Journal of Modern Applied Statistical Methods

Volume 14 | Issue 1 Article 9

5-1-2015

Modified Lilliefors Test

Achut Adhikari
University of Northern Colorado, adhi2939@gmail.com

Jay Schaffer
University of Northern Colorado, jay.schaffer@unco.edu

Journal of Modern Applied Statistical Methods
May 2015, Vol. 14, No. 1, 53-69.

ISSN 1538 − 9472

Modified Lilliefors Test

A. Adhikari
University of Northern Colorado
Greeley, Colorado

J. Schaffer
University of Northern Colorado
Greeley, Colorado

A new exponentiality test was developed by modifying the Lilliefors test of exponentiality. The proposed test considered the sum of all the absolute differences between the exponential cumulative distribution function (CDF) and the sample empirical distribution function (EDF). The proposed test is simple to understand and easy to compute.

Keywords: Cumulative distribution function, empirical distribution function, exponentiality test, critical value, significance level, and power

Introduction

Exponential distributions are quite often used in duration models and survival analysis, including several applications in macroeconomics, finance and labor economics (optimal insurance policy, duration of unemployment spell, retirement behavior, etc.). Quite often the data-generating process for estimating these types of models is assumed to behave as an exponential distribution. This calls for developing tests for distributional assumptions in order to avoid misspecification of the model (Acosta & Rojas, 2009). "The validity of estimates and tests of hypotheses for analyses derived from linear models rests on the merits of several key assumptions. The analysis of variance can lead to erroneous inferences if certain assumptions regarding the data are not satisfied" (Kuehl, 2000, p. 123).

As statistical consultants we should always consider the validity of the assumptions, be doubtful, and conduct analyses to examine the adequacy of the model. "Gross violations of the assumptions may yield an unstable model in the sense that different samples could lead to a totally different model with opposite conclusions" (Montgomery, Peck, & Vining, 2006, p. 122).

In this study we developed a new Goodness-of-Fit Test (GOFT) of exponentiality and compare it with four other existing GOFTs in terms of

Both authors are in the Applied Statistics and Research Methods. Email Dr. Adhikari at adhi2939@gmail.com. Email Dr. Schaffer at Jay.Schaffer@unco.edu.

computation and performance. This study also derived the critical values of the proposed test. The proposed test considered the sum of all the absolute differences between the empirical distribution function (EDF) and the exponential cumulative distribution function (CDF).

Methodology

To generate critical values, this study used data simulation techniques to mimic the desired parameter settings. Three different scale parameters (θ = 1, 5, and 10) were used to generate random samples from an exponential distribution. Sample sizes 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45 and 50 were used. The study considered three different significance levels (α) (0.01, 0.05 and 0.10). For each sample size and significance level, 50,000 trials were run from an exponential distribution which generated 50,000 test statistics. The 50,000 test statistics were then arranged in the order from smallest to largest. The proposed test is a right tail test. So, this study used the 99th, 95th, and 90th percentile of the test statistics as the critical values for the given sample size for the 0.01, 0.05, and 0.10 significance levels respectively.

To verify the accuracy of the intended significance levels and to compare the power of the proposed test with other four exponentiality tests, data were produced from varieties of 12 distributions (Weibull (1,0.50), Weibull (1,0.75), Gamma (4,0.25), Gamma (0.55,0.275), Gamma (0.55,0.412), Gamma (4,0.50), Gamma (4,0.75), Gamma (4,1), Chi-Square (1), Chi-Square (2), *t* (5) and log-normal (0,1)) to see how the proposed test statistic works. Fifty thousand replications were drawn from each distribution for sample sizes 5, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 1000, and 2000. For each sample size, the proposed test statistic and critical values were compared to make decisions about the null hypothesis. There were 50,000 trials for each sample size. The study tracked the number of rejections (rejection yes or no) in 50,000 trials to evaluate capacity of the proposed test to detect the departure from exponentiality.

The study used R 3.0.2 for most of the simulations to generate test statistics, critical values and power comparisons. Microsoft Excel 2010 was also used to make tables and charts. Monte Carlo simulation techniques were used to generate random numbers which were used to approximate the distribution of critical values for each test.

The proposed modified Lilliefors exponentiality test statistic (PML) takes the form,

$$PML = \sum_{i=1}^{n} \left| F^*(x_i) - S(x_i) \right|, \tag{1}$$

where $F^*(x_i)$ is the CDF of exponential distribution using the maximum likelihood estimator for the scale parameter θ and $S(x_i)$ is the sample cumulative distribution function. The estimator $\hat{\theta}$ is the uniformly minimum variance unbiased estimator (UMVUE) of the scale parameter θ.

The CDF, $F^*(x_i)$, is given by 2

$$F^*(x_i) = 1 - exp\left(-\frac{xi}{\bar{x}}\right), \tag{2}$$

where $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$. The EDF is given by equation 3

$$S(x_i) = i/n \tag{3}$$

Lilliefors test (LF-test) statistic (Lilliefors, 1969) is given by:

$$D = \underset{x}{Sup} \left| F^*(x_i) - S(x_i) \right|, \tag{4}$$

where, $F^*(x_i) = 1 -$ $\exp\left(-\frac{xi}{\bar{x}}\right)$, $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$, and $S(x_i)$ is the empirical distribution function (EDF). Finkelstein & Schafers test (S-test) statistics (Finkelstein & Schafer, 1971) is given by:

$$S = \sum_{i=1}^{n} \max\left\{ \left| F_0\left(X_{(i)}, \hat{\theta}\right) - \frac{i}{n} \right|, \left| F_0\left(X_{(i)}, \hat{\theta}\right) - \frac{i-1}{n} \right| \right\}, \tag{5}$$

where, $\hat{\theta} = \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$. Van-Soest test (VS-test) statistics (Soest, 1969) is given by:

$$W^2 = \frac{1}{12n} + \sum_{i=1}^{n}\left[t_i - \left(\frac{i-0.5}{n}\right)\right]^2, \tag{6}$$

