Maximum Penalized Likelihood Estimation for the Reliability Function

A.Mousa [a] , M.khalil [b], N. Said [b], A.Fathi [b,*]
[a] Institute of statistical studies, research,
Cairo University, Egypt
[b] Department of Mathematics, Faculty of Education,
Ain Shams University, Egypt

Abstract

This paper considers the problem of estimating probability density function based on maximum penalized likelihood estimation, we will review some of the previous studies on Maximum Penalized Likelihood Estimation (MPLE) approaches. Finally, a comparative study using maximum penalized likelihood estimation for reliability parameter based on two-parameter exponential distribution that supported by simulation study.

Keywords: maximum penalized likelihood estimation; reliability parameter; two-parameter exponential distribution.

1 Introduction

The maximum penalized likelihood approach was first applied to density estimation by Good and Gaskins (1971, 1972). They suggested enforcing nonnegative by operating on the square root of the density $\gamma = f^{1/2}$ and then squaring the result. They proposed the penalties

$$\Phi_1(f) = \alpha \int_{-\infty}^{\infty} \frac{\acute{f}(u)^2}{f(u)}\, du, \qquad (1)$$

and

$$\Phi_2(f) = 4\alpha \int_{-\infty}^{\infty} \acute{\gamma}(u)^2\, du + \beta \int_{-\infty}^{\infty} \gamma''(u)^2\, du, \qquad (2)$$

where α and β are the smoothing parameters.

De Montricher, Tapia, and Thompson (1975) examined in detail the properties of the estimators, including existence and uniqueness, and showed that the MPLE based on a penalty involving derivatives is a spline with knots at the order statistics. Klonias (1982, 1984) examined a general class of penalized likelihood estimators and suggested a smoothing parameter (S) based on cross-validation. Cox and O'Sullivan (1990) provided asymptotic analysis of penalized likelihood estimators.

Penalized likelihood estimation can be viewed as a Bayesian approach, with the prior for the density having the form $exp[-\Phi(f)]$ and the posterior mode being the final estimate. Good and Gaskins (1980) used this Bayesian framework to suggest a way to evaluate the importance of individual modes ("bump-hunting") through the logarithm of the Bayes factor on the odds that the bumps would be present in a sample of infinite size.

In order to avoid computational difficulties, Scott, Tapia, and Thompson (1980) converted the penalized likelihood to one on discrete data by binning the observations (see also Tapia and Thompson, 1978, reprinted in Thompson and Tapia, 1990). They called this the discrete maximum penalized likelihood estimator (DMPLE) and gave conditions where the DMPLE converges to the MPLE as the bins narrow. Granville and Rasson (1995) also proposed binning the observations, and they examined an approximation to the MPLE based on a Taylor series expansion around a uniform set of binned counts. Ghorai and Rubin (1979), Good and Gaskins (1980), Ishiguro and Sakamoto (1984), and Klonias and Nash (1987) discussed other MPLE computational methods.

The methods discussed so far are all derived in an ad hoc way from the definition of a density. It is interesting to ask whether it is possible to apply standard statistical techniques, like maximum likelihood, to density estimation. The likelihood of a curve (g) as density underlying a set of independent identically distributed observations is given by (see silverman (1986))

$$L\,(g|X_1,\dots,X_n)=\prod_{i=1}^{n} g(X_i)\,. \qquad (3)$$

This likelihood has no finite maximum over the class of all densities. To see this, let $\widehat{f_h}$ be the naive density estimate with window width $1/2h$; then, for each i,

$$\widehat{f_h}(X_i) \geq \frac{1}{nh},$$

and so

$$\prod \widehat{f_h}(X_i) \geq n^{-n}h^{-n} \to \infty \qquad \text{as} \quad h \to 0.$$

Thus the likelihood can be made arbitrarily large by taking densities approaching the sum of delta functions $\omega(\mathrm{x}) = \frac{1}{n}\sum_{i=1}^{n} \delta(x - X_i)$, and it is not possible to use maximum likelihood directly for density estimation without placing restrictions on the class of densities over which the likelihood is to be maximized.

There are, nevertheless, possible approaches related to maximum likelihood. One method is to incorporate into the likelihood a term which describes the roughness in some sense of the curve under consideration. Suppose $R\,(g)$ is a functional which quantifies the roughness of g. One possible choice of such a functional is

$$R(g) = \int_{-\infty}^{\infty} (g'')^2\,.$$

Define the penalized log likelihood by

$$L_\alpha(g) = \sum_{i=1}^{n} \log g(X_i) - \alpha\, R(g), \qquad (4)$$

where α is a positive smoothing parameter.

