Estimation of the scale parameter in truncated Rayleigh distribution

Jiexiang Li

Department of Mathematics, College of Charleston, U.S.A.

Abstract

The paper investigated maximum likelihood (ML) estimation of the scale parameter in truncated Rayleigh distribution (TRD). Some important statistical properties of MLE for the scale parameter in TRD are derived.

JEL classification numbers: C13, C46, C55

Key words and phrases. Truncated Rayleigh distribution (TRD), maximum likelihood, Fisher information.

Send Correspondence to:
Jiexiang Li
College of Charleston
Math Dept.
Robert Scott Small Room 339
Charleston, SC 29424

1 Introduction

Truncated data arise frequently in different situations. An example of truncated data that is relevant to almost everyone is given by Wikipedia: "If policyholders are subject to a policy limit T, then any loss amounts that are actually above T are reported to the insurance company as being exactly T because T is the amount the insurance company pays. The insurer knows that the actual loss is greater than T but they don't know what it is. On the other hand, left truncation occurs when policyholders are subject to a deductible. If policyholders are subject to a deductible D, any loss amount that is less than D will not even be reported to the insurance company. If there is a claim on a policy limit of T and a deductible of D, any loss amount that is greater than T will be reported to the insurance company as a loss of $T - D$ because that is the amount the insurance company has to pay." In statistical literature, Zhang and Xie (2011) investigated on upper truncated Weibull distribution. Wingo (1988) studied on fitting right-truncated Weibull distribution to life-test and survival data. Rayleigh distribution introduced by Lord Rayleigh in 1880 plays a crucial role in modeling and analyzing life time data such as project effort loadings modelling, life testing experiments, reliability analysis, communication theory, physical sciences, engineering, medical imaging science, applied statistics and clinical studies. Due to the importance of Raleigh distribution in a variety of fields, a wide range of investigations of Raleigh Distribution has been established. Siddiqui (1962) worked on some problems connected with Rayleigh distributions. Lalitha and Khan (1996) studied modified maximum likelihood estimation for Rayleigh distribution and Provost (2009) studied predictive densities from the Rayleigh life model in the presence of different censoring sampling schemes. In this paper, we will estimate the scale parameter in Rayleigh distribution. We will start with right truncated Rayleigh distribution, left truncated Rayleigh distribution then doubly truncated Rayleigh distribution. In the end, we will fit right truncated Rayleigh distribution to the data studied in Siddiqui (1962) example 3, and use the results derived in the paper to carry out statistical inference on the scale parameter.

2 Right Truncated Rayleigh Distribution

$$f(x) = \frac{2x}{\theta} \exp{(-\frac{x^2}{\theta})} \tag{1}$$

Here θ is a scale parameter. The characteristics of this function is well known. However, the characteristics of this function are different if some of the values of the r.v. Y are right truncated, which happens when the increasing hazard saturate at a time point. Consider the probability density function (pdf) of RTRD at T,

$$f(x) = \frac{\frac{2x}{\theta} \exp{(-\frac{x^2}{\theta})}}{1 - \exp{(-\frac{T^2}{\theta})}} \tag{2}$$

for $0 < x < T$ and $\theta > 0$. We will consider the statistical inference of scale parameter θ when truncation point T is known. Let X_1, X_2,...,X_n be a random sample from RTRD specified in (2), we will study the maximum likelihood estimator for θ. For notation purpose, let $\exp{(-\frac{T^2}{\theta})} = b$, it follows from (2),

$$\begin{aligned}
E(X^2) &= \frac{1}{1-b} \int_0^T \frac{2x^3}{\theta} \exp{(-\frac{x^2}{\theta})} dx \\
&= \frac{\theta}{1-b} \int_0^{T^2/\theta} y \exp{(-y)} dy \\
&= \frac{\theta}{1-b} [-b\frac{T^2}{\theta} + 1 - b] \\
&= \theta - \frac{bT^2}{1-b}.
\end{aligned}$$

Note

$$\begin{aligned}
F(x) &= \frac{1}{1-b} \int_0^x \frac{2t}{\theta} \exp{(-\frac{t^2}{\theta})} dt \\
&= \frac{1}{1-b} (1 - \exp{(-\frac{x^2}{\theta})}).
\end{aligned}$$

$$\begin{aligned}
u &= \frac{1 - \exp{(-\frac{x^2}{\theta})}}{1 - \exp{(\frac{-T^2}{\theta})}} \\
(1 - \exp{(\frac{-T^2}{\theta})})u &= 1 - \exp{(-\frac{x^2}{\theta})} \\
1 - (1 - \exp{(\frac{-T^2}{\theta})})u &= \exp{(-\frac{x^2}{\theta})} \\
x &= \sqrt{-\theta \ln(1 - (1 - \exp{(\frac{-T^2}{\theta})})u)}. \qquad (3)
\end{aligned}$$

