Thailand Statistician
January 2011; 9(1) : 37-49
http://statassoc.or.th
Contributed paper

A Power of Comparison of Goodness of Fit Tests for Exponential Distribution With Grouped Data

Ampai Thongteeraparp* [a] and Kanisa Chodjuntug [b]

[a] Department of Statistics, Faculty of Science, Kasetsart University, Bangkok 10900, Thailand.

[b] Department of Statistics, Faculty of Science, Ubon Ratchathani University, Ubon Ratchathani 34190, Thailand.

* Author for correspondence; e-mail: fsciamu@ku.ac.th

Received: 23 November 2009
Accepted: 22 November 2010

Abstract

This study compares the powers of Goodness of fit tests for grouped data from the exponential distribution. These tests include the Weighted Kolmogorov-Smirnov statistic, Anderson-Darling statistic and Cramer-von Mises statistic. The sample sizes were 30, 50 and 100. These samples were grouped in intervals of 6, 7 and 10. The specified significance level for the tests was 0.05. Each sample of data was generated by simulation and each combination of sample size and interval was run 1,000 times. The simulation studies show that the best test for control the type I error is the Weighted Kolmogorov-Smirnov statistic. Next in rank is Anderson-Darling statistic followed by the Cramer-von Mises statistic. The Weighted Kolmogorov-Smirnov statistic has the highest power of test followed by the Anderson-Darling statistic then the Cramer-von Mises statistic.

Keywords: exponential distribution, goodness of fit tests, grouped data.

1. Introduction

In today's global situation, statistical data is more important and involved in many kinds of education research such as a study of product's expiry date calculation, a trace of growth rate of living things and etc. The raw data might be incomplete because errors occurred during the data collection process. The error might be caused by human factors or equipment failures. In a situation like this, a grouped data method can be applied to handle the problem. The raw data will be arranged into groups and then gathered by counting the observed values that appear during the chosen time intervals throughout the duration of the data collection process. The exponential distribution is often concerned with the amount of time until some specific event occurs. Also, the exponential distribution can be used to model situations where certain events occur with a constant probability per unit length. Sometimes, it is importance that the distribution of the data is known before the data analysis is done. The data distribution testing is called "Goodness of Fit Test" [1]. Many researchers propose the statistical tests, the Kolmogorov- Smirnov test [2], the Cramer-von Mises test, and the Anderson-Darling test [3] which have been modified and applied to grouped data. In this study, these statistical tests will be compared determine whether the power of the test is suitable for grouped data.

2. Scope of the study

2.1 A simulation study was conducted with the following design;

The distributions used to generate the sample were

- Exponential distribution with parameter (λ) equal 1,
- Weibull distribution with parameter (α, β) equal (1.25, 0.75) and (0.75, 1),
- Gamma distribution with parameter (λ, α) equal (1.5, 1) and (1.5, 0.5),

2.2 Sample sizes were 30, 50 and 100.

2.3 The number of intervals was taken as 6, 7 and 10.

2.4 The significance level of the test was taken as 0.05.

3. Goodness of fit tests for the exponential distribution with grouped data

Suppose a random sample of size n from exponential distribution. Then the times of events, X, have the density function

$$f(x;\lambda) = \frac{1}{\lambda}\exp\left(-\frac{x}{\lambda}\right) \; ; \; x \geq 0, \; \lambda > 0$$

We assume that the units are observed at times $x_1, x_2, \ldots, x_{k-1}$, leading to observations in k groups $[0,x_1), [x_1,x_2), \ldots, [x_{k-1},\infty)$. The recorded data are $n_1, n_2, \ldots,$ n_k, where n_i, i = 1, 2, …, k-1 is the number of observations falling in the i^{th} interval. The purpose of this paper is to use the grouped data to test the null hypothesis

H_o : The data distribution is an exponential distribution

H_1 : The data distribution is not an exponential distribution.

The power of the statistics was calculated when the test can control the type I error.

