Advances and Applications in Statistics
Volume 17, Number 1, 2010, Pages 95-104
Published Online: September 28, 2010
This paper is available online at http://pphmj.com/journals/adas.htm

EDF TESTS FOR THE EXPONENTIATED GAMMA DISTRIBUTION UNDER TYPE II CENSORING

H. F. CORONEL-BRIZIO[1], A. I. SHAWKY[2], R. A. BAKOBAN[3] and A. R. HERNÁNDEZ-MONTOYA[1]

[1]Facultad de Física e Inteligencia Artificial
Universidad Veracruzana
Jalapa, Veracruz, México
e-mail: hcoronel@uv.mx

[2]Department of Statistics
Faculty of Sciences
King Abdulaziz University
P. O. Box: 80203, Jeddah 21589, Saudi Arabia

Permanent address

Faculty of Engineering at Shoubra
P. O. Box 1206, El Maadi 11728, Cairo, Egypt
e-mail: aishawky@yahoo.com

[3]Department of Mathematics
Faculty of Sciences
King Abdulaziz University
P. O. Box: 4269, Jeddah 21491, Saudi Arabia
e-mail: rbakoban@kau.edu.sa

Abstract

The so-called exponentiated gamma distribution has a wide range of practical applications such as modelling life time data. In this paper, the

2010 Mathematics Subject Classification: 62-07, 62F10, 62F12, 62N01, 62N02.

Keywords and phrases: exponentiated gamma distribution, goodness of fit, EDF-based tests, Anderson-Darling, asymptotic distributions, type II cessoring, maximum likelihood estimation.

Received April 30, 2010

asymptotic percentage points of the distributions of the EDF-based Anderson-Darling and Cramér-von Mises statistics are given for the case of complete and type II censored samples.

1. Introduction

Following [4], a random variable X is said to have an *exponentiated gamma distribution*, if its cumulative distribution function (CDF) is given by

$$F(x; \theta) = [1 - e^{-x}(x + 1)]^{\theta} \quad \text{for } \theta > 0, \quad x > 0. \tag{1}$$

This distribution has important applications in life testing; details can be found, for example, in [8-13]. In this paper, the percentage points of the asymptotic distributions for the Cramér-von Mises W^2 and the Anderson-Darling A^2 statistics are obtained for the problem of testing goodness-of-fit for the exponentiated gamma distribution when the parameters are estimated from a sample of size n censored at the right. The type of censoring considered here is known as type II, which corresponds to the situation in which, for fixed r, the $n - r$ largest observations are missing.

Table 1. Upper percentage points of the asymptotic distribution of the $A^2_{r,n}$ statistic, for selected censoring proportions $q = 1 - r/n$

Significance level

q	**0.15**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**	**0.001**
0.00	0.914	1.060	1.320	1.589	1.957	2.242	2.921
0.05	0.851	0.994	1.247	1.510	1.869	2.148	2.811
0.10	0.785	0.920	1.160	1.409	1.750	2.014	2.642
0.15	0.720	0.845	1.068	1.300	1.617	1.863	2.448
0.20	0.656	0.771	0.976	1.189	1.481	1.707	2.243
0.25	0.595	0.700	0.886	1.080	1.345	1.550	2.038
0.30	0.537	0.632	0.799	0.974	1.212	1.397	1.836
0.35	0.482	0.567	0.717	0.872	1.085	1.250	1.641
0.40	0.431	0.506	0.639	0.776	0.964	1.110	1.456
0.45	0.382	0.448	0.565	0.686	0.850	0.978	1.281
0.50	0.337	0.394	0.496	0.601	0.744	0.854	1.117
0.55	0.294	0.344	0.432	0.522	0.644	0.739	0.965
0.60	0.254	0.297	0.372	0.448	0.552	0.633	0.824
0.65	0.217	0.253	0.316	0.380	0.467	0.534	0.694
0.70	0.182	0.212	0.264	0.317	0.389	0.444	0.574
0.75	0.149	0.173	0.216	0.258	0.316	0.360	0.464
0.80	0.118	0.137	0.170	0.203	0.248	0.283	0.364
0.85	0.088	0.102	0.127	0.152	0.185	0.211	0.271
0.90	0.059	0.069	0.086	0.102	0.125	0.142	0.183