Expression (3) will be used to generate observations from RTRD. The log-likelihood function of a random sample X_1, X_2,...,X_n from RTRD specified in (2) is given by

$$L(\theta) \quad = \quad \sum_{i=1}^{n} \ln(2x_i) - n\ln\theta - \sum_{i=1}^{n} \frac{x_i^2}{\theta} - n\ln(1-\exp{(-\frac{T^2}{\theta})}). \tag{4}$$

Differentiation of (4) with respect to θ leads to

$$L_\theta = -\frac{n}{\theta} + \frac{\sum_{i=1}^n x_i^2}{\theta^2} + \frac{nT^2\exp{(-\frac{T^2}{\theta})}}{\theta^2[1-\exp{(-\frac{T^2}{\theta})}]}.$$

Set $L_\theta = 0$, which is equivalent to $m_2 + \frac{T^2}{\exp{(\frac{T^2}{\theta})}-1} = \theta$ with $\frac{\sum_{i=1}^n x_i^2}{n} = m_2$. The maximum likelihood estimator can be solved using uniroot function in software R:

$$f < -\text{function}(x)(m_2 + T^2/(\exp(T^2/x) - 1) - x),$$

$$\hat{\theta} < -\text{try}(\text{uniroot}(f, \text{c}(0.001, 10000)))\$\text{root}.$$

The Fisher information $I(\theta)$ is derived in the following

$$\begin{aligned} I(\theta) \quad &= \quad -E(\frac{\partial L_\theta}{\partial\theta}) \\ &= \quad -\frac{n}{\theta^2} + 2\frac{nE(X_1^2)}{\theta^3} \\ &- \quad \frac{nT^4\exp{(-\frac{T^2}{\theta})}[1-\exp{(-\frac{T^2}{\theta})}] - nT^2\exp{(-\frac{T^2}{\theta})}(2\theta[1-\exp{(-\frac{T^2}{\theta})}] - \exp{(-\frac{T^2}{\theta})}T^2)}{\theta^4[1-\exp{(-\frac{T^2}{\theta})}]^2} \\ &= \quad -\frac{n}{\theta^2} + 2\frac{nE(X_1^2)}{\theta^3} + \frac{2\theta nT^2 b - nT^4 b - 2\theta nT^2 b^2}{\theta^4(1-b)^2}. \end{aligned}$$

3 Left Truncated Rayleigh Distribution

Consider the Rayleigh density function (pdf) left truncated at D,

$$f(x) = \frac{\frac{2x}{\theta}\exp{(-\frac{x^2}{\theta})}}{\exp{(-\frac{D^2}{\theta})}} \tag{5}$$

for $0 < D < x < \infty$ and $\theta > 0$. We will consider the statistical inference of scale parameter θ when truncation point D is known. Let X_1, X_2,...,X_n be a random sample from LTRD specified in (5), we will study the maximum likelihood estimator for θ. For notation purpose, let $\exp(-\frac{D^2}{\theta}) = c$. Note

$$\begin{aligned}
E(X^2) &= \frac{1}{c}\int_D^\infty \frac{2x^3}{\theta}\exp(-\frac{x^2}{\theta})dx \\
&= \frac{\theta}{c}\int_{D^2/\theta}^\infty y\exp(-y)dy \\
&= \frac{\theta}{c}(\frac{D^2}{\theta}\exp(-\frac{D^2}{\theta}) + \exp(-\frac{D^2}{\theta})) \\
&= D^2 + \theta
\end{aligned}$$

$$\begin{aligned}
F(x) &= \frac{1}{c}\int_D^x \frac{2t}{\theta}\exp(-\frac{t^2}{\theta})dt \\
&= \frac{1}{c}(c - \exp(-\frac{x^2}{\theta})).
\end{aligned}$$

$$\begin{aligned}
u &= \frac{c - \exp(-\frac{x^2}{\theta})}{c} \\
cu &= c - \exp(-\frac{x^2}{\theta}) \\
c - cu &= \exp(-\frac{x^2}{\theta}) \\
x &= \sqrt{-\theta\ln(c - cu)} \qquad (6)
\end{aligned}$$

