Applied Mathematical Sciences, Vol. 6, 2012, no. 128, 6355 - 6363

Goodness-of-Fit for the Topp-Leone Distribution with Unknown Parameters

Bander Al-Zahrani

Statistics Department, Faculty of Science, King Abdulaziz University
P.O. Box 80203, Jeddah 21589, Saudi Arabia

Abstract

A class of goodness-of-fit tests for the Topp-Leone distribution with estimated parameters is proposed. The tests are based on the empirical distribution function. Kolomogorov-Sminrov, Cramer-von-Mises, Anderson-Darling, Watson, and Liao-Shimokawa are proposed. Numerical simulations are performed to calculate the critical values for the proposed tests. Finally, the power estimate is performed with the hypothesized Topp-Leone distribution versus widely used alternates.

Keywords: Topp-Leone distribution, Maximum likelihood estimation, Anderson-Darling, Cramer-von Mises, Kolomogorov-Sminrov, Watson test statistics

1 Introduction

Let X be a random variable with distribution function $F(x) = P(X \leq x)$, and consider a parametric class $F_\Theta = \{F_\vartheta, \vartheta \in \Theta\}$. The main problem is that of testing hypotheses about F of the form:

$$H_0: \; F = F_\vartheta, \quad \text{for some} \;\; \vartheta \in \Theta. \tag{1}$$

Here we consider the classical goodness of fit tests which are based on the empirical distribution function (EDF). The EDF is a step function with jumps at the order statistics $(X_1, X_2, \ldots, X_n)$ and defined as

$$F_n(x) = \frac{\text{number of observations} \leq x}{n} = \frac{1}{n}\sum_{i=1}^{n} I(X_i \leq x),$$

where I is an indicator function. By Glivenko-Cantelli theorem the EDF, $F_n(x)$ converges to $F(x)$ as $t \to \infty$ almost surely, $\sup_x |F_n(x) - F(t)| \to$

0. Many authors have addressed the problem of testing the null hypothesis in (1) when X follows a specified parametric model, for details and examples see Abd-Elfattah and Omima (2009) and Shawky and Bakoban (2009) and references therein. Kolmogorov-Smirnov (KS), Cramer-von-Mises (CvM), Anderson-Darling (AD), Watson (W), and Liao-Shimokawa (LS) tests are among the available goodness-of-fit (GOF) tests in the literature. Hassan (2005) obtained critical values for these tests based on a random sample from generalized exponential distribution. In this paper, we study the GOF for the Topp-Leone distribution with unknown parameters.

The rest of the paper is organized as follows. In section 2, we provide an overview of the Topp-Leone distribution and derive the maximum likelihood estimator of the involved parameters. The most popular nonparametric goodness-of-fit tests are reviewed in section 3. Simulation results on the GOF tests in section 3 are presented in section 4.

2 Model description

Topp-Leone distribution is a continuous unimodal distribution with bounded support. It is a two-parametric family continuous distribution proposed by Topp and Leone (1955). Such a distribution is useful for modelling lifetime phenomena, different aspect of this class of distributions have been studied e.g. by Nadarajah and Kotz (2003).

We say that a random variable X with range of values $(0, \lambda)$ has a Topp-Leone distribution, and write $X \sim \mathrm{TL}(\alpha, \lambda)$, if the cumulative distribution function (cdf) is

$$F(x) = \left(\frac{x}{\lambda}\right)^{\alpha}\left(2 - \frac{x}{\lambda}\right)^{\alpha}, \quad 0 < x < \lambda, \quad \alpha, \lambda > 0, \tag{2}$$

and the density function (pdf) is

$$f(x) = \frac{2\alpha}{\lambda}\left(\frac{x}{\lambda}\right)^{\alpha-1}\left(1 - \frac{x}{\lambda}\right)\left(2 - \frac{x}{\lambda}\right)^{\alpha-1} \quad 0 < x < \lambda, \quad \alpha, \lambda > 0, \tag{3}$$

where α and λ are the scale and shape parameters respectively.