4. The Test Statistics

4.1 Weighted Komogorov-Smirnov statistic

Baklizi [4] proposed the statistics, SW1, SW2 and SW3 for the exponential distribution with grouped data. The statistics are defined as follows:

$$SW1 = \sum_{j=1}^{k-1}\left[F\left(x_j,\hat{\lambda}\right)\left(1-F\left(x_j,\hat{\lambda}\right)\right)\right]^{-1/2} S_j \quad ; \; j = 1, 2, \ldots, k-1$$

$$SW2 = \sum_{j=1}^{k-1}\left(\frac{k}{2}-j\right)^2 S_j$$

$$SW3 = \sum_{j=1}^{k-1} S_j$$

Where $S_j = \left|F_n\left(x_j\right) - F\left(x_j,\hat{\lambda}\right)\right|$.

4.2 Anderson-Darling statistic

Choulakian *et al.* [3] developed the Anderson-Darling statistic for grouped data from a continuous distribution. The statistics are defined as follows:

$$A = n^{-1}\sum_{j=1}^{k}\frac{Z_j^2 p_j}{H_j\left(1-H_j\right)}$$

Where $p_j = \exp(-\frac{x_{j-1}}{\lambda})-\exp(-\frac{x_j}{\lambda})$; j = 1, 2, …, k

$$H_j = \sum_{i=0}^{j} p_i$$

$$Z_j = N_1+\ldots+N_j - n\left(p_1+p_2+\ldots+p_j\right)$$

N_j: Number of observation in j^{th} interval.

4.3 Cramer-von Mises statistic

Spinelli [5] proposed this statistic for testing the exponential distribution with grouped data. The statistic is defined as follows:

$$W = n^{-1}\sum_{j=1}^{k} Z_j^2 p_j .$$

Let $p_j = \exp(-\frac{x_{j-1}}{\lambda})-\exp(-\frac{x_j}{\lambda})$; j = 1, 2, …, k

$$Z_j = N_1+N_2+\ldots+N_j - n\left(p_1+p_2+\ldots+p_j\right).$$

5. Methodology

The simulation study to compare the statistics has 4 steps.

Step 1 : Find the empirical probability of type I error

- Simulate samples from the exponential distribution when the data are grouped using SAS 9.1 for windows for testing the null hypothesis
  H_o : The data distribution is an exponential distribution
  H_1 : The data distribution is not an exponential distribution.
- Calculate each statistic, then find the empirical type I error rate collected from 1,000 replications.

Step 2 : Test the controllability of the probability of type I error

- The Bradlay test was used to test the controllability of the probability of type I error. The criterion used was that the statistic can control the type I error at 0.05 level of significance if the empirical type I error rate in step 1 is between 0.025 and 0.075.

Step 3 : Find the power of each test

- Simulate the non- exponential distribution when the data are grouped using SAS 9.1 for windows for testing the null hypothesis

  H_o : The data distribution is an exponential distribution

  against the alternative

  H_1 : The data distribution is a non exponential distribution.

- Calculate each statistic, then find the empirical power of each test collected from 1,000 replications.

Step 4 : Compare the powers of tests.

6. Research Results and Discussions

6.1 The controllability of the probability of type I error

The empirical type I error rates at the nominal 0.05 level of significance are presented in Table1. Statistics SW1, SW2 and SW3 can control the type I error for all sample sizes and number of inspection intervals. Statistics A and W cannot control the type I error for the number of inspection intervals equal 6 for sample size equal to 50 and statistic W cannot control the type I error for number of inspection intervals equal to 6 for all sample sizes, number of inspection intervals equal 7 for sample size n equal 50 and 100 and number of inspection intervals equal 10 at sample size n equal 100.

Table 1. The empirical type I error rate for a nominal 0.05 level of significance.

Sample size (n)	Statistic	Number of inspection intervals		
		6	7	10
30	SW1	0.057	0.055	0.051
	SW2	0.062	0.056	0.053
	SW3	0.054	0.052	0.047
	W	0.079*	0.075	0.071
	A	0.071	0.064	0.054
50	SW1	0.062	0.058	0.058
	SW2	0.062	0.055	0.056
	SW3	0.058	0.062	0.062
	W	0.085*	0.083*	0.077*
	A	0.078*	0.073	0.070

Sample size (n)	Statistic	Number of inspection intervals		
		6	7	10
100	SW1	0.051	0.049	0.053
	SW2	0.058	0.051	0.058
	SW3	0.054	0.046	0.060
	W	0.092*	0.096*	0.082*
	A	0.092*	0.083*	0.079*

* uncontrolled type I error

6.2 Power comparisons

The powers of the test statistics are presented in table 2-3. The SW1 and SW1 statistics have the highest power. Next in rank is the Anderson-Darling statistic and Cramer-von Mises statistic.