Expression (6) will be used to generate observations from LTRD. The log-likelihood function of a random sample X_1, X_2,...,X_n from LTRD specified in (5) is given by

$$L(\theta) = \sum_{i=1}^n \ln(2x_i) - n\ln(\theta) - \frac{\sum_{i=1}^n x_i^2}{\theta} + \frac{nD^2}{\theta}. \qquad (7)$$

Differentiation of (7) with respect to θ leads to

$$L_\theta = -\frac{n}{\theta} + \frac{\sum_{i=1}^n x_i^2}{\theta^2} - \frac{nD^2}{\theta^2}.$$

Set $L_\theta = 0$ and the maximum likelihood estimator is derived as

$$\hat{\theta} = \frac{\sum_{i=1}^n x_i^2 - nD^2}{n} = m_2 - D^2.$$

Therefore the Fisher information $I(\theta)$ is given by

$$\begin{aligned}
I(\theta) &= -E(\frac{\partial L_\theta}{\partial\theta}) \\
&= -\frac{n}{\theta^2} + 2\frac{nE(X_1^2)}{\theta^3} - \frac{2nD^2}{\theta^3}.
\end{aligned}$$

4 Doubly Truncated Rayleigh Distribution

Consider the Rayleigh density function (pdf) left truncated at D, and right truncated at T.

$$f(x) = \frac{\frac{2x}{\theta}\exp\left(-\frac{x^2}{\theta}\right)}{c-b} \tag{8}$$

for $0 < D < x < T$ and $\theta > 0$. We will consider the statistical inference of scale parameter θ when truncation point D and T are known. Let X_1, X_2,...,X_n be a random sample from DTRD specified in (8), we will study the maximum likelihood estimator for θ. Note

$$\begin{aligned}
E(X^2) &= \frac{1}{c-b}\int_D^T \frac{2x^3}{\theta}\exp\left(-\frac{x^2}{\theta}\right)dx \\
&= \frac{\theta}{c-b}\int_{D^2/\theta}^{T^2/\theta} y\exp(-y)dy \\
&= \frac{\theta}{c-b}\left(\frac{D^2}{\theta}\exp\left(-\frac{D^2}{\theta}\right) + \exp\left(-\frac{D^2}{\theta}\right) - \frac{T^2}{\theta}\exp\left(-\frac{T^2}{\theta}\right) - \exp\left(-\frac{T^2}{\theta}\right)\right) \\
&= \theta + \frac{D^2 c - T^2 b}{c-b}
\end{aligned}$$

$$\begin{aligned}
F(x) &= \frac{1}{c-b}\int_D^x \frac{2t}{\theta}\exp\left(-\frac{t^2}{\theta}\right)dt \\
&= \frac{1}{c-b}\left(c - \exp\left(-\frac{x^2}{\theta}\right)\right).
\end{aligned}$$

$$\begin{aligned}
u &= \frac{c-\exp\left(-\frac{x^2}{\theta}\right)}{c-b} \\
(c-b)u &= c - \exp\left(-\frac{x^2}{\theta}\right) \\
c-(c-b)u &= \exp\left(-\frac{x^2}{\theta}\right) \\
x &= \sqrt{-\theta\ln(c-(c-b)u)}
\end{aligned} \tag{9}$$

Expression (9) will be used to generate observations from DTRD. The log-likelihood function of a random sample X_1, X_2,...,X_n from DTRD specified in (8) is given by

$$L(\theta) = \sum_{i=1}^n \ln(2x_i) - n\ln(\theta) - \frac{\sum_{i=1}^n x_i^2}{\theta} - n\ln(c-b). \tag{10}$$

Differentiation of (10) with respect to θ leads to

$$L_\theta = -\frac{n}{\theta} + \frac{\sum_{i=1}^{n} x_i^2}{\theta^2} - \frac{n(cD^2 - bT^2)}{\theta^2(c-b)}.$$

Set $L_\theta = 0$ and the maximum likelihood estimator is derived using

$$f < -\text{function}(x)(m_2 - x - (D^2 \exp(-D^2/x) - T^2 \exp(-T^2/x))/(\exp(-D^2/x) - \exp(-T^2/x))),$$

$$\hat{\theta} < -\text{try}(\text{uniroot}(f, \text{c}(0.001, 10000)))\$\text{root}$$

in R.

