A COMPARISON STUDY OF ESTIMATION METHODS FOR NON-REGULAR DISTRIBUTIONS

by

Nan Zhang

with Professors Yongcheng Qi and Kang James

Department of Mathematics and Statistics
University of Minnesota, Duluth
Duluth, Minnesota 55812

May 13, 2010
A Comparison Study of Estimation Methods for Non-regular Distributions

Nan Zhang
with Professors Yongcheng Qi and Kang James
Department of Mathematics and Statistics
University of Minnesota, Duluth
Duluth, Minnesota 55812
May 13, 2010
Acknowledgments

I would like to take this opportunity to give my sincere thanks to my advisors, Dr. Yongcheng Qi, and Dr. Kang L James, for their support and guidance, which make it possible for me to finish the project in such a limited time.

Thanks also to Dr. Guihua Fei for being on my degree committee.
Contents

1 Introduction 2

2 MPS Method 4

3 Some good properties of MPS estimates 6
   3.1 Sufficiency ......................................................... 6
   3.2 Consistency and Asymptotic efficiency .......................... 6

4 Three simple examples 9
   4.1 Example 1 .......................................................... 9
   4.2 Example 2 .......................................................... 10
   4.3 Example 3 .......................................................... 11

5 Simulation study 13
   5.1 Results ............................................................. 13
   5.2 Conclusion ......................................................... 18

6 Estimating the endpoint (Bayesian likelihood) 19
   6.1 Introduction ....................................................... 19
   6.2 Methodology ...................................................... 20
   6.3 Interpretation of the empirical prior ............................ 20
   6.4 Simulations ......................................................... 21

7 References 24

8 Appendix 26
   8.1 Advanced MPS vs MLE three parameters Weibull distribution 26
   8.2 Bayesian vs MPS endpoint estimates Weibull distribution 28
A Comparison Study of Estimation Methods for Non-regular Distributions

Nan Zhang
with Professor Yongcheng Qi and Professor Kang James
Department of Mathematics and Statistics
University of Minnesota, Duluth
Duluth, MN 55812

Abstract

This project introduces Maximum Product of Spacings method (MPS), a general method of estimating parameters in continuous univariate distributions especially for cases where Maximum Likelihood method fails. It gives consistent estimators with asymptotic efficiency equal to MLE estimators when these exist. Several examples and simulation studies are showing the advantages of this method. Moreover, the comparison between MPS method and Bayesian likelihood method on estimating the endpoints are also presented.

Keywords: Maximum likelihood estimation; Spacings; Weibull distribution; gamma distribution; Bayesian estimation; Simulations.
1 Introduction

Consider the common statistical inference problem of estimating unknown parameters $\Theta$ (maybe a vector) from observed data $x_1, x_2, \cdots, x_n$, where the $x_i$’s are mutually independent observations from a distribution function $F(x, \Theta_0)$. To estimate $\Theta_0$ (true value), one often uses traditional estimation methods such as the method of moments, and maximum likelihood estimation (MLE).

The MLE method has nice properties of being consistent, asymptotically efficient under very general conditions. However, in unbounded likelihood problem like estimation of mixtures of continuous distributions (Lindsay, 1995), heavy-tailed distributions with unknown location and scale parameters (Pitman, 1979), and certain three-parameter distributions where the density is positive only to the right of a shifted origin (Cheng and Amin, 1983), the MLE method does not always give satisfactory results. The problems of the MLE in model fitting were discussed by Weiss and Wolfowitz (1973). Related discussions in connection to the Weibull and the gamma distributions can be found in Cheng and Amin (1983), Smith (1985) and Cheng and Traylor (1995). In details, which cited from Cheng and Amin (1983), for certain three-parameter distributions where density is positive only to the right of a shifted origin, such as Weibull, gamma, and lognormal distributions, the critical difficulty is that there are paths in the parameter space, with location parameter $\theta$ tending to be the smallest observation $x_{1:n}$ as the (global) MLE estimator of $\theta$. Unfortunately, the estimators of the other parameters are then inconsistent.

Though Harter and Moore (1996) have suggested the alternative of seeking a local, as opposed to global, maximum point of the likelihood. This may not be an effective procedure as suffering for two weaknesses. The theoretical weakness is that when the underlying distribution is J-shaped, which occurs in Weibull and gamma cases when the shape parameter $\alpha$ is less than one, Huzurbazar (1948) has shown that no local maximum can yield a consistent estimator. Then the MLE is bound to fail. The practical weakness is that even if the distribution is not J-shaped, so that a parameter can in principle be consistently estimated by local MLE estimation as the sample size tends to infinity, it can happen that, with fixed sample size, a particular random sample gives rise to a likelihood function with no local maximum at all (Griffiths, 1980). The probability of this occurring is high when the shape parameter is close to one, which is often precisely the distribution of practical interest, corresponding as it does to the exponential distribution in both the Weibull and Gamma cases.
Thus, Cheng and Amin (1983) introduced the maximum product of spacings (MPS) method as an alternative to MLE for the estimation of parameters of continuous univariate distributions, which retains the desirable properties of MLE but does not suffer difficulties mentioned above. Ranneby (1984) independently developed the same method as an approximation to the Kullback-Leibler measure of information. A description of the MPS method is introduced in the following section.