The penalized log likelihood can be seen as a way of quantifying the conflict between smoothness and goodness-of-fit to the data, since the log likelihood term $\sum \log g(X_i)$ measures how well g fits the data. The probability density function $\hat{f}$ is said to be *a* maximum penalized likelihood density estimate if it maximizes $L_\alpha(g)$ over the class of all curves g which satisfy $\int_{-\infty}^{\infty} g = 1$, $g(x) \geq 0$ for all x, and $R(g) < \infty$. The parameter α controls the amount of smoothing since it determines the rate of exchange between smoothness and goodness-of-fit; the smaller the value of α the rougher - in terms of $R(\hat{f})$ - will be the corresponding maximum penalized likelihood estimator.

In this paper a review of some previous studies on maximum penalized likelihood estimation (MPLE) approaches in section 2, section 3 presented the maximum penalized likelihood estimator for the reliability parameter R based on two-parameter exponential distribution, simulation results are presented in section 4.

2 Maximum penalized likelihood estimation (MPLE)

A fundamental problem in statistics is to determine the unknown distribution of a random variable. Consider the following situation: let X be a real valued random variable with unknown probability density function f assumed to exist. It is required to construct an estimator $\hat{f}$ for f based on a sample of n independent observations of X. Since f is not assumed to be known up to the value of a finite dimensional parameter. Large number of papers on this topic, and many different methods are used to obtain a reasonable estimate for a nonparametric case.

Given independent observations $x_1, x_2, \dots, x_n$ of $X \sim F$, with density $f(x) = \frac{d}{dx} F(x)$ on $[a, b]$ (i.e., P $[a \leq X \leq b] = 1$) and therefore, $x_i \in [a, b]\ \forall i$, $-\infty \leq a < b \leq \infty$, the goal is to estimate the true f, or F. Sometimes, we assume that f is (twice continuously) differentiable, and define

$$l(x) = \log f(x),$$

and

$$\acute{l}(x) = \frac{d}{dx} l(x) = \frac{\acute{f}(x)}{f(x)}, \qquad (5)$$

where $\acute{l}(x)$ is the score function.

To estimate f (or F or l, equivalently), maximize the penalized likelihood criterion,

$$\max_{f \in \mathcal{F}} \sum_{i=1}^{n} \log f(x_i) - \lambda \widetilde{\Phi}(f), \quad \text{or} \quad \max_{l \in \mathcal{L}} \sum_{i=1}^{n} l(x_i) - \lambda \Phi(l), \qquad (6)$$

where $\mathcal{F}$ and $\mathcal{L}$ are appropriate function classes, $f\colon [a, b] \longrightarrow \mathbb{R}^+$, i.e., $f(x) \geq 0$ for all x, or $l\colon [a, b] \longrightarrow \mathbb{R}$, respectively, with the property

$$\int_a^b f(t)dt = \int_a^b e^{l(t)}\, dt = 1, \qquad (7)$$

$\sum_i l(x_i)$ is the log likelihood, $\Phi : \mathcal{L} \rightarrow \mathbb{R}^+$ or $\widetilde{\Phi}(f) \equiv \Phi(\log f) = \Phi(l)$ is the roughness penalty, and $\lambda \geq 0$ is the smoothing parameter.

Often, the null space of Φ, $\Phi_{\perp} := \{l \mid \Phi(l)= 0\}$, is finite dimensional. This is especially attractive, since, the limiting case, $\lambda \to \infty$ gives the "most smooth" solution, which is equivalent to classical Maximum likelihood estimation in $\Phi_{\perp}$, see, Silverman (1982,1986), Cox and O'Sullivan (1990) and Gu and Qiu (1993) whom developed a nice theory, deriving existence and uniqueness results for a wide class of MPL problems, including speed of convergence and consistency in various norms.

2.1 Penalizing $\sqrt{f}$

Good and Gaskins (1971) used the roughness penalty $\widetilde{\Phi}_1(f) = \int_a^b \frac{f'^2(t)}{f(t)} dt$, the Fisher information which can be written as $\widetilde{\Phi}_1(f) = 4\int_a^b u'^2(t)dt$ where $u = \sqrt{f}$. A second proposal was to generalize the problem to penalties $\widetilde{\Phi}_2(f) = \alpha \int_a^b u'^2(t)dt + \beta \int_a^b u''^2(t)dt$. De Montricher, Tapia and Thompson (1975) derived exact existence results for the proposed estimators of Good and Gaskins (1971), and were able to characterize the first one as "exponential spline", see also Thompson and Tapia (1990).