where, $t_i = 1 - \exp\left(-\frac{xi}{\bar{x}}\right)$, and $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$. Srinivasan test ($\tilde{D}_n$ - test) statistics (Srinivasan, 1970) is given by:

$$\tilde{D}_n = \max 1 \le i \le n \left| S_n(x_i) - \tilde{F}(x\,;\lambda) \right|, \tag{7}$$

where, λ is a scale parameter, $\tilde{F}(x\,;\lambda)$ = 1 - $\left\{1 - \frac{x_i}{(n\bar{x})}\right\}^{n-1}$, $S_n(x_i)$ is the EDF. According to Pugh (1963), the test statistic, $\tilde{D}_n$ -test, is based on the Rao-Blackwell and Lehman-Scheffe theorems which give the best unbiased estimate. Schafer, Finkelstein and Collins (1972) corrected the critical points of this test statistic originally proposed by Srinivasan (1970).

Results

Development of critical values

The critical values from the simulated data generated for the three different values of the scale parameters (θ = 1, 5, and 10) are exactly the same for the set of parameters. It appeared that the critical values for the proposed test are the functions of the sample size (n) and the significance levels (α) but invariant with the choice of the scale parameter (θ). Table 1 shows the critical values for the proposed test. Due to space limitations, only five digits are shown on Table 1.

Table 1. Critical Values for the Proposed Exponentiality Test (θ = 1)

n	α = 0.01	α = 0.05	α = 0.10
4	1.0567	0.8331	0.7409
5	1.1760	0.9315	0.8202
6	1.2703	1.0109	0.8931
7	1.3642	1.0856	0.9562
8	1.4647	1.1580	1.0189
9	1.5403	1.2209	1.0757
10	1.6274	1.2875	1.1310
15	1.9444	1.5561	1.3653
20	2.2271	1.7731	1.5636
25	2.4762	1.9682	1.7342
30	2.7097	2.1624	1.9066
35	2.9111	2.3291	2.0584
40	3.1062	2.4837	2.1904
45	3.3216	2.6331	2.3204
50	3.4557	2.7526	2.4309

Accuracy of significance levels

The simulated significance levels are presented on Table 2. Due to the limitations of the space, the simulated significance levels are rounded to three digits. The results showed that all five tests of exponentiality worked very well in terms of controlling the intended significance levels. The study found that the proposed test performs very closely to other four tests of exponentiality in terms of the accuracy of the intended significance levels (for each sample size and overall averages across the 19 different sample sizes). To allow for a better view of the five exponentiality tests across all sample sizes and significance levels, the columns for Lilliefors test are labelled by “LF”, Van-Soest test by “VS”, proposed modified Lilliefors test by “PML”, Srinivasan test by “D” and Finkelstein & Schafers test by “S” for the rest of the tables and figures presented in this study.

Table 2. Average Simulated Significance Levels

α	LF	D	CVM	S	PML
0.010	0.010	0.010	0.010	0.010	0.010
0.050	0.051	0.051	0.051	0.051	0.051
0.100	0.100	0.100	0.101	0.101	0.101

Power analysis

First, consider the relationship between the alternative distribution, Weibull (1, 0.50) and the simulated power. Figure 1 summarizes the power analysis for the Weibull (1, 0.50) alternative distribution. The PML-test outperformed the power for all other four exponentiality tests across all significance levels and sample sizes. The power of all four exponentiality tests exceeded the LF-test. The VS-test, the D-test, and the S-test showed similar performance in power. It appears that for sample sizes 40 or more, the powers for all five exponentiality tests close to 1.

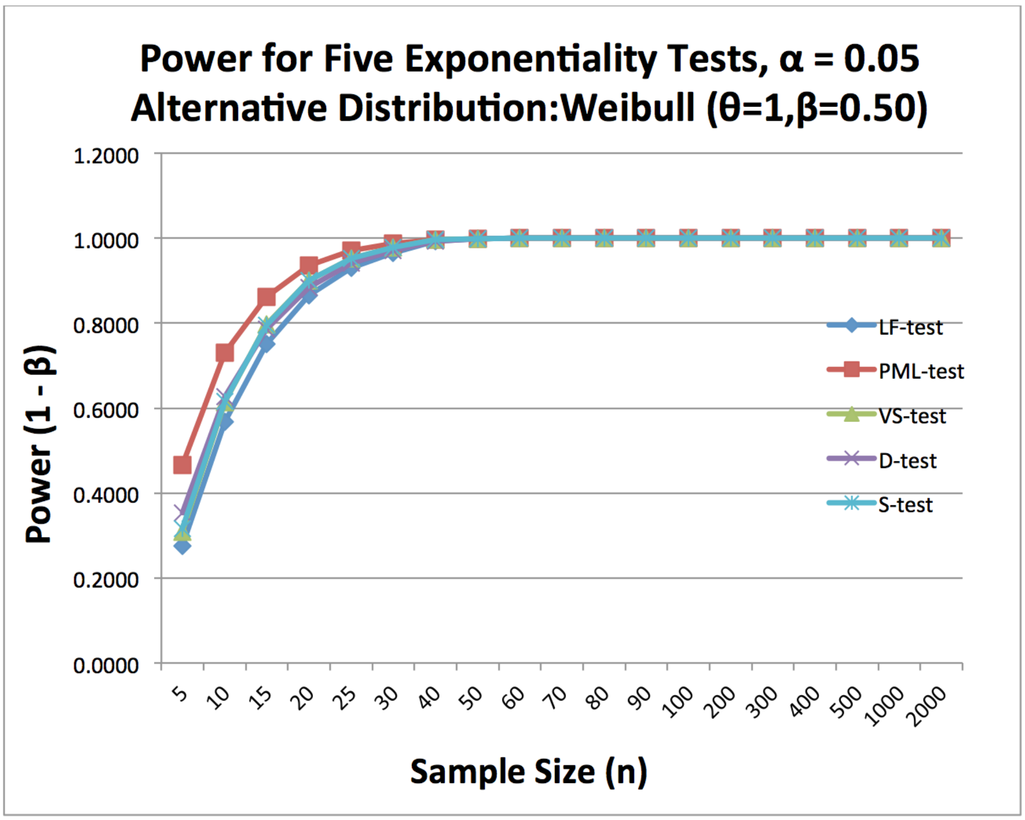

Figure 1. Power for Alternative Distribution: Weibull (1, 0.50)