The purpose of this project is to compare the finite sample properties of the MPS estimates and MLE estimates by three simple examples and two simulation studies. The distributions we use for comparison are three-parameter Weibull and gamma distributions. Sections 1 & 2 are introduction and methodology of MPS method. Section 3 discusses some good properties of MPS estimates. Examples and simulation studies will be discussed in Sections 4 & 5. After that, Section 6 studies Bayesian likelihood method on estimating endpoints of distributions, together with comparison studies with MPS method. In the end, Section 8 is the Appendix, where I post R codes in programming simulations.
2 MPS Method

Let $\Theta_0$ be the true parameter value and let $x_1 < x_2 < \cdots < x_n$ be an ordered random sample drawn from the distribution $F(x, \Theta_0)$. Define

\[ D_1 = F(x_{1:n}, \Theta), \ D_{n+1} = 1 - F(x_{n:n}, \Theta) \]

\[ D_i = F(x_{1:n}, \Theta) - F(x_{i-1:n}, \Theta), \ i = 2, 3, \cdots, n \]

as the uniform spacings of the sample. Clearly the spacings sum to unity: $\sum D_i = 1$. The MPS method is to choose $\Theta$ to maximize the geometric mean of the spacings

\[ G = \left\{ \prod_{i=1}^{n+1} D_i \right\}^{\frac{1}{n+1}}, \]

or, equivalently, its logarithm

\[ H = \log G. \]

The motivation for maximizing $G$ (or $H$) is that the maximum, which is bounded above because of the condition $\sum D_i = 1$, is obtained only when all the $D_i$’s are equal. Cheng and Amin (1983) showed that maximizing $H$ as a method of parameter estimation is as efficient as MLE estimation and the MPS estimators are consistent under more general conditions than the MLE estimators, especially when the underlying distribution is J-shape, situation where MLE is bound to fail. Additionally, they showed that ties present no problem in parameter estimation and so it seems that it is as good a method of estimation as MLE method.

Before looking in detail at the properties of MPS estimators, here is an intuitive comparison of MLE and MPS estimators to clarify their similarities and differences.

The log likelihood will be written as

\[ L = \frac{1}{n} \sum_{i=1}^{n} \log f(x_{i:n}, \Theta). \]

In comparison, the term in $H$ corresponding to $\log f(x_{i:n}, \Theta)$ is

\[ \log D_i = \log \int_{x_{i-1:n}}^{x_{i:n}} f(x, \Theta)dx = \log[f(x_{i:n}, \Theta)(x_{i:n} - x_{i-1:n})] + R(x_{i:n}, x_{i-1:n}, \Theta) \]
\[ = \log f(x_{i:n}, \Theta) + \log(x_{i:n} - x_{i-1:n}) + R(x_{i:n}, x_{i-1:n}, \Theta). \]

For standard situations, where the end points of the support of \( f \) are known, \( R \), though dependent on \( \Theta \), is essentially of order \( O(|x_{i:n} - x_{i-1:n}|) \). For most of terms, \( |x_{i:n} - x_{i-1:n}| \to 0 \) in probability as \( n \) increases so that the behavior of \( \log D_i \) with respect to \( \Theta \) is essentially the same as \( \log f(x_{i:n}, \Theta) \). Consequently, MPS and MLE estimation are asymptotically equal and have the same asymptotic sufficiency, consistency and efficiency properties.

In non-standard situations where the end points of the support of \( f \) are unknown, then there are a number of terms whose contribution to \( \log G \) is no longer negligible, and moreover, for which the above approximation is not useful because \( R \) is no longer small. This leads to a significant difference in the behavior of \( \log L \) and \( \log G \). It is now possible for \( \log L \) to become unbounded for fixed \( n \), while \( \log G \) is always bounded above by \( \log \frac{1}{n+1} \). Since \( \sum D_i = 1 \), the maxima is obtained only when \( D_i \)'s are equal. Then

\[ D_i = \frac{1}{n+1} \]

\[ \log G = \frac{1}{n+1} \log(\prod_{i=1}^{n+1} D_i) = \frac{1}{n+1} \log\left(\frac{1}{n+1}\right)^{n+1} = \log \frac{1}{n+1} \]

This allows consistent estimators to be obtained by MPS where MLE will fail.
3 Some good properties of MPS estimates

To help readers build a general idea of advantages of MPS estimation over MLE, we first list some good properties of MPS estimation, which were presented by Cheng and Amin (1983), including sufficiency, consistency and asymptotic efficiency.

3.1 Sufficiency

In general, MPS estimators will not necessarily be functions of sufficient statistics. However, for the case when the support of density functions are known, MPS estimators will have the same asymptotic properties as MLE estimators including asymptotic sufficiency. When the support of density functions are unknown, then sufficient statistics occur only under fairly strong restrictions on the form of the density. In this situation, MPS has better properties than MLE.

3.2 Consistency and Asymptotic efficiency

In certain cases, it is possible to obtain the distributional behavior of an MPS estimator for all sample sizes n. For the uniform distribution with unknown endpoints (example 2 which will discussed later), the MPS estimators are precisely the UMVUE. For a general distribution, however, the small sample behavior of MPS estimators, like MLE estimators, is usually difficult to obtain. However, asymptotic properties such as consistency and efficiency are obtainable.

A three-parameter Weibull distribution is given by

\[ f_W(x) = \alpha \beta^{-\alpha} (x - \theta)^{\alpha - 1} \exp\left[-\left(\frac{x - \theta}{\beta}\right)^\alpha\right], \quad x > \theta \]

and a gamma distribution is defined as

\[ f_G(x) = \left[\beta \Gamma(\alpha)\right]^{-1} (x - \theta)^{\alpha - 1} \exp\left[-\left(\frac{x - \theta}{\beta}\right)\right], \quad x > \theta. \]
Theorem 1 [Cheng and Amin (1983)]: For the Weibull and gamma distributions, consider the solution of the MPS equation

\[
\frac{\partial H}{\partial \Theta} = 0,
\]

where \( H \) is as defined as

\[
H = \log G
\]

and

\[
\Theta = (\alpha, \beta, \theta)'.
\]

Let \( L \) be the likelihood as

\[
L = \frac{1}{n} \sum_{i=1}^{n} \log f(x_i, \Theta).
\]

(i) If \( \alpha > 2 \) then there is, with a probability tending to 1, a solution \( \tilde{\Theta} = (\tilde{\alpha}, \tilde{\beta}, \tilde{\theta})' \) that is asymptotically normal with

\[
\sqrt{n}(\tilde{\Theta} - \Theta) \xrightarrow{L} N(\vec{0}, -E(\partial^2 L/\partial \Theta^2)^{-1}).
\]

(ii) If \( 0 < \alpha < 2 \) then there is, with a probability tending to 1, a solution \( \tilde{\Theta} \) where

\[
\tilde{\theta} - \theta = O_p(n^{-1/\alpha})
\]

and where \((\tilde{\alpha}, \tilde{\beta})' = \Phi, \) say, is asymptotically normal with

\[
\sqrt{n}(\tilde{\Phi} - \Phi) \xrightarrow{L} N(\vec{0}, -E(\partial^2 L/\partial \Phi^2)^{-1}).
\]

The theorem shows that when \( \alpha > 2 \) the behavior of all three estimators is regular in the sense of being asymptotically efficient with variance and covariances of order \( O(\frac{1}{n}) \). When \( 0 < \alpha < 2 \), the estimator \( \tilde{\theta} \) becomes "hyper-efficient" having variance smaller than order \( O(\tilde{\theta}) \); a remarkable consequence is that asymptotic distribution of \( \tilde{\alpha} \) and \( \tilde{\beta} \) is the same as if \( \theta \) were actually known.

