*STATISTICA, anno LXIX, n. 1, 2009*

# A CONSISTENT TEST FOR EXPONENTIALITY BASED ON THE EMPIRICAL MOMENT PROCESS

Simos G. Meintanis

## 1. INTRODUCTION

The exponential is one of the most frequently used distributions in survival analysis, reliability theory and other life–time phenomena. Due to its importance, a large number of goodness–of–fit tests have been designed for exponentiality. Many of these tests are reviewed by Spurrier (1984), Ascher (1990) and Henze and Meintanis (2005). When testing for goodness–of–fit, the spectrum of possible deviations from the null hypothesis of exponentiality is often restricted to a nonparametric class of life distributions. Such classes may be defined via monotonicity properties of the failure rate function, with the class of increasing failure rate average (IFRA) distributions being one of the most frequently encountered alternatives to the exponential distribution. It will be seen below that although the moments of an exponentiated random variable belonging to the IFRA class have a certain monotonicity property, the IFRA class is too narrow for this property to be characteristic. This will be the point of departure for considering a new class of life distributions and for constructing exponentiality tests which is consistent within this newly defined family of alternatives.

Let $X$ denote a nonnegative random variable with distribution function $F$ and finite mean

$$\mu = E(X) < \infty .$$

It can be shown that if $F$ belongs to the class of IFRA distributions, then

$$g(t) \text{ is increasing in } t > 0 \tag{1.1}$$

where

$$g(t) := [1 - E(Y^t)]/[tE(Y^t)], \textit{ and } Y = e^{-X} .$$

Condition (1.1) although necessary it is not sufficient for the IFRA class of distributions. Motivated by this fact, Bartoszewicz and Skolimowska (2007) enlarged the IFRA class by introducing the LIFRA as exactly that class of distributions for which (1.1) holds. The letter 'L' in the transition from the terminology IFRA to LIFRA stands for Laplace (transform) since $Y^t = e^{-tX}$ and hence (1.1) essentially refers to a monotonicuity property involving the Laplace transform of the original random variable $X$. Bartoszewicz and Skolimowska (2008) showed that $\text{NBU} \subset \text{LIFRA} \subset L$-class, and derived many interesting results for LIFRA distributions such as closure properties, connections with the notion of the total time on transform, and infinite divisibility.

In the present paper we develop tests for exponentiality which are appropriate against the class of LIFRA distributions. To do so notice that if $F$ is LIFRA, (1.1) is equivalent to

$$D(t) := m^2(t) - m(t) - m'(t) \geq 0, \ t > 0, \tag{1.2}$$

where

$$m(t) = E(Y^t),$$

and that in addition, the exponential distribution is a limiting member of the LIFRA class satisfying

$$D(t) = 0, \ t > 0. \tag{1.3}$$

To see (1.3) recall that under unit exponentiality, $Y = e^{-X}$ is uniformly distributed in (0, 1). Then $E(Y^t)$ is the moment of arbitrary order of the uniform (0,1)–distribution and therefore $m(t)=(1+t)^{-1}$. Hence our test statistic may be viewed as a test based on the moment process $E(Y^t)$ of arbitrary order $t > 0$ of the suitably transformed variate $Y = e^{-X}$, in comparison to the corresponding moment process of the uniform (0,1)–distribution. Since as it will be seen in the sequel, moments of all orders $t > 0$ will be taken into account, the new test is based on a continuum of moment conditions, as oppossed to simple moment–based tests which are based on the first few moments of integer order. The notion of a continuum of moment conditions has been recently exploited in the econometrics literature; see for instance Carrasco and Florens (2000, 2002).

In view of (1.2) and (1.3), it is reasonable to test

$$H_0 : F \text{ is exponential,}$$

against

$$H_1 : F \text{ is LIFRA and not exponential,}$$

by devising an empirical version, say $D_n(t)$, of $D(t)$, and reject $H_0$ in favor of $H_1$ for large values of some distance measure based on $D_n(t)$. Since however the exponential distribution is invariant under scale transformations of the type

$X \to cX$, $c > 0$, one has to standardize the original observations $X_1$, ..., $X_n$ by their sample mean

$$\bar{X}_n = n^{-1} \sum\nolimits_{j=1}^{n} X_j = 1 \ , j = 1,\ 2,\ ...,\ n \ .$$