Second, consider the relationship between the alternative distribution, Weibull (1, 0.75) and the simulated power. Figure 2 summarizes the power analysis for the Weibull (1, 0.75) alternative distribution. This distribution has the same scale parameter ($\theta = 1$) with the previous Weibull (1, 0.50) distribution but the shape parameter (β) is changed from 0.50 to 0.75. This caused the power to reduce substantially across all sample sizes and all significance levels under consideration.

The PML-test outperformed the power for all other four exponentiality tests across all sample sizes and significance levels. In all parameter settings under investigation, the powers for the LF-test were the lowest as compared to other four exponentiality tests. The powers of the S-test and VS-test were almost identical across all sample sizes and significance levels. For a fixed significance level, the powers for the D-test were greater than the S-test and VS-test for small sample sizes but this relationship was reversed for medium to large sample sizes. For all significance levels with sample sizes at least 200, the powers for all five exponentiality tests were almost equal and they approach 1.

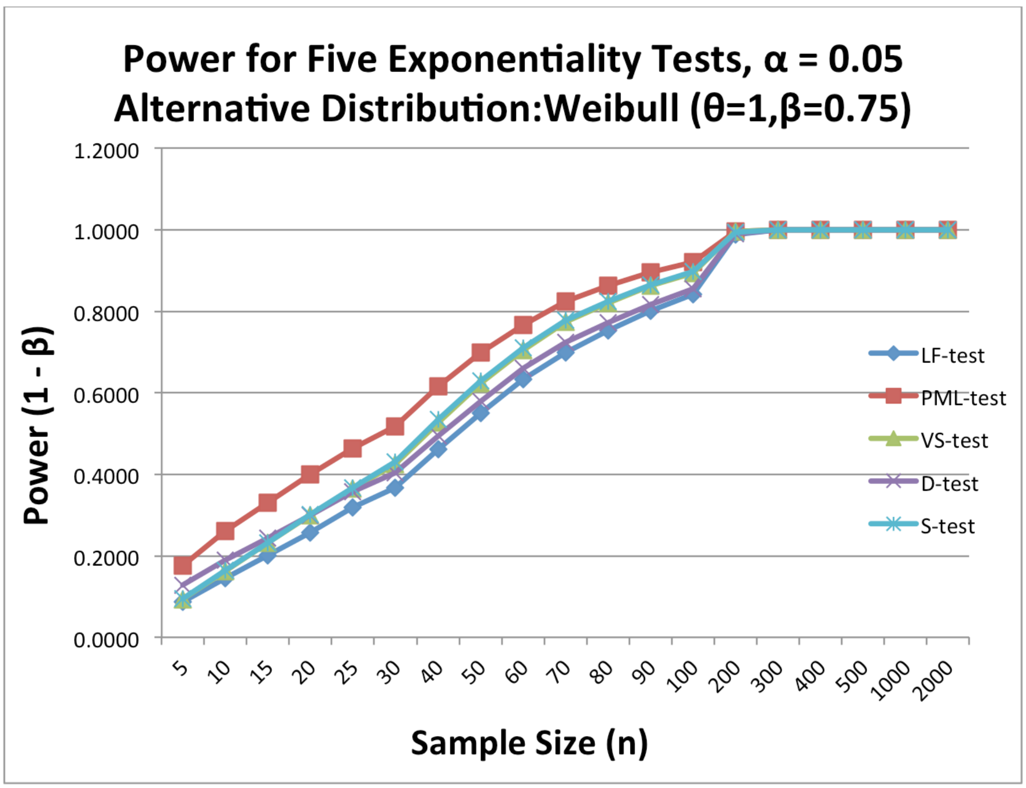

Figure 2. Power for Alternative Distribution: Weibull (1, 0.75)

Third, consider the relationship between the alternative distribution, Gamma (4, 0.25) and the simulated power. Figure 3 summarize the power analysis for the Gamma (4, 0.25) alternative distribution. According to Bain and Engelhardt (1992), the shape parameter, *k*, in the Gamma distribution determines the basic shape of the graph of the probability distribution function (PDF). The value of the shape parameter in null distribution is 1 and the shape parameter in this alternative distribution is 0.25 which are much different. The PML-test outperformed the powers of all other four exponentiality tests across all sample sizes and all significance levels under consideration. For a fixed significance level, the powers of the D-test, VS-test, and S-test exceeded the powers of the LF-test for small sample sizes. For medium to large sample sizes, the LF-test, D-test, S-test, and the VS-test exhibited the identical power across all significance levels. In all parameter settings, the powers of the D-test, the VS-test and the S-test were similar. For sample sizes at least 40, the powers of all five exponentiality tests were found almost equal which were close to 1 across all significance levels.

