

Goodness-of-Fit in Shifted Exponential Distribution

Mezbahur Rahman, Rebecca K. Sulley

Department of Mathematics and Statistics, Minnesota State University, Mankato, USA

Email: mezbahur.rahman@mnsu.edu, rebecca.sulley@mnsu.edu

How to cite this paper: Rahman, M. and Sulley, R.K. (2025) Goodness-of-Fit in Shifted Exponential Distribution. *Open Journal of Statistics*, 15, 243-250.

<https://doi.org/10.4236/ojs.2025.152012>

Received: March 27, 2025

Accepted: April 27, 2025

Published: April 30, 2025

Copyright © 2025 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The Shifted Exponential Distribution is widely used in engineering and industrial applications. Goodness-of-fit procedures are revisited. Shapiro-Wilk test, Shapiro-Francia test, Likelihood-ratio Anderson-Darling test, and Likelihood-ratio Kolmogorov-Smirnov test are implemented in shifted exponential distribution. A comparative study with usual Anderson-Darling, Chi-square, Cramer-von-mises, and Kolmogorov-Smirnov tests in testing for shifted exponential distribution is performed using simulation. The Likelihood-ratio Anderson-Darling test is found to be of most powerful irrespective of variant alternatives considered.

Keywords

Anderson-Darling Test, Chi-Square Test, Cramer-Von-Mises Test, Kolmogorov-Smirnov Test, Likelihood-Ratio Anderson-Darling Test, Likelihood-Ratio Kolmogorov-Smirnov Test, Shapiro-Francia Test, and Shapiro-Wilk Test

1. Introduction

The random variable X has a shifted exponential distribution if it has a probability density function of the form:

$$f(x) = \frac{1}{\beta} e^{-\frac{x-\alpha}{\beta}}; \quad x \geq \alpha, \beta > 0. \quad (1)$$

We will consider $X_{1:n}, X_{2:n}, \dots, X_{n:n}$ to be an ordered random sample from an exponential distribution (1).

Parameter estimation in exponential distributions is considered extensively, for example, Johnson and Kotz [1], Johnson *et al.* [2], and Balakrishnan and Basu [3]. Often, parameter estimation in exponential distributions is considered in a special

application scenario such as with survival functions as in Balakrishnan and Sandhu [4]. Variations of this scenario include censored samples, truncated populations, and situations where the shift parameter is assumed to be known. Here we treat exponential distributions of the form (1) and assume that both the parameters are unknown.

Rahman and Pearson [5] showed that the unbiased estimates which are functions of the maximum likelihood estimates, performances are superior compared to commonly used methods mentioned above, which are:

$$\hat{\alpha} = \frac{1}{n-1}(nX_{1:n} - \bar{X}) \quad \text{and} \quad \hat{\beta} = \frac{n}{n-1}(\bar{X} - X_{1:n})$$

with

$$V(\hat{\alpha}) = \frac{\beta^2}{n(n-1)}, \quad V(\hat{\beta}) = \frac{\beta^2}{n-1} \quad \text{and} \quad \text{Cov}(\hat{\alpha}, \hat{\beta}) = -\frac{\beta^2}{n(n-1)},$$

where \bar{X} is the sample mean. We intend to use these estimates in the process of testing the goodness-of-fit in shifted exponential distribution.

Here, we intend to test

H_0 : the sample is from the shifted exponential distribution (1).

H_1 : the sample is not from the shifted exponential distribution (1).

There are many tests to check goodness-of-fit for a specific density function. Recently, Rahman and Wu [6], compared a wide range of exponentiality tests, in that paper, they didn't consider shifted exponential distributions. In practice, people tend to use Chi-square goodness-of-fit as it is very easy to comprehend and perform necessary computation. Shapiro-Wilk test and Shapiro-Francia test are usually implemented for Normal Distribution. Here, we intend to implement the Shapiro-Wilk test and the Shapiro-Francia test along with other commonly used tests, such as, the Anderson-Darling, the Kolmogorov-Smirnov, the Cramer-von-Mises test and usual Chi-Square tests for comparison for the Shifted Exponential Distribution.