However, the resulting curve has "kinks", since the derivative f' is discontinuous at every data point (Silverman, 1986).Whereas the minimizer for the $\widetilde{\Phi}_2$ problem will be smoother, it is delicate to be computed, because $u(x) \geq 0$ is necessary (De Montricher, Tapia and Thompson (1975)).

The penalty $\widetilde{\Phi}_3(f) = \int_a^b {f^{(s)}}^2(t)dt$ (under $f^{(j)}(a) = f^{(j)}(b) = 0$ for $j = 0, 1, \ldots, s - 1$) is considered in De Montricher Tapia and Thompson (1975), where the authors proved the existence and uniqueness, and (for $s = 1$) provided an approximating solution, using discretization.

All these approaches have the drawback that the "most smooth" solution is problematic, since the space $\{f; \widetilde{\Phi}(f) = 0; f \geq 0\}$ is not well characterized or even degenerate.

2.2 Penalizing $\log f$

Silverman (1982, 1986) introduced the penalty $\Phi(l) = \int_a^b l'''^2(t)\, dt$ and proved the consistency in three different norms. Silverman also proved that the solution of the constrained MPL problem (6), (7) is equivalent to solving the unconstrained problem

$$\max_{l\in\mathcal{L}} \sum_{i=1}^{n} l(x_i) - \lambda\Phi(l) - n\int_a^b e^{l(t)}dt \tag{8}$$

for a very general class of penalties.

The choice of penalty here leads to the attractive feature that the smoothest limits $(\lambda \longrightarrow \infty)$ are in $\Phi_{\perp} = \{l''' = 0\} = \{l \ quadratic\}$ which are exactly the Gaussian distributions and $(\lambda \longrightarrow \infty)$corresponds to normal MLE. This feature is analogous to the cubic smoothing splines in regression which leads to least squares linear regression for$(\lambda \longrightarrow \infty)$, and is a property which vastly used kernel density estimates.

A related from several standpoints very appealing approach is to estimate (and penalize) the score function $\psi = -l' = \frac{-f'}{f}$(which is a straight line for a Gaussian).
Cox (1985) introduced and solved a penalized "mean square error" problem for the score function, and Ng (1994) provided further properties and computational algorithms.
The logspline approach of Kooperberg and Stone (1991) is an attractive practical approach for using cubic splines to model the log density. However, it is not an MPLE, but rather a MLE in carefully chosen space of "regression splines" (splines with knots determined by the data).
Many authors used MPLE such as Virginie, Daniel and Pierre (2003), used MPLE in a gamma frailty model, Adelchi and Reinaldo (2012), used MPLE for skew-normal and skew-t distributions and Mengjie (2013) used MPLE of two parameters exponential distribution.

3 The maximum penalized likelihood estimator for the reliability parameter *R* based on two-parameter exponential distribution

3.1 Two-parameter exponential distribution

Consider a random variable X having two-parameter exponential distribution $EXP(\theta, \mu)$, with probability density function (pdf) given by

$$f(x; \theta, \mu) = \begin{cases} \frac{1}{\theta} e^{-(x-\mu)/\theta} & x \geq \mu, \\ 0 & otherwise, \end{cases} \qquad (9)$$

where $\theta > 0$ is a scale parameter and $\mu \in \mathbb{R}$ is a location parameter, the cumulative distribution function (CDF) is

$$F(x) = P[X \leq x] = \int_{\mu}^{x} \frac{1}{\theta} e^{-\frac{x-\mu}{\theta}} dx = 1 - e^{-\frac{x-\mu}{\theta}} \;, \; x \geq \mu. \qquad (10)$$

The two-parameter exponential distribution has many real applications. It can be used to model the data such as the service time of agents in a system (Queuing Theory), the time it takes before your next telephone call, the time until a radioactive particle decays, the distance between mutations on a DNA strand, and the extreme values of annual snowfall or rainfall. According to Kotz, Lumelskii and Pensky (2003), the case of the two-parameter exponential distributions is of importance because it allows us to derive confidence limits for the reliability parameters involving Pareto distributions or power distributions by means of one-one transformations.

3.2 The stress-strength reliability parameter based on two-parameter exponential distribution

Let X ~Exponential(θ_1, μ_1) independently of Y ~Exponential(θ_2, μ_2). That is, the pdf of X is $f(x; \theta_1, \mu_1)$ and the pdf of Y is $f(y; \theta_2, \mu_2)$, where f is given in equation (9).