With this in mind, the obvious candidate for $D_n(t)$ results by replacing the moment process $m(t)$ in (1.2) by the empirical moment process

$$m_n(t) = n^{-1} \sum\nolimits_{j=1}^{n} Y_j^t \ \ where\ \ Y_j = e^{-X_j / \bar{X}_n} \ .$$

Then $D(t)$ in (1.2) is naturally estimated by

$$D_n(t) = m_n^2(t) - m_n(t) - tm'_n(t) \ ,$$

and the null hypothesis is rejected for large values of,

$$T_{n,a} = \sqrt{n} \int_0^\infty D_n(t) e^{-at} dt \ . \tag{1.4}$$

In other words we suggest the Laplace transform (LT) with its well known uniqueness properties as a appropriate distance measure based on $Dn(t)$. This choice leads to an interesting limiting interpretation. Specifically we consider the behavior of the LT of $\sqrt{n} D_n(t)$ as the argument $a > 0$ of the LT goes to infinity, i.e. as $e^{-at}$ approaches a Dirac type function. To this end rewrite the test statistic in (1.4) as

$$T_{n,a} = \int_0^\infty g(t) e^{-at} dt \ , \tag{1.5}$$

where

$$g(t) = \sqrt{n} D_n(t) \ .$$

Using the expansion

$$e^{-x} = 1 - x + (x^2/2) + o(x^2),\ x \to 0,\ in\ Y_j^t = e^{-t} Xj / \bar{X}_n$$

we have by straightforward algebra

$$g(t) = \sqrt{n} \left[ \frac{1}{n^2} \sum_{j=1,k}^{n} \left( \frac{X_j + X_k}{\bar{X}_n} \right)^2 - \frac{3}{n} \sum_{j=1}^{n} \left( \frac{X_j}{\bar{X}_n} \right)^2 \right] \frac{t^2}{2} + o(t^2) \ ,$$

as t → 0. The previous equation along with an Abelian theorem for the LT (see Zayed, section 5.11), yields after some further algebra

$$\lim_{a\to\infty} a^3 T_{n,a} = \frac{\sqrt{n}}{\bar{X}_n^2} [\bar{X}_n^2 - S_n^2] := T_{n,\infty},$$

where

$$S_n^2 = n^{-1}\sum_{j=1}^{n}(X_j - \bar{X}_n)^2 .$$

Consequently the test statistic when properly standardized possesses a limit value, as $a \to \infty$. Interestingly the limit statistic $\mathrm{T}_{\mathrm{n},\infty}$ involves the distance between the sample mean squared and the sample variance, a distance that vanishes under $\mathrm{H}_0$ (exponentiality), as $\mathrm{n} \to \infty$.

## 2. CONSISTENCY AND LIMIT DISTRIBUTION

By straightforward algebra we have from (1.4) that

$$\frac{T_{n,a}}{\sqrt{n}} = \frac{1}{n^2}\sum_{j,k=1}^{n}\frac{1}{Y_j + Y_k + a} - \frac{1}{n}\sum_{j=1}^{n}\frac{1}{Y_j + a} + \frac{1}{n}\sum_{j=1}^{n}\frac{Y_j}{(Y_j + a)^2} . \tag{2.1}$$

The consistency of the test statistic may be proved as follows. By application to (2.1) of the Law of Large Numbers for statistics with estimated parameters, e.g. Randles (1982), we have

$$\frac{T_{n,a}}{\sqrt{n}} \xrightarrow{P} \varepsilon_a(\mu), \text{ as } n \to \infty ,$$

where

$$\varepsilon_a(\mu) = \mu\left[E\left(\frac{1}{X_1 + X_2 + a\mu}\right) - E\left(\frac{1}{X_1 + a\mu}\right) + E\left(\frac{X_1}{(X_1 + a\mu)^2}\right)\right]. \tag{2.2}$$

On the other hand Fubini's Theorem yields that for each $\mu > 0$,

$$\varepsilon_a(\mu) = \int_0^\infty \Delta(t) e^{-at} dt , \tag{2.3}$$

where

$$\Delta(t) = \mu^2(t) - \mu(t) - t\mu'(t) \quad with \quad \mu(t) := E[(e^{-X/\mu})^t]$$

denotes the moment process of $e^{-X/\mu}$.