MLE estimators in the Weibull and gamma cases, found by solving \( \frac{\partial L}{\partial \Theta} = 0 \), do not perform as well. For comparison we give here the result for MLE estimators corresponding to Theorem 1. Details of the proof are given by Cheng and Amin (1982).
Theorem 2 [Cheng and Amin (1983)]: For the Weibull and gamma distributions, consider the solution of the MLE equation

$$\frac{\partial L}{\partial \Theta} = 0,$$

where $L$ is as defined as

$$L = \frac{1}{n} \sum_{i=1}^{n} \log f(x_i, \Theta)$$

and

$$\Theta = (\alpha, \beta, \theta)'.$$

(i) If $\alpha > 2$ then there is, with a probability tending to 1, a solution $\hat{\Theta}$ with the same asymptotic normal distribution as $\tilde{\Theta}$, the MPS estimator.

(ii) If $1 < \alpha < 2$, then there is, with a probability tending to 1, a solution $\hat{\Theta}$ where

$$\hat{\theta} - \theta = O_p(n^{-1/\alpha})$$

and where $\hat{\Phi} = (\hat{\alpha}, \hat{\beta})'$ has the same asymptotic normal distribution as $\tilde{\Phi}$, the MPS estimator.

(iii) If $\alpha < 1$, then there is no consistent solution of $\frac{\partial L}{\partial \Theta} = 0$.

The theorem shows that when the distribution is J-shaped (i.e. $\alpha < 1$), MLE estimation fails while MPS estimation does not.

The above two theorems assume all three parameters are unknown. They can be modified directly to cover the cases when some of the parameters are unknown; the known parameters simply have to be deleted from the vector of the unknowns $\Theta$ (and $\Phi$), with obvious adjustment to the statement of each theorem to take this deletion into account.

In summary, asymptotically MPS are at least as efficient as MLE estimation when they exist. For distributions where the endpoints are unknown and the density is J-shaped MLE is bound to fail, but MPS still gives asymptotically efficient estimators.
4 Three simple examples

4.1 Example 1

Suppose \( x_1, x_2, \ldots, x_n \) is a random sample from the uniform distribution \( UNIF(0, \theta) \) with unknown endpoint \( \theta \in (0, \infty) \). And let \( x_{1:n} < x_{2:n} < \cdots < x_{n:n} \) be the order statistics derived from \( x_1, x_2, \ldots, x_n \). Its well-known MLE estimator \( \hat{\theta} \) is \( x_{n:n} \). Then the procedure to find its MPS estimator should be as follows:

\[
D_i = F(x_{i:n}, \theta) - F(x_{i-1:n}, \theta) = \int_{x_{i-1:n}}^{x_{i:n}} \frac{1}{\theta} \, dx = \frac{1}{\theta} (x_{i:n} - x_{i-1:n}), \quad i = 2, 3, \ldots, n.
\]

\[
D_1 = F(x_{1:n}, \theta) = \frac{1}{\theta} x_{1:n},
\]

\[
D_{n+1} = 1 - F(x_{n:n}, \theta) = 1 - \frac{1}{\theta} x_{n:n}.
\]

\[
H = \log G = \frac{1}{n+1} \sum_{i=1}^{n+1} \log D_i
\]

\[
= \frac{1}{n+1} (\log x_{1:n} + \sum_{i=1}^{n} \log(x_{i:n} - x_{i-1:n}) + \log(\theta - x_{n:n}))) - \log \theta.
\]

Solve for Parameter:

\[
\frac{dH}{d\theta} = 0 \Rightarrow \frac{1}{n+1} \cdot \frac{1}{\theta} - \frac{1}{\theta} = 0 \Rightarrow \hat{\theta} = \frac{n+1}{n} x_{n:n}.
\]

Estimators Comparison: \( \hat{\theta} = x_{n:n} \) and \( \tilde{\theta} = \frac{n+1}{n} x_{n:n} \).

(i) \( E(\tilde{\theta}) = E(x_{n:n}) = n \int_{0}^{\theta} \frac{x^n}{\theta^n} \, dx = \frac{n}{n+1} \theta, \)

while

\[
E(\hat{\theta}) = E\left(\frac{n+1}{n} x_{n:n}\right) = \frac{n+1}{n} E(x_{n:n}) = \theta.
\]

So, \( \tilde{\theta} \) is an unbiased estimator, while \( \hat{\theta} \) is not.

(ii) \( E(x_{n:n}) = \frac{n}{n+1} \theta, \quad E(x_{n:n}^2) = \frac{n}{n+2} \theta^2, \quad Var(x_{n:n}) = \frac{n}{(n+2)(n+1)^2} \theta^2. \)
\[ MSE(\hat{\theta}) = Var(\hat{\theta}) + [E(\hat{\theta}) - \theta]^2 = Var(x_{n:n}) + \left(\frac{n}{n+1}\theta - \theta\right)^2 \]

\[ = \frac{n}{n+2}\theta^2 - \frac{n^2}{(n+1)^2}\theta^2 + \frac{1}{(n+1)^2}\theta^2 = \frac{2}{(n+1)(n+2)}\theta^2. \]

while

\[ MSE(\tilde{\theta}) = Var(\tilde{\theta}) = \frac{(n+1)^2}{n^2}Var(x_{n:n}) \]

\[ = \frac{1}{n(n+2)}\theta^2 \]

So, \( \tilde{\theta} \) has a smaller \( MSE \). i.e. \( \tilde{\theta} \) is a better estimator.