For the comparison of the powers of all scenarios at each specific sample size and number of inspection intervals, the results are as following (Figure. 1-12.):

1. The tests SW1 and SW3 have more power than the SW2, W, and A tests.
2. For the data have a Weibull distribution with parameters equal (0.75, 1) or a Gamma distribution with parameters equal (1.5, 0.5) which are similar to an Exponential distribution with parameter equal 1, then the W test has more power than the others. When we consider the weighted test, the SW2 is more powerful than the SW1 and SW3 tests.
3. When we consider the number of group intervals, the data from a Weibull distribution with parameters equal (1.25, 0.75) and a Gamma distribution with parameters equal (1.5, 1), then the powers of the tests for all numbers of group intervals are similar. When data have a Weibull distribution with parameters equal (0.75, 1) at sample size 100 and a Gamma distribution with equal (1.5, 0.5), then the number of group intervals effects the power of test. The power of the test gets larger when the number of group intervals increases.
4. When the sample size was considered, it was found that the sample size affects the power of test. The power of test increases as the sample size increases.

Table 2. Powers of the test statistics for grouped data from the Weibull distribution with parameters (α,β) equal (1.25, 0.75) and (0.75, 1).

Alternate distribution	Sample size	Statistic	Number of inspection intervals		
			6	7	10
Weibull (1.25, 0.75)	30	SW1	0.765	0.755	0.783
		SW2	0.707	0.709	0.722
		SW3	0.754	0.743	0.679
		W	0.709	0.687	0.682
		A	0.731	0.697	0.712
	50	SW1	0.931	0.926	0.942
		SW2	0.891	0.898	0.871
		SW3	0.925	0.919	0.918
		W	0.884	0.877	0.873
		A	0.914	0.893	0.902
	100	SW1	0.999	0.998	0.999
		SW2	0.999	0.998	0.998
		SW3	0.999	0.999	0.999
		W	0.996	0.998	0.999
		A	0.999	0.999	0.999
Weibull (0.75, 1)	30	SW1	0.251	0.272	0.250
		SW2	0.259	0.275	0.285
		SW3	0.235	0.236	0.245
		W	0.314	0.320	0.303
		A	0.299	0.293	0.281
	50	SW1	0.446	0.452	0.440
		SW2	0.459	0.456	0.440
		SW3	0.391	0.410	0.390
		W	0.491	0.478	0.460
		A	0.477	0.462	0.449
	100	SW1	0.549	0.621	0.659
		SW2	0.705	0.735	0.750
		SW3	0.645	0.653	0.673
		W	0.741	0.760	0.762
		A	0.779	0.779	0.771

Table 3. Powers of test statistics for the grouped data from a Gamma distribution with parameters (λ, α) equal (1.5, 1) and (1.5, 0.5).

Alternate distribution	Sample size	Statistic	Number of inspection intervals		
			6	7	10
Gamma (1.5, 1)	30	SW1	0.584	0.580	0.603
		SW2	0.525	0.533	0.553
		SW3	0.570	0.570	0.514
		W	0.552	0.509	0.514
		A	0.567	0.533	0.542
	50	SW1	0.781	0.780	0.804
		SW2	0.730	0.718	0.736
		SW3	0.767	0.765	0.705
		W	0.732	0.711	0.717
		A	0.754	0.727	0.746
	100	SW1	0.975	0.975	0.978
		SW2	0.940	0.950	0.948
		SW3	0.983	0.977	0.976
		W	0.945	0.953	0.953
		A	0.978	0.975	0.980
Gamma (1.5, 0.5)	30	SW1	0.290	0.337	0.336
		SW2	0.506	0.568	0.629
		SW3	0.332	0.385	0.48
		W	0.629	0.659	0.692
		A	0.586	0.628	0.671
	50	SW1	0.434	0.506	0.512
		SW2	0.717	0.797	0.849
		SW3	0.527	0.579	0.693
		W	0.814	0.843	0.888
		A	0.777	0.814	0.877
	100	SW1	0.320	0.432	0.554
		SW2	0.707	0.870	0.951
		SW3	0.522	0.676	0.809
		W	0.803	0.923	0.980
		A	0.775	0.899	0.968