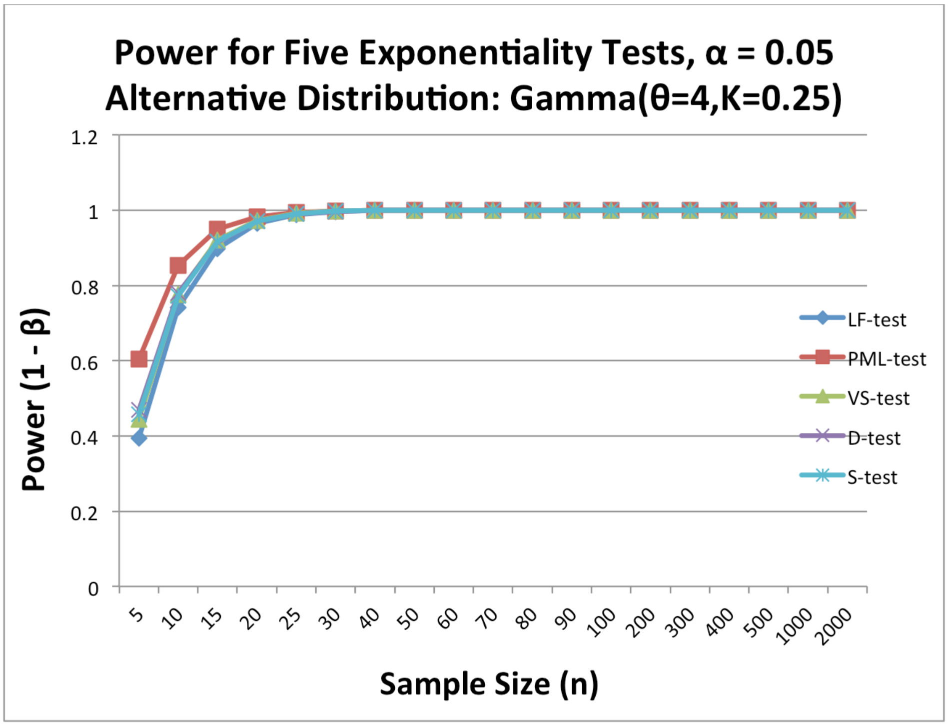

Figure 3. Power for Alternative Distribution: Gamma (4, 0.25)

Fourth, consider the relationship between the alternative distribution, Gamma (0.55, 0.275) and the simulated power. Figure 4 summarizes the power analysis for the Gamma (0.55, 0.275) alternative distribution. The PML-test outperformed other four exponentiality tests across all sample sizes and significance levels. The LF-test exhibited the lowest power across all sample sizes and significance levels. For sample sizes at least 50, the powers for all five tests were found almost equal which were close to 1 across all significance levels. In all parameter settings, the powers for the VS-test, the D-test, and the S-test were identical but all these three tests outperformed the LF-test across all sample sizes and significance levels.

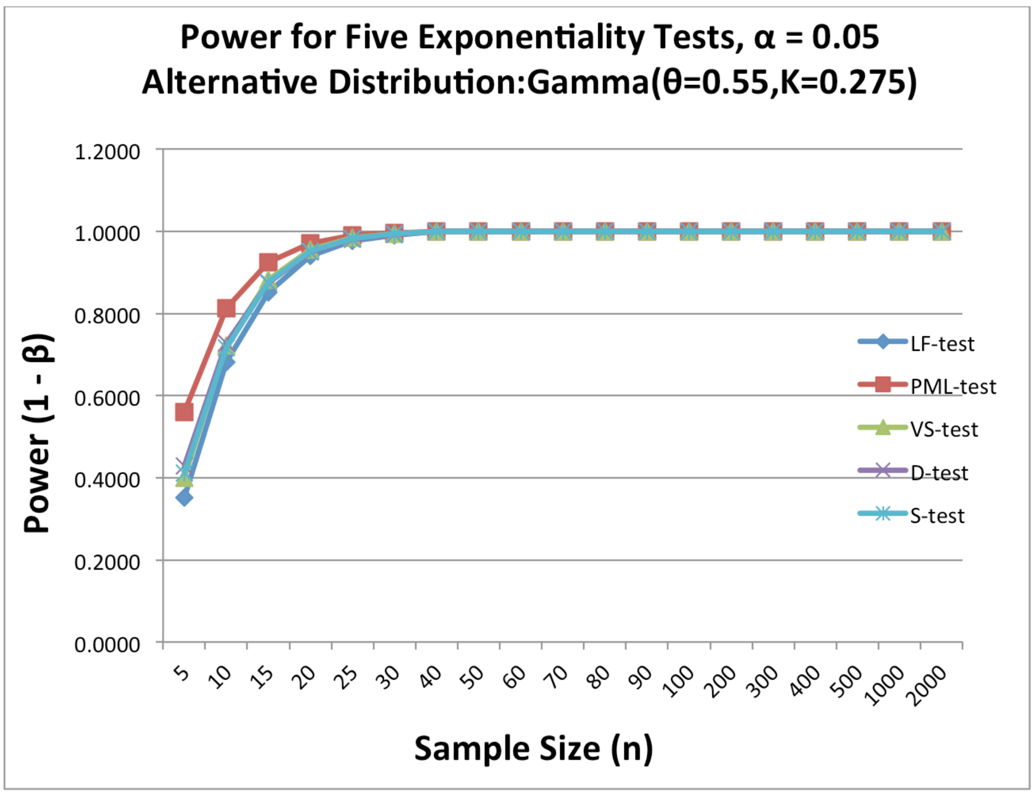

Figure 4. Power for Alternative Distribution: Gamma (0.55, 0.275)

Although the overall power trends in the previous alternative distribution (Gamma (4, 0.25)) and this distribution were similar among five exponentiality tests, the powers for this distribution was lower than the previous alternative distribution across all sample sizes and significance levels. In the previous alternative distribution, the value of the shape parameter (K) is 0.25 which is 0.275 in this alternative distribution.

Fifth, consider the relationship between the alternative distribution, Gamma (0.55, 0.412) and the simulated power. Figure 5 summarizes the power analysis for the Gamma (0.55, 0.412) alternative distribution. The PML-test outperformed other four exponentiality tests across all sample sizes and significance levels. The LF-test exhibited the lowest power across all sample sizes and significance levels. For sample sizes at least 80, the powers for all five tests were found almost equal which were close to 1 across all significance levels. In all parameter settings, the powers for the VS-test, the D-test, and the S-test were identical but all three tests outperformed the LF-test across all sample sizes and significance levels. Comparing the powers for this alternative distribution with the previous alternative distribution (Gamma (0.55, 0.275)), the powers were reduced in this alternative distribution across all sample sizes and significance levels. This is due to only the change in shape parameter (k) from 0.275 to 0.412. The scale parameters (θ) were the same on these two alternative distributions. It is relevant to argue that for Gamma alternative distribution, the powers for these five exponentiality tests depend only on the shape parameter (k). It is also important to note that the shape parameter (k) in the null distribution was 1. So, this study showed that as the shape parameter in the alternative distribution is close to the shape parameter of the null distribution, the simulated powers would be decreased.