### 4.2 Example 2

Suppose \( x_1, x_2, \ldots, x_n \) is a random sample from the uniform distribution \( UNIF(\theta_1, \theta_2) \) with unknown endpoints \( \theta_1, \theta_2 \in (0, \infty) \). And let \( x_{1:n} < x_{2:n} < \cdots < x_{n:n} \) be the order statistics derived from \( x_1, x_2, \ldots, x_n \). Its well-known MLE estimators are \( \hat{\theta}_1 = x_{1:n}, \hat{\theta}_2 = x_{n:n} \). Following a similar procedure to Example 1, we see that:

\[ H = \log G = \frac{1}{n+1} \sum_{i=1}^{n+1} \log D_i \]

\[ = \frac{1}{n+1} \left( \log(x_{1:n} - \theta_1) + \sum_{i=1}^{n} \log(x_{i:n} - x_{i-1:n}) + \log(\theta_2 - x_{n:n}) \right) - \log(\theta_2 - \theta_1). \]

Solve for Parameters: \( \frac{dH}{d\theta_1} = \frac{dH}{d\theta_2} = 0 \Rightarrow \tilde{\theta}_1 = \frac{n x_{1:n} - x_{n:n}}{n-1}, \tilde{\theta}_2 = \frac{n x_{n:n} - x_{1:n}}{n-1}. \)

which happened to be the UMVUE for \( \theta_1 \) and \( \theta_2 \). (By Cheng and Amin (1983))

(UMVUE: Uniform Minimum Variance Unbiased Estimators.)
4.3 Example 3

Suppose \(x_1, x_2, \cdots, x_n\) is a random sample from the exponential distribution \(\text{EXP}(\theta, \eta)\) with unknown endpoint \(\eta \in (0, \infty)\). And let \(x_{1:n} < x_{2:n} < \cdots < x_{n:n}\) be the order statistics derived from \(x_1, x_2, \cdots, x_n\). It’s well-known that MLE estimators are given as \(\hat{\theta} = \bar{x} - x_{1:n}\), \(\hat{\eta} = x_{1:n}\). Then the procedure to find its MPS estimator should be as follows:

\[
D_i = F(x_{i:n}; \theta, \eta) - F(x_{i-1:n}; \theta, \eta) = \int_{x_{i-1}}^{x_i} \frac{1}{\theta} e^{-(x-\eta)/\theta} \, dx, \quad i = 2, 3, \cdots, n.
\]

\[
D_1 = F(x_{1:n}; \theta, \eta), \quad D_{n+1} = 1 - F(x_{n:n}; \theta, \eta).
\]

\[
H = \log G = \frac{1}{n+1} \sum_{i=1}^{n+1} \log D_i
\]

\[
= \frac{1}{n+1} (\log(1 - e^{-(x_{1:n}-\eta)/\theta}) + \sum_{i=2}^{n} \log D_i + \log(e^{-(x_{n:n}-\eta)/\theta} - 0))
\]

\[
= \frac{1}{n+1} (\log e^{\eta/\theta} (e^{-\eta/\theta} - e^{-x_{1:n}/\theta}) + \cdots \log e^{\eta/\theta} (e^{-x_{n-1:n}/\theta} - e^{-x_{n:n}/\theta} + \frac{\eta - x_{n:n}}{\theta}))
\]

\[
= \frac{1}{n+1} ((n+1) \frac{\eta}{\theta} + \log(e^{-\eta/\theta} - e^{-x_{1:n}/\theta}) + \cdots + \log(e^{-x_{n-1:n}/\theta} - e^{-x_{n:n}/\theta}) - \frac{x_{n:n}}{\theta})
\]

Solve for Parameter:

\[
\frac{dH}{d\eta} = 0 \Rightarrow \frac{1}{\theta} + \frac{1}{n+1} \cdot \frac{1}{e^{-\eta/\theta} - e^{-x_{1:n}/\theta}} \cdot e^{-\eta/\theta} \cdot (-\frac{1}{\theta}) = 0
\]

\[
\Rightarrow \tilde{\eta} = x_{1:n} + \hat{\theta} \log \frac{n}{n+1}
\]

In special cases, for example, when shape parameter \(\theta = 1\), i.e. \(x_i \sim \text{EXP}(1, \eta)\), it would be much easier to evaluate the advantages of MPS estimator \(\tilde{\eta} = x_{1:n} + \log \frac{n}{n+1}\) over MLE estimator \(\hat{\eta} = x_{1:n}\).

Estimators Comparison: \(\hat{\eta}\) and \(\tilde{\eta}\).

(pdf of \(x_{1:n}\): \(g_1(x) = ne^{-n(x_{1:n}-\eta)}, \quad \eta < x_{1:n} < \infty\))

(i) \(E(\tilde{\eta}) = E(x_{1:n}) = n \int_{\eta}^{\infty} x e^{-n(x-\eta)} \, dx = \eta + \frac{1}{n}\),

while

\[
E(\hat{\eta}) = E(x_{1:n}) + \log \frac{n}{n+1} = \eta + \frac{1}{n} + \log \frac{n}{n+1}.
\]

So, neither \(\hat{\eta}\) nor \(\tilde{\eta}\) is unbiased estimator.
(ii) \( E(x_{1:n}) = \eta + \frac{1}{n}, \quad E(x_{1:n}^2) = \eta^2 + \frac{2\eta}{n} + \frac{2}{n^2}, \quad \text{Var}(x_{1:n}) = \frac{1}{n^2}. \)

\[
MSE(\hat{\eta}) = \text{Var}(\hat{\eta}) + [b(\hat{\eta})]^2 = \text{Var}(\hat{\eta}) + [E(\hat{\eta}) - \eta]^2
\]

\[
= \text{Var}(x_{1:n}) + \left(\frac{1}{n}\right)^2 = \frac{2}{n^2}.
\]

\[
MSE(\tilde{\eta}) = \text{Var}(x_{1:n} + \log \frac{n}{n+1}) + \left(\frac{1}{n} + \log \frac{n}{n+1}\right)^2
\]

\[
= \frac{2}{n^2} + \left(\log \frac{n}{n+1}\right)^2 + \frac{2}{n} \log \frac{n}{n+1}.
\]

