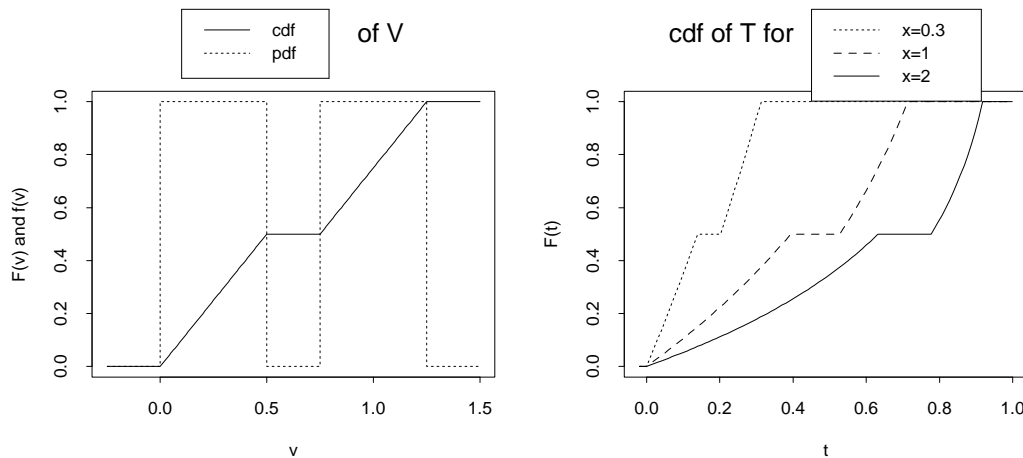

Figure 2:   Left: cdf of random variable $\mathcal{V}$ in Example B and its corresponding pdf. Right: cdf of random variable $T = T(\mathcal{V}; x) = F_{X|\mathcal{V}}(x|\mathcal{V}) = 1 - e^{-\mathcal{V}x}$ in Example B.

Theorem 3 *Let $F_{\boldsymbol{X}|\mathcal{V}}(\boldsymbol{x}|\nu)$ be conditional cdf of $\boldsymbol{X} = (X_1, \cdots, X_n)'$ given $\mathcal{V} = \nu$ and $L_{(k_1, \cdots, k_r)}(\nu) = F_{(X_{k_1}, \cdots, X_{k_r})|\mathcal{V}}(x_{k_1}, \cdots, x_{k_r}|\nu)$ be monotone function of ν for each $\{k_1, \cdots, k_r\} \subseteq \{1, \cdots, n\}$. Let also $\mathcal{V}$ have a unique median $F_{\mathcal{V}}^{-1}(\frac{1}{2})$. If for each $\{k_1, \cdots, k_r\} \subseteq \{1, \cdots, n\}$,*

$$F_{(X_{k_1}, \cdots, X_{k_r})|\mathcal{V}}(x_{k_1}, \cdots, x_{k_r}|\nu) = \prod_{i=1}^{r} F_{X_{k_i}|\mathcal{V}}(x_{k_i}|\nu),$$

i.e. $\boldsymbol{X} \mid \mathcal{V} = \nu$ *has independent components, then*

$$\widetilde{F}_{(X_{k_1}, \cdots, X_{k_r})}(x_{k_1}, \cdots, x_{k_r}) = \prod_{i=1}^{r} \widetilde{F}_{X_{k_i}}(x_{k_i}).$$

Proof:
Conditions of Theorem 2 hold and so, for each $\{k_1, \cdots, k_r\} \subseteq \{1, \cdots, n\}$,

$$
\begin{aligned}
\widetilde{F}_{(X_{k_1}, \cdots, X_{k_r})}(x_{k_1}, \cdots, x_{k_r}) &= L_{(k_1, \cdots, k_r)}(F_{\mathcal{V}}^{-1}(\tfrac{1}{2})) \\
&= F_{(X_{k_1}, \cdots, X_{k_r})|\mathcal{V}}(x_{k_1}, \cdots, x_{k_r}|F_{\mathcal{V}}^{-1}(\tfrac{1}{2})) \\
&= \prod_{i=1}^{r} F_{X_{k_i}|\mathcal{V}}(x_{k_i}|F_{\mathcal{V}}^{-1}(\tfrac{1}{2})) \\
&= \prod_{i=1}^{r} L_{k_i} F_{\mathcal{V}}^{-1}(\tfrac{1}{2}) \\
&= \prod_{i=1}^{r} \widetilde{F}_{X_{k_i}}(x_{k_i}).
\end{aligned}
$$

$\square$

Remark 5 *If $\boldsymbol{X} \mid \mathcal{V} = \nu$ has independent components, then the marginal distribution of $\boldsymbol{X}$ cannot have independent components. For example, in general case,*

$$F_{\boldsymbol{X}}(\boldsymbol{x}) = \int_{-\infty}^{+\infty} F_{\boldsymbol{X}|\mathcal{V}}(\boldsymbol{x}|\nu)\, dF_{\mathcal{V}}(\nu) \neq \prod_{i=1}^{n} F_{X_i}(x_i) = \prod_{i=1}^{n} \int_{-\infty}^{+\infty} F_{X_i|\mathcal{V}}(x_i|\nu)\, dF_{\mathcal{V}}(\nu).$$

It can be shown that, if $\boldsymbol{X} \mid \mathcal{V} = \nu$ has Independent and Identically Distributed (iid) components, then the marginal distribution of $\boldsymbol{X}$ is exchangeable, see Example 1. We recall that for identically distributed random variables exchangeability is a weaker condition than independence.

In the following we show that some families of distributions (e.g. [6]) have a monotone distribution function wrt their parameters and so, calculation of $\widetilde{F}_{\boldsymbol{X}}(\boldsymbol{x})$ is very easy by using Theorem 2.

