Singh-Maddala Distribution: Parameter estimation by L- and TL-moments for extreme value data

Mirza Naveed Shahzad

Department of Mathematics & Statistics, University of Gujrat, Gujrat, Pakistan;
Corresponding author Email: nvd.shzd@uog.edu.pk

Abstract

Modeling, accurate inference, and prediction of extreme events by probabilistic models are very important in every field to minimize the damage due to extremes as much as possible. To secure this useful purpose, Singh-Maddala distribution is considered in this article for the analysis of extreme events. The L- and TL-moments and L- and TL-moments ratios are derived for this distribution. The method of MLE is commonly used for parameter estimation but the extreme value data are frequently heavy-tailed data. In this situation method of L-moments (MoLM) and method of TL-moments (MoTLM) are proposed alternatively to estimate the parameters of the Singh-Maddala distribution. Monte Carlo simulations and real data studies have been analyzed for illustrative purposes.

1. Introduction

Modeling, accurate inference, and prediction of extreme events are very important in every field to minimize the damage due to extremes as much as possible. Probabilistic models secure a useful purpose to model and predict such extreme events. And the fitting of probabilistic models heavily depends on parameter estimation method, as every method of estimation is not suitable for every probability distribution. The method of maximum likelihood estimation (MLE) is commonly used for parameter estimation but the extreme value data are frequently heavy-tailed data and in this situation MoLM and MoTLM are considered alternatively. MoLM and MoTLM usually provides robust results in the field of extreme value analysis (e.g., see, Asquith, 2007, 2014 and Shahzad and Asghar 2013, 2015).

In this study, parameter estimation methods such as the MoLM and MoTLM are developed for Singh-Maddala distribution. These estimation methods have been compared with the method of moments (MoM) and MLE to find out the most reliable and accurate estimation method for extreme value data analysis. L- and TL-moments and their parameter estimation approach also provide the precise results with the good summarization and description of the observed sample data. Therefore, we estimated the parameters of the Singh-Maddala distribution through the MoLM and MoTLM.
The main focus of this study is to model the extreme value data through Singh-Maddala distribution. Second is to develop the MoLM and MoTLM of this density to estimate their parameters, to avoid the problems those are associated with MoM and method of MLE in the presence of heavy-tailed data. Specifically, the rest of the article is organized as follows. In Section 2, Singh-Maddala distribution is introduced. The L- and TL-moments and moment ratios of the Singh-Maddala distribution are derived in Section 3. To compare the four considered estimation methods and highlight the properties of the estimates a comprehensive simulation study is conducted in Section 4. A real data application is discussed in Section 5 observing the monthly maximum temperature data of Jacobabad, Pakistan. Finally, the study is concluded in Section 6.

2. Singh-Maddala distribution

Singh and Maddala (1976) introduced the Singh-Maddala (SM) distribution and it appeared in Econometrica and soon after, it was frequently used for the analysis of the income, wealth, consumption, expenditure and related data. McDonald and Ransom (1979) compared the Lognormal, Beta, Gamma and SM distribution for family income data for the year 1960 and 1969 to 1975 by three method of estimation and found that the fitting of SM distribution better even than Beta distribution. McDonald (1984) considered many three and four parameters probability distributions to model the grouped income distribution data and concluded that the performance of the SM distribution is best among all other distributions. Atoda, Suruga and Tachibanaki (1988) concluded that the SM distribution is more favorable than other candidate distributions for the Japanese income survey grouped data of 1975.

Henniger and Schmitz (1989) considered various distributions to model the United Kingdom family expenditure data for the period 1968–1983, but none of them accepted for the whole data set except SM and Fisk distributions. Brachmann and Trede (1996) analyzed the German household income data for 1984 – 1993 by the distributional approach and found the SM and generalized Beta-II distributions best to model such data. Dastrup et al. (2007) studied the disposable income data and found that in three parameter distributions, the best-fitted distributions are the SM and Dagum distributions. Guessous et al. (2014) compared six probability distributions to model travel time and validated the supremacy of the SM distribution in many aspects. Brzezinski (2014) modeled the empirical impact factor distribution and observed that the performance of
SM distribution much better than the other models, those were considered previously for this type of data. Sakulski et al. (2014) quantified several statistical distributions for the analysis of rainfall such as Extreme Value, Frechet, Log-normal, Log-logistic, Rice, SM and Rayleigh probability distributions for summer, autumn, winter and spring seasons and finally they stated that for all seasons, SM distribution fits quite well. Shao et al. (2004) proposed the extended three-parameter Burr XII distribution to model the flood frequency data, and method of MLE was investigated for parameter estimation of this distribution.