Next, we want to prove \((\log \frac{n}{n+1})^2 + \frac{2}{n} \log \frac{n}{n+1} < 0:\)

Let \( t(x) = (\log \frac{x}{x+1})^2 + \frac{2}{x} \log \frac{x}{x+1}, \) with \( x > 0. \)

By taking the first derivative, we got

\[
\frac{dt}{dx} = \frac{2}{x} \cdot \log \left(\frac{x}{x+1}\right) \left(\frac{1}{x+1} - \frac{1}{x}\right) + \frac{2}{x} \cdot \frac{1}{x(x+1)} > 0.
\]

\( t(x) \) is an increasing function. Together with limit of \( t(x) \) goes to zero, we know \((\log \frac{n}{n+1})^2 + \frac{2}{n} \log \frac{n}{n+1} \) less than zero.

Thus, \( \hat{\eta} \) has a smaller \( MSE \) compared to \( \hat{\eta} \).

i.e the MPS estimator for location parameter \( \eta \) is \( x_{1:n} + \hat{\theta} \log \frac{n}{n+1} \) which is a better estimator compared to MLE estimator because of a smaller \( MSE \).
5 Simulation study

In order to compare the finite sample properties of the MPS method and the MLE method in parameter estimates, a set of simulations was performed based on the three-parameter Weibull distribution:

\[ f_W(x) = \alpha \beta^{-\alpha} (x - \theta)^{\alpha-1} \exp \left( -\frac{(x - \theta)}{\beta} \right)^\alpha, \quad x > \theta \]

and Gamma distribution:

\[ f_G(x) = \left[ \beta^\alpha \Gamma(\alpha) \right]^{-1} (x - \theta)^{\alpha-1} \exp \left[ -\frac{(x - \theta)}{\beta} \right], \quad x > \theta. \]

We generate \( N = 1000 \) samples of size \( n \) from each distribution for \( n = 10, 30, 50 \). Since the functions are singular, the minimizations are performed using method "CG", a conjugate gradients method by Fletcher and Reeves (1964), and method "CG" has already been proved as a more stable and successful iteration in larger optimization problems than downhill simplex method due to Nelder and Mead (1965). The results are presented in the following tables.

5.1 Results

Weibull (Tables 1 & 2)

Gamma (Tables 3 & 4)
Table 1.
Simulation results of MPS estimates, MLE estimates on three-parameter Weibull distribution. Shown are means of estimated parameters from $N = 1000$ simulations of sample sizes $n = (10, 30, 50)$.

Numbers in the bracket are mean squared errors of estimates.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\alpha_0$</th>
<th>$\beta_0$</th>
<th>$\theta_0$</th>
<th>MPS estimates</th>
<th>MLE estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\hat{\alpha}$</td>
<td>$\tilde{\beta}$</td>
</tr>
<tr>
<td>10</td>
<td>0.5</td>
<td>3</td>
<td>0</td>
<td>0.48 (0.05)</td>
<td>3.26 (0.93)</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>3</td>
<td>0</td>
<td>0.90 (0.23)</td>
<td>3.33 (1.60)</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>3</td>
<td>0</td>
<td>1.54 (0.69)</td>
<td>3.21 (1.43)</td>
</tr>
<tr>
<td>30</td>
<td>0.5</td>
<td>3</td>
<td>0</td>
<td>0.50 (0.00)</td>
<td>3.06 (0.07)</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>3</td>
<td>0</td>
<td>0.86 (0.03)</td>
<td>3.12 (0.41)</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>3</td>
<td>0</td>
<td>1.46 (0.14)</td>
<td>3.08 (0.36)</td>
</tr>
<tr>
<td>50</td>
<td>0.5</td>
<td>3</td>
<td>0</td>
<td>0.50 (0.00)</td>
<td>3.01 (0.00)</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>3</td>
<td>0</td>
<td>0.90 (0.23)</td>
<td>3.33 (1.60)</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>3</td>
<td>0</td>
<td>1.49 (0.07)</td>
<td>3.08 (0.19)</td>
</tr>
<tr>
<td></td>
<td>2.0</td>
<td>3</td>
<td>0</td>
<td>2.00 (0.18)</td>
<td>3.09 (0.24)</td>
</tr>
</tbody>
</table>
Table 2.
Simulation results of MLE estimates compared to MPS estimates. Shown are means of estimated parameters from $N = 1000$ simulations of sample sizes $n = (10, 30, 50)$.