Lemma 1 *Let $L(\nu) = F_{\boldsymbol{X}|\mathcal{V}}(\boldsymbol{x}|\nu)$. If ν is a real location parameter then $L(\nu)$ is decreasing wrt ν.*

Proof: Let $\nu_1 < \nu_2$ and ν be a location parameter. Then

$$
\begin{aligned}
L(\nu_1) &= F_{\boldsymbol{X}|\mathcal{V}}(\boldsymbol{x}|\nu_1) \\
&= F_{\boldsymbol{X}|\mathcal{V}}(\boldsymbol{x} - \nu_1), \quad \text{where} \ \ \boldsymbol{x} - \nu_1 = (x_1 - \nu_1, \cdots, x_n - \nu_1)' \\
&\geq F_{\boldsymbol{X}|\mathcal{V}}(\boldsymbol{x} - \nu_2), \quad \text{because } x_i - \nu_1 > x_i - \nu_2, \ i = 1, \cdots, n \\
&= F_{\boldsymbol{X}|\mathcal{V}}(\boldsymbol{x}|\nu_2) \\
&= L(\nu_2).
\end{aligned}
$$

$\square$

Lemma 2 *Let $L(\nu) = F_{X|\mathcal{V}}(x|\nu)$. If ν is a scale parameter then $L(\nu)$ is monotone wrt ν.*

Proof: Let $\nu_1 < \nu_2$. If ν is a scale parameter, $\nu > 0$, then

$$
\begin{aligned}
L(\nu_1) &= F_{X|\mathcal{V}}(\frac{x}{\nu_1}) \\
&\begin{array}{c} \leq \\ \geq \end{array} F_{X|\mathcal{V}}(\frac{x}{\nu_2}), \quad \begin{array}{l} \text{if } x < 0 \\ \text{if } x > 0 \end{array} \\
&= L(\nu_2).
\end{aligned}
$$

Therefore, $L(\nu)$ is an increasing function if $x < 0$ and is a decreasing function if $x > 0$, i.e. $L(\nu)$ is a monotone function wrt ν. $\square$

The proof of the following lemma is straightforward.

Lemma 3 *Let $X_1, \cdots X_n$ given $\mathcal{V} = \nu$ be iid random variables and $\boldsymbol{X} = (X_1, \cdots, X_n)'$. If $L(\nu) = F_{X_1|\mathcal{V}}(x|\nu)$ is an increasing (a decreasing) function then $L^*(\nu) = F_{\boldsymbol{X}|\mathcal{V}}(\boldsymbol{x}|\nu)$ is an increasing (a decreasing) function of ν.*

In some cases we can show directly that $L(.)$ is a monotone function. For example, in the exponential family this property can be proved by using differentiation. Let $\boldsymbol{X}|\boldsymbol{\eta}$ be distributed according to an exponential family with pdf

$$
f_{\boldsymbol{X}|\boldsymbol{\eta}}(\boldsymbol{x}|\boldsymbol{\eta}) = h(\boldsymbol{x}) \exp\left(\boldsymbol{\eta}' \boldsymbol{T}(\boldsymbol{x}) - A(\boldsymbol{\eta})\right),
$$

where $\boldsymbol{\eta} = (\eta_1, \cdots, \eta_n)'$ and $\boldsymbol{T} = (T_1, \cdots, T_n)'$. It can be shown that $L(\boldsymbol{\eta}) = F_{\boldsymbol{X}|\boldsymbol{\eta}}(\boldsymbol{x}|\boldsymbol{\eta})$ is a monotone function wrt each of its arguments in many cases by the following method: Let

$I_{\boldsymbol{y}\leq\boldsymbol{x}} = 1$ if $y_1 \leq x_1, \cdots, y_n \leq x_n$ and 0 elsewhere; and note that the differentiation under the integral sign is true for exponential family. Then

$$
\begin{aligned}
\frac{\partial}{\partial\eta_i}L(\boldsymbol{\eta}) &= \frac{\partial}{\partial\eta_i}F_{\boldsymbol{X}|\boldsymbol{\eta}}(\boldsymbol{x}|\boldsymbol{\eta}) \\
&= \frac{\partial}{\partial\eta_i}\int_{R^n} I_{\boldsymbol{y}\leq\boldsymbol{x}}\, h(\boldsymbol{y})\,\exp\left(\boldsymbol{\eta}'\boldsymbol{T}(\boldsymbol{y}) - A(\boldsymbol{\eta})\right)\, d\boldsymbol{y} \\
&= \int_{R^n} I_{\boldsymbol{y}\leq\boldsymbol{x}}\, h(\boldsymbol{y})\,\frac{\partial}{\partial\eta_i}\exp\left(\boldsymbol{\eta}'\boldsymbol{T}(\boldsymbol{y}) - A(\boldsymbol{\eta})\right)\, d\boldsymbol{y} \\
&= \int_{R^n} I_{\boldsymbol{y}\leq\boldsymbol{x}}\, h(\boldsymbol{y})\,\left(T_i(\boldsymbol{y}) - \frac{\partial}{\partial\eta_i}A(\boldsymbol{\eta})\right)\exp\left(\boldsymbol{\eta}'\boldsymbol{T}(\boldsymbol{y}) - A(\boldsymbol{\eta})\right)\, d\boldsymbol{y} \\
&\underset{\geq}{\leq} \int_{R^n} h(\boldsymbol{y})\,\left(T_i(\boldsymbol{y}) - \frac{\partial}{\partial\eta_i}A(\boldsymbol{\eta})\right)\exp\left(\boldsymbol{\eta}'\boldsymbol{T}(\boldsymbol{y}) - A(\boldsymbol{\eta})\right)\, d\boldsymbol{y} \\
&= E_{\boldsymbol{X}|\boldsymbol{\eta}}\left(T_i(\boldsymbol{X}) - \frac{\partial}{\partial\eta_i}A(\boldsymbol{\eta})\right) = 0.
\end{aligned}
$$

The last equality follows from $E_{\boldsymbol{X}|\boldsymbol{\eta}}(T_i(\boldsymbol{X})) = \frac{\partial}{\partial\eta_i}A(\boldsymbol{\eta})$, e.g. [6] page 27.

