ISSN 0974-3235
ProbStat Forum, Volume 4, January 2011, Pages 01-11

ESTIMATION OF $P(X > Y)$ FOR THE DOUBLE LOMAX DISTRIBUTION

Bindu Punathumparambath
Department of Statistics
St.Thomas College, Pala
Kerala, India
e-mail: *ppbindukannan@gmail.com*

Paper received on 02 July 2010; revised, 15 August 2010; accepted, 17 August 2010.

Abstract. Currently there is lot of interest in the area of stress-strength models, especially in the estimation of $R = P(X > Y)$, when X and Y are independent random variables belonging to the same univariate family of distributions. $P(X > Y)$ is of greater significance than just in reliability since it provides a general measure of the difference between two populations and has applications in many areas. For instance, if Y is the response for a control group, and X refers to a treatment group, R is a measure of the effect of the treatment. In this paper, we introduce a new family of distribution referred to as the double Lomax distribution, which is the ratio of two independent and identically distributed classical Laplace distributions and estimate $P(X > Y)$. Also we derive the pdf and the expression for the reliability R for the double Lomax distribution truncated below zero. The maximum likelihood estimate of the reliability parameter is calculated using an algorithm in R package. Finally, Robert's data dealing with Otis IQ scores is analyzed to illustrate the procedure.

Mathematics Subject Classification. 62N01, 62N02, 62N05, 62P15.

Keywords. Double Lomax distribution, Laplace distribution, Reliability, stress-strength.

1 Introduction

In the context of reliability, the stress-strength model describes the life of a component which has a random strength X and is subjected to a random stress Y. The component fails at the instant that the stress applied to it exceeds the strength, and the component will function satisfactorily whenever $X > Y$. Thus, $R = P(X > Y)$ is a measure of component reliability. The parameter R is referred to as the reliability parameter. This type of functional can be of practical importance in many applications. For instance, if X is the response for a control group, and Y refers to a treatment group, $P(X < Y)$

is a measure of the effect of the treatment. $R = P(X > Y)$ can also be useful when estimating heritability of a genetic trait. For more applications of R, see Halperin et al. (1987), Simonoff et al. (1986), Reiser and Farragi (1994) and Bamber (1975). In fact, Bamber gives a geometrical interpretation of $A(X, Y) = P(X < Y) + \frac{1}{2}P(X = Y)$ and demonstrates that $A(X, Y)$ is a useful measure of the size of the difference between two populations.

Weerahandi and Johnson (1992) proposed inferential procedures for $P(X > Y)$ assuming that X and Y are independent normal random variables. Gupta and Brown (2001) illustrated the application of skew normal distribution to stress-strength model. Nadarajah (2004) studied the reliability for Laplace distribution and its generalizations. The present work introduce the double Lomax distribution and present its application to the stress-strength model. The double Lomax distribution is the ratio of independent and identically classical Laplace distributions. For given random variables X and Y, the distribution of the ratio X/Y is of interest in biological and physical sciences, econometrics, and ranking and selection. Examples include Mendelian inheritance ratios in genetics, mass to energy ratios in nuclear physics, target to control precipitation in meteorology, inventory ratios in economics, and the stress-strength model in the context of reliability.

The functional $R = P(X > Y)$ or $\lambda = P(X > Y) - P(X < Y)$ is of practical importance in many situations, including clinical trials, genetics, and reliability. In this paper we are interested in applying $R = P(X > Y)$ as a measure of the difference between two populations, in particular where X and Y refer to Otis IQ scores of 87 white males and 52 non-white males hired by a large insurance company in 1971(Roberts (1988)). This article is organized as follows. In section 2, the univariate double Lomax distribution and its density truncated below zero are derived. In section 3 we derived the expression for the $P(X > Y)$ for double Lomax and double Lomax distribution truncated below at zero. The estimation of the $P(X > Y))$ using the Maximum Likelihood Estimators of the parameters obtained from the simulated data is presented in section 4. In section 5, Robert's data dealing with Otis IQ scores are analyzed to illustrate the procedure. Finally, we summarize in section 6.

2 Double Lomax distribution

The Laplace model is an alternative to the normal model in situations where the normality assumption do not hold. In a similar way the the ratio of independent and identically distributed Laplace distribution is an alternative to the Cauchy distribution. The double Lomax distribution studied by Bindu et al. (2010), is shown to be a suitable model for a microarray gene expression data.