In view of (2.2) and (2.3), $\varepsilon_a(\mu)$ is zero under exponentiality but positive under $H_1$ which implies the consistency of the test which rejects $H_0$ for large values of $T_{n,a}$. Hence we have implicitly proved the following characterization: Among all non–negative $X$ with finite mean $\mu$, the exponentially distributed random variable is the only one which satisfies $\varepsilon_a(\mu)=0$ for each $a>0$.

For the asymptotic null distribution we will write ≈ when two statistics are asymptotically equivalent. Under the null hypothesis $H_0$, the random variable $X$ follows an exponential distribution with mean $\mu$. Without loss of generality let $\mu=1$. We will consider the symmetrized version $S_{n,a}$ of $T_{n,a}$, where,

$$S^*_{n,a}=\frac{\sqrt{n}}{n^2}\sum_{j,k=1}^{n}\frac{2}{Y_j+Y_k+a}-\left[\frac{1}{Y_j+a}+\frac{1}{Y_k+a}\right]+\left[\frac{Y_j}{(Y_j+a)^2}+\frac{Y_k}{(Y_k+a)^2}\right] \tag{2.4}$$

A typical Taylor expansion yields $S_{n,a}\approx S^*_{n,a}$, where

$$S_{n,a}=\frac{\sqrt{n}}{n^2}\sum_{j,k=1}^{n}\frac{2}{X_j\ +X_k\ +a}-\left[\frac{1}{X_j+a}+\frac{1}{X_k+a}\right]+\left[\frac{X_j}{(X_j+a)^2}+\frac{X_k}{(X_k+a)^2}\right]$$
$$+\sqrt{n}(\bar{X}_n-1)\frac{1}{n^2}\sum_{j,k=1}^{n}\{2h(X_j+X_k)-[h(X_j)+h(X_k)]+[X_jh_1(X_j)+X_kh_1(X_k)]\}$$

where

$$h(x)=h(x;\mathrm{a})=\frac{1}{x+a}-\frac{a}{(x+a)^2}\text{, and}$$

$$h_1(x)=h_1(x;\mathrm{a})=\frac{1}{(x+a)^2}-\frac{2a}{(x+a)^3}.$$

Since $\sqrt{n}(\bar{X}_n-1)$ *is a* $Op(1)$ sequence, an application of the Law of Large Numbers to the second term in the right–hand side of $S^*_{n,a}$ yields

$$S^*_{n,a}\approx\frac{\sqrt{n}}{n^2}\sum_{j,k=1}^{n}\frac{2}{X_j+X_k+a}-\left[\frac{1}{X_j+a}+\frac{1}{X_k+a}\right]+\left[\frac{X_j}{X_j+a}+\frac{X_j}{X_k+a}\right]+\sqrt{n}(\bar{X}_n-1)2\delta_a \tag{2.5}$$

where

$$\delta_a = E[h(X_1+X_2)-h(X_1)+X_1h_1(X_1)]$$
$$= E\left(\frac{1}{X_1+X_2+a}\right) - E\left(\frac{1}{X_1+a}\right) + E\left(\frac{X_1}{(X_1+a)^2}\right) \quad (2.6)$$
$$+\alpha\left[E\left(\frac{1}{(X_1+a)^2}\right) - E\left(\frac{1}{(X_1+X_2+a)^2}\right) - 2E\left(\frac{X_1}{(X_1+a)^3}\right)\right]$$

However by recalling equation (2.2) it follows that the first term of $\delta_a$ coincides with $\varepsilon_a(1)$, while the second term is equal to $a\left(\frac{\delta\varepsilon_a(1)}{\delta a}\right)$. Consiquently under unit exponentiality we have $\delta_a = 0$, and substitution in the right–hand side of (2.5) implies that

$$S^*_{n,a} \approx \left(\frac{n-1}{n}\right)\left[\sqrt{n}\,\frac{2}{n(n-1)}\sum_{j<k} W(X_J, X_K; a)\right],$$

where

$$W(x_1, x_2; a) = \frac{2}{x_1+x_2+a} - \left[\frac{1}{x_1+a} + \frac{1}{x_2+a}\right] + \left[\frac{x_1}{(x_1+a)^2} + \frac{x_2}{(x_2+a)^2}\right].$$