Then the reliability parameter $R = P(X > Y)$ can be expressed as, (see Krishnamoorthy and Mukherjee (2006))

$$R = \left(1 - \frac{\theta_2 e^{\frac{\mu_2-\mu_1}{\theta_2}}}{\theta_1+\theta_2}\right) I(\mu_1 > \mu_2) + \left(\frac{\theta_1 e^{\frac{\mu_1-\mu_2}{\theta_1}}}{\theta_1+\theta_2}\right) I(\mu_1 \leq \mu_2) \tag{11}$$

where $I(\cdot)$ is the indicator function.

There are three cases of the previous equation:

i. If $\mu_1 = \mu_2$ then the reliability parameter $R = \frac{\theta_1}{\theta_1+\theta_2}$

ii. If $\mu_1 > \mu_2$ then the reliability parameter $R = 1 - \frac{\theta_2 e^{\frac{\mu_2-\mu_1}{\theta_2}}}{\theta_1+\theta_2}$

iii. If $\mu_1 < \mu_2$ then the reliability parameter $R = \frac{\theta_1 e^{\frac{\mu_1-\mu_2}{\theta_1}}}{\theta_1+\theta_2}$

3.3 A maximum penalized likelihood estimation for reliability parameter R

Let X be an exponential random variable with pdf $f(x; \theta_1, \mu_1)$ and Y be an exponential random variable with pdf $f(y; \theta_2, \mu_2)$, where the pdf's are as defined in equation (9). Assume that X and $Y$are independent. Let $x_1, \dots, x_n$ be a sample of observations on X and $y_1, \dots, y_m$ be a sample of observations on Y. Furthermore, let $\hat{\theta}_1, \hat{\mu}_1$ denote the MPLE of θ_1, μ_1 respectively based on X observations, and let $\hat{\theta}_2, \hat{\mu}_2$ denote the MPLE of θ_2, μ_2 respectively based on Y observations.

To find the MPLE of R, we use the following algorithm:

1) For i.i.d sample $x_1, \dots, x_n$ and i.i.d sample $y_1, \dots, y_m$ with pdf's as (9), the penalized likelihood function is (see Mengjie (2013))

$$L(\theta_1, \theta_2, \mu_1, \mu_2) = (x_{(1)} - \mu_1)(y_{(1)} - \mu_2) \prod_{i=1}^{n} f(x_i | \theta_1, \mu_1) \prod_{j=1}^{m} f(y_j | \theta_2, \mu_2)$$

$$= (x_{(1)} - \mu_1)(y_{(1)} - \mu_2) \frac{1}{\theta_1^{\,n}} e^{-\frac{1}{\theta_1}\sum_{i=1}^{n}(x_i-\mu_1)} \frac{1}{\theta_2^{\,m}} e^{-\frac{1}{\theta_2}\sum_{j=1}^{m}(y_j-\mu_2)}$$

$$, x_{(1)} \geq \mu_1 ,\ y_{(1)} \geq \mu_2 , \tag{12}$$

where $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ are order statistics based on $x_1, x_2, \dots, x_n$ and $x_{(1)}$ is the minimum of the sample, $y_{(1)} \leq y_{(2)} \leq \dots \leq y_{(m)}$ are order statistics based on $y_1, y_2, \dots, y_m$ and $y_{(1)}$ is the minimum of the sample.

2) Take the logarithm of likelihood function, we have,

$$\ln L(\theta_1,\theta_2,\mu_1,\mu_2) = \ln\left(x_{(1)}-\mu_1\right) - n\ln\theta_1 - \frac{1}{\theta_1}\sum_{i=1}^{n}(x_i-\mu_1)$$
$$+\ln\left(y_{(1)}-\mu_2\right) - m\ln\theta_2 - \frac{1}{\theta_2}\sum_{j=1}^{m}(y_j-\mu_2)$$
$$,x_{(1)} \geq \mu_1\,,\ y_{(1)} \geq \mu_2 \qquad (13)$$

3) Differentiate the logarithm of the likelihood function with respect to $\theta_1,\theta_2,\mu_1,\mu_2$ respectively, and set the derivatives equal to 0.