On the other hand, we can use stochastic ordering property of a family of distributions for showing that $L(.)$ is a monotone function. A family of cdfs

$$\mathcal{F} = \{F_{X|\mathcal{V}}(x|\nu),\ \nu \in V\} \tag{4}$$

where V is an interval on the real line, is said to have Monotone Likelihood Ratio (MLR) property if, for every $\nu_1 < \nu_2$ in V the likelihood ratio

$$\frac{f_{X|\mathcal{V}}(x|\nu_2)}{f_{X|\mathcal{V}}(x|\nu_1)},$$

is a monotone function of x. The property of MLR defines a very strong ordering of a family of distributions.

Lemma 4 *If $\mathcal{F}$ is an MLR family wrt x then $F_{X|\mathcal{V}}(x|\nu)$ is an increasing (or a decreasing) function of ν for all x.*

Proof: See e.g. [12] page 124.   □

A family of cdfs in (4) is said to be stochastically increasing (SI) if $\nu_1 < \nu_2$ implies $F_{X|\mathcal{V}}(x|\nu_1) \geq F_{X|\mathcal{V}}(x|\nu_2)$ for all x. For stochastically decreasing (SD) the inequality is reversed. This definition is a weaker property than MLR property (by Lemma 4), but is a stronger property than monotonicity of $L(\nu) = F_{X|\mathcal{V}}(x|\nu)$ (because $L(\nu)$ is monotone for each fixed x). Therefore, we have

$$\text{MLR} \implies \text{SI or SD} \implies L(\nu) \text{ is monotone}$$

It can be shown that the converse of the above relations are not true.

Remark 6 *In Theorem 1, we prove that $\widetilde{F}_X(x)$ is an increasing function. In the proof of this theorem we do not use the monotonicity property of $L(\nu) = F_{X|\mathcal{V}}(x|\nu)$ wrt ν. For example (called Example C), assume that*

$$F_{X|\mathcal{V}}(x|\nu) = \frac{1}{2}(1 - e^{-\nu x}) \; I_{\{x>0\}}(x) + \frac{1}{2}\int_{-\infty}^{x} \mathcal{C}\,(t;\; \nu, 1)\; dt$$

be mixture cdf of an exponential and a Cauchy cdf with parameter $\nu > 0$. Figure 3-left shows the graphs of $L(\nu) = F_{X|\mathcal{V}}(x|\nu)$ for different x. $L(\nu)$ is not monotone for some of x values in this figure. If we assume that the prior pdf of $\mathcal{V}$ is known and is also exponential with parameter 1, then, still median of random variable T is a cdf, see Figure 3-right.
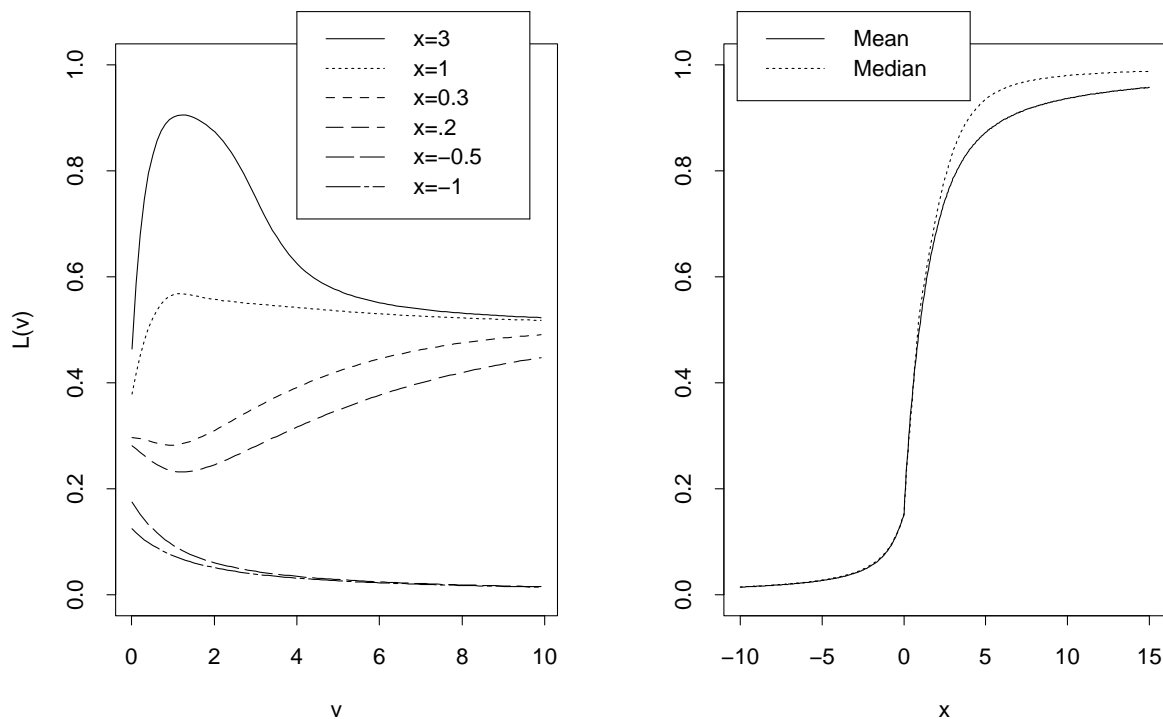

Figure 3: Left: the graphs of $L(\nu) = F_{X|\mathcal{V}}(x|\nu)$ for different x in Example C. Right: the mean and median of random variable T in Example C.