Numbers in the bracket are mean squared errors of estimates.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\alpha_0$</th>
<th>$\beta_0$</th>
<th>$\theta_0$</th>
<th>$\hat{\alpha}$</th>
<th>$\hat{\beta}$</th>
<th>$\hat{\theta}$</th>
<th>$\tilde{\alpha}$</th>
<th>$\tilde{\beta}$</th>
<th>$\tilde{\theta}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.5</td>
<td>3</td>
<td>0</td>
<td>0.48</td>
<td>3.40</td>
<td>0.02</td>
<td>0.67</td>
<td>3.84</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(1.11)</td>
<td>(0.02)</td>
<td>(0.08)</td>
<td>(4.58)</td>
<td>(0.02)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.9</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0.90</td>
<td>3.39</td>
<td>-0.07</td>
<td>1.11</td>
<td>3.21</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>(0.21)</td>
<td>(1.71)</td>
<td>(0.26)</td>
<td>(0.17)</td>
<td>(1.56)</td>
<td>(0.13)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>0.5</td>
<td>3</td>
<td>0</td>
<td>0.49</td>
<td>3.10</td>
<td>0.00</td>
<td>0.54</td>
<td>3.39</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.13)</td>
<td>(0.00)</td>
<td>(0.01)</td>
<td>(0.13)</td>
<td>(0.00)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.9</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0.85</td>
<td>3.09</td>
<td>0.00</td>
<td>0.95</td>
<td>3.06</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.38)</td>
<td>(0.01)</td>
<td>(0.02)</td>
<td>(0.44)</td>
<td>(0.01)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>0.5</td>
<td>3</td>
<td>0</td>
<td>0.50</td>
<td>3.03</td>
<td>0.00</td>
<td>0.54</td>
<td>3.04</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.03)</td>
<td>(0.00)</td>
<td>(0.01)</td>
<td>(0.60)</td>
<td>(0.00)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.9</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0.87</td>
<td>3.07</td>
<td>0.01</td>
<td>0.94</td>
<td>3.05</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.19)</td>
<td>(0.00)</td>
<td>(0.01)</td>
<td>(0.24)</td>
<td>(0.01)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3.
Simulation results of MPS estimates, MLE estimates on three-parameter gamma distribution. Shown are means of estimated parameters from $N = 1000$ simulations of sample sizes $n = (10, 30, 50)$.

Numbers in the bracket are mean squared errors of estimates.

<table>
<thead>
<tr>
<th>$n$</th>
<th>True parameters</th>
<th>MPS estimates</th>
<th>MLE estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\alpha_0$</td>
<td>$\beta_0$</td>
<td>$\theta_0$</td>
</tr>
<tr>
<td>10</td>
<td>0.5</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>30</td>
<td>0.5</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>50</td>
<td>0.5</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2.0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
Table 4. 
Simulation results of MLE estimates compared to MPS estimates. Shown are means of estimated parameters from $N = 1000$ simulations of sample sizes $n = (10, 30, 50)$. 

Numbers in the bracket are mean squared errors of estimates.

<table>
<thead>
<tr>
<th>$n$</th>
<th>True parameters</th>
<th>MPS estimates</th>
<th>MLE estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\alpha_0$</td>
<td>$\beta_0$</td>
<td>$\theta_0$</td>
</tr>
<tr>
<td>10</td>
<td>0.5</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>(0.11)</td>
<td>(0.09)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>0.9</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>(0.27)</td>
<td>(0.16)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>30</td>
<td>0.5</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.02)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>0.9</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.04)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>50</td>
<td>0.5</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>0.9</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.03)</td>
<td>(0.00)</td>
</tr>
</tbody>
</table>
5.2 Conclusion

When the shape parameter $\alpha$ is less than one in three-parameter Weibull distribution and gamma distribution, the underlying distribution is J-shaped, and no local maximum can yield a consistent estimator. As shown in Table 1 and Table 3, MLE method breaks down when $\alpha \leq 1$, but MPS method still gives efficient estimators. However, Smith (1985) discussed that when $\alpha \leq 1$, the likelihood function for MLE method has no local maximum but is globally maximized at the sample minimum $x_{1,n}$. And $x_{1,n}$ itself is a consistent estimator of location parameter $\theta$ and has a property of asymptotic sufficiency (Weiss 1979). So without violating the basic idea of solving MLE estimators, the alternative is to use $x_{1:n}$ as estimator of $\theta$, reduce the model to two-parameter Weibull and gamma distributions, and estimate the remain parameters by conventional MLE method based on $x_{2:n}, x_{3:n}, \hat{x}_{n:n}$. The results are presented in Table 2 and Table 4. Though the alternative method gives MLE estimators, and, to some extent fixes those NaNs (no solutions) in Table 1 and Table 3, it’s still very unstable. Especially, when $\hat{\theta}$ quite often happens to have big bias compared to the true value $\theta_0$.

When the shape parameter $\alpha > 1$, as shown in Table 1 and Table 3, both MLE and MPS methods give efficient and relatively accurate estimators. And as sample size $n$ gets larger (from 10 to 50), both MLE and MPS estimators become hyper-efficient (estimators are getting closer to the real parameter value, and mean squared errors are getting closer to zero). In short, MPS estimation gives consistent estimators under much more general conditions than MLE estimators. In particular, it gives consistent estimators when the underlying distribution is J-shaped, a situation where MLE estimation is bound to fail.
6 Estimating the endpoint (Bayesian likelihood)

While I was reading papers about using MPS method in non-regular estimating problems, another method called Bayesian likelihood methods intrigued my interests. So I did some extensive studies on Bayesian likelihood methods, which is widely used on estimating the endpoint of a distribution and was proven effective to provide a better performance compared to MPS method and MLE method. I will first introduce the background and methodology of Bayesian likelihood methods, then do a simulation study on Weibull and gamma distributions to illustrate the performance.

6.1 Introduction

Let a probability density be of the form \( f(x) = (x - \theta)^{\alpha - 1}g(x \mid \theta, \omega), x > \theta, \) where \( \theta \) is a single unknown location parameter, \( \omega \) is a vector of parameters other than \( \theta \) and \( g(x \mid \theta, \omega) \) converges to a strictly positive constant as \( x \) decreases to \( \theta \). Throughout, \( \alpha \) will be taken to be an unknown shape parameter. Distributions of this type include the Weibull and gamma distributions. In extreme value theory the generalized extreme value distribution and the generalized Pareto distribution are closely related to this form.

Since the support of \( f \) depends on \( \theta \), the classical maximum likelihood method fail to apply in this case. The problem of non-regular estimation emerges when the shape parameter \( \alpha \) satisfies \( \alpha \leq 1 \) and the likelihood function fails to produce a solution other than \( \hat{\theta} = x_{1:n} \), the smallest observation. Various methods have been sought to overcome the problem. In general the likelihood approach is preferred, mainly because the likelihood function and related inference procedures, such as the likelihood ratio test and the Bayesian method, are well established and available for practical use.