$$\frac{d\ln L(\theta_1,\theta_2,\mu_1,\mu_2)}{d\theta_1} = -\frac{n}{\theta_1} + \frac{\sum_{i=1}^{n}(x_i-\mu_1)}{{\theta_1}^2} = 0$$

$$\frac{d\ln L(\theta_1,\theta_2,\mu_1,\mu_2)}{d\theta_2} = -\frac{m}{\theta_2} + \frac{\sum_{i=1}^{m}(y_i-\mu_2)}{{\theta_2}^2} = 0$$

$$\frac{d\ln L(\theta_1,\theta_2,\mu_1,\mu_2)}{d\mu_1} = \frac{n}{\theta_1} - \frac{1}{x_{(1)}-\mu_1} = 0$$

$$\frac{d\ln L(\theta_1,\theta_2,\mu_1,\mu_2)}{d\mu_2} = \frac{m}{\theta_2} - \frac{1}{y_{(1)}-\mu_2} = 0$$

4) The MPLEs for $\theta_1,\theta_2,\mu_1,\mu_2$ are

$$\hat{\theta}_1 = \frac{n(\bar{x}-x_{(1)})}{n-1} \qquad (14)$$
$$\hat{\theta}_2 = \frac{m(\bar{y}-y_{(1)})}{m-1} \qquad (15)$$
$$\hat{\mu}_1 = \frac{nx_{(1)}-\bar{x}}{n-1} \qquad (16)$$
$$\hat{\mu}_2 = \frac{my_{(1)}-\bar{y}}{m-1} \qquad (17)$$

5) The MPLE of the reliability parameter R can be obtained by replacing the parameters $\theta_1,\theta_2,\mu_1,\mu_2$ in R by their MPLE's. That is, the MPLE of R is given by

$$\hat{R} = \left(1-\frac{\hat{\theta}_2 e^{\frac{\hat{\mu}_2-\hat{\mu}_1}{\hat{\theta}_2}}}{\hat{\theta}_1+\hat{\theta}_2}\right) I(\hat{\mu}_1 > \hat{\mu}_2) + \left(\frac{\hat{\theta}_1 e^{\frac{\hat{\mu}_1-\hat{\mu}_2}{\hat{\theta}_1}}}{\hat{\theta}_1+\hat{\theta}_2}\right) I(\hat{\mu}_1 \leq \hat{\mu}_2) \qquad (18)$$

3.4 A maximum likelihood estimation for reliability parameter R

Let X be an exponential random variable with pdf $f(x;\theta_1,\mu_1)$ and Y be an exponential random variable with pdf $f(y;\theta_2,\mu_2)$, where the pdf's are as defined in equation (9). Assume that X and $Y$are independent. Let $x_1,\dots,x_n$ be a sample of observations on X and $y_1,\dots,y_m$ be a sample of observations on Y. Furthermore, let $\hat{\theta}_1^*,\hat{\mu}_1^*$ denote the MLE of θ_1,μ_1 respectively based on X observations, and let $\hat{\theta}_2^*,\hat{\mu}_2^*$ denote the MLE of θ_2,μ_2 respectively based on Y observations.

To find the MLE of R, we use the following algorithm:

1) For i.i.d sample $x_1,\dots,x_n$ and i.i.d sample $y_1,\dots,y_m$ with pdf's as (9), the likelihood function is (see Mengjie (2013))

$$L^*(\theta_1,\theta_2,\mu_1,\mu_2)=\prod_{i=1}^{n} f(x_i|\,\theta_1,\mu_1)\prod_{j=1}^{m} f(y_j|\,\theta_2,\mu_2)$$
$$=\frac{1}{\theta_1{}^n}e^{-\frac{1}{\theta_1}\sum_{i=1}^{n}(x_i-\mu_1)}\frac{1}{\theta_2{}^m}e^{-\frac{1}{\theta_2}\sum_{j=1}^{m}(y_j-\mu_2)}$$
$$,x_{(1)}\geq\mu_1\,,\;y_{(1)}\geq\mu_2 \tag{19}$$

2) Take the logarithm of likelihood function, we have,

$$\ln L^*(\theta_1,\theta_2,\mu_1,\mu_2)=-\,n\ln\theta_1-\frac{1}{\theta_1}\sum_{i=1}^{n}(x_i-\mu_1)-m\ln\theta_2-\frac{1}{\theta_2}\sum_{j=1}^{m}(y_j-\mu_2)$$
$$,x_{(1)}\geq\mu_1\,,\;y_{(1)}\geq\mu_2 \tag{20}$$

3) Differentiate the logarithm of the likelihood function with respect to $\theta_1,\theta_2,\mu_1,\mu_2$ respectively, by taking $\hat{\mu}_1^*=x_{(1)}$, $\hat{\mu}_2^*=y_{(1)}$ and set the derivatives equal to 0.