Before considering the power for next two alternative distributions, it is imperative to discuss that the Chi-Square distribution is a special case of Gamma distribution. According to Bain and Engelhardt (1992), if a variable Y is a special Gamma distribution with scale parameter ($\theta = 2$) and shape parameter ($k = \nu/2$), the variable Y is said to follow a Chi-Square distribution with ν degrees of freedom. So, if $Y \sim$ Gamma ($\theta = 2$, $k = \nu/2$), a special notation for this distribution can be written as:

$$Y \sim \chi^2(\nu) \tag{8}$$

Using equation 8, the Gamma (4, 0.5) and the Chi-Square (1) distributions are equivalent. This study previously showed that the power for the Gamma distribution depends only on the shape parameter (k). So, the powers of the Gamma (4, 0.5) and Chi-Square (1) alternative distributions must be equivalent.

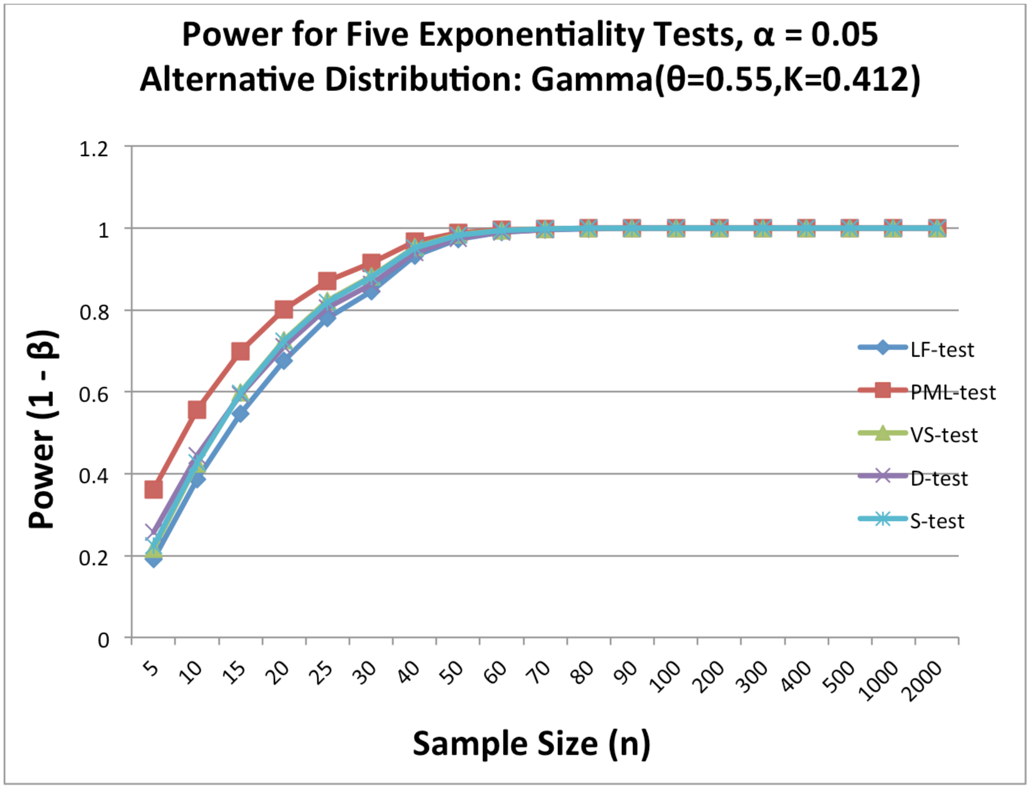

Figure 5. Power for Alternative Distribution: Gamma (0.55, 0.412)

Sixth, consider the relationship between the alternative distributions, Gamma (4, 0.5), Chi-Square (1) and the simulated power. Figure 6 summarizes the power analysis for the Gamma (4, 0.5) and Chi-Square (1) alternative distributions. For a fixed sample size and a significance level, powers for these two alternative distributions were exactly the same. As in the previous alternative distributions, the PML-test outperformed all other four exponentiality tests across all sample sizes and significance levels. The LF-test was in the last place on the power curve. The powers for the VS-test and S-test were identical for a fixed sample size and a significance level. The D-test demonstrated the superior power than the VS-test and the S-test for small sample sizes across all significance levels but this relationship was reversed for medium to large sample sizes. For sample sizes at least 200, the powers for all five tests were equivalents which were close to 1. As compare with the previous alternative distribution (Gamma (0.55, 0.412)), powers for these two alternative distributions decrease across all sample sizes and significance levels. It is relevant to note that the shape parameter (k) was changed from 0.412 to 0.50 which caused the decrease in power. It appears that as the value of the shape parameter (k) approaches that of the null distribution (k = 1), the simulated powers decreases.