1.1. Anderson-Darling Test

The Anderson Darling test assesses whether a sample comes from a specified distribution. It makes use of the fact that, when given a hypothesized underlying distribution and assuming the data does arise from this distribution, the cumulative distribution function (CDF) of the data can be assumed to follow a uniform distribution. Let us consider X_1, X_2, \dots, X_n be a random sample. Anderson-Darling statistic A^2 (here we denote as TAD) is given by Anderson and Darling [7] as follows:

$$TAD = -n - \frac{1}{n} \sum_{i=1}^n (2i-1) \left[\ln(F(Y_i)) + \ln(1-F(Y_{n+1-i})) \right], \quad (2)$$

where Y_1, Y_2, \dots, Y_n be the ordered measurements and F is the CDF (Cumulative distribution function) of (1). Zhang and Wu [8] proposed Likelihood-ratio Anderson-Darling test for exponentiality test as follows:

$$LAD = -\sum_{i=1}^n \left\{ \frac{\log F(Y_i)}{n-i+0.5} + \frac{\log[1-F(Y_i)]}{i-0.5} \right\} \quad (3)$$

Extensive research has been conducted on the asymptotic distributions of these statistics. But here we are proposing simulation distribution under the null hypothesis to obtain the upper tail p -value for the tests (2 & 3).

1.2. Kolmogorov-Smirnov Test

Kolmogorov-Smirnov test (Kolmogorov [9] and Smirnov [10]) is a nonparametric test of the equality of continuous or discontinuous, one-dimensional probability distributions that can be used to test whether a sample came from a given reference probability distribution.

The Kolmogorov-Smirnov statistic quantifies a distance between the empirical distribution function of the sample and the cumulative distribution function of the reference distribution. The empirical distribution function F_n for n independent and identically distributed (i.i.d.) ordered observations X_i is defined as

$$TKS = \sup_x |F_n(x) - F(x)| \quad (4)$$

where $F(x)$ is the CDF of the null hypothesis distribution. Zhang and Wu [8] proposed Likelihood-ratio Kolmogorov-Smirnov test for exponentiality test as follows:

$$LKS = \max_{i \in \{1, 2, \dots, n\}} \left\{ (i-0.5) \log \frac{i-0.5}{nF(Y_i)} + (n-i+0.5) \log \frac{n-i+0.5}{n[1-F(Y_i)]} \right\}, \quad (5)$$

A wide range of research is done in obtaining asymptotic distributions of this statistic. But here we are proposing simulation distribution under the null hypothesis to obtain the upper tail p -value.

1.3. Shapiro-Wilk Test

The Shapiro-Wilk test is a statistical test for the normality of a population, based on sample data. It was introduced by Shapiro and Wilk [11] in testing for normality. Here, we are proposing to implement the test for testing shifted exponential distribution as follows: Let $X_{(i)}$ be the i^{th} ordered values from a sample size n .

$$TSW = \frac{\sum_{i=1}^n (a_i X_{(i)})^2}{\sum_{i=1}^n (X_{(i)} - \bar{X})^2} \quad (6)$$

where \bar{X} is the mean of the sample,

$$(a_1, a_2, \dots, a_n) = \frac{\mathbf{m}^T \mathbf{V}^{-1}}{\mathbf{C}},$$

where $m_i = \sum_{r=1}^i 1/(n-r+1)$, $1 \leq i \leq n$, $V_{ii} = \sum_{r=1}^i 1/(n-r+1)^2$, $1 \leq i \leq n$, $V_{ij} = \sum_{r=1}^i 1/(n-r+1)^2$, $1 \leq i < j \leq n$ Balakrishnan and Basu [3], and

$$C = \|V^{-1}m\| = (m^T V^{-1} V^{-1} m)^{1/2}.$$

Note that this is a left tailed test.

1.4. Shapiro-Francia Test

The Shapiro-Francia test is a statistical test for the normality of a population, based on sample data. It was introduced by S. S. Shapiro and R. S. Francia in 1972 as a simplification of the Shapiro-Wilk test [12].

$$TSF = \frac{\sum_{i=1}^n (X_{(i)} - \bar{X})(m_i - \bar{m})}{\sqrt{\sum_{i=1}^n (X_{(i)} - \bar{X})^2 \sum_{i=1}^n (m_i - \bar{m})^2}} \quad (1.4)$$

where \bar{X} is the mean of the sample and \bar{m} is the mean of m_i 's, given in section 1.3. Note that this is a left tailed test.

1.5. Cramer-von Mises Test

The test statistic is as follows:

$$TLC = \frac{1}{12n} + \sum_{i=1}^n \left[\frac{2i-1}{2n} - F(x_i) \right]^2 \quad (1.5)$$

Note that this is a right tailed test.

1.6. Chi-Square Goodness-of-Fit Test

Standard Chi-Square Goodness-of-fit test is computed as

$$\chi^2 = \sum_{k=1}^g \left(\frac{O_k - E_k}{E_k} \right)^2 \quad (1.6)$$

where g stands for the number of groups, O_k stands for the observed counts in the k^{th} group, and E_k stands for the expected counts under H_0 in the k^{th} group. Note that χ^2 will follow approximate Chi-square distribution with $g - 2 - 1$ degrees of freedom as both the parameters in the Beta distribution are assumed to be unknown.