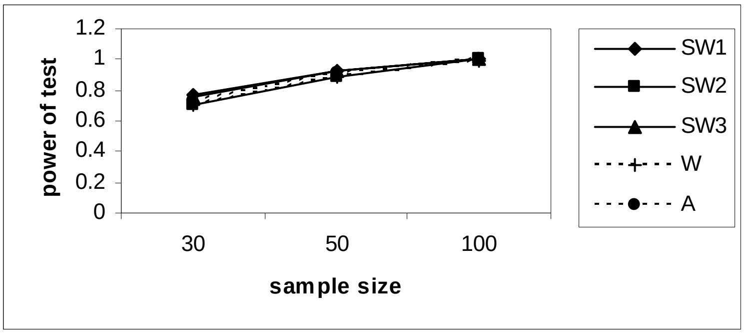

Figure 1. The powers of the test statistics for grouped data from the Weibull distribution with parameters equal (1.25, 0.75) and number of inspection intervals equal to 6.

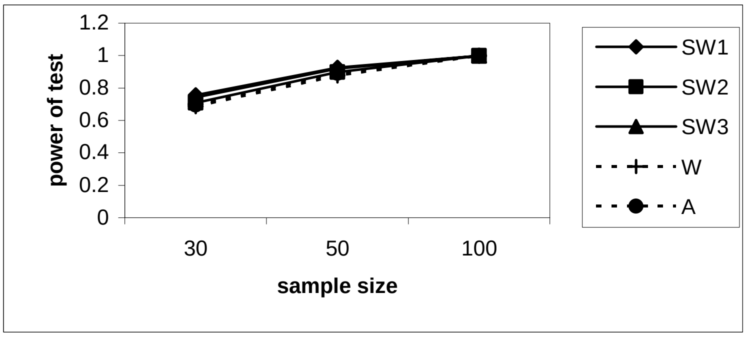

Figure 2. The powers of the test statistics for grouped data from the Weibull distribution with parameters equal (1.25, 0.75) and number of inspection intervals equal to 7

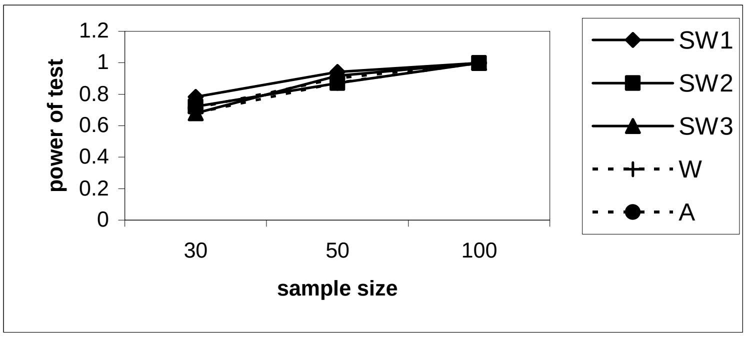

Figure 3. The powers of the test statistics for grouped data from the Weibull distribution with parameters equal (1.25, 0.75) and number of inspection intervals equal to 10.

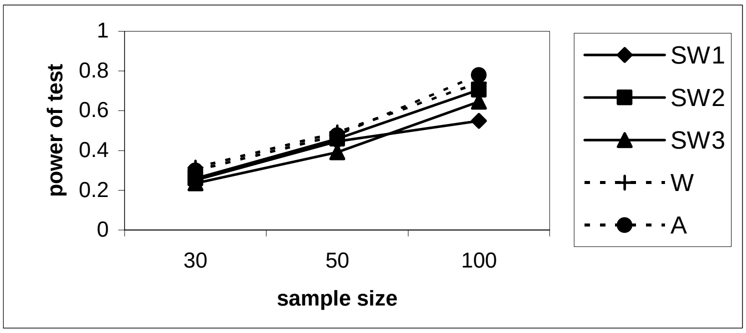

Figure 4. The powers of the test statistics for grouped data from the Weibull distribution with parameters equal (1.25, 0.75) and number of inspection intervals equal to 6.