Also the double Lomax distribution can be obtained by compounding classical Laplace distribution with exponential density. The compound distribution is of interest for the study of production/inventory problems, since it provides a flexible description of the stochastic properties of the system. These distributions play a central role in insurance and other areas of applied probability modelling such as queuing theory, reliability etc.

Definition 2.1 *Let X_1 and X_2 be two independent and identically distributed (i.i.d.) standard classical Laplace random variables. Then the corresponding probability distribution of $Y = X_1/X_2$ is given by*

$$f(y) = \frac{1}{2(1+|y|)^2}, \quad -\infty < y < \infty. \tag{2.1}$$

The Laplace distribution can also be expressed as the difference of two i.i.d exponentials and hence,
$X_i \overset{d}{=} I_i W_i$, for $i = 1, 2$ where, $W_i \sim Exp(1)$, for $i = 1,\ 2$ and $I_i's$ are independent of W_i and takes values ± 1 with equal probabilities.
Then the cdf can be given by

$$F(y) = \begin{cases} \frac{1}{2(1-y)}, & \text{for } y \le 0 \\ \\ (1 - \frac{1}{2(1+y)}), & \text{for } y > 0 \end{cases} \tag{2.2}$$

Remark 2.1 *Let X_1 and X_2 be two independent and identically distributed (i.i.d.) standard classical Laplace random variables. Then the corresponding probability distribution of $Z = |X_1/X_2|$ is given by*

$$f(z) = \frac{1}{(1+z)^2}, \quad 0 < z < \infty. \tag{2.3}$$

which is the pdf of the Lomax distribution. Hence the random variable Y having pdf given by (2.1) may be referred to as the double Lomax distribution (DLD).

Then the double Lomax distribution, which is the extension of the Lomax distribution over the real line is defined as follows.

Definition 2.2 *A random variable Y is said to have a double Lomax distribution with parameters μ and σ, $DLD(\mu, \sigma)$ if its probability distribution function is*

$$F_Y(y) = \begin{cases} \frac{1}{2[1+\frac{(\mu-y)}{\sigma}]}, & y \le \mu \\ \\ 1 - \frac{1}{2[1+\frac{(y-\mu)}{\sigma}]}, & y > \mu. \end{cases} \tag{2.4}$$

and the probability density function is

$$f_Y(y) = \frac{1}{2\,\sigma[1 + |\frac{(y-\mu)}{\sigma}|]^2}, -\infty < y < \infty \tag{2.5}$$

If $Y \sim DLD(0,1)$ then it is referred as the *standard double Lomax distribution*. The corresponding probability distribution and density functions are given by (2.1) and (2.2), respectively.

If Y has the standard double Lomax distribution and $q = F(y)$, then $y = F^{-1}(q)$ and quantiles ξ_q can be written explicitly as follow

$$\xi_q = \begin{cases} (1 - \frac{1}{2q}), & \text{for } 0 < q \leq \frac{1}{2} \\ \\ (\frac{1}{2(1-q)} - 1), & \text{for } \frac{1}{2} \leq q < 1 \end{cases} \tag{2.6}$$

Thus if Y has the double Lomax distribution with cdf given by (2.4) then quantiles $\acute{\xi_q}$ is obtained as $\acute{\xi_q} = \mu + \sigma\xi_q$. In particular, the first and the third quartiles are given by

$$Q_1 = \xi_{1/4} = \mu - \sigma, \qquad Q_3 = \xi_{3/4} = \mu + \sigma$$

Evidently, the second quartile Q_2-median is μ. The shape of the density (2.5) is given in Figure 1 along with the normal and Laplace density functions.