To estimate the parameters of SM distribution Singh and Maddala (1976) used regression method, Shah and Gokhale (1993) considered the maximum product of spacing and Stoppa (1995) derived its maximum likelihood. Shahzad and Asghar (2013) estimated only its shape parameter using L- and TL-moments by considering two shape parameters known. Herein, we have used the method of MLE, MoM, MoLM and MoTLM to estimate all the three parameters of the SM distribution.

Let $X$ be a random sample of size $n$, follows the SM distribution then its pdf is given by

$$g(x; \alpha, \beta, \delta) = \frac{a \delta x^{a-1}}{\beta^a (1 + (x/\beta)^a) \delta + 1}, \quad 0 \leq x \leq \infty,$$

where $\alpha$ and $\delta$ are the shape parameters ($\alpha, \delta > 0$) and $\beta$ is the scale parameter ($\beta > 0$).

The corresponding cdf is given by

$$G(x; \alpha, \beta, \delta) = 1 - \frac{1}{(1 + (x/\beta)^a)^\delta},$$

and the $r$th moment of the SM distribution is as follows

$$m(sm)_r = \beta^r \frac{\Gamma(1 + r/\alpha) \Gamma(\delta - r/\alpha)}{\Gamma(\delta)}.$$

3. **L- and TL-moments for Singh-Maddala distribution**

In this section, the population L- and TL-moment for the SM distribution are derived.

3.1 **L-moment**

L-moments and parameter estimation through the method of L-moments were introduced by Hosking (1990). These moments exist for all random variables whose mean can be defined in close form. Mathematically L-moments are defined as, let $X_{1:n}, X_{2:n}, ..., X_{r:n}, ..., X_{n:n}$ denote the order statistics with pdf, cdf and quantile
function, \(g(x), G(x)\) and \(x(G)\), respectively. Then, Hosking (1990) proposed the population the \(r\)th L-moment as follows

\[
T_r = \sum_{t=0}^{r-1} \binom{r-1}{t} \frac{(-1)^t}{(r-t-1)!} \int_{-\infty}^{\infty} x[G(x)]^{r-t-1} [1 - G(x)]^{t} g(x) dx.
\]

It is easy to establish the expression for a particular order of L-moments using (1).

The crosspointing \(r\)th sample L-moment \(\tilde{T}_r\) defined as

\[
\tilde{T}_r = \frac{1}{r} \sum_{1 \leq k_1 < k_2 < \ldots < k_r \leq n} \sum_{t=0}^{r-1} \frac{1}{r} \sum_{t=0}^{r-1} (-1)^t \binom{r-1}{t} X_{k_r-t:n}.
\]

The \(r\)th population L-moment for the SM distribution is derived using (1) and (2) in (4) and obtained in the following form

\[
T_{sr} = \frac{1}{r} \sum_{t=0}^{r-1} \sum_{j=0}^{r-t-1} \frac{(-1)^j}{\Gamma(r-t-j)} \frac{r!}{(r-t-1)!} \frac{\beta \delta}{t!} B \left( \frac{1}{\alpha} + 1, \delta(t+j+1) - \frac{1}{\alpha} \right),
\]

where \(B(., .)\) is the beta type-II function.

Taking \(r = 1, 2, 3\) and 4 in (5) to obtain the first four L-moments as follows

\[
T_{s1} = \beta \frac{\Gamma(1 + 1/\alpha) \Gamma(\delta - 1/\alpha)}{\Gamma(\delta)},
\]

\[
T_{s2} = \beta \Gamma \left( 1 + \frac{1}{\alpha} \right) \frac{\Gamma(\delta - 1/\alpha)}{\Gamma(\delta)} - \frac{\Gamma(2\delta - 1/\alpha)}{\Gamma(2\delta)},
\]

\[
T_{s3} = \beta \Gamma \left( 1 + \frac{1}{\alpha} \right) \left[ \frac{\Gamma(\delta - 1/\alpha)}{\Gamma(\delta)} - \frac{3\Gamma(2\delta - 1/\alpha)}{\Gamma(2\delta)} + \frac{2\Gamma(3\delta - 1/\alpha)}{\Gamma(3\delta)} \right],
\]

\[
T_{s4} = \beta \Gamma \left( 1 + \frac{1}{\alpha} \right) \left[ \frac{\Gamma(\delta - 1/\alpha)}{\Gamma(\delta)} - \frac{6\Gamma(2\delta - 1/\alpha)}{\Gamma(2\delta)} + \frac{10\Gamma(3\delta - 1/\alpha)}{\Gamma(3\delta)} \right]
\]

Equating the population and corresponding sample L-moments, we can estimate the parameter of the distribution.