$$\frac{d\ln L^*(\theta_1,\theta_2,\mu_1,\mu_2)}{d\theta_1}=-\frac{n}{\theta_1}+\frac{\sum_{i=1}^{n}(x_i-\mu_1)}{\theta_1{}^2}=0$$

$$\frac{d\ln L^*(\theta_1,\theta_2,\mu_1,\mu_2)}{d\theta_2}=-\frac{m}{\theta_2}+\frac{\sum_{i=1}^{m}(y_i-\mu_2)}{\theta_2{}^2}=0$$

4) The MLEs for $\mu_1,\mu_2,\theta_1,\theta_2$ are

$$\hat{\mu}_1^*=x_{(1)} \tag{21}$$
$$\hat{\mu}_2^*=y_{(1)} \tag{22}$$
$$\hat{\theta}_1^*=\bar{x}-x_{(1)} \tag{23}$$
$$\hat{\theta}_2^*=\bar{y}-y_{(1)} \tag{24}$$

5) The MLE of the reliability parameter R can be obtained by replacing the parameters $\theta_1, \theta_2, \mu_1, \mu_2$ in R by their MLE's. That is, the MLE of R is given by

$$\hat{R}^* = \left(1 - \frac{\hat{\theta}_2^{\,*} e^{\frac{\hat{\mu}_2^{\,*}-\hat{\mu}_1^{\,*}}{\hat{\theta}_2^{\,*}}}}{\hat{\theta}_1^{\,*}+\hat{\theta}_2^{\,*}}\right) I(\hat{\mu}_1^{\,*} > \hat{\mu}_2^{\,*}) + \left(\frac{\hat{\theta}_1^{\,*} e^{\frac{\hat{\mu}_1^{\,*}-\hat{\mu}_2^{\,*}}{\hat{\theta}_1^{\,*}}}}{\hat{\theta}_1^{\,*}+\hat{\theta}_2^{\,*}}\right) I(\hat{\mu}_1^{\,*} \le \hat{\mu}_2^{\,*}) \tag{25}$$

4 Simulation Study

A simulation study was carried out and designed to investigate the performance of the MPLE of the reliability parameter R when $\mu_1 > \mu_2$ by compare it with the MLE of the reliability parameter R.

The data were generated according to two-parameter exponential distribution for sample sizes (n, m) where $n = 30,70,90,120$ and $m = 30,70,90,120$. All computations are performed using Mathematica10 as follows:

1. For given parameters compute the true value of R, (see equation 11)
2. Compute MLE and MPLE for parameters.
3. From the invariance property of the maximum estimator, compute MLE and MPLE for R.
4. Repeat step 2 and 3 where the simulation is repeated 1000 times and biases and mean square errors are calculated for each estimator.

Table (1), Table (2), Table (3), Table (4), Table (5) and Table (6) shows the results of a simulation study.

Note that, $R.B(1) = \frac{Bias}{Estimator}$ is Relative Bias with respect to estimator ,

$R.B(2) = \frac{Bias}{true\ value}$ is Relative Bias with respect to true value and

$R.E = \frac{MSE\ of\ MPLE}{MSE\ of\ MLE}$ is Relative Efficiency with respect to MLE.

The results from this simulation according to biases and mean square errors of MPLE of R is smaller than biases and mean square errors of MLE of R for all sizes (n, m), so the MPLE of R is better than the MLE of R.

Table (1)

Biases and MSE when ($\mu_1 = 0.5,\ \mu_2 = 0,\ \theta_1 = 1,\ \theta_2 = 1$ and true value of R=0.696735) for MLE of R

Sample Sizes (n, m)	MLE of *R*	Bias	R.B(1)	R.B(2)	MSE
(30,30)	0.99979	0.30306	0.30312	0.43497	0.09184
(30,70)	0.99940	0.30266	0.30284	0.43440	0.09161
(30,90)	0.99973	0.30299	0.30307	0.43487	0.09181
(30,120)	0.99918	0.30244	0.30269	0.43408	0.09147
(70,70)	0.99927	0.30253	0.30275	0.43421	0.09153
(70,90)	0.99965	0.30291	0.30302	0.43476	0.09176
(70,120)	0.99777	0.30103	0.30170	0.43206	0.09062
(90,90)	0.99853	0.30179	0.30223	0.43315	0.09108
(90,120)	0.99911	0.30237	0.30264	0.43398	0.09143
(120,120)	0.99849	0.30175	0.30221	0.43309	0.09106

Table (2)

Biases and MSE when ($\mu_1 = 0.5,\ \mu_2 = 0,\ \theta_1 = 1,\ \theta_2 = 1$ and true value of R=0.696735) for MPLE of R