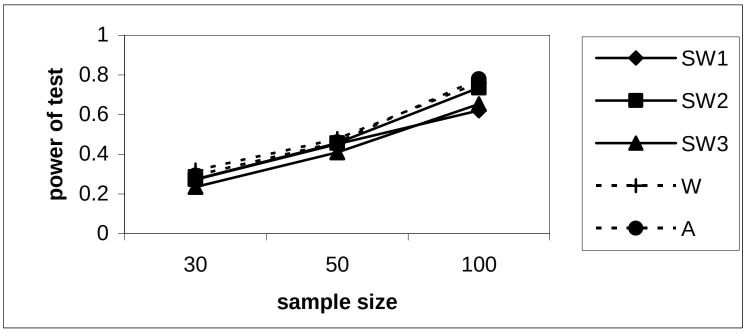

Figure 5. The powers of the test statistics for grouped data from the Weibull distribution with parameters equal (1.25, 0.75) and number of inspection intervals equal to 7.

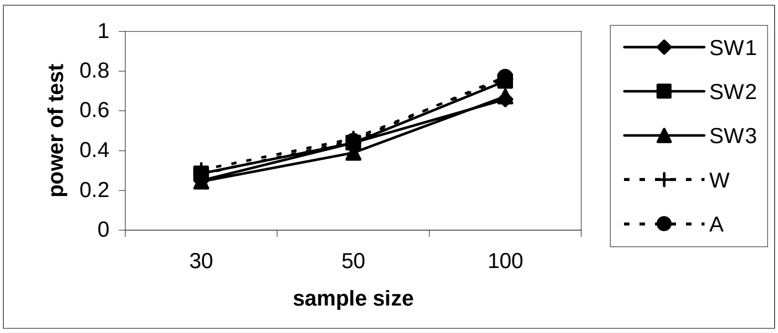

Figure 6. The powers of the test statistics for grouped data from the Weibull distribution with parameters equal (1.25, 0.75) and number of inspection intervals equal to 10.

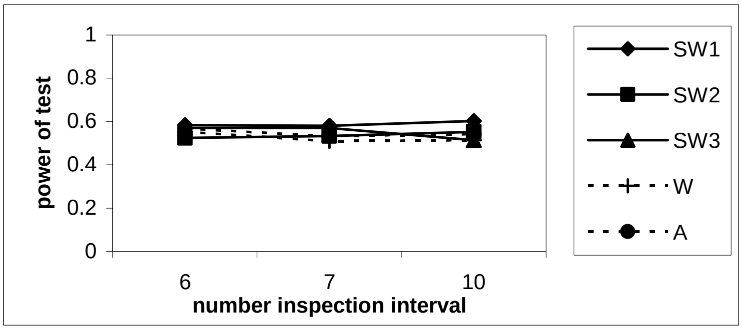

Figure 7. The powers of the test statistics for grouped data from the Gamma distribution with parameters equal (1.5, 1) and sample sizes equal to 30.

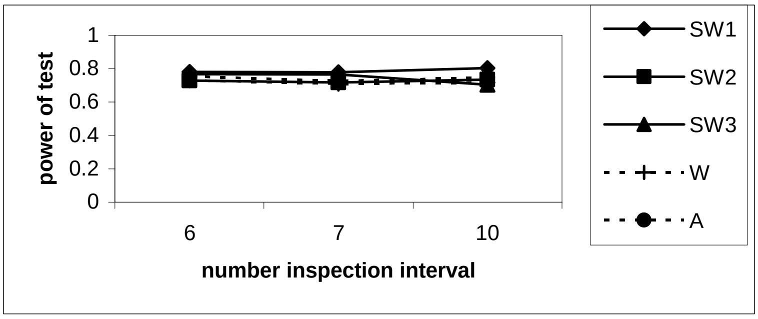

Figure 8. The powers of the test statistics for grouped data from the Gamma distribution with parameters equal (1.5, 1) and sample sizes equal to 50.

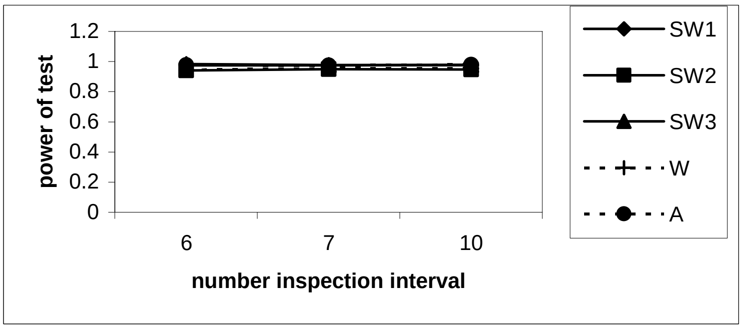

Figure 9. The powers of the test statistics for grouped data from the Gamma distribution with parameters equal (1.5, 1) and sample sizes equal to 100.

