Estimation of multi-truncated rayleigh and Weibull distributions with application

Kawther Alhasan*, Kareema Abad Al-Kadim

*Department of Mathematics, College of Education for Pure Science, University of Babylon, Iraq

(Communicated by Madjid Eshaghi Gordji)

Abstract

In this paper, we introduced the estimation of parameters for distributions that have multi-double truncation by the maximum likelihood method. We depended in the applied study on data of disease of Covid-19 by using one double truncation. Some statistical data were analyzed and unknown parameters and were compared with some distribution.

Keywords: estimation, Covid-19, double truncation, Weibull distributions

1. Introduction

Many researchers are very interested in statistical models for analyzing and studying phenomena in the real world. Several distributions have been widely used over the past decades for modeling data in many phenomena medical, engineering, economic fields, etc.

Later, the researchers focus on adding new distributional models depending on (truncated or mixture) classical distributions, which are considered more flexible in modeling data. The truncation in data is a delete data (intervals, or points) from the original data.

In statistic distribution, the definition of truncation is the cut apart from the curve of the distribution if it is continuous, that is, cut off the subinterval from a domain of the random variable.

within $[0,1]$ represented by the truncated Fréchet -Pareto distribution. A New family of upper truncated distributions was suggested by H. Amal, A. Mohamed in (2020) [3].

In (2021) Kawther Alhasan and Kareema have suggested a new truncation: multi–double truncated on continuous distribution [1], and studied the MDTRD [12, 11]. In this paper, we introduced the estimation of parameters for distributions with multi-truncated which have the type of truncation is double by using maximum likelihood estimation for real data that taken of infected with COVID-19 in the city of Babylon-Iraq. Real data set is taken to appreciate the fitness and applicability of the intended distribution. It is observed that it can be used quite effectively to analyze lifetime data and do better as compared to the other distributions.

2. The Maximum Likelihood Estimation For The Multi-Double Truncation Distribution

A random variable $X_T$ is said to have the multi-double truncated distribution if its probability density function is given by [2]:

$$g_N(X_T) = \frac{1}{n} \sum_{i=1}^{n} f(x_T) \left( F(b_i) - F(a_i) \right) \text{ when } a_i < x_T < b_i$$

$$L = \prod_{i=1}^{n} g_N(x_i; \theta)$$

Then the log-likelihood function of multi-double truncated continuous distribution is:

$$\ell = -n \log(n) - r_1 \log(F_1) - (r_2 - r_1) \log(F_1) - ... - (n - \sum_{i=1}^{n-1} r_i) \log(F_n) + \sum_{i=1}^{n} \log(f(x_T))$$

3. One-Double Truncated Rayleigh Distribution(ODTR)

The probability distribution function of one interval truncated of Rayleigh distribution is given by [1]:

$$f(x_T) = \begin{cases} \frac{x_T^2}{\sigma^2} e^{-\frac{x_T^2}{2\sigma^2}} & b_0 < x_T < a_1 \\
\frac{x_T^2}{\sigma^2} e^{-\frac{x_T^2}{2\sigma^2}} \left( e^{-\frac{a_1^2}{2\sigma^2}} - e^{-\frac{b_0^2}{2\sigma^2}} \right) & b_1 < x_T < a_2 \\
\frac{x_T^2}{\sigma^2} e^{-\frac{x_T^2}{2\sigma^2}} \left( e^{-\frac{a_2^2}{2\sigma^2}} - e^{-\frac{a_1^2}{2\sigma^2}} \right) & b_1 < x_T < a_2 \end{cases}$$

that is, the interval $(a_1, b_1)$, is truncated.