Therefore the Fisher information $I(\theta)$ is given by

$$\begin{aligned} I(\theta) &= -E(\frac{\partial L_\theta}{\partial \theta}) \\ &= -\frac{n}{\theta^2} + 2\frac{nE(X_1^2)}{\theta^3} \\ &+ \frac{n((cD^4 - bT^4)(c-b) - (2\theta(c-b) + (cD^2 - bT^2))(cD^2 - bT^2))}{\theta^4(c-b)^2} \end{aligned} \quad (11)$$

5 Applications

In Siddiqui (1962) example 3, a systematic sample of 80 observations of received field intensity in (microvolts)^2 were investigated and the data were shown to be consistent with the hypothesis of exponential distribution. The observed values of received power in $(\mu v)^2$ are given below:

0.20, 0.71, 0.06, 0.05, 0.76, 0.32, 0.96, 0.63, 0.09,

0.18, 0.25, 0.45, 0.26, 0.10, 0.95, 0.01, 0.50, 1.26,

1.99, 0.32, 0.51, 0.01, 0.16, 0.56, 3.16, 1.27, 2.24,

1.00, 0.81, 1.29, 0.28, 0.21, 0.35, 0.20, 0.39, 0.89,

1.24, 0.08, 0.98, 1.01, 0.49, 0.90, 1.90, 1.42, 1.56,

1.32, 1.20, 1.59, 2.40, 2.24, 0.80, 0.56, 1.45, 0.18,

0.02, 0.28, 0.81, 0.18, 1.31, 0.64, 1.95, 0.48, 0.55,

0.44, 0.28, 0.07, 0.71, 0.48, 0.40, 0.06, 0.79, 1.01, 0.51, 0.70, 0.14, 0.16, 0.01, 0.06, 0.03, 0.01

Note that the square root of the observations will be consistent with Rayleigh distribution and we will examine our MLE derived in this paper on the square root of the observations from Siddiqui (1962). Impose right truncation at $T = \sqrt{2}$ and we derive $\hat{\theta} = 0.99$ and 95% confidence interval is given by (0.58,1.40). To see how well RTRD with $\hat{\theta} = 0.99$ fit the truncated data, we provide a plot of empirical distribution function paired with fitted RTRD. From the figure with heading "ECDF with RTRD Fit" we can see that the fit is quite reasonable.

References

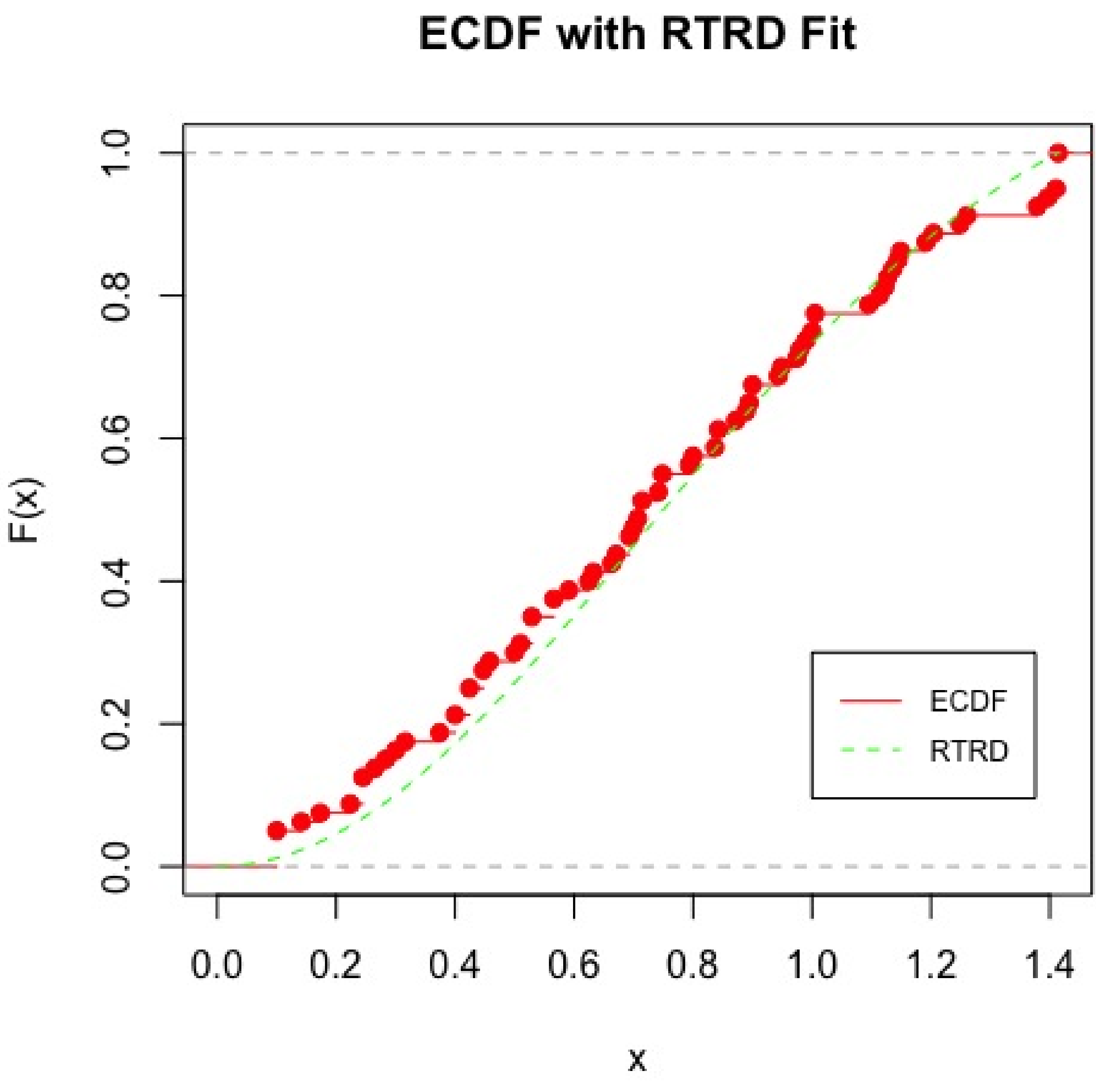
[1] H. M. R. Khan, S. B. Provost and A. Amparo (2009). Predictive densities from the Rayleigh life mdoel under type II censored samples. *J. Stat. Manag. Syst.* **12,** 305–317.