Dr. Peter Hall believes that a piece of vital information is missed in the conventional likelihood. In other words, the information that is provided by the product is incomplete and conditional. Therefore a conditional likelihood seems to be more appropriate. The condition can be represented from a Bayesian point of view, in the form of a prior distribution. Hall and Wang (2005) proposed a sample spacing-based empirical prior for \( \theta \). The prior is empirical, in that it relies on empirical spacings to determine shape and scale, and has simple monotone shape near the smallest order statistic \( x_{1:n} \). Away from \( x_{1:n} \) the prior plays a negligible role in inference. Therefore, the prior might fairly be said to be informative, and
empirical, only in a near neighborhood of $x_{1:n}$, and uninformative elsewhere.

Combining the prior with the product of the densities Hall and Wang (2005) obtained a penalized likelihood function that provides solutions as estimates of the parameters. The prior proposed provides an approach to Bayesian data analysis in non-regular cases.

6.2 Methodology

Suppose that data are drawn by sampling randomly from the distribution with density $f(x - \theta_0)$, where $f$ is supported on $(0, \infty)$ and satisfies

$$f(x) \sim \beta x^{\alpha - 1}$$

with $0 < \alpha < 1$ and $\beta > 0$. Denote by $x_{1:n} \leq \ldots \leq x_{n:n}$ the ordered values of the sample. If $f$ is known but $\theta_0$ is not, the maximum likelihood estimator $\hat{\theta}$ of $\theta_0$ is identical to $x_{1:n}$, which is also the quantity that maximizes the ‘semiparametric likelihood’, involving the empirical prior for $\theta$, specifically

$$p(\theta) = \frac{x_{1:n} - \theta}{x_{2:n} - \theta}.$$  

This gives penalized forms of likelihood, i.e.

$$L(\theta) = p(\theta) \prod_{i=1}^{n} f(x_{i:n}, \theta).$$

The factor $p(\theta)$ shifts the estimator of $\theta$ from its value, under $L$ of $x_{1:n}$, which we know to be an overestimate, to a more plausible value that is strictly less than $x_{1:n}$.

6.3 Interpretation of the empirical prior

The empirical prior $p(\theta)$ has a general form and holds the information about $\theta$ that is contained in the first two sample spacings. Intuitively we may see the following.

- No estimates of $\theta$ can equal $x_{1:n}$, since $p(x_{1:n}) = 0$.
- For $\theta$ distant from $x_{1:n}$, $p(\theta) \approx 1$, indicating that those values are equally likely. Hence the prior is non-informative.
6.4 Simulations

In this section, we examine the performance of the Bayesian likelihood method on estimating the endpoint $\theta$. The two distributions that were studied are the Weibull and gamma distributions, which contain three unknown parameters: $\theta$, $\alpha$ and $\beta$. We are interested in estimating the location parameter $\theta$, so we set $\theta = 0$ and $\beta = 1$, with various values of $\alpha$. The proposed Bayesian likelihood method (BL) is compared with the method of maximum product of spacings (MPS).

Simulation results are summarized in Table 5 and Table 6. It can be seen from those results that MPS estimates has in general lower mean-squared error in most of the cases (17 out of 24) compared to BL estimates. Similar situation is also shown in the absolute bias, i.e. MPS estimates has less absolute bias than BL estimates in 17 out of 24 cases. Thus, MPS method has its certain advantages to Bayesian likelihood methods on estimating endpoints, at least for three parameters Weibull and gamma distributions.
Table 5.
Simulation results of BL estimates, MPS estimates on three-parameter Weibull distribution. Shown are means and mean-squared errors of estimated parameters from $N = 1000$ simulations of sample sizes $n = (30, 100)$ with $\beta = 1$ and $\theta = 0$.

Numbers in the bracket are mean squared errors of estimates.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\alpha$</th>
<th>BL</th>
<th>MPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>0.5</td>
<td>0.001274</td>
<td>0.001169</td>
</tr>
<tr>
<td></td>
<td></td>
<td>($1.79 \times 10^{-5}$)</td>
<td>($1.76 \times 10^{-5}$)</td>
</tr>
<tr>
<td></td>
<td>0.75</td>
<td>0.007128</td>
<td>0.003430</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.000369)</td>
<td>(0.000325)</td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>0.013770</td>
<td>-0.000722</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.001175)</td>
<td>(0.000951)</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>0.015623</td>
<td>-0.012380</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.003905)</td>
<td>(0.003729)</td>
</tr>
<tr>
<td></td>
<td>2.0</td>
<td>0.004634</td>
<td>-0.015616</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.003912)</td>
<td>(0.004118)</td>
</tr>
<tr>
<td></td>
<td>3.0</td>
<td>-0.0001967</td>
<td>-0.005615</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.003302)</td>
<td>(0.003365)</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.000121</td>
<td>0.000111</td>
</tr>
<tr>
<td></td>
<td></td>
<td>($1.73 \times 10^{-7}$)</td>
<td>($1.69 \times 10^{-7}$)</td>
</tr>
<tr>
<td></td>
<td>0.75</td>
<td>0.001436</td>
<td>0.000680</td>
</tr>
<tr>
<td></td>
<td></td>
<td>($1.38 \times 10^{-5}$)</td>
<td>($1.20 \times 10^{-5}$)</td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>0.004627</td>
<td>-0.000208</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.000115)</td>
<td>(0.000093)</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>0.008370</td>
<td>-0.004854</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.000818)</td>
<td>(0.000754)</td>
</tr>
<tr>
<td></td>
<td>2.0</td>
<td>0.002433</td>
<td>-0.008989</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.001061)</td>
<td>(0.001136)</td>
</tr>
<tr>
<td></td>
<td>3.0</td>
<td>-0.002100</td>
<td>-0.005836</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.000921)</td>
<td>(0.000950)</td>
</tr>
</tbody>
</table>
Table 6.
Simulation results of BL estimates, MPS estimates on three-parameter gamma distribution. Shown are means and mean-squared errors of estimated parameters from $N = 1000$ simulations of sample sizes $n = (30, 100)$ with $\beta = 1$ and $\theta = 0$