### 3.2 TL-moments

Elamir and Seheult (2003) introduced some robust modification in L-moments and proposed TL-moments, which is a generalization of L-moments. TL-moments are more
robust towards outliers than L-moments. TL-moments provide the opportunity to trim any possible the number of observations from the data.

Elamir and Seheult (2003) defined the $r$th TL-moment of the random variable $X$ as

$$T_r^{(1)} = \frac{1}{r} \sum_{t=0}^{r-1} (-1)^t \left( \begin{array}{c} r - 1 \\ t \end{array} \right) \frac{(r + 2)!}{(r - t)! (t + 1)!} \int_0^\infty F^{r-t} (1 - F)^{t+1} dF. \quad (2)$$

It is easy to obtain the particular TL-moments for specific value of $r$.

The sample TL-moments $J_r^{(1)}$, corresponding to the population L-moments $T_r^{(1)}$ are as follows

$$J_r^{(1)} = \frac{1}{r} \sum_{j=2}^{r-1} \left[ \frac{\sum_{t=0}^{r-1} (-1)^k \left( \begin{array}{c} r - 1 \\ t \end{array} \right) \left( \begin{array}{c} j - 1 \\ r - t \end{array} \right) \left( \begin{array}{c} n - j \\ t + 1 \end{array} \right)}{(n + 2)} \right] X_{i:n}.$$  

The $r$th TL-moment for SM distribution is derived using **Error! Reference source not found.**, (1) and (2) and obtained in the following form

$$T_r^{(1)} = \frac{1}{r} \sum_{t=0}^{r-1} \sum_{j=0}^{t-1} \left( \begin{array}{c} r - t \\ j \end{array} \right) \left( \begin{array}{c} r - 1 \\ t \end{array} \right) (-1)^{t+j} \frac{(r + 2)!}{(r - t)! (t + 1)!} \beta \delta$$

$$\times B \left( \frac{1}{\alpha} + 1, \delta(t + j + 2) - \frac{1}{\alpha} \right).$$

Taking $r = 1, 2, 3$ and $4$ in (5) to obtain the first four TL-moments as follows

$$T_{S1}^{(1)} = \beta \Gamma \left( 1 + \frac{1}{\alpha} \right) \left[ \frac{3 \Gamma(2\delta - 1/\alpha)}{\Gamma(2\delta)} - \frac{2 \Gamma(3\delta - 1/\alpha)}{\Gamma(3\delta)} \right], \quad (12)$$

$$T_{S2}^{(1)} = 3 \beta \Gamma \left( 1 + \frac{1}{\alpha} \right) \left[ \frac{\Gamma(2\delta - 1/\alpha)}{\Gamma(2\delta)} - \frac{2 \Gamma(3\delta - 1/\alpha)}{\Gamma(3\delta)} + \frac{\Gamma(4\delta - 1/\alpha)}{\Gamma(4\delta)} \right], \quad (13)$$

$$T_{S3}^{(1)} = \frac{10 \beta}{3} \Gamma \left( 1 + \frac{1}{\alpha} \right) \left[ \frac{\Gamma(2\delta - 1/\alpha)}{\Gamma(2\delta)} - \frac{4 \Gamma(3\delta - 1/\alpha)}{\Gamma(3\delta)} + \frac{5 \Gamma(4\delta - 1/\alpha)}{\Gamma(4\delta)} - \frac{2 \Gamma(5\delta - 1/\alpha)}{\Gamma(5\delta)} \right], \quad (14)$$
\[ T_{S4}^{(1)} = 15 \beta \Gamma \left(1 + \frac{1}{\alpha}\right) \left[\frac{\Gamma(2\delta - 1/\alpha)}{4\Gamma(2\delta)} - \frac{5\Gamma(3\delta - 1/\alpha)}{3\Gamma(3\delta)} + \frac{15\Gamma(4\delta - 1/\alpha)}{4\Gamma(4\delta)} \right. \\
\left. - \frac{7\Gamma(5\delta - 1/\alpha)}{2\Gamma(5\delta)} + \frac{7\Gamma(6\delta - 1/\alpha)}{6\Gamma(6\delta)} \right]. \]  

Equating the population and corresponding sample TL-moments, we can estimate the parameter of the SM distribution through simultaneous equation solution.