The Topp-Leone distribution is known as the J-shaped distribution. This is due to the fact that $f(x) > 0$, $f'(x) < 0$, and $f''(x) > 0$ for all $0 < x < \lambda$, where f' and f'' are the first and second derivatives of f respectively. Nadarajah and Kotz (2003) and Ghitany et al. (2005) derived and studied the properties of the failure rate, mean residual lifetime, reversed failure rate, and mean inactivity time for a random variable X, where $X \sim \mathrm{TL}(\alpha, \lambda)$. Kotz and Seier (2007) explore the kurtosis of the Topp-Leone distributions. Genc (2010) derive explicit algebraic expressions for both of the single and product moments of order statistics from Topp-Leone distribution.

2.1 Maximum likelihood estimates

Suppose that $X_1, X_2, \cdots, X_n$ is a random sample from TL(α, λ). Then the log-likelihood function of the observed sample is

$$L(x, \alpha, \lambda) = n \ln \frac{2}{\lambda} + n \ln \alpha + \sum_i^n \ln \left(1 - \frac{x_i}{\lambda}\right) + (\alpha - 1) \left\{ \sum_i^n \ln \left(\frac{x_i}{\lambda}\right) + \sum_i^n \ln \left(2 - \frac{x_i}{\lambda}\right) \right\}.$$

The MLEs of α and λ say $\hat{\alpha}$ and $\hat{\lambda}$, respectively, can be obtained as the solutions of the following equations

$$\begin{aligned} \frac{\partial L}{\partial \alpha} &= \frac{n}{\alpha} + \sum_i^n \ln \left(\frac{x_i}{\lambda}\right) + \sum_i^n \ln \left(2 - \frac{x_i}{\lambda}\right) = 0. \\ \frac{\partial L}{\partial \lambda} &= \frac{-2n\alpha}{\lambda} + \sum_i^n \frac{1}{\lambda - x_i} + (\alpha - 1) \sum_i^n \frac{2}{2\lambda - x_i} = 0. \end{aligned}$$

We obtain

$$\hat{\alpha} = \frac{-n}{\sum_i^n \ln \left(\frac{x_i}{\lambda}\right) + \sum_i^n \ln \left(2 - \frac{x_i}{\lambda}\right)}. \tag{4}$$

and $\hat{\lambda}$ can be obtained as the solution of the nonlinear equation $f(\lambda) = 0$, where

$$f(\lambda) = \frac{-2n\alpha}{\lambda} + \sum_i^n \frac{1}{\lambda - x_i} + \left\{ \frac{-n}{\sum_i^n \ln \left(\frac{x_i}{\lambda}\right) + \sum_i^n \ln \left(2 - \frac{x_i}{\lambda}\right)} - 1 \right\} \sum_i^n \frac{2}{2\lambda - x_i}.$$

Therefore, $\hat{\lambda}$ can be obtained as solution of the nonlinear equation of the form $g(\lambda) = \lambda$, where

$$g(\lambda) = 2n\alpha \left[\sum_i^n \frac{1}{\lambda - x_i} + \left\{ \frac{-n}{\sum_i^n \ln \left(\frac{x_i}{\lambda}\right) + \sum_i^n \ln \left(2 - \frac{x_i}{\lambda}\right)} - 1 \right\} \sum_i^n \frac{2}{2\lambda - x_i} \right]^{-1}. \tag{5}$$

Since, $\hat{\lambda}$ is a fixed point solution of the non-linear equation (5), therefore, it can be obtained using an iterative scheme as $g(\lambda_j) = \lambda_{j+1}$, where λ_j is the jth iterate of $\hat{\lambda}$. The iteration procedure should be stopped when $|\lambda_j - \lambda_{j+1}|$ is sufficiently small. Once we obtain $\hat{\lambda}$, then $\hat{\alpha}$ can be obtained from (4).

3 Goodness-of-Fit Tests

The goodness of fit tests are used to measure how compatible a random sample with a theoretical probability distribution function. We offer a revision of the most popular nonparametric goodness-of-fit tests, namely; the Kolmogorov-Smirnov, Cramer-von-Mises, Anderson-Darling, Watson, and Liao-Shimokawa test statistics. A goodness of fit test based on the empirical distribution function (EDF), when the parameters are estimated, is called a modified goodness of fit test.