2. Maximum Likelihood Estimation

The log-likelihood for a right-censored sample $x_{(1)}, \ldots, x_{(r)}$ from the distribution (1) is given by

$$l(\theta) = r\ln(\theta) - \sum_{i=1}^{r} x_{(i)} + \sum_{i=1}^{r} \ln x_{(i)} + (\theta - 1)\sum_{i=1}^{r} \ln V_{(i)} + (n - r)\ln(1 - V_{(r)}^{\theta}),$$

where $V_{(i)} = 1 - (1 + x_{(i)})\exp(-x_{(i)}),\ i = 1, \ldots, r$. The maximum likelihood estimator (MLE) of the parameter θ, is the value $\hat{\theta}$ which satisfies

$$\frac{\partial l(\theta)}{\partial \theta} = \frac{r}{\theta} + \sum_{i=1}^{r} \ln V_{(i)} - \frac{(n - r)V_{(r)}^{\theta} \ln V_{(r)}}{1 - V_{(r)}^{\theta}} = 0. \tag{2}$$

For $r < n$, the solution of (2) can be found using an iterative procedure such as the Newton-Raphson method. In order to obtain the asymptotic variance of $\hat{\theta}$, we also find

$$\frac{\partial^2 l(\theta)}{\partial \theta^2} = -\frac{r}{\theta^2} - \frac{(n - r)V_{(r)}^{\theta}(\ln V_{(r)})^2}{1 - V_{(r)}^{\theta}}\left[1 + \frac{V_{(r)}^{\theta}}{1 - V_{(r)}^{\theta}}\right]. \tag{3}$$

Thus,

$$-E\left[\frac{\partial^2 l(\theta)}{\partial \theta^2}\right] = \frac{r}{\theta^2} + (n - r)E\left[\frac{V_{(r)}^{\theta}(\ln V_{(r)})^2}{1 - V_{(r)}^{\theta}}\right] + (n - r)E\left[\frac{(V_{(r)}^{\theta})^2(\ln V_{(r)})^2}{(1 - V_{(r)}^{\theta})^2}\right].$$

As it can be seen, expression (3) involves the sample through the rth order statistic only, so the above expected values are computed using its distribution. Here, we obtain

$$-E\left[\frac{\partial^2 l(\theta)}{\partial \theta^2}\right] = \frac{H(r, n)}{\theta^2},$$

where

$$H(r, n) = -[r^2 + 2nr\Psi(n)\Psi(r) + r - nr(\Psi(r))^2 - rn - nr\Psi(1, r)$$
$$- nr(\Psi(n))^2 + nr\Psi(1, n) - 2n\Psi(r) + 2n\Psi(n)]/(n - r - 1).$$

In the above expression, $\Psi(x)$ denotes the digamma function and $\Psi(k, x)$ its kth derivative. A more detailed treatment of these functions can be found in [1]. Using the above results, assuming that the value of q remains constant as n tends to infinity, we have

$$\lim_{n\to\infty} -\frac{\theta^2}{n} E\left[\frac{\partial^2 l(\theta)}{\partial\theta^2}\right] = \frac{(1-q)\left[q+(\ln(1-q))^2\right]}{q}.$$

3. Test Procedures

3.1. Right-censored samples

Suppose that we are interested in testing the null hypothesis that the random sample $x_1, ..., x_n$, was drawn from the distribution (1), based on the r smallest observations. The test can be performed as follows:

1. Find the maximum likelihood estimator $\hat{\theta}$ of θ in (1).

2. Obtain the order statistics $x_{(1)} \leq \cdots \leq x_{(r)}$ and compute $z_{(i)} = F(x_{(i)}, \hat{\theta})$ for $i = 1, ..., n.$

Table 2. Upper percentage points of the asymptotic distribution of the $W^2_{r,n}$ statistic, for selected censoring proportions $q = 1 - r/n$