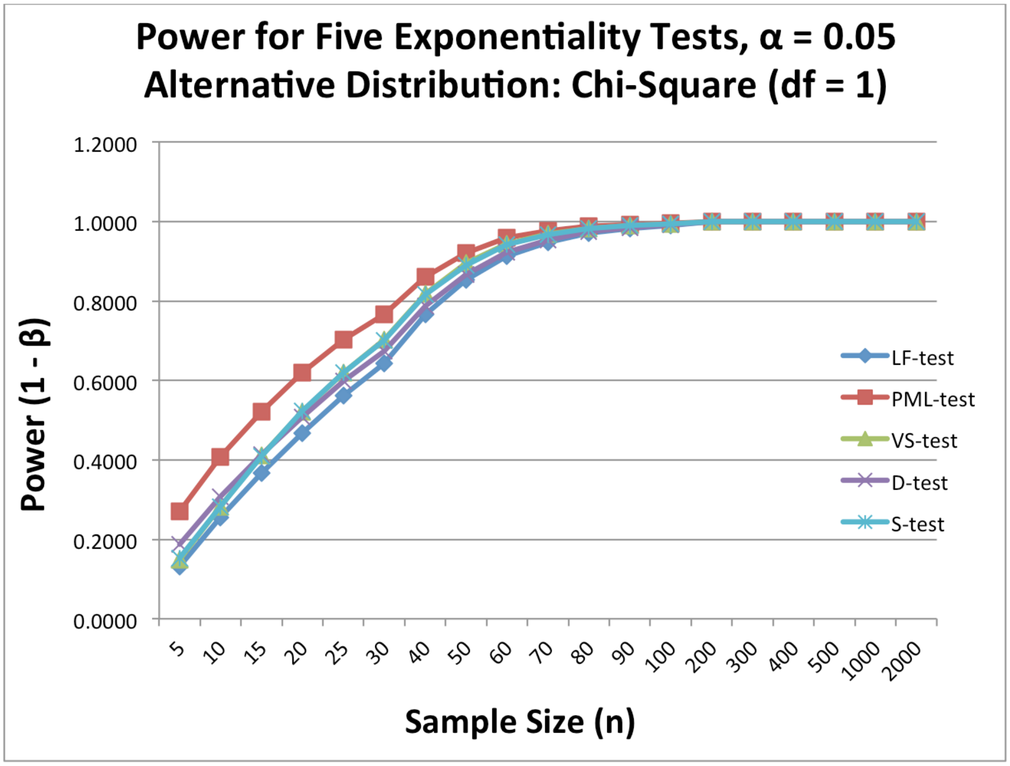

Figure 6. Power for Alternative Distribution: Chi-Square (1)

Seventh, consider the relationship between the alternative distribution Gamma (4, 0.75) and the simulated power. Figure 7 summarizes the power analysis for the Gamma (4, 0.75) alternative distribution. The PML-test outperformed all other four exponentiality tests across all sample sizes and significance levels. The LF-test was in the last place on the power curve. The powers for the VS-test and S-test were identical for a fixed sample size and significance level. The D-test demonstrated the superior power than the VS-test and the S-test for small sample sizes across all significance levels but this relationship was reversed for medium to large sample sizes. For sample size at least 1,000, the powers of all five tests were equivalents which were close to 1. As compare with the previous alternative distribution (Gamma (4, 0.5)), powers of this alternative distributions were significantly decrease across all sample sizes and significance levels. It is relevant to note that the shape parameter (k) was changed from 0.5 to 0.75 which caused the decrease in power. Among five Gamma alternative distributions discussed in this chapter, this alternative distribution exhibited the lowest power across all sample sizes and significance levels.

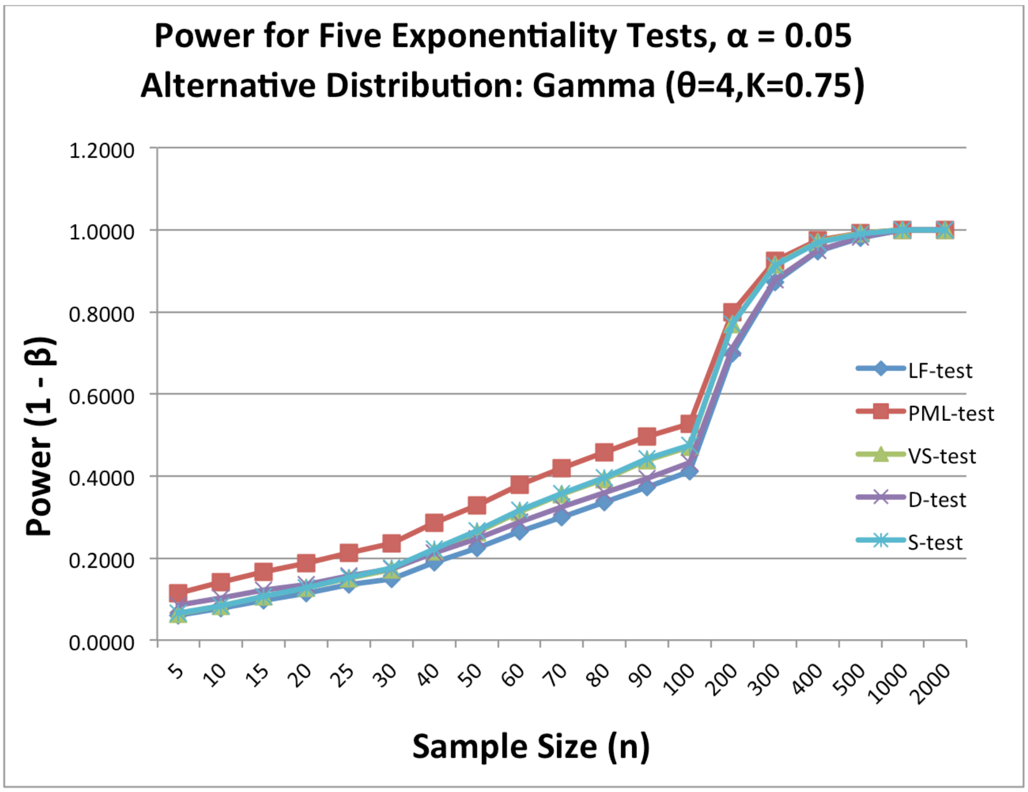

Figure 7. Power for Alternative Distribution: Gamma (4, 0.75)

Before considering the power for next two alternative distributions, it is indispensable to revisit that the Chi-Square distribution is a special case of Gamma distribution (equation 8). This study previously showed that the power for the Gamma distribution depends only on the shape parameter (k). Null distributions were generated using the exponential ($\theta = 5$) for power simulation.