Suppose that $X_1, X_2, \ldots, X_n$ are independent observations from the random variable X, on the basis of which we wish to test the null hypothesis in (1). Suppose further $X \sim TL(\alpha, \lambda)$, and the parameters α and λ are unknown to us. In what follows we overview some issues associated with the implementation of goodness-of-fit tests when fitting the Topp-Leone distribution.

The most popular GOF test is the Kolmogorov-Smirnov test. This test compares the EDF based on a sample of size n, with a theoretical cumulative distribution function under the null hypothesis H_0. The statistic of the KS test is denoted by D_n. The test statistic is defined as

$$D = \sup_x |F(x) - F_n(x)|$$

For an ordered random sample, $X_1, X_2, \ldots, X_n$, from $TL(\alpha, \lambda)$ with distribution function given in (2), The KS test statistic is given by

$$D_n = \max_i \left[\frac{i}{n} - F_{TL}(x_i, \hat{\alpha}, \hat{\lambda}), F_{TL}(x_i, \hat{\alpha}, \hat{\lambda}) - \frac{i-1}{n}\right]. \quad (6)$$

The Cramér-von Mises tests have been shown to be more powerful than KS test against a large class of alternatives hypotheses. It is denoted by W^2 and is defined as

$$W^2 = n \int_0^\infty [F_n(x) - F(x)]^2 \, w(x) dF(x), \quad (7)$$

where $w(x)$ is a weight function such that $w(x) = 1$. For given ordered sample, $X_1, X_2, \ldots, X_n$, from $TL(\alpha, \lambda)$, the relation (7) can be written as follows:

$$W_n^2 = \frac{1}{12n} + \sum_{i=1}^n \left[F_{TL}(x_i, \hat{\alpha}, \hat{\lambda}) - \frac{2i-1}{2n}\right]^2. \quad (8)$$

The AD test statistic was developed by Anderson and Darling (1954) as a limiting distribution of the Cramér-von Mises test as $n \to \infty$. Here the weight function $w(x)$ in the relation (7) is defined to be $w(x) = [F(x)(1 - F(x))]^{-1}$, therefore, the test statistics AD is given by

$$A^2 = n \int_0^\infty \frac{[F_n(x) - F(x)]^2}{F(x)[1 - F(x)]} \, dF(x). \quad (9)$$

The formula for the test statistic A^2 to assess whether the ordered sample, $X_1, X_2, \ldots, X_n$, comes from Topp-Leone distribution F_{TL} is:

$$A_n^2 = -n - \frac{1}{n} \sum_{i=1}^{n} (2i - 1) \left[\log(F_{TL}(x_i, \hat{\alpha}, \hat{\lambda})) + \log(1 - F_{TL}(x_i, \hat{\alpha}, \hat{\lambda})) \right]^2. \quad (10)$$

Watson test statistic was developed for distributions which are cyclic. It is a generalization of the CvM test statistic. It is defined by

$$U^2 = n \int_0^\infty \left[F_n(x) - F(x) - W^2 \right]^2 dF(x). \quad (11)$$

For given ordered sample from $TL(\alpha, \lambda)$ and using (11), we can write the test as follows:

$$U_n^2 = W_n^2 + \sum_{i=1}^{n} \left[\frac{F_{TL}(x_i, \hat{\alpha}, \hat{\lambda})}{n} - \frac{1}{2} \right]^2. \quad (12)$$

The Liao-Shimokawa statistic measures the average of all weighted distances over the entire range of the data. For more details we refer to Liao and Shimokawa (1999). The test statistic is given by

$$L_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\max_i \left[\frac{i}{n} - F_{TL}(x_i, \hat{\alpha}, \hat{\lambda}), F_{TL}(x_i, \hat{\alpha}, \hat{\lambda}) - \frac{i-1}{n} \right]}{\sqrt{F_{TL}(x_i, \hat{\alpha}, \hat{\lambda})[1 - F_{TL}(x_i, \hat{\alpha}, \hat{\lambda})]}}. \quad (13)$$

4 Implementation

4.1 Critical values for the test statistics

To numerically illustrate the test, we have simulated the lower percentile points for the significance level $\alpha = 0.01, , 0.02, 0.05, 0.10, 0.20$ and 0.25. The calculation of the tests is based on $10,000$ simulated samples from the standard Topp-Leone distribution. Table 1 shows the critical values for the tests statistic.