[2] S. Lalitha and A. Mishra (1996). Modified maximum likelihood estimation for Rayleigh distribution *Comm. Statist. Theory Methods* **25,** 389-401

[3] M. M. Siddiqui (1962). Some problems connected with Rayleigh distributions. *J. Res. Nat. Bur. Stand.* **66D,** 167-174.

[4] D. R. Wingo(1988). Methods for fitting the right-truncated Weibull distribution to life-test and survival data *Biometrical Journal* **30,** 5, 545-551.

[5] T. Zhang and M. Xie (2011). On the upper truncated Weibull distribution and its reliability implications *Reliability Engineering and Eystem Safety* , 96, 194-200.


ECDF with RTRD Fit
F(x)
x
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
ECDF
RTRD

```
# application original observation from exp
#JOURNAL OF RESEARCH of the National Bureau of Standards-D. Radio Propagation
#Vol. 66D,    No.2,    March-April 1962
#sqrt(original observation) from Rayleigh
#Make right truncation at T=sqrt(2)
#Some Problems Connected With Rayleigh Distributions
#Code for calculation of the MLE and 95% CI  for scale parameter in RTRD
original<-c(
0.20,  0.71,  0.06,  0.05,  0.76,  0.32,  0.96,  0.63,  0.09,
0.18,  0.25,  0.45,  0.26,  0.10,  0.95,  0.01,  0.50,  1.26,
1.99,  0.32,  0.51,  0.01,  0.16,  0.56,  3.16,  1.27,  2.24,
1.00,  0.81,  1.29,  0.28,  0.21,  0.35,  0.20,  0.39,  0.89,
1.24,  0.08,  0.98,  1.01,  0.49,  0.90,  1.90,  1.42,  1.56,
1.32,  1.20,  1.59,  2.40,  2.24,  0.80,  0.56,  1.45,  0.18,
0.02,  0.28,  0.81,  0.18,  1.31,  0.64,  1.95,  0.48,  0.55,
0.44,  0.28,  0.07,  0.71,  0.48,  0.40,  0.06,  0.79,  1.01,
0.51,  0.70,  0.14,  0.16,  0.01,  0.06,  0.03,  0.01)
obs_truncated<-c(
0.20,  0.71,  0.06,  0.05,  0.76,  0.32,  0.96,  0.63,  0.09,
0.18,  0.25,  0.45,  0.26,  0.10,  0.95,  0.01,  0.50,  1.26,
1.99,  0.32,  0.51,  0.01,  0.16,  0.56,  2.00,  1.27,  2.00,
1.00,  0.81,  1.29,  0.28,  0.21,  0.35,  0.20,  0.39,  0.89,
1.24,  0.08,  0.98,  1.01,  0.49,  0.90,  1.90,  1.42,  1.56,
1.32,  1.20,  1.59,  2.00,  2.00,  0.80,  0.56,  1.45,  0.18,
0.02,  0.28,  0.81,  0.18,  1.31,  0.64,  1.95,  0.48,  0.55,
0.44,  0.28,  0.07,  0.71,  0.48,  0.40,  0.06,  0.79,  1.01,
0.51,  0.70,  0.14,  0.16,  0.01,  0.06,  0.03,  0.01)