Using 8, Gamma (4, 1) and Chi-Square (2) alternative distributions must produce similar powers for the set of parameters (n and α). In other words Gamma (4, 1) and Chi-Square (2) alternative distributions can be used for the simulation of significance levels.

Eighth, consider the relationship between the alternative distributions, Gamma (4, 1), Chi-Square (2) and the simulated power. Figure 8 summarizes the power analysis for the Gamma (4, 1) and Chi-Square (2) alternative distributions. The powers of all five exponentiality tests across all sample sizes and significance levels were too low which were pretty close to their significance levels. It is due to the fact that the power of these five exponentiality tests depends only on the shape parameter (k). It appears that the scale parameter (θ) does not have any role on the simulated powers.

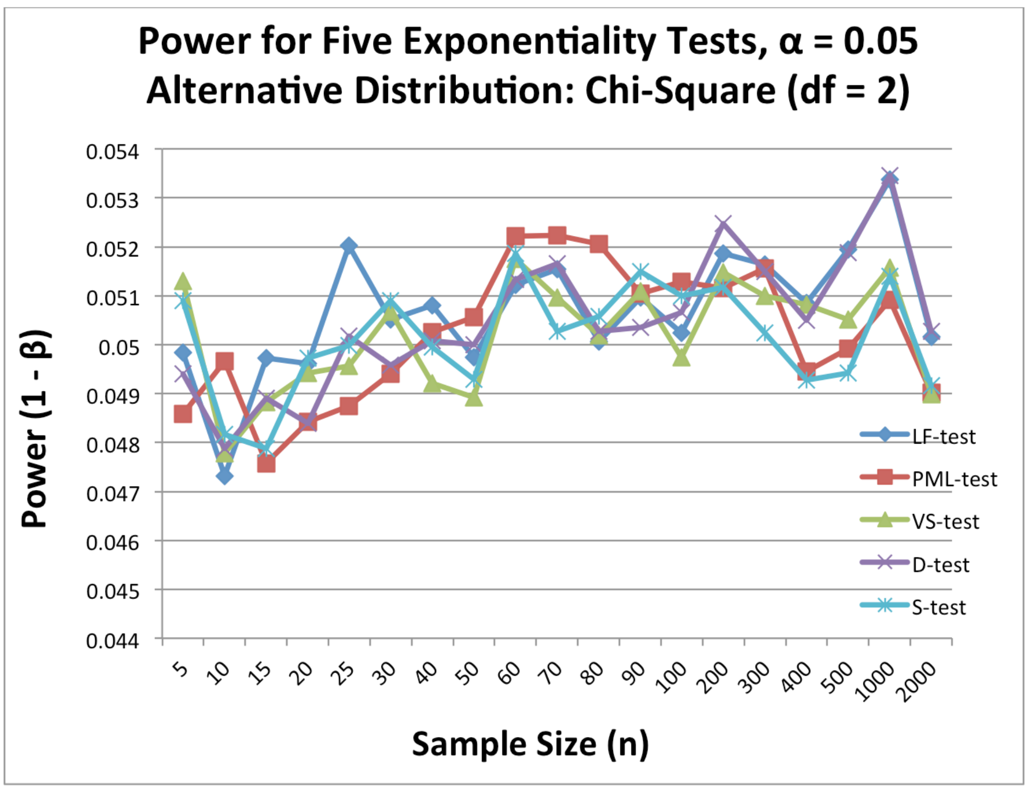

Figure 8. Power for Alternative Distribution: Chi-Square (2)

Ninth, consider the relationship between the alternative distribution *t* (5) and the simulated power. Figure 9 summarizes the power analysis for the *t* (5) alternative distribution. This is the only one symmetric distribution used in the power analyses. All five exponentiality tests quickly detected non-exponentiality. For sample sizes at least 15, the powers for all five tests were almost identical which were close to 1. The range of the powers was found to be very narrow across all sample sizes for a fixed significance level.

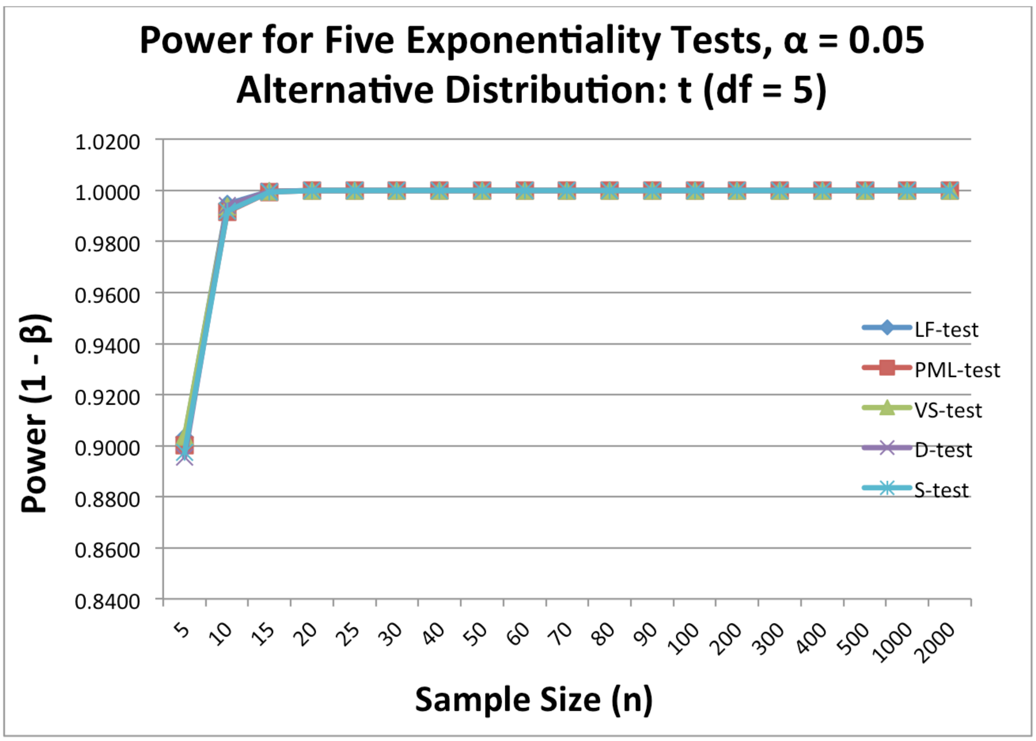

Figure 9. Power for Alternative Distribution: *t* (5)