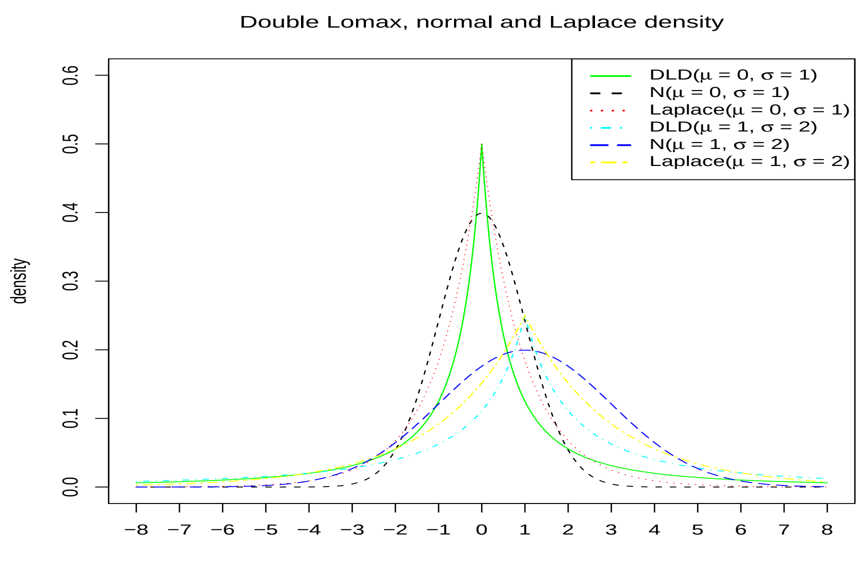

Figure 1: Double Lomax density functions
for various values of parameters

From figure 1, we can see that for double Lomax density (DLD) more area is concentrated towards the center and has heavier tails than normal and Laplace distribution.

2.1 Double Lomax distribution truncated below at zero

If a random variable Y has a standard double Lomax distribution , $DLD(0, 1)$ then the double Lomax density truncated below zero is given by

$$f(y) = K\frac{1}{(1+y)^2}I_{[0,\infty)}(y), \quad y > 0. \tag{2.7}$$

where K is the normalizing constant. Then the double Lomax distribution, $DLD(0, \sigma)$ truncated below at zero is given by

$$f(y) = \frac{1}{\sigma\left[1+\left(\frac{y}{\sigma}\right)\right]^2}, \quad y > 0,\ \sigma > 0. \tag{2.8}$$

The cdf is given by

$$F(y) = 1 - \frac{1}{\left[1+\left(\frac{y}{\sigma}\right)\right]}, \quad y > 0,\ \sigma > 0. \tag{2.9}$$

3 $P(X > Y)$ for the double Lomax distribution

Let X and Y are two continuous and independent random variables. Let f_2 denote the probability density function (pdf) of Y and F_1 denote the cumulative distribution function (cdf) X. Then $P(X > Y)$ can be given as,

$$P(X > Y) = \int_{-\infty}^{\infty} F_2(z)f_1(z)dz. \tag{3.1}$$

To evaluate $P(X > Y)$ we can use the generalized hypergeometric functions. The generalized hypergeometric function is given by a hypergeometric series, in which the ratio of successive terms can be written

$$\frac{c_{k+1}}{c_k} = \frac{(k+a_1)(k+a_2)\cdots(k+a_p)}{(k+b_1)(k+b_2)\cdots(k+b_q)(k+1)}$$

Hence the generalized hypergeometric function is written as,

$$_pF_q(a_1, a_2, \cdots, a_p; b_1, b_2, \cdots, b_q; x) = \sum_{k=0}^{\infty}\frac{(a_1)_k(a_2)_k\cdots(a_n)_k}{(b_1)_k(b_2)_k\cdots(b_n)_k}\frac{x^k}{k!} \tag{3.2}$$

where $(a_j)_k$ and $(b_j)_k$ are the Pochammer symbols given by

$$(a)_k = a(a+1)\cdots(a+k-1),\ (a)_0 = 1,\ a \neq 0$$

also $(a)_k = \frac{\Gamma(a+k)}{\Gamma(a)}$, whenever the gamma function exists, that is $\mathbb{R}(a) > 0$. For $p = 2$ and $d = 1$ we get the Gauss hypergeometric function.

$$ {}_2F_1(a_1, a_2; b; x) = \sum_{k=0}^{\infty} \frac{(a_1)_k (a_2)_k}{(b)_k} \frac{x^k}{k!} \tag{3.3} $$