In figure 1 we show the curve of one double truncated Rayleigh distribution for different values of the parameters $\sigma = \{5, 4.5, 4\}$, that is the interval which deletes is $(4, 8)$. So if we have two intervals $(b_0, a_1) = (0, 4), (b_1, a_2) = (5, 10)$, then the interval which deleting is $(a_1, b_1) = (3, 5)$ we see it in the figure 2 for different values of the parameters $\sigma = \{3, 4.5, 4\}$. While, in figure 3, the same interval which deleting is $(5, 10)$ for different values for parameter $\sigma = \{7, 6.5, 4\}$ and $\sigma = \{6.5, 6, 5.4\}$ respectively [1].
Figure 1: One interval (4,8) is cut

Figure 2: One interval (3,5) is cut

Figure 3: One interval (5,10) is cut

Figure 4: One interval (5,10) cut
4. Maximum Likelihood Estimates of ODTRD

If \( X_1, X_2, ..., X_n \) denotes a random sample from ODTRD, then the likelihood function is given by:

\[
L = \prod_{i=1}^{n} g(x_i; \sigma^2)
\]

\[
= \prod_{i=1}^{r} \left( \frac{x_i^2 e^{-x_i^2/2\sigma^2}}{2 (e^{x_i^2/2\sigma^2} - e^{-x_i^2/2\sigma^2})} \right) \prod_{i=r+1}^{n} \left( \frac{x_i^2 e^{-x_i^2/2\sigma^2}}{2 (e^{x_i^2/2\sigma^2} - e^{-x_i^2/2\sigma^2})} \right)
\]

\[
= \left( \frac{\prod_{i=1}^{r} x_i^2 e^{-x_i^2/2\sigma^2}}{2^r (e^{x_i^2/2\sigma^2} - e^{-x_i^2/2\sigma^2})^{r}} \right) \left( \frac{\prod_{i=r+1}^{n} x_i^2 e^{-x_i^2/2\sigma^2}}{2^{n-r} (e^{x_i^2/2\sigma^2} - e^{-x_i^2/2\sigma^2})^{n-r}} \right)
\]

The log-likelihood function is

\[
\ell = \sum_{i=1}^{n} \log x_i - \frac{\sum_{i=1}^{n} x_i^2}{2\sigma^2} - n \log(\sigma^2) - n \log 2 - r \left[ \log \left( \exp \left( \frac{-b_1^2}{2\sigma^2} \right) - \exp \left( \frac{-a_1^2}{2\sigma^2} \right) \right) \right] - (n - r) \left[ \log \left( \exp \left( \frac{-b_1^2}{2\sigma^2} \right) - \exp \left( \frac{-a_1^2}{2\sigma^2} \right) \right) \right]
\]

By taking the derivatives of \( \ell \) with respect to the parameter \( \sigma^2 \) and the result equal zero,

\[
\frac{d\ell}{d\sigma^2} = -r \left[ \left( \frac{a_1^2}{2(\sigma^2)^2} \right) \exp \left( \frac{-a_1^2}{2\sigma^2} \right) - \left( \frac{b_1^2}{2(\sigma^2)^2} \right) \exp \left( \frac{-b_1^2}{2\sigma^2} \right) \right] + \frac{\sum_{i=1}^{n} x_i^2}{2\sigma^2} + \frac{n}{\sigma^2}
\]

\[
- (n - r) \left[ \left( \frac{a_1^2}{2(\sigma^2)^2} \right) \exp \left( \frac{-a_1^2}{2\sigma^2} \right) - \left( \frac{b_1^2}{2(\sigma^2)^2} \right) \exp \left( \frac{-b_1^2}{2\sigma^2} \right) \right]
\]

\[
= 0.
\]

We can obtain the MLE of the parameter \( \sigma^2 \), by solving the above equation numerically for \( \sigma^2 \).

5. Real Data Analysis

In this section, we illustrate the applicability of ODTRD distribution by considering real data of the infected with COVID-19 in the city of Babylon-Iraq. Data was collected for the period of their infection with the virus until their death for the interval (1/10/2020-1/12/2020). We also fit Maxwell distribution(MAWE) and Chi-Square Distribution.