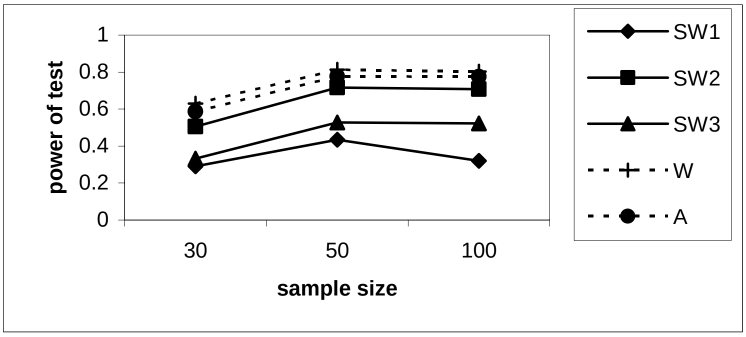

Figure 10. The powers of the test statistics for grouped data from the Gamma distribution with parameters equal (1.5, 0.5) and number inspection intervals equal to 6.

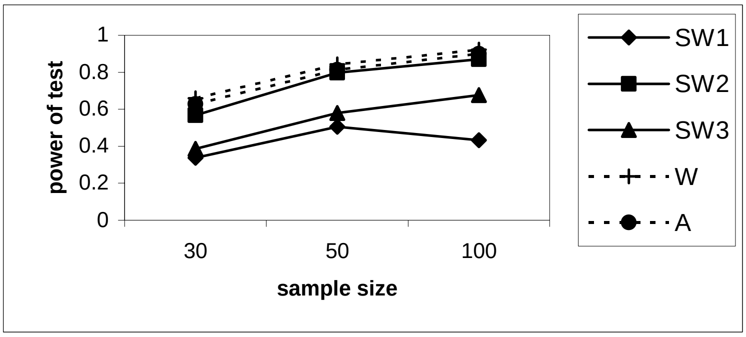

Figure 11. The powers of the test statistics for grouped data from the Gamma distribution with parameters equal (1.5, 0.5) and number of inspection intervals equal to 7.

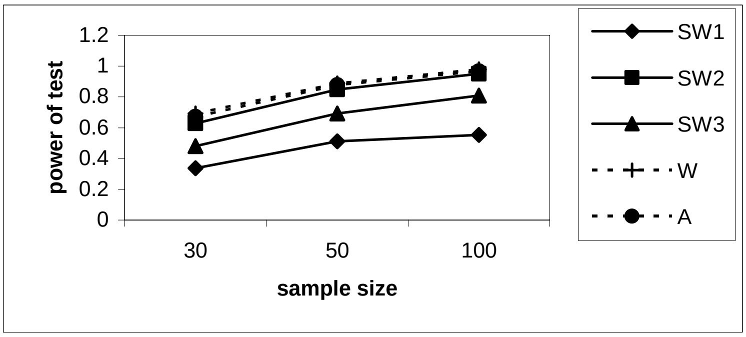

Figure 12. The powers of the test statistics for grouped data from the Gamma distribution with parameters equal (1.5, 0.5) and number of inspection interval equal to 10.

References

[1] Pettitt, A.N. and M.A. stephens. The Kolmogrov-Simirnov Goodness of fit statistic with discrete and grouped data. *Technometrics*, 1977;19: 205-210.

[2] Gulati, S. and J. Neus. Goodness of fit statistics for the Exponential distribution when the data are grouped. *Communication in Statistics Theory and Methods,* 2003;32: 681-700.

[3] Choulakian, V., R.A. Lockhart and M.A. Stephens. Cramer-von Mises test of discrete distribution. *The Canadian Journal of statistics,* 1994;22: 125-137.

[4] Baklizi, A. Weighted Kolmogrov-Simirnov type tests for grouped Rayleigh data. *Applied Mathematical Modelling,* 2006;30: 437-445.

[5] Spinelli, J.J. Testing fit for the grouped Exponential distribution. *The Canadian Journal of Statistics*, 2001; 29: 451-458.