Numbers in the bracket are mean squared errors of estimates.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\alpha$</th>
<th>BL</th>
<th>MPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>0.5</td>
<td>0.000750</td>
<td>0.000669</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$(8.57 \times 10^{-6})$</td>
<td>$(8.41 \times 10^{-6})$</td>
</tr>
<tr>
<td></td>
<td>0.75</td>
<td>0.006119</td>
<td>0.002880</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$(0.000282)$</td>
<td>$(0.000252)$</td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>0.015770</td>
<td>0.001033</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$(0.001406)$</td>
<td>$(0.001136)$</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>0.026891</td>
<td>-0.016300</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$(0.007267)$</td>
<td>$(0.006691)$</td>
</tr>
<tr>
<td></td>
<td>2.0</td>
<td>0.032141</td>
<td>-0.029619</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$(0.0177473)$</td>
<td>$(0.0174271)$</td>
</tr>
<tr>
<td></td>
<td>3.0</td>
<td>0.024249</td>
<td>-0.058878</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$(0.042740)$</td>
<td>$(0.045517)$</td>
</tr>
<tr>
<td>100</td>
<td>0.5</td>
<td>$7.69 \times 10^{-5}$</td>
<td>$6.91 \times 10^{-5}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$(7.99 \times 10^{-8})$</td>
<td>$(7.82 \times 10^{-8})$</td>
</tr>
<tr>
<td></td>
<td>0.75</td>
<td>0.001248</td>
<td>0.000562</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$(9.72 \times 10^{-6})$</td>
<td>$(8.29 \times 10^{-6})$</td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>0.004801</td>
<td>0.000250</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$(0.000129)$</td>
<td>$(0.000103)$</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>0.011617</td>
<td>-0.006814</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$(0.001338)$</td>
<td>$(0.001238)$</td>
</tr>
<tr>
<td></td>
<td>2.0</td>
<td>0.011958</td>
<td>-0.017405</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$(0.004108)$</td>
<td>$(0.004228)$</td>
</tr>
<tr>
<td></td>
<td>3.0</td>
<td>0.012452</td>
<td>-0.028206</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$(0.011752)$</td>
<td>$(0.012411)$</td>
</tr>
</tbody>
</table>
7 References


8 Appendix

In this section, I shall post key R codes in programming simulations. All the codes below are used for Weibull distribution. Those codes for gamma distribution should be very similar.

8.1 Advanced MPS vs MLE three parameters Weibull distribution

z1=NULL; z2=NULL;

mlogl1=function(theta,x) {
  z1 = dweibull(x-cc1, shape=theta[1], scale=theta[2], log=FALSE)
  z1[z1<1e-16]=1e-16
  -mean(log(z1))
}

mlogl2=function(beta,x) {
  z2 = diff(c(0, pweibull(x-beta[3], shape=beta[1], scale=beta[2], lower.tail=TRUE,
                      log.p=FALSE),1))
  z2[z2<1e-16]=1e-16
  -mean(log(z2))
}

s1=0;s2=0;s3=0
t1=0;t2=0;t3=0
r1=0;r2=0;r3=0
h1=0;h2=0;h3=0

n=50
N=1000
a=0.9; b=1; cc=0; cc1=0
for (i in 1:N) {

}
```R
y = rweibull(n, a, b)
x = y + cc
x = sort(x)
cc1 = x[1]
t = x[2:n]
theta.start = c(a, b, cc1)
out1 = optim(theta.start, mlogl1, x = t, method = "CG")
theta.hat = out1$par
beta.start = c(a, b, cc)
out2 = optim(beta.start, mlogl2, x = x, method = "CG")
beta.hat = out2$par
s1 = c(s1, (theta.hat[1] - a)^2)
s2 = c(s2, (theta.hat[2] - b)^2)
s3 = c(s3, (cc1 - cc)^2)
t1 = c(t1, (beta.hat[1] - a)^2)
t2 = c(t2, (beta.hat[2] - b)^2)
t3 = c(t3, (beta.hat[3] - cc)^2)
r1 = c(r1, theta.hat[1])
r2 = c(r2, theta.hat[2])
r3 = c(r3, cc)
h1 = c(h1, beta.hat[1])
h2 = c(h2, beta.hat[2])
h3 = c(h3, beta.hat[3])
}

mean11 = mean(r1); mean12 = mean(r2); mean13 = mean(r3)
mean21 = mean(h1); mean22 = mean(h2); mean23 = mean(h3)
 mse11 = mean(s1); mse12 = mean(s2); mse13 = mean(s3)
mse21 = mean(t1); mse22 = mean(t2); mse23 = mean(t3)
print("MLE - estimate&MSE")
c(mean11, mean12, mean13)
c(mse11, mse12, mse13)
print("MPS - estimate&MSE")
c(mean21, mean22, mean23)
c(mse21, mse22, mse23)
```
8.2 Bayesian vs MPS endpoint estimates Weibull distribution

```r
z2=NULL;
z3=NULL; z31=NULL; z32=1;
a=NULL;

mlogl2=function(beta,x) {
z2 = diff(c(0, pweibull(x-beta, shape=a, scale=1, lower.tail=TRUE, log.p=FALSE),1))
z2[z2<1e-16]=1e-16
-mean(log(z2))
}

mlogl5=function(eta,x) {
z31 = (x[1]-eta)/(x[2]-eta)
for (k in 1:length(x)) { z32=z32*(dweibull(x[k]-eta,a,1,log=FALSE)) }
z3 = z31*z32
-mean(log(z3))
}

t=NULL; s=NULL;
beta=NULL;
eta=NULL;
n=30
N=1000
a=2; b=1; cc=0
for (i in 1:N) {
y = rweibull(n,a,b)
x = y + cc
x = sort(x)
out1 = optim(cc, mlogl5, x = x)
eta.hat = out1$par
out2 = optim(cc, mlogl2, x = x)
```

28
beta.hat = out2$par

t = c(t, (eta.hat - cc)^2)
s = c(s, (beta.hat - cc)^2)
eta = c(eta, eta.hat)
beta = c(beta, beta.hat)
}
print("End points - estimate&MSE")
mean(eta)
mean(t)
mean(beta)
mean(s)