The data set are:

11, 22, 16, 17, 6, 12, 7, 29, 11, 7, 10, 15, 8, 1, 21, 8, 1, 17, 0, 1, 1, 10, 21, 16, 12, 1, 2, 7, 4, 11, 5, 5, 8, 4, 15, 1, 23, 16, 8, 10, 10, 11, 27, 9, 18, 1, 6, 3, 26, 4, 1, 3, 11, 2, 12, 3, 32, 12, 1, 5, 10, 14, 3, 0, 0, 66, 51, 20, 8, 1, 15, 18, 22, 16, 0, 19, 8, 21, 19, 12, 8, 3, 9, 17, 7, 12, 12, 44, 11, 23, 12, 10, 8, 11, 27, 10, 4, 5, 6, 11, 1, 28, 22, 6, 7, 31, 6, 16, 0, 2, 12, 21, 19, 2, 19, 8, 18, 39. With size is 118.
5.1. Maxwell distribution \[^7\]

The probability density function of (MAWE) introduced by (James Maxwell and Boltzmann), is:

\[
f(X, \theta) = \frac{4}{\sqrt{\pi}} \frac{X^2 e^{-\frac{X^2}{\theta^2}}}{\theta^{3/2}} ; \quad X, \theta > 0
\]

5.2. Chi-Square Distribution

The probability density function of (CH-SQ) was introduced by (James Maxwell and Boltzmann). The pdf of (CH-SQ) is:

\[
f(X, \alpha) = \frac{X^{\alpha - 1} e^{-\frac{X}{2}}}{2^{\frac{\alpha}{2}} \Gamma\left(\frac{\alpha}{2}\right)} ; \quad X > 0, \; \alpha \text{ is degrees of freedom.}
\]

Each parameter of the above distributions is estimated by using the maximum likelihood method. Some criteria are used for comparison, Akaike information criterion (AIC), Bayesian information criterion (BIC), Corrected Akaike Information criterion (CAIC), and Hannan-Quinn information criterion (HQIC), which are used to select the best model among several models. The expressions to compute AIC, BIC, CAIC, and HQIC are given below:

\[
\text{AIC} = -2\hat{\ell} + 2q,
\]
\[
\text{BIC} = -2\hat{\ell} + q \log(n),
\]
\[
\text{CAIC} = -2\hat{\ell} + \frac{2q}{n-q-1},
\]
\[
\text{HQIC} = -2\hat{\ell} + 2q \log(\log(n)),
\]

By using these criteria, we can note the distribution which has the least criteria is the better fit for the data, as shown in table \[^2\].

The maximal likelihood estimation of all models parameters for the above data is shown in table \[^1\].

And we can see in Table \[^2\] that the model ODTRD gives the least values for the criteria AIC, BIC, CAIC, and HQIC than the other models (MAWE, CH-SQ).

### Table 1: Estimations parameters of data for ODTRD

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameter</th>
<th>Parameter estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>ODTR ((\sigma^2))</td>
<td>(\sigma^2)</td>
<td>4.753</td>
</tr>
<tr>
<td>MAWE ((\theta))</td>
<td>(\theta)</td>
<td>8.996</td>
</tr>
<tr>
<td>CH-SQ ((\alpha))</td>
<td>(\alpha)</td>
<td>1.935</td>
</tr>
</tbody>
</table>

This table \[^1\] shows the values of the estimator parameter of (ODTRD) and all the distributions compared to it, using the maximum likelihood method.

In this table \[^2\], the numerical values of statistics for all models are given and show the model ODTRD is the better fit of data than other models.

6. One-Double Truncated Weibull Distribution (ODTWD)

The probability distribution function of one interval truncated of Weibull distribution \[^2\]:

\[
\text{Table 1: Estimations parameters of data for ODTRD}
\]
Table 2: the values of statistics $\ell$, AIC, BIC, CAIC, HQIC

<table>
<thead>
<tr>
<th>Model</th>
<th>$\ell$</th>
<th>AIC</th>
<th>BIC</th>
<th>CAIC</th>
<th>HQIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>ODTR</td>
<td>-533.079</td>
<td>3.066158</td>
<td>3.138.038</td>
<td>1.100158</td>
<td>1.698858</td>
</tr>
<tr>
<td>MAWE</td>
<td>-1.631.689</td>
<td>5.263378</td>
<td>5.335258</td>
<td>3.297378</td>
<td>3.896078</td>
</tr>
<tr>
<td>CH-SQ</td>
<td>-794.687</td>
<td>3.589374</td>
<td>3.661254</td>
<td>1.623374</td>
<td>2.2222074</td>
</tr>
</tbody>
</table>