Significance level

q	**0.15**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**	**0.001**
0.00	0.1480	0.1744	0.2215	0.2706	0.3376	0.3897	0.5130
0.05	0.1466	0.1730	0.2200	0.2689	0.3358	0.3878	0.5108
0.10	0.1423	0.1684	0.2147	0.2630	0.3291	0.3814	0.5019
0.15	0.1354	0.1606	0.2056	0.2524	0.3165	0.3625	0.4840
0.20	0.1264	0.1503	0.1929	0.2374	0.2981	0.3375	0.4571
0.25	0.1157	0.1379	0.1775	0.2187	0.2751	0.3184	0.4226
0.30	0.1041	0.1242	0.1600	0.1974	0.2486	0.2879	0.3823
0.35	0.0918	0.1097	0.1414	0.1746	0.2199	0.2548	0.3384
0.40	0.0795	0.0949	0.1224	0.1511	0.1903	0.2208	0.2928
0.45	0.0674	0.0805	0.1037	0.1278	0.1609	0.1861	0.2473
0.50	0.0559	0.0666	0.0857	0.1056	0.1327	0.1537	0.2036
0.55	0.0452	0.0538	0.0690	0.0848	0.1064	0.1230	0.1628
0.60	0.0354	0.0421	0.0539	0.0660	0.0826	0.0933	0.1260
0.65	0.0268	0.0318	0.0406	0.0496	0.0618	0.0705	0.0938
0.70	0.0195	0.0230	0.0292	0.0355	0.0441	0.0508	0.0665
0.75	0.0133	0.0157	0.0198	0.0240	0.0297	0.0341	0.0443
0.80	0.0084	0.0099	0.0124	0.0150	0.0184	0.0210	0.0272
0.85	0.0046	0.0055	0.0068	0.0082	0.0101	0.0114	0.0147
0.90	0.0021	0.0024	0.0030	0.0036	0.0044	0.0050	0.0064

3. Compute the Anderson-Darling or the Cramér-von Mises statistics in their version for a type II right-censored sample [2],

$$A_{r,n}^2 = -\frac{1}{n}\sum_{i=1}^{r}(2i-1)\{\ln z_{(i)} - \ln[1-z_{(i)}]\} - 2\sum_{i=1}^{r}\ln[1-z_{(i)}]$$

$$-\frac{1}{n}[(r-n)^2\ln\{1-z_{(r)}\} - r^2\ln z_{(r)} + n^2 z_{(r)}], \tag{4}$$

$$W_{r,n}^2 = \sum_{i=1}^{r}\left(z_{(i)} - \frac{2i-1}{2n}\right)^2 + \frac{r}{12n^2} + \frac{n}{3}\left(z_{(r)} - \frac{r}{n}\right)^3. \tag{5}$$

4. Using the value $q = 1 - r/n$, the proportion of right-censoring, refer to Table 1 or Table 2, according to the statistic used. If the value of the test statistic exceeds the value in the table, then for a given significance level, reject the null hypothesis.

3.2. Left-censored samples

A test for the left-censored case, based on $x_{(n-r+1)}, \ldots, x_{(n)}$, the largest r observations, can be carried out by taking $z_{(i)}^* = 1 - z_{(n-i+1)}$, $i = 1, \ldots, r$, and computing the appropriate statistic for the right-censored case using the values $z_{(1)}^*, \ldots, z_{(r)}^*$. For the left-censored case, the z-values must be computed using the appropriate maximum likelihood estimator of the parameter θ, namely,

$$\hat{\theta} = -\frac{r}{(n-r)\ln V_{(n-r+1)} + \sum_{i=1}^{r}\ln V_{(n-r+i)}}.$$

3.3. Complete samples

For the case of a complete sample, i.e., $r = n$, the maximum likelihood estimator of θ reduces to $\hat{\theta} = -n\left[\sum_{i=1}^{n}\ln V_{(i)}\right]^{-1}$ and that expressions (4) and (5) also apply.

Table 3. Empirical percentage points of the statistic $A^2_{r,n}$ for selected censoring proportions $q = 1 - r/n$

Significance level

q	n	**0.15**	**0.10**	**0.05**	**0.01**
0.00	60	0.875	1.024	1.272	1.895
	80	0.895	1.046	1.283	1.932
	300	0.910	1.048	1.324	1.974
	∞	0.914	1.060	1.320	1.957
0.20	60	0.634	0.745	0.939	1.493
	80	0.642	0.749	0.992	1.568
	300	0.651	0.768	0.976	1.480
	∞	0.656	0.771	0.976	1.481
0.40	60	0.398	0.470	0.606	0.984
	80	0.411	0.484	0.627	1.003
	300	0.427	0.502	0.620	0.969
	∞	0.431	0.506	0.639	0.964
0.60	60	0.240	0.281	0.356	0.578
	80	0.246	0.290	0.372	0.627
	300	0.250	0.291	0.364	0.563
	∞	0.254	0.297	0.372	0.552
0.65	60	0.194	0.231	0.300	0.488
	80	0.199	0.234	0.298	0.487
	300	0.217	0.254	0.318	0.464
	∞	0.217	0.253	0.316	0.467