Now we evaluate the $P(X > Y)$ for two independent double Lomax distributions. Let X and Y are continuous and independent variables having double Lomax distribution with parameters θ_i *and* σ_i, $1 = 1, 2$ respectively.
From equation (3.1) we get the reliability R for the type II compound Laplace distribution as follows

$$ R = 1 - F_Y(\theta_2) + \frac{1}{4\sigma_1} \left(I_1 + I_2 - I_3 \right). \tag{3.4} $$

where the integrals I_1, I_2 *and* I_3 given below and F_Y is obtained by substituting $\sigma = \sigma_2$, $\theta = \theta_2$ in (2.4)

$$ I_1 = \int_{-\infty}^{min(\theta_1,\theta_2)} \left[1 + (\theta_1 - z)/\sigma_1\right]^{-2} \left[1 + (\theta_2 - z)/\sigma_2\right]^{-1} dz $$

$$ I_2 = \int_{min(\theta_1,\theta_2)}^{max(\theta_1,\theta_2)} \left[1 + |z - \theta_1|/\sigma_1\right]^{-2} \left[1 + |z - \theta_2|/\sigma_2\right]^{-1} dz $$

$$ I_3 = \int_{max(\theta_1,\theta_2)}^{\infty} \left[1 + (z - \theta_1)/\sigma_1\right]^{-2)} \left[1 + (z - \theta_2)/\sigma_2\right]^{-1} dz $$

The integrals I_1 and I_3 can be expressed in terms of the Gauss hyper geometric function. The integral I_2 cannot be simplified further unless $\theta_1 = \theta_2$, in this case we get $R = 1/2$. For instance, if $\theta_1 < \theta_2$,

$$ I_1 = \frac{1}{8} \frac{\sigma_2}{\sigma_1} \, {}_2F_1 \left(2, 1; 3; 1 - \frac{\sigma_2}{\sigma_1} - \frac{(\theta_2 - \theta_1)}{\sigma_1} \right) \tag{3.5} $$

$$ I_3 = \frac{1}{8} \frac{1}{\sigma_1 \left(1 + \frac{\theta_2 - \theta_1}{\sigma_1}\right)} \, {}_2F_1 \left(1, 1; 3; 1 - \frac{\sigma_1}{\sigma_2} - \frac{(\theta_2 - \theta_1)}{\sigma_2} \right) \tag{3.6} $$

3.1 $P(X > Y)$ for the double Lomax distribution truncated below at zero

Let X and Y are two independent double Lomax random variables truncated below zero. Let f_2 denote the probability density function (pdf) of Y and F_1 denote the cumulative distribution function (cdf) X. Then $P(X > Y)$ can be given as,

$$ P(X > Y) = 1 - \frac{1}{\sigma_1} \int_0^{\infty} \left[1 + z/\sigma_1\right]^{-2} \left[1 + z/\sigma_2\right]^{-1} dz. \tag{3.7} $$

4 Maximum Likelihood Estimate of $P(X > Y)$

The MLE of the $P(X > Y)$ can be obtained by replacing the parameters θ_1, θ_2, σ_1 and σ_2 in (3.4) by their MLE's. Then the MLE of R is given by

$$\widehat{R} = 1 - F_Y(\widehat{\theta_2}) + \frac{1}{4\widehat{\sigma_1}}\left(\widehat{I_1} + \widehat{I_2} - \widehat{I_3}\right). \tag{4.1}$$

where

$$\widehat{I_1} = \int_{-\infty}^{min(\widehat{\theta_1},\widehat{\theta_2})} \left[1 + (\widehat{\theta_1} - z)/\widehat{\sigma_1}\right]^{-2} \left[1 + (\widehat{\theta_2} - z)/\widehat{\sigma_2}\right]^{-1} dz$$

$$\widehat{I_2} = \int_{min(\widehat{\theta_1},\widehat{\theta_2})}^{max(\widehat{\theta_1},\widehat{\theta_2})} \left[1 + |z - \widehat{\theta_1}|/\widehat{\sigma_1}\right]^{-2} \left[1 + |z - \widehat{\theta_2}|/\widehat{\sigma_2}\right]^{-1} dz$$

$$\widehat{I_3} = \int_{max(\widehat{\theta_1},\widehat{\theta_2})}^{\infty} \left[1 + (z - \widehat{\theta_1})/\widehat{\sigma_1}\right]^{-2)} \left[1 + (z - \widehat{\theta_2})/\widehat{\sigma_2}\right]^{-1} dz$$

Using Maple program we can evaluate the integrals for the Maximum Likelihood Estimates of the parameters.