$$g_2(\lambda_T) = \begin{cases} 
\frac{\frac{\lambda}{x_T^{\delta-1}}e^{-\left(\frac{x_T}{\lambda}\right)\delta}}{2\left(e^{-\left(\frac{a_1}{\lambda}\right)\delta} - e^{-\left(\frac{b_1}{\lambda}\right)\delta}\right)} & -\infty < x_T < b_1 \\
\frac{\frac{\lambda}{x_T^{\delta-1}}e^{-\left(\frac{x_T}{\lambda}\right)\delta}}{2\left(e^{-\left(\frac{a_2}{\lambda}\right)\delta} - e^{-\left(\frac{b_2}{\lambda}\right)\delta}\right)} & a_2 < x_T < \infty 
\end{cases}$$

that is, a deleted interval is $(b_1, a_2)$.

The Fig. 5,6,7 show that, the curve of one double truncated Weibull distribution for different values of the parameters $\delta, \omega$. Note that, if there exist two intervals $(a_1, b_1), (a_2, b_2)$, then the interval which deleting is $(b_1, a_2) = (6, 10)$ [2].

![Figure 5](image1)

Figure 5: One interval [6,10] is truncated

![Figure 6](image2)

Figure 6: One interval [6,10] is truncated
7. Likelihood Estimates of ODTWD

If \( X_1, X_2, \ldots, X_n \) denotes a random sample from ODTRD, then the likelihood function is given by:

\[
L = \prod_{i=1}^{n} g(x_i; \delta, \omega) = \prod_{i=1}^{r} \left( \frac{\delta x_i^{\delta-1} e^{-\left(\frac{x_i}{\omega}\right)^\delta}}{2 \left(e^{-\left(\frac{n_1}{\omega}\right)^\delta} - e^{-\left(\frac{n_2}{\omega}\right)^\delta}\right)} \right) \prod_{i=r+1}^{n} \left( \frac{\delta x_i^{\delta-1} e^{-\left(\frac{x_i}{\omega}\right)^\delta}}{2 \left(e^{-\left(\frac{n_3}{\omega}\right)^\delta} - e^{-\left(\frac{n_4}{\omega}\right)^\delta}\right)} \right)
\]

The log-likelihood function is:

\[
\ell = n \log(\delta) - n \delta \log(\omega) + \sum_{i=1}^{n} \log x_i^{\delta-1} - \frac{\sum_{i=1}^{n} x_i^{\delta}}{\delta} - n \log(2) - r \left[ \log \left( \exp \left( -\left(\frac{b_1}{\omega}\right)^\delta \right) \right) - \log \left( \exp \left( -\left(\frac{a_1}{\omega}\right)^\delta \right) \right) \right] - (n - r) \left[ \log \left( \exp \left( -\left(\frac{b_2}{\omega}\right)^\delta \right) \right) - \log \left( \exp \left( -\left(\frac{a_2}{\omega}\right)^\delta \right) \right) \right].
\]

By taking the derivatives of \( \ell \) with respect to the parameter \( \sigma^2 \) and the result equal zero,

\[
\frac{\partial \ell}{\partial \delta} = n - r \left[ \log \left( \exp \left( -\left(\frac{b_1}{\omega}\right)^\delta \right) \right) - \log \left( \exp \left( -\left(\frac{a_1}{\omega}\right)^\delta \right) \right) \right]
\]

\[
- n \log(\omega) - 2 \log(\delta) \sum_{i=1}^{n} x_i^{\delta} - r \left[ \log \left( \exp \left( -\left(\frac{b_2}{\omega}\right)^\delta \right) \right) - \log \left( \exp \left( -\left(\frac{a_2}{\omega}\right)^\delta \right) \right) \right].
\]

8.3. Rayleigh with two parameters distribution [5]

8.1. Gamma Rayleigh distribution [8]

In this section, we used the real data of the infected with COVID-19 in the city of Babylon-Iraq. Data were presented for the period of their infection with the virus until their death for the interval (1/10/2020-1/12/2020). Compared the results of fitting the importance of the ODTRD with the Gamma Rayleigh distribution (GARA), Modified Weibull distribution (MOWE), Rayleigh with two parameters (RATO), and Weibull distribution (WEBU), the distributions are:

8. Discussion of the real data on model ODTWD

We can obtain the MLE of the parameters \( \delta, \omega \), by solving the above equation numerically for \( \delta, \omega \).