3   Examples

In what follows, we use the following notations and expressions, [2] pages 427-422:

Normal:	$\mathcal{N}(x;\ \mu,\sigma^2)$	$=$	$(2\pi\sigma^2)^{-\frac{1}{2}}\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\};\ \ \sigma > 0$
Double Exponential:	$\mathcal{DE}(x;\ \lambda)$	$=$	$\frac{1}{2\lambda}\exp\left\{-|x|/\lambda\right\};\ \ \lambda > 0$
Inverse Gamma:	$\mathcal{IG}(x;\ \alpha,\beta)$	$=$	$\frac{\beta^\alpha}{\Gamma(\alpha)}x^{-(\alpha+1)}\exp\left\{-\beta/x\right\};\ \ x,\ \alpha,\ \beta > 0$
Student:	$\mathcal{S}(x;\ \mu,\lambda,\alpha)$	$=$	$\frac{\Gamma((\alpha+1)/2)\lambda^{1/2}}{\Gamma(\alpha/2)(\pi\alpha)^{1/2}}\left(1+\frac{\lambda}{\alpha}(x-\mu)^2\right)^{\frac{-(\alpha+1)}{2}};\lambda,\alpha > 0$
Cauchy:	$\mathcal{C}(x;\ \mu,\sigma)$	$=$	$\frac{1}{\pi\sigma}\left(1+(\frac{x-\mu}{\sigma})^2\right)^{-1};\ \ \sigma > 0$

Exchangeable Normal: The random vector $\boldsymbol{X} = (X_1,\cdots,X_n)'$ is said to have an exchangeable normal distribution, $\mathcal{EN}(\boldsymbol{x};\ \mu,\sigma^2,\rho)$, if its distribution is multivariate normal with the following mean vector and variance-covariance matrix

$$\begin{pmatrix} \mu \\ \mu \\ . \\ \mu \end{pmatrix}_{n\times 1},\quad \sigma^2\begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ . & . & \cdots & . \\ \rho & \rho & \cdots & 1 \end{pmatrix}_{n\times n},\quad \sigma > 0,\ \rho \in [0,1).$$

It can be shown that $\mathcal{EN}(\boldsymbol{x};\ \mu,\sigma^2,\rho) =$

$$k_n(\rho,\sigma^2)\exp\left\{-\frac{1}{2}\left(\frac{\sum_{i=1}^n(x_i-\mu)^2}{\sigma^2(1-\rho)} - \frac{\rho(\sum_{i=1}^n(x_i-\mu))^2}{\sigma^2(1+(n-1)\rho)(1-\rho)}\right)\right\},$$

where $k_n(\rho,\sigma^2) = (\sqrt{2\pi}\sigma)^{-n}(1-\rho)^{-(n-1)/2}(1+(n-1)\rho)^{-1/2}$, and $\boldsymbol{x} = (x_1,\cdots,x_n)'$.

3.1   Example 1

The first example we consider is

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

where we assume that the mean value ν is the nuisance parameter. Let $X_1,\cdots X_n$ be an iid copy of X (i.e. $X|\mathcal{V} = \nu,\theta$) and $\boldsymbol{X} = (X_1,\cdots,X_n)'$, then:

- Prior pdf case $f_{\mathcal{V}}(\nu) = \mathcal{N}(\nu;\ \nu_0,\theta_0)$:
  Then we have

$$f_{X|\theta}(x|\theta) = \int_{-\infty}^{+\infty} f_{X|\mathcal{V},\theta}(x|\nu,\theta)\ f_{\mathcal{V}}(\nu)\ \mathrm{d}\nu = \mathcal{N}(x;\ \nu_0,\theta+\theta_0),$$

and

$$f_{\boldsymbol{X}|\theta}(\boldsymbol{x}|\theta) = \int_{-\infty}^{+\infty} f_{\boldsymbol{X}|\mathcal{V},\theta}(\boldsymbol{x}|\nu,\theta)\, f_{\mathcal{V}}(\nu)\, \mathrm{d}\nu = \mathcal{EN}\left(\boldsymbol{x};\ \nu_0, \theta + \theta_0, \frac{\theta_0}{\theta + \theta_0}\right).$$

- Unique median knowledge case Median $\{\mathcal{V}\} = \nu_0$:
  Then, as we could see, by using Lemma 1 and Theorem 2, we have

$$\widetilde{F}_{X|\theta}(x|\theta) = F_{X|\mathcal{V},\theta}(x|\nu_0,\theta),$$

or equivalently,

$$\widetilde{f}_{X|\theta}(x|\theta) = \mathcal{N}\left(x;\ \nu_0, \theta\right).$$

Now we can use Theorem 3 for calculating $\widetilde{F}_{\boldsymbol{X}|\theta}(\boldsymbol{x}|\theta)$ (because $F_{\boldsymbol{X}|\mathcal{V},\theta}(\boldsymbol{x}|\nu,\theta)$ is a decreasing function wrt ν by Lemma 1), therefore,

$$\widetilde{f}_{\boldsymbol{X}|\theta}(\boldsymbol{x}|\theta) = \prod_{i=1}^{n} \mathcal{N}\left(x_i;\ \nu_0, \theta\right) = \mathcal{EN}\left(\boldsymbol{x};\ \nu_0, \theta, 0\right). \tag{5}$$

Note that, if $f_{\mathcal{V}}(\nu) = \mathcal{N}\left(\nu;\ \nu_0, \theta_0\right)$ or $f_{\mathcal{V}}(\nu) = \mathcal{C}\left(\nu;\ \nu_0, \theta_0\right)$ then $\widetilde{f}_{\boldsymbol{X}|\theta}(\boldsymbol{x}|\theta)$ is given by (5), because the median of these two distributions are equal to ν_0 (see Remark 3).

- Moments knowledge case $E(|\mathcal{V}|) = \nu_0$:
  Then the ME pdf is given by $\mathcal{DE}\left(\nu;\ \nu_0\right)$. In this case we cannot obtain an analytical expression for

$$
\begin{aligned}
f_{X|\theta}(x|\theta) &= \int_{-\infty}^{+\infty} (2\pi\theta)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\theta}(x-\nu)^2\right\} \frac{1}{2\nu_0} \exp\left\{-|\nu|/\nu_0\right\}\, \mathrm{d}\nu \\
&= \exp\left\{\frac{\theta - 2\nu_0 x}{2\nu_0^2}\right\}\left(1 + \mathrm{erf}\left\{\frac{\nu_0 x - \theta}{\nu_0\sqrt{2\theta}}\right\} + \exp\left\{\frac{2x}{\nu_0}\right\}\mathrm{erfc}\left\{\frac{\nu_0 x + \theta}{\nu_0\sqrt{2\theta}}\right\}\right),
\end{aligned}
$$

where $\mathrm{erf}(y) = \frac{2}{\sqrt{\pi}}\int_0^y \exp(-t^2)\, \mathrm{d}t$ and $\mathrm{erfc}(y) = \frac{2}{\sqrt{\pi}}\int_y^\infty \exp(-t^2)\, \mathrm{d}t$. We recall that, if we know that $E(\mathcal{V}) = \nu_0$ or Median $\{\mathcal{V}\} = \nu_0$ and the support of $\mathcal{V}$ is R the ME pdf does not exist.