4. Quadratic Statistics and Asymptotic Theory

The Anderson-Darling A^2 and the Cramér-von Mises W^2 statistics belong to a class of discrepancy measures of the form

$$Q_n = n \int_{-\infty}^{\infty} [F_n(x) - F(x; \theta)]^2 \psi(x) dF(x; \theta)$$

known as *quadratic statistics*, where F_n denotes the Empirical Distribution Function (EDF) of a random sample of size *n* from an absolutely continuous distribution *F*, θ denotes (in general) a vector parameter and ψ is a weighting function. The well-known Cramér-von Mises W^2 and the Anderson-Darling A^2 statistics are obtained taking $\psi(x) = 1$ and $\psi(x) = \{[F(x; \theta)][1 - F(x; \theta)]\}^{-1}$, respectively. For details on the asymptotic theory of the EDF statistics, the reader is referred, for example, to [3, 5, 6, 7, 14], where the main results used here to obtain numerically the asymptotic percentage points for the exponentiated gamma distribution under type II censoring, can be found. In the following, it will be assumed that the proportion censored, $q = 1 - r/n,$ remains constant as n tends to infinity. Let $\hat{\theta}$ denote the maximum likelihood estimator of the vector parameter $\theta,$ with estimates where necessary, q is the censoring proportion and $p = 1 - q.$ For the case of a singly right-censored sample, the process $\sqrt{n}\{F_n(x) - F(x; \hat{\theta})\},$ evaluated at $t = F(x; \hat{\theta}),$ converges weakly to a Gaussian process $\{Y(t) : t \in (0, p)\}$ with certain covariance function $\rho(s, t)$ which depends both, the functional form of F, and on which parameters have been estimated. The statistics W^2 and A^2 are asymptotic functionals of the process $\{Y(t) : t \in (0, p)\};$ namely, W^2 converges in distribution to $\int_0^p Y^2(t)dt$ and A^2 converges in distribution to $\int_0^p a^2(t)dt,$ where $a(t) = Y(t)[\sqrt{t(1-t)}]^{-1}.$ $Y(t)$ and $a(t)$ are both Gaussian processes defined in $(0, p),$ with covariance functions $\rho(s, t)$ and $\rho_a(s, t) = \rho(s, t)[(s - s^2)(t - t^2)]^{-1},$ respectively, for $0 \leq s, \ t \leq p.$

According to [3], in both cases, the limiting distribution is known to be that of $\sum_{i=1}^{\infty} \lambda_i^* \nu_i,$ where $\nu_1, \ldots$ are independent chi-square random variables with one degree of freedom, and $\lambda_1^*, \ldots$ are the eigenvalues of the integral equation

$$\int_0^p \rho^*(s, t) f_i(s)ds = \lambda_i^* f_i(t), \tag{6}$$

where ρ^* denotes the covariance function corresponding to the limiting process on which the test statistic is based; in our case $\rho(s, t)$ or $\rho_a(s, t).$ In samples from the exponentiated gamma distribution defined in (1), with a right-censored proportion q, the limiting covariance function given of the process $Y(t)$ is given by

$$\rho(s, t) = \min(s, t) - st - \frac{qst\ln(s)\ln(t)}{(1-q)[q + \ln^2(1-q)]}.$$

Thus, the asymptotic distributions of the EDF statistics will not depend on the particular value of the shape parameter.

In order to obtain numerically the asymptotic percentage points, the integral equation (6) was solved using 400 points in $(0, p)$ to approximate the integral and the appropriate covariance function was evaluated in the 400×400 grid, for values of $p = 0.05(0.05)0.90$. The eigenvalues were used to calculate the asymptotic percentage points using Imhof's [5] method. The results are shown in Tables 1 and 2. The row corresponding to $q = 0$ denotes the asymptotic percentage points for complete samples.