Sample Sizes (n, m)	MPLE of R	Bias	R.B(1)	R.B(2)	MSE
(30,30)	0.99973	0.30299	0.30307	0.43487	0.09181
(30,70)	0.99934	0.30260	0.30280	0.43431	0.09157
(30,90)	0.99971	0.30297	0.30306	0.43484	0.09179
(30,120)	0.99912	0.30239	0.30266	0.43401	0.09144
(70,70)	0.99920	0.30247	0.30271	0.43412	0.09149
(70,90)	0.99962	0.30289	0.30301	0.43473	0.09174
(70,120)	0.99766	0.30093	0.30164	0.43191	0.09056
(90,90)	0.99843	0.30169	0.30216	0.43301	0.09102
(90,120)	0.99906	0.30233	0.30261	0.43392	0.09140
(120,120)	0.99842	0.30168	0.30216	0.43299	0.09101

Table (3)

Comparison between MLE and MPLE of R when ($\mu_1 = 0.5,\ \mu_2 = 0,$ $\theta_1 = 1,\ \theta_2 = 1$ and true value of R=0.696735)

MLE R.B(1)	MPLE R.B(1)	MLE R.B(2)	MPLE R.B(2)	R.E
0.30312	0.30307	0.43497	0.43487	0.99967
0.30284	0.30280	0.43440	0.43431	0.99956
0.30307	0.30306	0.43487	0.43484	0.99978
0.30269	0.30266	0.43408	0.43401	0.99967
0.30275	0.30271	0.43421	0.43412	0.99956
0.30302	0.30301	0.43476	0.43473	0.99978
0.30170	0.30164	0.43206	0.43191	0.99934
0.30223	0.30216	0.43315	0.43301	0.99934
0.30264	0.30261	0.43398	0.43392	0.99967
0.30221	0.30216	0.43309	0.43299	0.99945

Table (4)

Biases and MSE when ($\mu_1 = 1,\ \mu_2 = 0,\ \theta_1 = 0.5,\ \theta_2 = 1$ and true value of R=0.754747) for MLE of R

Sample Sizes (n, m)	MLE of R	Bias	R.B(1)	R.B(2)	MSE
(30,30)	0.99977	0.24502	0.24508	0.32464	0.06004
(30,70)	0.99923	0.24449	0.24468	0.32394	0.05977
(30,90)	0.99904	0.24429	0.24452	0.32367	0.05968
(30,120)	0.99852	0.24377	0.24413	0.32298	0.05942
(70,70)	0.99930	0.24456	0.24473	0.32403	0.05981
(70,90)	0.99938	0.24464	0.24479	0.32414	0.05985
(70,120)	0.99898	0.24424	0.24449	0.32361	0.05965
(90,90)	0.99805	0.24330	0.24378	0.32236	0.05920
(90,120)	0.99842	0.24367	0.24406	0.32285	0.05937
(120,120)	0.99798	0.24323	0.24372	0.32227	0.05916

Table (5)

Biases and MSE when ($\mu_1 = 1,\ \mu_2 = 0,\ \theta_1 = 0.5,\ \theta_2 = 1$ and true value of R=0.754747) for MPLE of R

Sample Sizes (n, m)	MPLE of R	Bias	R.B(1)	R.B(2)	MSE
(30,30)	0.99971	0.24496	0.24503	0.32455	0.06000
(30,70)	0.99916	0.24441	0.24462	0.32383	0.05974
(30,90)	0.99897	0.24422	0.24447	0.32358	0.05965
(30,120)	0.99844	0.24370	0.24408	0.32289	0.05939
(70,70)	0.99924	0.24449	0.24468	0.32394	0.05978
(70,90)	0.99934	0.24459	0.24475	0.32407	0.05982
(70,120)	0.99893	0.24418	0.24444	0.32353	0.05963
(90,90)	0.99793	0.24318	0.24368	0.32220	0.05914
(90,120)	0.99834	0.24359	0.24400	0.32274	0.05934
(120,120)	0.99789	0.24314	0.24365	0.32215	0.05912

Table (6)

Comparison between MLE and MPLE of R when ($\mu_1 = 1,\ \mu_2 = 0$, $\theta_1 = 0.5$, $\theta_2 = 1$ and true value of R=0.754747)