### 3.3 L- and TL-moments ratios

The L- and TL-moments ratios are analogues to the C-moments ratio and have the same interpretation but L- and TL-moments summarize the probability distribution more accurately than the conventional measures (Hosking, 1990). The first L-moment ratio lies in the range \( 0 < Y < 1 \). The \( Y_3 \) is used to measure the asymmetry and it lies between 0 and 1. Hosking and Wallis (2005) proved the range of the \( Y_4 \) and it lies between \((5Y_3^2 - 1)/4\) and 1. The population L-moments ratios for SM distribution are obtained as follows

\[ Y_{S1} = 1 - \frac{\Gamma(2\delta - 1/\alpha)}{\Gamma(2\delta)} \frac{\Gamma(\delta - 1/\alpha)}{\Gamma(\delta)} \frac{\Gamma(\delta)}{\Gamma(\delta - 1/\alpha)} \]

\[ Y_{S3} = \frac{\Gamma(\delta - 1/\alpha)}{\Gamma(\delta)} - \frac{3\Gamma(2\delta - 1/\alpha)}{\Gamma(2\delta)} + \frac{2\Gamma(3\delta - 1/\alpha)}{\Gamma(3\delta)} - \frac{\Gamma(2\delta - 1/\alpha)}{\Gamma(2\delta)} \]

\[ Y_{S4} = \frac{\Gamma(\delta - 1/\alpha)}{\Gamma(\delta)} - \frac{6\Gamma(2\delta - 1/\alpha)}{\Gamma(2\delta)} + \frac{10\Gamma(3\delta - 1/\alpha)}{\Gamma(3\delta)} - \frac{5\Gamma(4\delta - 1/\alpha)}{\Gamma(4\delta)} - \frac{\Gamma(2\delta - 1/\alpha)}{\Gamma(2\delta)} \]

The population TL-moments ratios are derived and obtained as follow

\[ \gamma_{S1}^{(1)} = \frac{3}{9} \left[ \frac{\Gamma(2\delta - 1/\alpha)}{\Gamma(2\delta)} - \frac{2\Gamma(3\delta - 1/\alpha)}{\Gamma(3\delta)} + \frac{\Gamma(4\delta - 1/\alpha)}{\Gamma(4\delta)} \right] \]

\[ \gamma_{S3}^{(1)} = \frac{10}{9} \left[ \frac{\Gamma(2\delta - 1/\alpha)}{\Gamma(2\delta)} - \frac{2\Gamma(3\delta - 1/\alpha)}{\Gamma(3\delta)} + \frac{\Gamma(4\delta - 1/\alpha)}{\Gamma(4\delta)} \right] \]
The general range of TL-moments is still not available in the literature.

4. Simulation Study

A simulations study has been carried out for two purposes. First, is to investigate and compare the performance of the MLE, MoM, MoLM and MoTLM estimation techniques. Second, is to explore the impact of sample size on estimation techniques. Keeping it in mind, simulation study is presented, to compare the properties of the estimation methods for the SM distribution on their bias, mean square error of estimates (MSE). The data is simulated using the R-language assuming different sample sizes, \( n \in (25, 50 \text{ and } 200) \) and different value of each parameter and each sample is repeated 1000 times, to obtain the precision and accuracy. The summary of the results for Maximum likelihood estimates (MLEs), Method of moments estimates (MMEs), L-moments estimates (LMEs) and TL-moments estimates (TLMEs) are presented in Table 1 and it is self-explanatory. In general the TL-moments trim the outlier values from both sides of the data and the data become less skewed, in this way its estimates are better than L-moments estimates. The parameter estimates of L-moments are accurate and efficient than the estimates attained using MLEs, particularly for small sample and approximately equal in large sample. All three types of moment ratios also computed to summarize the data.

\[
Y_{s4}^{(1)} = \frac{5}{\Gamma(2\delta - 1/\alpha)} \times \left[ \frac{\Gamma(2\delta - 1/\alpha)}{\Gamma(2\delta)} - \frac{2\Gamma(3\delta - 1/\alpha)}{\Gamma(3\delta)} + \frac{\Gamma(4\delta - 1/\alpha)}{\Gamma(4\delta)} - \frac{7\Gamma(5\delta - 1/\alpha)}{2\Gamma(5\delta)} \right] + \frac{\Gamma(6\delta - 1/\alpha)}{6\Gamma(6\delta)}.
\]

The general range of TL-moments is still not available in the literature.

Table 1: Summary of the average bias and MSEs of all the estimators (\( \hat{\theta} \)) of the Singh-Maddala distribution for different sample size.
5. Application

In this section, we have compared the four considered estimation methods using monthly maximum temperature data of Jacobabad, Pakistan. Jacobabad is one of the hottest cities of Pakistan, and its highest recorded temperature is 52.8 °C. The geographical location of this city has Latitude 28.28 North and Longitude 68.45 South and it is famous due to the consistently highest temperature in South Asia. The data of the monthly maximum temperature of Jacobabad retrieved from the Pakistan Meteorological Department (PMD) Islamabad. The length of data is 391 records from January 1981 to December 2013 excluding an unobserved or unreported month.