3.2 *Example 2*

The second example we consider is

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

where, this time, we assume that ν is the variance and the nuisance parameter. Then:

- Prior pdf case $f_{\mathcal{V}}(\nu) = \mathcal{IG}(\nu;\ \alpha, \beta)$:
  Then, it is easy to show that,

$$f_{X|\theta}(x|\theta) = \mathcal{S}(x;\ \theta, \alpha/\beta, 2\alpha), \tag{6}$$

  but $f_{\boldsymbol{X}|\theta}(\boldsymbol{x}|\theta)$ cannot be calculated analytically.

- Unique median knowledge case Median $\{\mathcal{V}\} = \nu_0$:
  Then, as we could see, by using Lemma 2 and Theorem 2, we have

$$\widetilde{f}_{X|\theta}(x|\theta) = \mathcal{N}(x;\ \theta, \nu_0).$$

  It can be shown that $F_{\boldsymbol{X}|\mathcal{V},\theta}(\boldsymbol{x}|\nu, \theta)$ is a monotone function wrt ν (by using derivative) and by Theorem 3 we have

$$\widetilde{f}_{\boldsymbol{X}|\theta}(\boldsymbol{x}|\theta) = \prod_{i=1}^{n} \mathcal{N}(x_i;\ \theta, \nu_0) = \mathcal{EN}(\boldsymbol{x};\ \theta, \nu_0, 0).$$

- Moments knowledge case $E(1/\mathcal{V}) = 1/\nu_0$:
  Then, knowing that the variance is a positive quantity, the ME pdf $f_{\mathcal{V}}(\nu)$ is an $\mathcal{IG}(\nu;\ 1, \nu_0)$. In this case we have

$$f_{X|\theta}(x|\theta) = \mathcal{S}(x;\ \theta, 1/\nu_0, 2),$$

  and $f_{\boldsymbol{X}|\theta}(\boldsymbol{x}|\theta)$ cannot be calculated analytically.

3.3 *Example 3*

In this example we consider is $\mathcal{EN}(\boldsymbol{x};\ \nu, \sigma^2, \rho)$, where ν is a nuisance parameter. Noting that, we can write $\mathcal{EN}(\boldsymbol{x};\ \nu, \sigma^2, \rho)$, as follows (exponential family),

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = q_n(\theta_1, \theta_2, \theta_3) \exp\{\theta_1 t_1 + \theta_2 t_2 + \theta_3 t_3\},$$

where $\theta_1 = \rho/l$, $t_1 = \sum_{i<j} x_i x_j$, $\theta_2 = -(1 + (n-2)\rho)/(2l)$, $t_2 = \sum_{i=1}^{n} x_i^2$, $\theta_3 = (1-\rho)\nu/l$, $t_3 = \sum_{i=1}^{n} x_i$, $l = \sigma^2(1 + (n-1)\rho)(1-\rho)$, and $q_n(\theta_1, \theta_2, \theta_3)$ can be determined. This pdf is a monotone function wrt θ_3 and so $L(\nu)$ is a monotone function. Let $\theta = (\sigma^2, \rho)$ and the median of prior pdf be ν_0, then

$$\widetilde{f}_{\boldsymbol{X}|\theta}(\boldsymbol{x}|\theta) = \mathcal{EN}(\boldsymbol{x};\ \nu_0, \sigma^2, \rho).$$

3.4 Comparison of Estimators in Example 1

Suppose we are interested in estimating θ in Example 1. In the case that $n = 1$

$$f_{X|\theta}(x|\theta) = \mathcal{N}\left(x;\ \nu_0, \theta + \theta_0\right) \quad \text{and} \quad \widetilde{f}_{X|\theta}(x|\theta) = \mathcal{N}\left(x;\ \nu_0, \theta\right),$$

and so the ML estimator (MLE) of θ based on these two pdfs are equal to

$$\widehat{\theta} = \max\{(X - \nu_0)^2 - \theta_0, 0\} \quad \text{and} \quad \widetilde{\theta} = (X - \nu_0)^2$$

respectively. For $n > 1$ the MLE of θ based on

$$f_{\boldsymbol{X}|\theta}(\boldsymbol{x}|\theta) = \mathcal{EN}\left(\boldsymbol{x};\ \nu_0, \theta + \theta_0, \frac{\theta_0}{\theta + \theta_0}\right),$$

can be calculated numerically by the following simplified likelihood function,

$$l(\theta) = \exp\left\{-\frac{1}{2}\left(\frac{\sum_{i=1}^{n}(x_i - \nu_0)^2}{\theta} - \frac{(\sum_{i=1}^{n}(x_i - \nu_0))^2}{\theta(\theta + n)}\right)\right\} / \sqrt{(2\pi)^n \theta^{n-1}(\theta + n)},$$

where we assume that $\theta_0 = 1$. The MLE of θ based on $\widetilde{f}_{\boldsymbol{X}|\theta}(\boldsymbol{x}|\theta) = \mathcal{EN}\left(\boldsymbol{x};\ \nu_0, \theta, 0\right)$, is equal to $\widetilde{\theta} = \sum_{i=1}^{n}(X_i - \nu_0)^2/n$.

Before comparing these two estimators (by considering normal prior for ν), one can predict that, $\widehat{\theta}$ is better than $\widetilde{\theta}$, because $\widehat{\theta}$ uses more information (i.e. known normal prior) than $\widetilde{\theta}$ which uses only the median of prior distribution. We may also recall that, $f_{\boldsymbol{X}|\theta}(\boldsymbol{x}|\theta)$ is the true pdf of observations obtained using the full prior knowledge on the nuisance parameter, while $\widetilde{f}_{\boldsymbol{X}|\theta}(\boldsymbol{x}|\theta)$ is a pseudo pdf which includes only prior knowledge of the median of nuisance parameter.