MLE R.B(1)	MPLE R.B(1)	MLE R.B(2)	MPLE R.B(2)	R.E
0.24508	0.24503	0.32464	0.32455	0.99933
0.24468	0.24462	0.32394	0.32383	0.99950
0.24452	0.24447	0.32367	0.32358	0.99949
0.24413	0.24408	0.32298	0.32289	0.99949
0.24473	0.24468	0.32403	0.32394	0.99950
0.24479	0.24475	0.32414	0.32407	0.99950
0.24449	0.24444	0.32361	0.32353	0.99966
0.24378	0.24368	0.32236	0.32220	0.99899
0.24406	0.24400	0.32285	0.32274	0.99949
0.24372	0.24365	0.32227	0.32215	0.99932

References

[1] Adelchi, A. and Reinaldo, B. (2012): Maximum Penalized Likelihood Estimation for Skew-Normal and Skew-t Distributions. Stat. ME., 1203. 2376v1.

[2] Cox, D.D. and Sullivan, F. (1990): Asymptotic Analysis of Penalized Likelihood and Related Estimators. Annals of Statistics, 18, 1676-1695.

[3] De Montricher, G.M., Tapia, R.A., and Thompson, J.R. (1975): Nonparametric Maximum Likelihood Estimation of Probability Densities by Penalty Function Methods. Annals of Statistics, 3, 1329-1348.

[4] Ghorai, J.K. and Rubin, H. (1979): Computational Procedure for Maximum Penalized Likelihood Estimate. Journal of Statistical Computation and Simulation, 10, 65-78.

[5] Good, I.J. and Gaskins, R.A. (1971): Nonparametric Roughness Penalties for Probability Densities. Biometrika, 58, 255-277.

[6] Good, I.J. and Gaskins, R.A. (1972): Global Nonparametric Estimation of Probability Densities. Virginia Journal of Science, 23(4), 171-193.

[7] Good, I.J. and Gaskins, R.A. (1980): Density Estimation and bumb- hunting by the Penalized Likelihood Method Exemplified by Scattering and Meteorite Data (with Discussion). J.Amer.Statist.Assoc., 75, 42-73.

[8] Granville, V. and Rasson, J.P. (1995): Multivariate Discriminant Analysis and Maximum Penalized Likelihood Density Estimation. Journal of the Royal Statistical Society, Ser. B. 57, 501-518.

[9] Gu, C. and Qiu, C. (1993): Smoothing Spline Density Estimation: Theory, Annals of Statistics 21(1), 217–234.

[10] Ishiguro, M. and Sakamoto, Y. (1984): A Bayesian Approach to the Probability Density Estimation. Ann. Inst. Statist. Math., 36, 523-538.

[11] Krishnamoorthy, K. Shubhabrata, M. and Huizhen, G. (2006): Inference on Reliability in Two-Parameter Exponential Stress–Strength Model. Springer- New York.

[12] Klonias, V.K. (1982): Consistency of Two Nonparametric Penalized Likelihood Estimators of the Probability Density Function. Annals of Statistics, 10, 811-824.

[13] Klonias, V.K. and Nash, S.G. (1987): Numerical Techniques in Nonparametric Estimation. Journal of Statistical Computation and Simulation, 28, 97-126.

[14] Kooperberg, C. and Stone, C. J. (1991): A Study of Logspline Density Estimation, Computat. Statist. Data Anal., 12, 327–347.

[15] Kotz, S., Lumelskii, Y. and Pensky, M. (2003): The Stress–Strength Model and its Generalizations: Theory and Applications. World Scientific, Singapore.

[16] Mengjie, Z. (2013): Penalized Maximum Likelihood Estimation of Two-Parameter Exponential Distributions. Annals of Statistics.

[17] Ng, P. T. (1994): Smoothing Spline Score Estimation, SIAM J. Sci. Statist. Comput. 15, 1003–1025.
[18] Scott, D.W., Tapia, R.A. and Thompson, J.R. (1980): Nonparametric Probability Density Estimation by Discrete Maximum Penalized Likelihood Criteria. Annals of Statistics, 8, 820-832.
[19] Silverman, B. W. (1986): Density Estimation for Statistics and Data Analysis, Vol. 26 of Monographs on Statistics and Applied Probability, Chapman & Hall, London.
[20] Tapia, R.A. and Thompson, J.R. (1978): Nonparametric Probability Density Estimation, Johns Hopkins University Press, Baltimore, MD.
[21] Thompson, J.R. and Tapia, R.A. (1990): Nonparametric Function Estimation, Modeling, and Simulation, SIAM, Philadelphia.
[22] Virginie, R., Daniel, C. and Pierre, J. (2003): Maximum Penalized Likelihood Estimation in A Gamma-Frailty Model. Lifetime Data Anal, 9 (2), 139-53.