2. Simulation Results

One thousand samples are generated when H_0 is true, that is, from Exponential distribution with $\alpha = 4.0$ and $\beta = 1.5$. Then one thousand samples are selected from shifted Laplace distribution, Normal distribution with mean 12 and standard deviation 2, shifted Beta distribution with parameters 2 and 4, from shifted Gompertz distribution with parameters 1 and 0.01, when H_0 is false.

Sample sizes are considered 20, 40, 60, and 100. In all tests except the approximate Chi-square test, p -values are computed using simulation, the algorithm is given below. Proportions of rejections are computed for $\alpha = 0.01$, $\alpha = 0.05$, and $\alpha = 0.10$, here α denotes the levels of significance.

In **Tables 1-2**, tests are represented as *TAD* for Anderson-Darling test, *LAD* for Likelihood-Ratio Anderson-Darling test, *TKS* for Kolmogorov-Smirnov test, *LKS*

for Likelihood-Ratio Kolmogorov-Smirnov test, TSW for Shapiro-Wilk test, TSF for Shapiro-Francia test, TLC for Cramer-von Mises test, TCS for Chi-square test using approximate Chi-square distribution and SCS for Chi-square test using simulation.

All tests, except TCS , critical values are determined using the following algorithm.

- Step 1: Generate a sample from a distribution mentioned above.
- Step 2: Estimate parameters α and β as if H_0 is true.
- Step 3: Compute the test statistic and save.
- Step 4: Generate 1000 samples from a shifted exponential distribution with estimated parameter values in Step 2. Compute the respective test statistic to construct the simulated distribution.
- Step 5: Obtain p -value by comparing test statistic value in Step 3 and the simulated distribution in Step 4 and save.
- Step 6: Repeat Step 1 through Step 5 to generate 1000 p -values.
- Step 7: Count number of p -values in Step 6 below 0.01, 0.05, and 0.10 then the proportions of rejections are displayed in **Tables 1-2**.

Note that in TCS and SCS computation, $g = 4$ is used for $n = 20$, $g = 6$ is used for $n = 40$, $g = 8$ is used for $n = 60$, and $g = 10$ is used for $n = 100$, in addition equal probability maintained for each groups in deciding groups.

MATLAB software is used in all computations and the codes are readily available from the primary author.

Table 1. Samples are from shifted exponential distribution.

n	TKS	TAD	TSW	TSF	LKS	TLC	LAD	TCS	SCS
Proportions of rejections of H_0 at $\alpha = 0.01$									
20	0.010	0.008	0.008	0.013	0.003	0.010	0.013	0.042	0.008
40	0.016	0.020	0.009	0.011	0.012	0.017	0.013	0.020	0.011
60	0.009	0.008	0.013	0.017	0.014	0.013	0.010	0.019	0.010
100	0.008	0.013	0.014	0.012	0.010	0.005	0.008	0.017	0.008
Proportions of rejections of H_0 at $\alpha = 0.05$									
20	0.046	0.045	0.044	0.059	0.054	0.052	0.049	0.178	0.043
40	0.061	0.059	0.052	0.055	0.064	0.057	0.058	0.105	0.058
60	0.061	0.072	0.052	0.047	0.051	0.052	0.040	0.091	0.058
100	0.060	0.061	0.053	0.051	0.041	0.042	0.052	0.076	0.055
Proportions of rejections of H_0 at $\alpha = 0.10$									
20	0.117	0.109	0.078	0.106	0.095	0.099	0.100	0.309	0.078
40	0.091	0.088	0.086	0.103	0.090	0.099	0.095	0.177	0.116
60	0.103	0.092	0.086	0.103	0.097	0.102	0.095	0.151	0.088
100	0.108	0.093	0.111	0.098	0.104	0.101	0.099	0.146	0.102
Samples are from shifted Laplace Distribution									
Proportions of rejections of H_0 at $\alpha = 0.01$									
20	0.000	0.000	0.000	0.424	0.807	0.870	0.847	0.834	0.727
40	0.000	0.000	0.000	0.630	0.993	0.996	0.997	0.991	0.984
60	0.000	0.000	0.000	0.764	1.000	1.000	1.000	1.000	0.999
100	0.000	0.000	0.000	0.940	1.000	1.000	1.000	1.000	1.000