Table 4. Empirical percentage points of the statistic $W^2_{r,n}$ for selected censoring proportions $q = 1 - r/n$

Significance level

q	n	**0.15**	**0.10**	**0.05**	**0.01**
0.00	60	0.145	0.170	0.219	0.335
	80	0.147	0.174	0.219	0.337
	300	0.147	0.174	0.224	0.344
	∞	0.148	0.174	0.222	0.338
0.20	60	0.120	0.143	0.182	0.289
	80	0.122	0.143	0.189	0.299
	300	0.125	0.150	0.193	0.293
	∞	0.126	0.150	0.193	0.298
0.40	60	0.070	0.085	0.113	0.189
	80	0.074	0.088	0.115	0.195
	300	0.078	0.092	0.119	0.189
	∞	0.079	0.095	0.122	0.190
0.60	60	0.033	0.039	0.052	0.094
	80	0.033	0.041	0.055	0.104
	300	0.035	0.041	0.053	0.088
	∞	0.035	0.042	0.054	0.083
0.65	60	0.023	0.027	0.038	0.071
	80	0.023	0.028	0.037	0.069
	300	0.027	0.031	0.041	0.063
	∞	0.027	0.032	0.040	0.062

5. Small Sample Distributions

In order to investigate the appropriateness of a test of fit based on the asymptotic percentage points, a simulation study was performed for both statistics considering different censoring proportions. For selected values of q and n, ten thousand pseudo-random samples from the distribution (1) were generated and the statistics $W^2_{r,n}$ and $A^2_{r,n}$ were calculated to estimate the empirical percentage points. The results are shown in Tables 3 and 4.

These results suggest that the speed of convergence of the empirical, to the asymptotic percentage points, does not depend heavily on the proportion of censoring; therefore, the asymptotic percentage points can be used with very good accuracy, for moderately large samples, especially for the case of the statistic $W_{r,n}^2$.

6. Conclusions

The percentage points of the asymptotic distributions of the Cramér-von Mises W^2 and the Anderson-Darling A^2 statistics were calculated numerically and tables for testing goodness-of-fit for the exponentiated gamma distribution, when its parameter is estimated from a complete or censored sample, were provided.

Results from a simulation study showed that for moderately large samples, a goodness-of-fit test can be performed with good accuracy using the asymptotic percentage points of these statistics.

Acknowledgement

H. F. Coronel-Brizio and A. R. Hernandez-Montoya acknowledge support from the Sistema Nacional de Investigadores (SNI), CONACyT, México.

References

[1] M. Abramowitz and I. A. Stegun, Handbook of Mathematical Functions, National Bureau of Standards, 1964.

[2] R. B. D'Agostino and M. A. Stephens, Goodness-of-fit Techniques, Marcel Dekker, New York, 1986.

[3] J. Durbin, Distribution theory for tests based on the sample distribution function, Regional Conference Series in Appl. Math., 9, SIAM, Philadelphia, 1973.

[4] R. C. Gupta, R. D. Gupta and P. L. Gupta, Modelling failure time data by Lehman alternatives, Comm. Statist. Theory Methods 27(4) (1998), 887-904.

[5] J. P. Imhof, Computing the distribution of quadratic forms in normal variables, Biometrika 48 (1961), 419-426.

[6] A. N. Pettitt and M. A. Stephens, Modified Cramér-von Mises statistics for censored data, Biometrika 63 (1976), 291-298.

[7] A. N. Pettitt, Cramér-von Mises statistics for testing normality with censored samples, Biometrika 63 (1976), 475-481.

[8] A. I. Shawky and R. A. Bakoban, Bayesian and non-Bayesian estimations on the exponentiated gamma distribution, Appl. Math. Sci. 2(51) (2008), 2521-2530.

[9] A. I. Shawky and R. A. Bakoban, Characterization from exponentiated gamma distribution based on record values, JSTA 7(3) (2008), 263-278.

[10] A. I. Shawky and R. A. Bakoban, Modified goodness-of-fit tests for the exponentiated gamma distribution with unknown shape parameter, Interstat, July, 3, 2009, http://interstat.statjournals.net/.

[11] A. I. Shawky and R. A. Bakoban, Order statistics from exponentiated gamma distribution and associated inference, Int. J. Contemp. Math. Sci. 4(2) (2009), 71-91.

[12] A. I. Shawky and R. A. Bakoban, Conditional expectation of certain distributions of record values, Int. J. Math. Anal. 3(17) (2009), 829-838.

[13] A. I. Shawky and R. A. Bakoban, On finite mixture of two-component exponentiated gamma distribution, J. Appl. Sci. Res. 5(10) (2009), 1351-1369.

[14] M. A. Stephens, Asymptotic results for goodness-of-fit statistics with unknown parameters, Ann. Statist. 4 (1976), 357-369.