Now we can find the Maximum Likelihood estimates of the parameters from the simulated sample of 100 observations form the double Lomax distribution. Let $X = (X_1, \cdots, X_n)$ be independent and identically distributed samples from a double Lomax distribution. The log-likelihood function takes the form

$$logL(\theta_1, \sigma_1; X) = -nlog2 - nlog\sigma_1 - 2\sum_{i=1}^{n} log\left[1 + |(x_i - \theta_1)/\sigma_1|\right]$$

Let $Y = (Y_1, \cdots, Y_n)$ be independent and identically distributed samples from a double Lomax distribution. The log-likelihood function takes the form

$$logL(\theta_2, \sigma_2; Y) = -nlog2 - nlog\sigma_2 - 2\sum_{i=1}^{n} log\left[1 + |(y_i - \theta_2)/\sigma_2|\right]$$

The maximum likelihood estimates of the parameters can be obtained by using the approach of the maximum likelihood estimation of Laplace distribution (see, Kotz et al. (2001)). Maximum likelihood estimators (MLE) of the parameters (θ_1, σ_1) and (θ_2, σ_2) are obtained by solving two sets of score equations. Numerical methods are needed to solve these score equations. In our illustration, the maximisation of the likelihood is implemented using the `optim` function of the R statistical software, applying the BFGS algorithm (See R Development Core Team, 2006). Estimates of the standard errors were obtained by inverting the numerically differentiated information matrix at the maximum likelihood point.

4.1 Simulation

In this section we use the simulation study of the double Lomax distribution to validate the estimation algorithm developed in R. Since we can express the distribution function of the double Lomax distribution as well as its inverse in closed form, the inversion method of simulation is straightforward to implement. We simulated a data set of size 100 from the double Lomax distribution with parameters $(\theta_1, \sigma_1) = (0.5, 0.1)$ and $(\theta_2, \sigma_2) = (1, 0.5)$ by inverting the distribution function (2.4) in R and then applied the algorithm to obtain the MLEs of the parameters. Table 1 gives the estimated value of the parameters using MLE, standard error (SE) and 90% confidence limits (LCL & UCL) for the parameters.

Table 1: Simulation study - parameter values used for simulation (TRUE), MLEs, standard errors (SE) and 90% confidence limits (LCL & UCL) for the parameters

	TRUE	MLE	SE	LCL	UCL
$\widehat{\theta_1}$	0.5	0.487	0.002	0.48372	0.49028
$\widehat{\sigma_1}$	0.1	0.026	0.001	0.02436	0.02764
$\widehat{\theta_2}$	1.0	0.950	0.004	0.94344	0.95656
$\widehat{\sigma_2}$	0.5	0.128	0.006	0.11816	0.13784

Replacing the parameters by the estimates we get the MLE of the reliability parameter R as 0.6035. Then the MLE's are $\widehat{\theta_1} = 0.487$, $\widehat{\theta_2} = 0.950$, $\widehat{\sigma_1} = 0.026$, $\widehat{\sigma_2} = 0.128$ and $\widehat{R} = 0.6035$.

Now we estimate R for simulated samples of size n=50, 25 and 15. The maximum likelihood estimators for the simulated data from double Lomax distributions are reported in Table 2.

Table 2: MLE of $P(X > Y)$

n	$\widehat{\theta_1}$	$\widehat{\sigma_1}$	$\widehat{\theta_2}$	$\widehat{\sigma_2}$	$\widehat{R}$
100	0.487	0.026	0.950	0.128	0.604
50	0.468	0.020	0.873	0.110	0.6027
25	0.523	0.022	0.804	0.094	0.617
15	0.484	0.022	1.163	0.087	0.556

5 Analysis of the Roberts data

In this section, we apply the double Lomax distribution to IQ score data set from Roberts (1988). The Roberts IQ data gives the Otis IQ scores for 87 white males and 52 non-white males hired by a large insurance company in 1971. We now let X represent the score for whites and let Y represent the scores for non-whites. We now assume that X and Y have independent double Lomax distribution to estimate the probability that the IQ score for a white employee is greater than the IQ score for a non-white employee. The two data sets are analyzed using the double Lomax distribution and the estimates are given Table 3 and Table 4. Then we calculate the maximum likelihood estimate for $P(X > Y)$.