Continued

Proportions of rejections of H_0 at $\alpha = 0.05$									
20	0.001	0.001	0.000	0.644	0.920	0.953	0.944	0.940	0.866
40	0.000	0.000	0.000	0.857	0.997	0.999	0.998	0.997	0.993
60	0.000	0.000	0.000	0.938	1.000	1.000	1.000	1.000	1.000
100	0.000	0.000	0.000	0.991	1.000	1.000	1.000	1.000	1.000
Proportions of rejections of H_0 at $\alpha = 0.10$									
20	0.000	0.000	0.000	0.739	0.936	0.957	0.946	0.962	0.896
40	0.000	0.000	0.000	0.929	0.999	1.000	1.000	0.999	0.999
60	0.000	0.000	0.000	0.984	1.000	1.000	1.000	1.000	1.000
100	0.000	0.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000
Samples are from Normal (12, 2) Distribution									
Proportions of rejections of H_0 at $\alpha = 0.01$									
20	0.000	0.000	0.000	0.325	0.602	0.741	0.781	0.575	0.442
40	0.000	0.000	0.000	0.609	0.968	0.982	0.990	0.957	0.934
60	0.000	0.000	0.000	0.825	0.998	1.000	1.000	0.999	0.998
100	0.000	0.000	0.000	0.974	1.000	1.000	1.000	1.000	1.000
Proportions of rejections of H_0 at $\alpha = 0.05$									
20	0.000	0.000	0.000	0.645	0.815	0.892	0.925	0.788	0.621
40	0.000	0.000	0.000	0.921	0.996	1.000	1.000	0.994	0.980
60	0.000	0.000	0.000	0.988	1.000	1.000	1.000	0.999	0.999
100	0.000	0.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000
Proportions of rejections of H_0 at $\alpha = 0.10$									
20	0.001	0.001	0.001	0.753	0.868	0.925	0.951	0.895	0.714
40	0.000	0.000	0.000	0.959	0.998	0.999	1.000	0.992	0.986
60	0.000	0.000	0.000	0.999	1.000	1.000	1.000	1.000	1.000
100	0.000	0.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000

MATLAB software is used in all computations and the codes are readily available from the primary author.

Table 2. Samples are from Beta (2, 4) + 4 Distribution.

n	TKS	TAD	TSW	TSF	LKS	TLC	LAD	TCS	SCS
Proportions of rejections of H_0 at $\alpha = 0.01$									
20	0.000	0.000	0.000	0.087	0.209	0.318	0.395	0.217	0.120
40	0.000	0.000	0.010	0.155	0.589	0.787	0.871	0.513	0.427
60	0.000	0.000	0.001	0.270	0.858	0.969	0.992	0.805	0.751
100	0.000	0.000	0.000	0.606	0.997	1.000	1.000	0.986	0.974
Proportions of rejections of H_0 at $\alpha = 0.05$									
20	0.003	0.001	0.000	0.305	0.444	0.572	0.660	0.420	0.228
40	0.000	0.000	0.046	0.566	0.850	0.930	0.971	0.779	0.665
60	0.000	0.000	0.003	0.785	0.983	0.996	0.999	0.947	0.909
100	0.000	0.000	0.000	0.975	1.000	1.000	1.000	0.999	0.998
Proportions of rejections of H_0 at $\alpha = 0.10$									
20	0.002	0.002	0.000	0.452	0.577	0.684	0.756	0.662	0.348
40	0.000	0.000	0.101	0.775	0.928	0.969	0.991	0.839	0.764
60	0.000	0.000	0.002	0.922	0.994	0.999	0.999	0.974	0.939
100	0.000	0.000	0.000	0.997	1.000	1.000	1.000	0.997	0.996

Continued

Samples are from Gompertz (1, 0.01) + 4 Distribution									
Proportions of rejections of H_0 at $\alpha = 0.01$									
20	0.002	0.002	0.000	0.051	0.076	0.127	0.155	0.104	0.045
40	0.000	0.000	0.000	0.055	0.137	0.266	0.385	0.159	0.113
60	0.000	0.000	0.000	0.081	0.270	0.503	0.683	0.269	0.206
100	0.000	0.000	0.000	0.201	0.589	0.799	0.919	0.493	0.426
Proportions of rejections of H_0 at $\alpha = 0.05$									
20	0.007	0.009	0.000	0.161	0.183	0.264	0.321	0.242	0.109
40	0.000	0.003	0.000	0.301	0.403	0.546	0.655	0.364	0.270
60	0.000	0.000	0.000	0.474	0.600	0.732	0.868	0.517	0.423
100	0.000	0.000	0.000	0.732	0.856	0.941	0.986	0.754	0.675
Proportions of rejections of H_0 at $\alpha = 0.10$									
20	0.020	0.027	0.001	0.300	0.314	0.402	0.462	0.465	0.187
40	0.007	0.002	0.000	0.503	0.522	0.672	0.761	0.476	0.346
60	0.001	0.001	0.000	0.684	0.748	0.841	0.924	0.649	0.534
100	0.000	0.000	0.000	0.896	0.932	0.975	0.997	0.822	0.758

In **Table 1**, we notice that proportions of rejections are close to α , the level of significance, when H_0 is true. In **Tables 1-2**, for all alternatives, tests *TKS*, *TKD*, and *TSW*, proportions of rejections are close to zero irrespective of alternatives or sample sizes.