The following steps are used in calculating critical values for the proposed tests at different significance levels and sample sizes:

Step 1 Generate n independent $U(0,1)$ random variables $U_1, U_2, \cdots, U_n$.

Step 2 For given values of parameters α and λ, we set $x_i = F^{-1}(U_i)$. Then, $(x_1, x_2, \cdots, x_n)$ is the required sample of size n from the TL distribution.

Step 3 Use the generated sample to estimate the unknown parameters using maximum likelihood estimators given by (4) and (5).

Step 4 Take sample size as $n = 5(5)50$ and 100. Calculate the appropriate test statistics for the given values of n, as given in relations (6), (8), (10), (12) and (13).

Step 5 Repeat the Steps (1)-(4), $10,000$ times (i.e., 10,000 simulated samples).

Step 6 Arrange the critical values of the test statistics in a ascending increasing order.

Step 7 Select the the percentile values of these test statistics for the sample size n at the significance levels $\alpha = 0.01$, 0.02, 0.05, 0.10, 0.20, 0.25.

4.2 Power tests

As an illustration we have estimated the power of GOF tests when alternatives are the exponential, the Weibull and Pareto of the second type (generalized Pareto) distributions. Our findings are summarized in Table 2. The power was determined by generating $10,000$ random sample of size $n = 5$, 15 and 30 from each of the alternatives and for each test at the significance level $\alpha = 0.05$. Table 2 shows that the AD statistic is generally superior to other test statistic. The power of the test statistic increases as the sample size increases.

Table 1: Critical values for KS, CvM, AD, W and LS tests

Sample size: n	Test Statistics	Significance level: α 0.01	0.02	0.05	0.10	0.20	0.25
5	D_n	0.5602	0.5380	0.5158	0.4837	0.4447	0.4306
	W_n^2	0.3686	0.3349	0.3031	0.2674	0.2235	0.2076
	A_n^2	1.7221	1.5776	1.4398	1.2921	1.1031	1.0357
	U_n^2	0.2869	0.2560	0.2282	0.1987	0.1630	0.1501
	L_n	1.5086	1.4520	1.4002	1.3477	1.2820	1.2552
10	D_n	0.4686	0.4454	0.4250	0.3999	0.3702	0.3581
	W_n^2	0.5267	0.4745	0.4279	0.3783	0.3182	0.2965
	A_n^2	2.5471	2.3185	2.1055	1.8991	1.6301	1.5364
	U_n^2	0.3853	0.3418	0.3068	0.2671	0.2195	0.2034
	L_n	1.6068	1.5456	1.4877	1.4250	1.3401	1.3072
15	D_n	0.4202	0.3978	0.3810	0.3611	0.3350	0.3258
	W_n^2	0.6513	0.5871	0.5322	0.4748	0.4068	0.3825
	A_n^2	3.2087	2.9250	2.6922	2.4357	2.1275	2.0212
	U_n^2	0.4630	0.4120	0.3692	0.3241	0.2717	0.2536
	L_n	1.7028	1.6424	1.5805	1.5106	1.4272	1.3927
20	D_n	0.3948	0.3748	0.3570	0.3388	0.3159	0.3075
	W_n^2	0.7702	0.6963	0.6397	0.5705	0.4941	0.4677
	A_n^2	3.8432	3.5105	3.2655	2.9604	2.6246	2.5010
	U_n^2	0.5362	0.4798	0.4339	0.3821	0.3241	0.3048
	L_n	1.7978	1.7350	1.6776	1.6052	1.5131	1.4806
30	D_n	0.3588	0.3429	0.3289	0.3124	0.2932	0.2864
	W_n^2	1.0015	0.9113	0.8387	0.7641	0.6730	0.6356
	A_n^2	5.0779	4.6769	4.3706	4.0274	3.6152	3.4508
	U_n^2	0.6811	0.6133	0.5561	0.4975	0.4322	0.4053
	L_n	1.9912	1.9224	1.8563	1.7838	1.6929	1.6542
40	D_n	0.3349	0.3213	0.3083	0.2960	0.2806	0.2746
	W_n^2	1.1998	1.1002	1.0197	0.9322	0.8330	0.7964
	A_n^2	6.1647	5.7337	5.3808	4.9699	4.5295	4.3626
	U_n^2	0.7960	0.7225	0.6656	0.5965	0.5252	0.4990
	L_n	2.1565	2.0755	2.0109	1.9279	1.8365	1.8010
50	D_n	0.3195	0.3083	0.2975	0.2862	0.2716	0.2669
	W_n^2	1.4071	1.2935	1.2118	1.1131	0.9967	0.9582
	A_n^2	7.2920	6.7806	6.4361	5.9767	5.4613	5.2781
	U_n^2	0.9335	0.8431	0.7805	0.7077	0.6240	0.5962
	L_n	2.3109	2.2263	2.1595	2.0792	1.9794	1.9434
100	D_n	0.2856	0.2776	0.2703	0.2618	0.2519	0.2488
	W_n^2	2.3836	2.2290	2.1067	1.9742	1.8228	1.7671
	A_n^2	12.6842	11.9593	11.4024	10.8026	10.1133	9.8596
	U_n^2	1.5303	1.4027	1.3174	1.2204	1.1133	1.0727
	L_n	2.9220	2.8316	2.7615	2.6772	2.5807	2.5431