Finally, consider the relationship between the alternative distribution log-normal (0, 1) and the simulated power. Figure 10 summarizes the power analysis for the log-normal (0, 1) alternative distribution. For small sample sizes, all five exponentiality tests demonstrated similar power across all significance levels. For medium to large sample sizes, the PML-test and S-test were in the top, the VS-test was in the middle and the D-test and LF-test were in the bottom of the power curve. It appears that the PML-test exhibited equal or better power among five exponentiality tests in the set of parameters considered in this study. For sample sizes at least 1000, the powers for all five tests were almost identical which were close to 1.

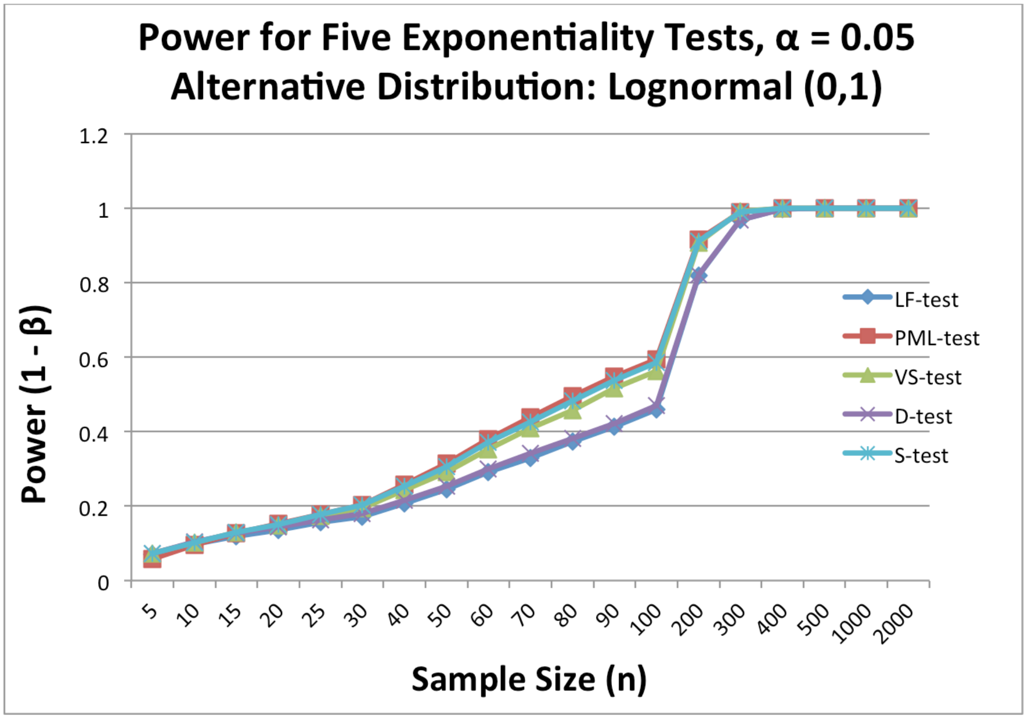

Figure 10. Power for Alternative Distribution: log-normal (0, 1)

Conclusion

This study claimed that the PML-test demonstrated consistently superior power over the S-test, LF-test, VS-test, and D-test for most of the alternative distributions presented in this study. The D-test, VS-test, and S-test exhibited similar power for a fixed sample size and a significance level. The LF-test consistently showed the lowest power among five exponentiality tests. So, practically speaking the proposed test can hope to replace the other four exponentiality tests discussed throughout this study while maintaining a very simple form for computation and easy to understand for those people who have limited knowledge of statistics.

References

Acosta, P., & Rojas, G. M. (2009). A simple IM test for exponential distributions. *Applied Economics Letters, 16*(2), 109-112. doi:10.1080/13504850601018221

Bain, L. & Engelhardt, M. (1992). *Introduction to Probability and Mathematical Statistics* (2nd ed.). MA: PWS-KENT Publishing Company.

Finkelstein, J. M. & Schafer, R. E. (1971). Improved goodness-of-fit tests. *Biometrika, 58*(3), 641-645. doi:10.1093/biomet/58.3.641

Kuehl, R. (2000). *Design of Experiments: Statistical Principles of Research Design and Analysis* (2nd ed.). CA: Duxbury.

Lilliefors, H. W. (1969). On the Kolmogorov-Smirnov test for the exponential distribution with mean unknown. *Journal of the American Statistical Association*, *64*(325), 387-389. doi:10.1080/01621459.1969.10500983

Montgomery, D. C., Peck, E. A. & Vining, G. G. (2006). *Introduction to Linear Regression Analysis* (4th ed.). NJ: John Wiley and Sons, Inc.

Pugh, E. L. (1963). The best estimate of reliability in the exponential case. *Operations Research, 11*(1), 57-61.doi:10.1287/opre.11.1.57

Schafer, R. E., Finkelstein, J. M. & Collins, J. (1972). On a goodness-of-fit test for the exponential distribution with mean unknown. *Biometrika, 59*(1), 222-224. doi:10.1093/biomet/59.1.222

Soest, J. v. (1969). Some goodness of fit tests for the exponential distribution. *Statistica Neerlandica*, *23*(1), 41-51. doi:10.1111/j.1467-9574.1969.tb00072.x

Srinivasan, R. (1970). An approach to testing the goodness of fit of incompletely specified distributions. *Biometrika*, *57*(3), 605-611. doi:10.1093/biomet/57.3.605