LAD test has overall higher power except the Laplace alternative *TLC* test has competitive powers.

3. Application

We demonstrate the four different parameter estimation procedures given above using real-life data. The data given in **Table 3** below is obtained from Bain and Engelhardt [13] and represents the times between successive failures. It is assumed that the times are exponentially distributed while successive failures are assumed to be from a Poission process.

Table 3. Times between system failures data.

5.2	8.4	0.9	0.1	5.9	17.9	3.6	2.5	1.2	1.8	1.8	6.1	5.3
1.2	1.2	3.0	3.5	7.6	3.4	0.5	2.4	5.3	1.9	2.8	0.1	

The respective p -values for *TKS* is 0.251, for *TAD* is 0.367, for *TSW* is 0.229, for *TSF* is 0.234, for *LKS* is 0.631, for *TLC* is 0.649, for *LAD* is 0.541, for *TCS* is 0.449 and for *SCS* is 0.765.

4. Conclusion and Remarks

Likelihood-ratio Anderson-Darling test has higher power irrespective of alternative distribution. Cramer-von Mises test is the next best test. Between Shapiro-Wilk and Shapiro-Francia tests, the Shapiro-Francia test has higher power.

Kolmogorov-Smirnov, Anderson-Darling, and Shapiro-Wilk tests have poor

performances as they have very low powers irrespective of alternative distributions.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Johnson, N.L. and Kotz, S. (1970) Continuous Univariate Distributions-1. Houghton Mifflin Company.
- [2] Johnson, N.L., Kotz, S. and Balakrishnan, N. (1994) Continuous Univariate Distributions-1. 2nd Edition, John Wiley & Sons, Inc.
- [3] Balakrishnan, N. and Basu, A.P. (1995) The Exponential Distribution: Theory, Methods and Applications. Gordon and Breach Publishers.
- [4] Balakrishnan, N. and Sandhu, R.A. (1996) Best Linear Unbiased and Maximum Likelihood Estimation for Exponential Distributions under General Progressive Type-II Censored Samples. *Sankhya: The Indian Journal of Statistics, Series B, Part I*, **58**, 1-9.
- [5] Rahman, M. and Pearson, L.M. (2001) Estimation in Two-Parameter Exponential Distributions. *Journal of Statistical Computation and Simulation*, **70**, 371-386. <https://doi.org/10.1080/00949650108812128>
- [6] Rahman, M. and Wu, H. (2017) Tests for Exponentiality: A Comparative Study. *American Journal of Applied Mathematics and Statistics*, **5**, 125-135. <https://doi.org/10.12691/ajams-5-4-3>
- [7] Anderson, T.W. and Darling, D.A. (1954) A Test of Goodness of Fit. *Journal of the American Statistical Association*, **49**, 765-769. <https://doi.org/10.1080/01621459.1954.10501232>
- [8] Zhang, J. and Wu, Y. (2005) Likelihood-Ratio Tests for Normality. *Computational Statistics & Data Analysis*, **49**, 709-721. <https://doi.org/10.1016/j.csda.2004.05.034>
- [9] Kolmogorov, A. (1933) Sulla determinazione empirica di una legge di distribuzione. *Giornale dell'Istituto Italiano degli Attuari*, **4**, 83-91.
- [10] Smirnov, N. (1948) Table for Estimating the Goodness of Fit of Empirical Distributions. *The Annals of Mathematical Statistics*, **19**, 279-281. <https://doi.org/10.1214/aoms/1177730256>
- [11] Shapiro, S.S. and Wilk, M.B. (1965) An Analysis of Variance Test for Normality (Complete Samples) *Biometrika*, **52**, 591-611. <https://doi.org/10.1093/biomet/52.3-4.591>
- [12] Shapiro, S.S. and Francia, R.S. (1972) An Approximate Analysis of Variance Test for Normality. *Journal of the American Statistical Association*, **67**, 215-216. <https://doi.org/10.1080/01621459.1972.10481232>
- [13] Bain, L.J. and Engelhardt, M. (1992) Introduction to Probability and Mathematical Statistics. PWS-KENT Publishing Company.