In order to compare the estimation methods, the Kolmogorov-Smirnov (KS) goodness of fit test is considered and fitting of SM distribution is also displayed graphically. The smaller value of the KS-test, better the method will be. The SM-distribution is a three-parameter distribution; therefore, the first three sample moments are equated with the population moments to compute the parameter estimates. Herein Newton-Raphson iterative estimation is employed for the solution of the system of nonlinear equations and finally found out the estimates for all four considered estimation methods. The results are reported in Table 2.

Table 2: Parameter estimates of the SM-distribution using different parameter estimation methods and the result of the goodness of fit test.

<table>
<thead>
<tr>
<th>Estimation Method</th>
<th>Parameter Estimate</th>
<th>KS-test</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>TL-moments</td>
<td>( \hat{\delta} = 5.17328 )</td>
<td>0.0700</td>
<td>0.0417</td>
</tr>
<tr>
<td>( \hat{\beta} = 51.6708 )</td>
<td>( \hat{\delta} = 5.93200 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Method</td>
<td>$\hat{\alpha}$</td>
<td>$\hat{\beta}$</td>
<td>$\hat{\delta}$</td>
</tr>
<tr>
<td>---------------------</td>
<td>----------------</td>
<td>--------------</td>
<td>---------------</td>
</tr>
<tr>
<td>L-moments</td>
<td>5.48775</td>
<td>61.7679</td>
<td>17.1670</td>
</tr>
<tr>
<td>C-moments</td>
<td>6.07706</td>
<td>36.0512</td>
<td>1.49854</td>
</tr>
<tr>
<td>Maximum Likelihood</td>
<td>5.46013</td>
<td>117.570</td>
<td>611.647</td>
</tr>
</tbody>
</table>

It has been observed that the method of TL-moments has provided better fitting and estimates as compared to the other methods. As the KS-test has minimum value for the method of TL-moments. The fitting of the SM distribution using estimates of the considered methods are presented graphically as follows:
It is obvious from empirical results and Figure 1 that the method of TL-moments has produced more accurate fitting on the real data set and the second good fitted estimation method is the method of L-moments. Same conclusion is drawn from the presentation of the PP-plots.

Finally, it is observed that the MoM does not produce satisfactory results. MoLM and MoTLM provide better results than the method of MLE and MoM.
According to the statistical and graphical presentations, TL-moments estimation method is better as it has provided the superior fit on the data set. The moments and moment ratios are commonly used to find the characteristics of the probability distribution of the observed data set. These moments are calculated using Conventional (C), L- and TL-moments and presented in Table 3.

Table 3: C-, L- and TL-moments and moments ratios for monthly maximum temperature of Jacobabad

<table>
<thead>
<tr>
<th>rth</th>
<th>C-moments</th>
<th>L-moments</th>
<th>TL-moments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>34.168350</td>
<td>34.168350</td>
<td>34.411185</td>
</tr>
<tr>
<td>2nd</td>
<td>1219.9000</td>
<td>4.1591960</td>
<td>2.4469623</td>
</tr>
<tr>
<td>3rd</td>
<td>45186.230</td>
<td>-0.2428311</td>
<td>-0.1919081</td>
</tr>
<tr>
<td>4th</td>
<td>1724723.0</td>
<td>0.0809250</td>
<td>0.0596961</td>
</tr>
<tr>
<td>CV</td>
<td>0.2121728</td>
<td>0.1217266</td>
<td>0.0711095</td>
</tr>
<tr>
<td>Sk</td>
<td>-0.2060238</td>
<td>-0.0583841</td>
<td>-0.0784271</td>
</tr>
</tbody>
</table>

6. Conclusion

It is a standard statistical practice to use the moments to summarize the behaviour of an observed data. Therefore, the L- and TL-moments for the SM distribution are derived to summarize the data accurately. These derived moments are also used to estimate the parameters of the density and then MoLM and MoTLM are established by equating the population and sample moments. To compare these estimation methods with MLE and MoM, a simulation study has been carried. The results of the simulation study are indicated that the estimates of the L- or TL-moments are the least bias with minimum MSE than the other considered estimation methods. It has become more obvious as the sample size increased. These conclusions are also justified in real dataset application.
Additionally, the shape of the distributions has been investigated through L- and TL-moments numerically.

References