Table 3: MLEs, standard errors (SE) and 90% confidence limits (LCL & UCL) for the parameters

	MLE	SE	LCL	UCL
$\widehat{\theta_1}$	103.000	0.021	102.966	103.034
$\widehat{\sigma_1}$	3.988	0.915	2.487	5.489

Table 4: MLEs, standard errors (SE) and 90% confidence limits (LCL & UCL) for the parameters

	MLE	SE	LCL	UCL
$\widehat{\theta_2}$	112.001	0.017	111.973	112.029
$\widehat{\sigma_2}$	4.447	0.784	3.161	5.733

Replacing the parameters by the estimates we get the MLE of the reliability parameter R as 0.6097. Then the MLE's are $\widehat{\theta_1} = 103$, $\widehat{\theta_2} = 112.001$, $\widehat{\sigma_1} = 3.988$, $\widehat{\sigma_2} = 4.447$ and $\widehat{R} = 0.6097$.

6 Summary

$P(X > Y)$ is of greater interest than just in reliability since it provides a general measure of the difference between two populations and has applications in many areas. One of the interesting application is the relationship between the stress-strength models and the quality control concept, such as process capability indices. We are planning to consider the problem of hypothesis testing and interval estimation of the reliability parameter in a stress-strength model involving two-parameter double Lomax distributions based on the novel concept of generalized p-value and generalized confidence limits. The practical applications of the stress-strength reliability no means confined to engineering or to military problems but also to medical statistics and other areas.

Acknowledgement

The author is grateful to the Department of Science & Technology, Government of India, New Delhi, for financial support under the Women Scientist Scheme (WOS-A (2008)), Project No: SR/WOS-A/MS-09/2008 . Also the author is grateful to Dr. Sebastian George (Department of Statistics, St. Thomas college, Pala, Kerala, India) and to Dr.Sangita Kulathinal (The Indic Society for Education and Development, Nashik, India) for valuable comments. I am also very thankful to the referee for the valuable suggestions and comments.

References

[1] Bamber, D. (1975). The area above the ordinal dominance graph and the area below the receiver operating characteristic graphs. *Journal of Mathematical Psychology*, **12**, 387–415.

[2] Bindu, P. P., Sebastian, G. and Sangita, K. (2010). On the ratio of two independent and identically distributed Classical Laplace random variables and applications, (Unpublished).

[3] Enis, P. and Geisser, S. (1971). Estimation of probability that $Y > X$, *Jour. Amer. Statist. Assoc.*, **66**, 162–68.

[4] Halperin, M., Gilbert, P. R. and Lachin, J. M. (1987). Distribution free confidence intervals for $P(X1 < X2)$. *Biomterics*, **43**, 71–80.

[5] Kotz, S., Kozubowski, T.J. and Podgorski, K. (2001). The Laplace Distribution and Generalizations: A Revisit with Applications to Communications, Economics, Engineering and Finance. Birkhäuser, Boston.

[6] Nadarajah, S. (2004). Reliability of Laplace distributions, *Math. Problems in Eng.*, **2**, 169–183.

[7] R Development Core Team. (2006). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing: Vienna, Austria, `http://www.R-project.org/`.

[8] Ramesh C. Gupta and Nicole Brown (2001). Reliability studies of the skew-normal distribution and its application to a strength-stress model, *Commun. statist.-theory meth.*, **30**(11), 2427–2445.

[9] Reiser, B. and Farragi, D. (1994). Confidence bounds for $P(\acute{a}x > \acute{b}y)$, *Statistics*, **25**, 107–111.

[10] Roberts, H. V. (1988). *Data Analysis for Managers with Minitab.*, Scientific Press: Redwood City, CA.

[11] Simonoff, J. S., Hochberg, Y. and Reiser, B. (1986). Alternative estimation procedures for $Pr : (X < Y)$ in categorized data. *Biometrics*, **42**, 895–907.

[12] Weerahandi, S. and Johnson, R. A. (1992). Testing reliability in a stress-strength model when X and Y are normally distributed, *Technometrics*, **34**, 83-91.