The empirical Mean Square Error (MSE) of 4 estimators are plotted in Figure 3.4 for different sample sizes n. We note by T the MLE of θ when $\nu = \nu_0$, and we note by T_{MaxEnt} the MLE of θ when the prior mean and variance are known.

In Figure 3.4-left we plot the graphs of MSE of $\widehat{\theta}$, $\widetilde{\theta}$, T and T_{MaxEnt}. In Table 1 we classify these 4 estimators and corresponding assumptions for $n = 1$. We see that, in Figure 3.4-left, $\widehat{\theta}$ is better than $\widetilde{\theta}$, especially for large sample size n, and T is the best.

In Figure 3.4-right we plot the graphs of MSE wrt median, ν_0. This is useful for checking robustness of estimators wrt false prior information. We see that $\widehat{\theta}$ is more robust than $\widetilde{\theta}$ relative to ν_0, but both of them dominated by T. In this case, samples are generated from a normal distribution with random normal mean (median ν_0) when $\theta = 2$, however, we assume that ν has a standard normal prior distribution.

The simulations confirm the following logic: more we have information better will be the estimation. In fact for calculating T we have not nuisance parameter; for $\widehat{\theta}$, we use all prior distribution information; for T_{MaxEnt} we use prior mean and prior variance information; and for $\widetilde{\theta}$ we use only the median value of prior distribution.
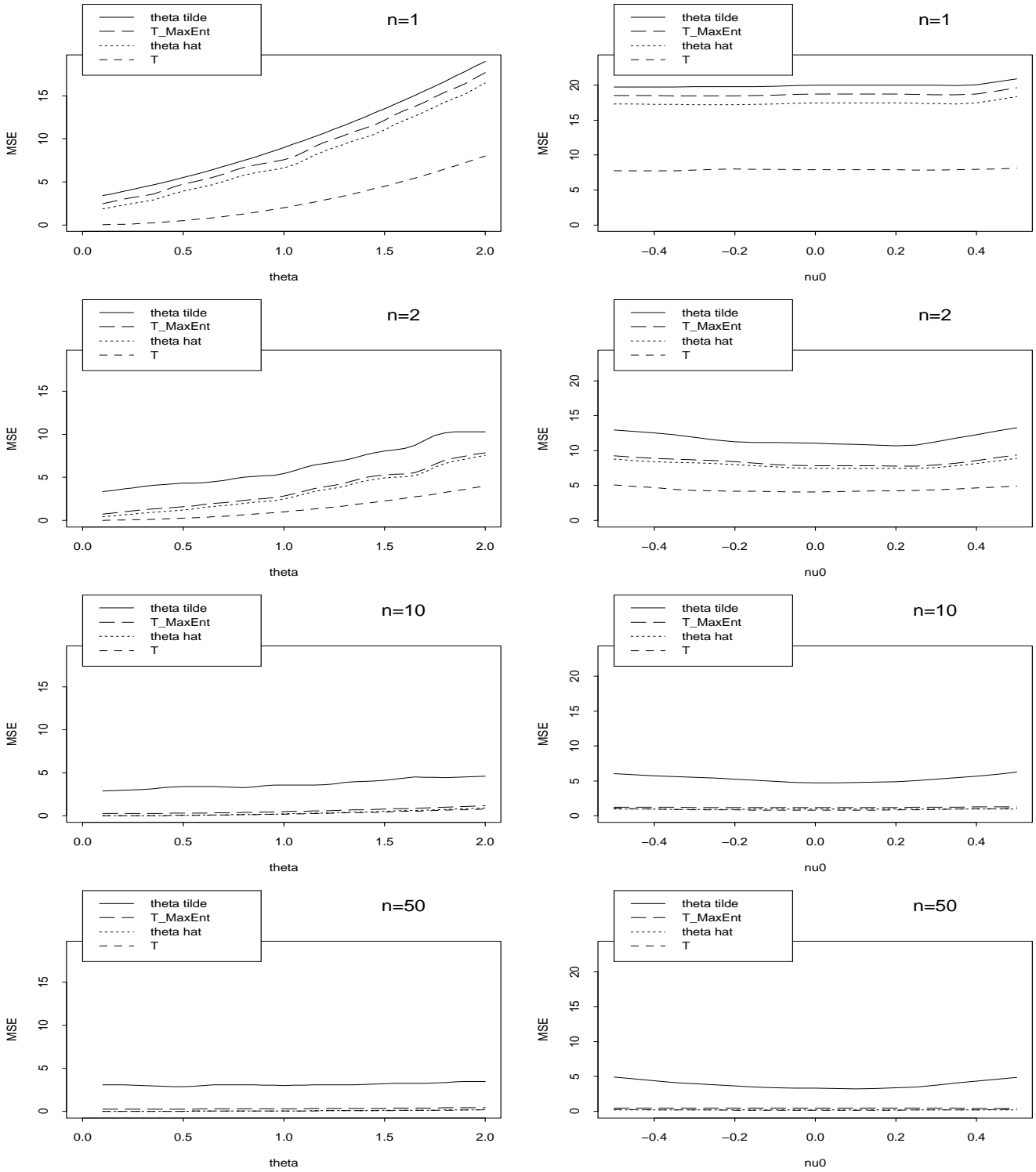
Figure 4: The empirical MSEs of $\widetilde{\theta}$, T_{MaxEnt}, $\widehat{\theta}$, and T wrt θ (left) and ν_0 (right, for $\theta = 2$) for different sample sizes n.