Table 2: Power estimates for test statistics

Alternatives	Test statistics	Sample size n		
		5	15	30
Exponential $Exp(1)$	D_n	0.7936	0.9786	1.0000
	W_n^2	0.8944	0.9624	1.0000
	A_n^2	0.8971	0.9698	1.0000
	U_n^2	0.8896	0.9532	1.0000
	L_n	0.8059	0.8728	0.9993
Weibull $Wei(1,2)$	D_n	0.2872	0.9968	1.0000
	W_n^2	0.3248	0.9934	1.0000
	A_n^2	0.3302	0.9958	1.0000
	U_n^2	0.3245	0.9912	1.0000
	L_n	0.3213	0.9425	1.0000
Generalized Pareto $Pa(1,2)$	D_n	0.1830	0.9004	1.0000
	W_n^2	0.1776	0.6761	0.9784
	A_n^2	0.1779	0.6701	0.9815
	U_n^2	0.1796	0.6811	0.9805
	L_n	0.1659	0.4447	0.8161

References

[1] Abd-Elfattah, A.M. and Omima A.M. (2009), Estimation of the Unknown Parameters of the Generalized Frechet Distribution, *Journal of Applied Sciences Research*, 5, 10, 1398–1408.

[2] Anderson, T.W., and Darling, D.A. (1954), A test of goodness-of-fit, *J. Am. Stat. Assoc.*, Vol. 49, 765-769.

[3] Genc, A.I. (2010), Moments of order statistics of Topp-Leone distribution, *Statistical Papers*, 51, doi:10.1007/s00362-010-0320-y.

[4] Ghitany, M., Kotz, S. and Xie, M. (2005), On some reliability measures and their stochastic orderings for the Topp-Leone distribution, *Journal of Applied Statistics*, 32, 715–722.

[5] Hassan, A.S., (2005), Goodness-of-fit for the generalized exponential distribution, *Interstat Electronic Journal*, July, 1–15.

[6] Kotz, S. and Seier, E. (2007), Kurtosis of the Topp-Leone distributions, *Interstat*, 4 July.

[7] Liao, M. and Shimokawa, T. (1999), A new goodness-of-fit test for Type-I extreme-value and 2-parameter Weibull distributions with estimated parameters, *Journal of Statistical Computation and Simulation*, 64, 23–48.

[8] Nadarajah, S. and Kotz, S. (2003), Moments of some J-shaped distributions, *Journal of Applied Statistics*, 30, 311–317.

[9] Shawky, A.I. and Bakoban, R.A. (2009), Modified Goodness-of-Fit Tests for Exponentiated gamma Distribution with Unknown Shape Parameter, *Interstate Electronic Journal*, July.

[10] Topp, C.W. and Leone, F.C. (1955), A family of J-shaped frequency functions, *Journal of the American Statistical Association*, 50, 313–321.

Received: August, 2012