The probability density function of (GARA) introduced by (Aliya Syed Malik and S.P. Ahmad 2019), is:

\[
f(x; \alpha, \theta, \omega) = \frac{X}{\theta^2 \Gamma(\alpha) \omega^\alpha} \frac{X^2 e^{-\frac{X^2}{2\omega^2}}}{\theta^2} \left( \frac{X^2}{2\theta^2} \right)^{\alpha - 1} ; \quad X > 0, \quad \alpha, \theta, \omega > 0
\]


The probability density function of (MOWE) introduced by (M. Ammar and Z. Mazen), is:

\[
f(x; \alpha, \omega, \gamma) = (\alpha + \omega \gamma x^{\gamma - 1}) e^{-\alpha x + \omega x^{\gamma - 1}} ; \quad X > 0, \quad \alpha, \gamma, \omega > 0
\]

8.3. Rayleigh with two parameters distribution [5]

The probability density function of (RATO) introduced by (Sanku, Tanujit, and Debasis 2003), is:

\[
f(x; \omega, \gamma) = 2\omega (x - \gamma) e^{-\omega(x - \gamma)^2} ; \quad X > 0, \quad \gamma, \omega > 0
\]
8.4. Weibull distribution

Weibull distribution is named after Swedish mathematician Waloddi Weibull, who described it in detail in 1951. The probability density function of (WEBU) is:

\[ f(x; \omega, \gamma) = \frac{\omega}{\gamma} \left( \frac{x}{\gamma} \right)^{\omega - 1} e^{- \left( \frac{x}{\gamma} \right)^{\omega - 1}}; \quad X > 0, \quad \alpha, \gamma, \omega > 0 \]

By using the above criteria, we can note the distribution which has the least criteria is the better fit for the data, as shown in table 4.

The maximal likelihood estimation of a model (ODTW) and (WEBU), (GARA), (MOWE), and (RATW) parameters for the above data is shown in table 3, where \( r = 76 \).

And we can see in the table 4, that the model ODTWD gives the least values for the criteria than the other models, where the iteration of a program by MatLab is 405.

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameters Estimates</th>
<th>( \hat{\ell} )</th>
<th>AIC</th>
<th>BIC</th>
<th>CAIC</th>
<th>HQIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>ODTW(( \delta, \omega ))</td>
<td>( \hat{\delta} = 9.973 )</td>
<td>( \hat{\omega} = 0.85 )</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WEBU(( \omega, \gamma ))</td>
<td>( \hat{\omega} = 0.94 )</td>
<td>( \hat{\gamma} = 8.939 )</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GARA(( \alpha, \theta, \omega ))</td>
<td>( \hat{\alpha} = 9.974 )</td>
<td>( \hat{\theta} = 9.974 )</td>
<td>( \hat{\omega} = 0.791 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MOWE(( \alpha, \omega, \gamma ))</td>
<td>( \hat{\alpha} = 0.94 )</td>
<td>( \hat{\omega} = 1.569 )</td>
<td>( \hat{\gamma} = 0.01 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RATW(( \omega, \gamma ))</td>
<td>( \hat{\omega} = 0.94 )</td>
<td>( \hat{\gamma} = 0.988 )</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This table 4 show the statistics \( \hat{\ell}, AIC, BIC, CAIC, \) and HQIC for (ODTR), (WEBU), (GARA), (MOWE), and (RATW), and the model ODTWD is the better fit of data than other models.

9. Conclusion

In this study, the general formula for maximum likelihood estimation to multi-double truncated continuous distribution has been derived. An estimation of two new distributions (one double truncated Rayleigh and one double truncated Weibull) was introduced. We have taken real data set to explain the methodology of multi-double truncated. We used the MatLab program to calculate the maximum likelihood estimates. We considered the models (ODTRD) and (ODTWD) with other distributions for comparison based on various information criteria such as AIC, BIC, CAIC, HQIC, and it showed the models (ODTRD) and (ODTWD) are better than other distributions with which they were compared.
References