Table 1: Comparing estimators of variance in four different situations

Assumptions	pdf of $X\|\theta$ based on prior information MLE of θ	Simulated data pdf $MSE(\theta) = E(MLE - \theta)^2$
Known parameter $\nu = \nu_0$	$\mathcal{N}(x;\ \nu_0, \theta)$ $T = (X - \nu_0)^2$	$\mathcal{N}(x;\ 0, \theta)$ $2\theta^2$
Known prior $f_\mathcal{V}(\nu) = \mathcal{N}(\nu;\ \nu_0, \theta_0)$	$\mathcal{N}(x;\ \nu_0, \theta + \theta_0)$ $\widehat{\theta} = \max\{(X - \nu_0)^2 - \theta_0, 0\}$	$\mathcal{N}(x;\ 0, \theta + 1)$ $E(\widehat{\theta} - \theta)^2$
Known moments $E(\mathcal{V}) = \nu_0,\ V(\mathcal{V}) = \frac{\theta_0}{2}$	$\mathcal{N}\left(x;\ \nu_0, \theta + \frac{\theta_0}{2}\right)$ $T_{MaxEnt} = \max\{(X - \nu_0)^2 - \frac{\theta_0}{2}, 0\}$	$\mathcal{N}(x;\ 0, \theta + 1)$ $E(T_{MaxEnt} - \theta)^2$
Known unique median $Median(\mathcal{V}) = \nu_0$	$\mathcal{N}(x;\ \nu_0, \theta)$ $\widetilde{\theta} = (X - \nu_0)^2$	$\mathcal{N}(x;\ 0, \theta + 1)$ $2(\theta + 1)^2 + 1$

4 Extensions

In this section, we show that the suggested new tool can be extended to other functions such as quantiles instead of median, but not to other functions such as mode. For example, mode of the random variable $T = T(\mathcal{V}; \boldsymbol{x}) = F_{\boldsymbol{X}|\mathcal{V}}(\boldsymbol{x}|\mathcal{V})$ in Definition 1, i.e.,

$$Mod(T) = \arg\max_t f_T(t), \tag{7}$$

is not a cdf in Example A. The mode of T is: (see Figure 1 top)

$$Mod(T) = \begin{cases} 0 & 0 < x < 1 \\ \forall\ k \in [0,1] & x = 1 \\ 1 & x > 1 \end{cases},$$

which is not a distribution function. If we assume $k = 1$, then $Mod(T)$ is a degenerate cdf. In Figure 5 we plot the mean, median and mode of the random variable T. We see that they are cdfs. However, the cdf based on mode is the extreme case of the two others.

As noted by one of the referees, the mode of prior pdf is useful for introducing a pseudo cdf similar to our new inference tool, $\widetilde{F}_{\boldsymbol{X}}(\boldsymbol{x})$. That is, instead of using the result of Theorem 2: $\widetilde{F}_{\boldsymbol{X}|\theta}(\boldsymbol{x}|\theta) = F_{\boldsymbol{X}|\nu,\theta}(\boldsymbol{x}|Med(\mathcal{V}), \theta)$, using $\widetilde{F}_{\boldsymbol{X}|\theta}^{Mod}(\boldsymbol{x}|\theta) = F_{\boldsymbol{X}|\nu,\theta}(\boldsymbol{x}|Mod(\mathcal{V}), \theta)$. This method was used for eliminating the nuisance parameter ν. In this case, Theorem 3, i.e. separability property of pseudo marginal distribution, also holds for $\widetilde{F}_{\boldsymbol{X}|\theta}^{Mod}(\boldsymbol{x}|\theta)$. Note that, the mode of the random variable T, defined in (7) is not equal to $\widetilde{F}_{\boldsymbol{X}|\theta}^{Mod}(\boldsymbol{x}|\theta)$ and may not be a cdf similar to the above
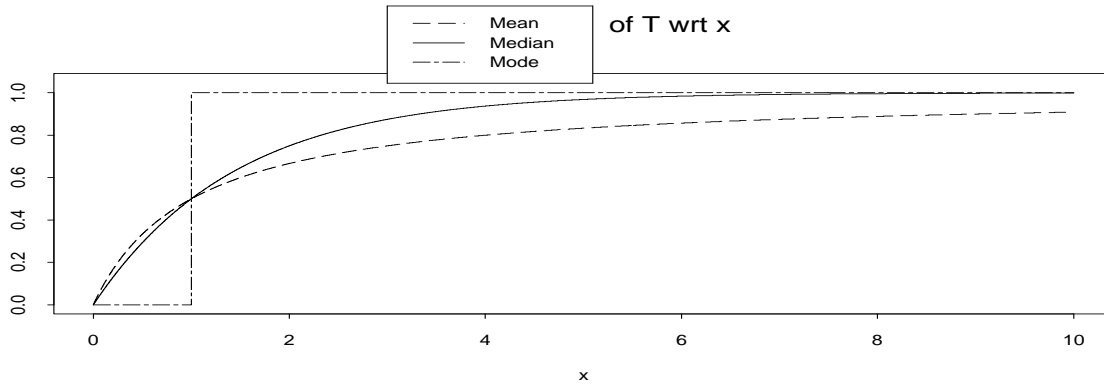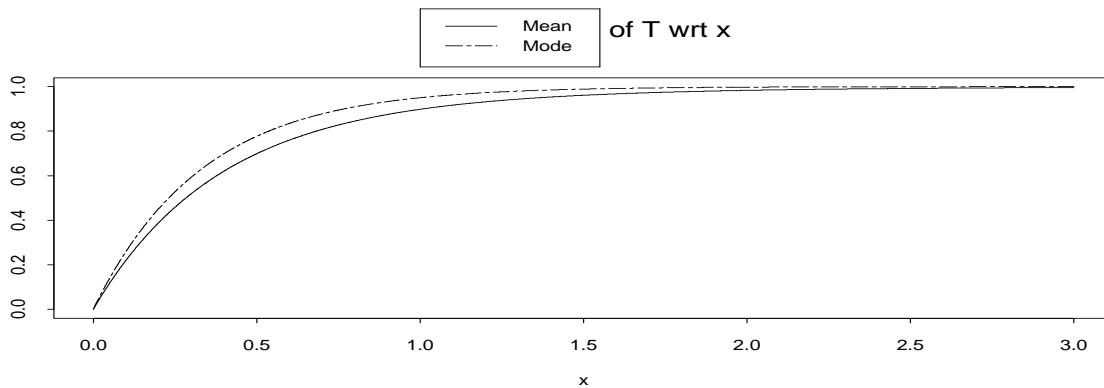
Figure 5: Mean, median and mode of random variable $T = T(\mathcal{V}; x) = F_{X|\mathcal{V}}(x|\mathcal{V}) = 1 - e^{-\mathcal{V}x}$ wrt x.

illustration. However, it may be a cdf similar to the following example pointed out by the referee. In Example A, let $\mathcal{V} - 1$ be a binomial distribution with parameters $\mathcal{B}(2, \frac{3}{4})$, i.e. $\mathcal{V}$ is a discrete random variable with support $\{1, 2, 3\}$. Then $E(T) = 1 - (e^{-x} + 6e^{-2x} + 9e^{-3x})/16$ and $Mod(T) = 1 - e^{-3x}$ are cdfs see Figure 6.