obs<-sqrt(obs_truncated)
T<-sqrt(2)
n<-length(obs)
M2<-mean(obs^2)
f<- function(x) (M2-x + T^2/(exp(T^2/x)-1))
Theta_MLE<-try(uniroot(f,    c(0.001,    10000)))$root
b<-exp(-T^2/Theta_MLE)
Third_num<- -n*T^4*b+2* Theta_MLE*n*T^2*b-2* Theta_MLE*n*b^2*T^2
Third_den<- Theta_MLE^4*(1-b)^2
I_theta<- -n/ Theta_MLE^2+2*n*M2/ Theta_MLE^3+Third_num/Third_den
Var_MLE<-1/I_theta
Lower<- Theta_MLE -1.96*sqrt(Var_MLE )
Upper<- Theta_MLE +1.96*sqrt(Var_MLE )
plot(ecdf(obs), xlim=c(0, T),    ylab="F(x)",
col="red", main="ECDF with RTRD Fit", lty=1)
curve((1-exp(-x^2/ Theta_MLE))/(1-exp(-T^2/ Theta_MLE)), add=TRUE, lty=2, col="green")
legend(x = 1.0, y=0.3,             # Position
      legend = c("ECDF",  "RTRD"),     # Legend texts
      lty = c(1, 2),              # Line types
      col = c("red", "green") ,
      cex=0.75 )
```

```
#Code for calculation of the MLE and 95% CI  for scale parameter in RTRD
T<-
obs<-c()
n<-length(obs)
M2<-mean(obs^2)
f1<- function(x) (M2-x + T^2/(exp(T^2/x)-1))
Theta_MLE<-try(uniroot(f1, c(0.001,10000)))$root
	b<-exp(-T^2/Theta_MLE)
Third_num<- -n*T^4*b+2* Theta_MLE*n*T^2*b-2* Theta_MLE*n*b^2*T^2
Third_den<- Theta_MLE^4*(1-b)^2
I_theta<- -n/ Theta_MLE^2+2*n*M2/ Theta_MLE^3+Third_num/Third_den
Var_MLE<-1/I_theta
Lower<- Theta_MLE -1.96*sqrt(Var_MLE )
Upper<- Theta_MLE +1.96*sqrt(Var_MLE )

#Code for calculation of the MLE and 95% CI  for scale parameter in LTRD
D<-
obs<-c()
n<-length(obs)
M2<-mean(obs^2)
Theta_MLE<- M2-D^2
I_theta<- -n/Theta_MLE^2+2* n*M2/Theta_MLE^3-2*n*D^2/Theta_MLE^3
Var_MLE<-1/ I_theta
Lower<- Theta_MLE-1.96*sqrt(Var_MLE)
Upper<- Theta_MLE+1.96*sqrt(Var_MLE)

#Code for calculation of the MLE and 95% CI  for scale parameter in DTRD
T<-
D<-
obs<-c()
n<-length(obs)
M2<-mean(obs^2)
f1<- function(x) (M2-x - (D^2*exp(-D^2/x)-T^2*exp(-T^2/x))/( exp(-D^2/x)- exp(-T^2/x)))
Theta_MLE <-try(uniroot(f1, c(0.1,10000)))$root
ce<-exp(-D^2/ Theta_MLE)
be<-exp(-T^2/ Theta_MLE)
	I_theta<- -n/ Theta_MLE^2+2* n*M2/ Theta_MLE^3+n*((ce*D^4-be*T^4)*(ce-be)-(2*
	Theta_MLE*(ce-be)+(D^2*ce-T^2*be))*(ce*D^2-be*T^2))/( Theta_MLE^4*(ce-be)^2)
	Var_MLE<-1/ I_theta
Lower<- Theta_MLE-1.96*sqrt(Var_MLE)
Upper<- Theta_MLE+1.96*sqrt(Var_MLE)
```