Figure 6: Mean and mode of random variable $T = T(\mathcal{V}; x) = F_{X|\mathcal{V}}(x|\mathcal{V}) = 1 - e^{-\mathcal{V}x}$ wrt x.

On the other hand, we may extend the method presented in this paper to the class of quantiles (e.g., quartiles or percentiles). To make our point clear we consider the first and third quartiles of random variable T in Example A (instead of median, which is the second quartile). We denote the new inference tools based on first and third quartiles by $\widetilde{F}_{\boldsymbol{X}|\theta}^{Q_1}(\boldsymbol{x})$ and $\widetilde{F}_{\boldsymbol{X}|\theta}^{Q_3}(\boldsymbol{x})$ respectively. They can be calculated such as (2) by

$$P(F_{\boldsymbol{X}|\mathcal{V}}(\boldsymbol{x}|\mathcal{V}) \leq \widetilde{F}_{\boldsymbol{X}}^{Q_1}(\boldsymbol{x})) = \frac{1}{4} \quad \text{and} \quad P(F_{\boldsymbol{X}|\mathcal{V}}(\boldsymbol{x}|\mathcal{V}) \leq \widetilde{F}_{\boldsymbol{X}}^{Q_3}(\boldsymbol{x})) = \frac{3}{4}.$$

It can be shown that, in Example A, $\widetilde{F}_X^{Q_1}(x) = 1 - e^{x \ln 0.75}$ and $\widetilde{F}_X^{Q_3}(x) = 1 - e^{x \ln 0.25}$. In Figure 7
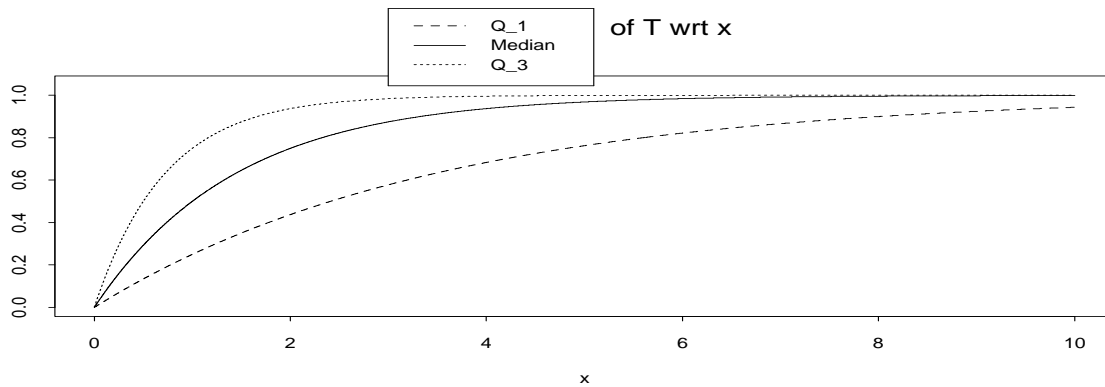
we plot them.

Figure 7: Q_1, median and Q_3 of random variable $T = T(\mathcal{V}; x) = F_{X|\mathcal{V}}(x|\mathcal{V}) = 1 - e^{-\mathcal{V}x}$ wrt x.

In conclusion, it seems that the method can be extended to any quantiles instead of median, but its extension to other functions may need more care.

5   Conclusion

In this paper we considered the problem of inference on one set of parameters of a continuous probability distribution when we have some partial information on a nuisance parameter. We considered the particular case when this partial information is only the knowledge of the median of the prior and proposed a new inference tool which looks like the marginal cdf (or pdf) but its expression needs only the median of the prior. We gave precise definition of this new tool, studied some of its main properties, compared its application with classical marginal likelihood in a few examples, and finally gave an example of its usefulness in parameter estimation.

References

[1] Berger, J. O., 1980. Statistical Decision Theory: Foundations, Concepts, and Methods. Springer, New York.

[2] Bernardo, J. M., Smith, A. F. M., 1994. Bayesian Theory. Wiley, Chichester, UK.

[3] Hernández Bastida, A., Martel Escobar, M. C., Vázquez Polo, F. J., 1998. On maximum entropy priors and a most likely likelihood in auditing. Qüestiió 22, 2, 231–242.

[4] Jaynes, E. T., 1957. Information theory and statistical mechanics I,II. Physical review 106, 620–630 and 108, 171–190.

[5] Jaynes, E. T., 1968. Prior probabilities. IEEE Transactions on Systems Science and Cybernetics SSC-4 (3), 227–241.

[6] Lehmann, E. L., Casella, G., 1998. Theory of point estimation. 2nd ed, Springer, New York.

[7] Mohammad-Djafari, A., Mohammadpour, A., 2004. On the estimation of a parameter with incomplete knowledge on a nuisance parameter. Vol. 735. AIP, pp. 533–540.

[8] Mohammadpour, A., Mohammad-Djafari, A., 2004. An alternative criterion to likelihood for parameter estimation accounting for prior information on nuisance parameter. In: Soft methodology and Random Information Systems. Springer, Berlin, pp. 575–580.

[9] Mohammadpour, A., Mohammad-Djafari, A., 2004. An alternative inference tool to total probability formula and its applications. Vol. 735. AIP, pp. 227–236.

[10] Robert, C. P., Casella, G., 2004. Monte Carlo statistical methods. 2nd ed, Springer, New York.

[11] Rohatgi, V. K., 1976. An Introduction to Probability Theory and Mathematical Statistics. Wiley, New York.

[12] Zacks, S., 1981. Parametric statistical inference. Pergamon, Oxford.