The (common) asymptotic distribution of $S^*_{n,a}$ and $T_{n,a}$ follows then by the theory of U–statistics, e.g. Severini (2005), §13.4. Specifically, define $g(x_1) =: \int_0^\infty W(x_1,\, x_2)\, e^{-x_2}\, dx_2$, *and let* $\sigma_a^2 := Var[g(X_1)]$. Then

$$\mathrm{T_{n,a}} \xrightarrow{D} N(0, 4\sigma_a^2). \quad (2.7)$$

By straightforward algebra we have

$$g(x) = 2e^{a+x}\,\Gamma(0,\, a+x) + \frac{x}{(a+x)^2} - \frac{1}{(a+x)} + ae^a\Gamma(0,a) - 1,$$

where

$$\Gamma(c,\, x) = \int_x^\infty t^{c-1}e^{-t}\,dt$$

denotes the incomplete gamma function. Although an analytic expression for $\sigma_a^2$ may be tedious to derive, numerical values of the limit variance may be easily calculated in the computer as a function of the Laplace parameter a alone. Such values are shown in Table 1.

TABLE 1

*Asymptotic variance for the test statistic $S_{n,a}$ with parameter a*

| a → | 0.5 | 1.0 | 1.5 | 2.0 | 4.0 | 5.0 |
|---|---|---|---|---|---|---|
| Asympt. variance | 0.143583 | 0.0217205 | 0.00614425 | 0.00232429 | 0.000170416 | 0.0000676787 |

3. SIMULATIONS

This section presents the results of a Monte Carlo study conducted to assess the finite–sample behavior of the new test. The asymptotic test has $(1-\alpha)\times 100\%$ rejection region $Z_{n,a} > z_\alpha$, where $Z_{n,a} = S_{n,a} / 2\sigma_a$ and $z_\alpha$ denotes the $(1-\alpha)\times 100\%$ quantile of the standard normal diststribution. For comparison purposes corresponding results are also shown for the classical goodness–of–fit tests based on the empirical distribution function (EDF). The Kolmogorov–Smirnov (KS) statistic is

$$KS = \max\{D^+,\ D^-\},\ D^+ = \max_j \left\{\frac{j}{n} - U_{(j)}\right\},\ D^- = \max_j \left\{U_{(j)} - \frac{j-1}{n}\right\}$$

the Crámer–von Mises (CM) statistic is

$$CM = \frac{1}{12n} + \sum_{j=1}^{n}\left(U_{(j)} - \frac{2j-1}{2n}\right)^2,$$

and the Anderson–Darling (AD) statistic is

$$AD = -n - \frac{1}{n}\sum_{j=1}^{n}(2j-1)\log\ U_{(j)} + (2(n-j)+1)\ \log(1-U_{(j)}),$$

where $U_j = 1 - Y_j$ *and* $U_{(j)}$ denote the corresponding order statistics, $j = 1,\ 2,\ \ldots,\ n$.

In fact we have used the modified statistics

$$KS^* = \left(KS - \frac{0.2}{n}\right)\left(\sqrt{n} + 0.26 + \frac{0{,}5}{\sqrt{n}}\right),\ \ CM^* = CM\left(1 + \frac{0.16}{n}\right),$$

$$\textit{and}\ AD^* = AD(1 + (0.6/n))$$

which were proposed by D'Agostino and Stephens (1986) in order to accommodate differencies in sample size. Percentage points for the modified EDF–statistics may be found in D'Agostino and Stephens (1986), Table 4.11.

Distributions considered are the Gamma with density

$$\Gamma(\theta)^{-1} x^{\theta-1} e^{-x},$$

the Weibull with distribution function $(1-e^{-x^{\theta}})$, the Linear failure rate with density

$$(1+\theta \mathrm{x})\mathrm{e}^{-\mathrm{x}-\frac{1}{2}\theta x^{2}},$$

the half–Normal distribution with density

$$(2/\pi)^{1/2}\ \exp(-\mathrm{x}^2/2),$$

the Lognormal with density

$$(\theta x\sqrt{2\pi})^{-1}\exp[-\log^2 x/2\theta^2],$$

and the Inverge Gaussian distribution with density

$$(\theta/2\pi)^{1/2} x^{-3/2}\exp[-\theta(x-1)^2/2x].$$

These distributions are denoted by G, W, LF, HN, LN and IG, respectively.

Table 2 shows results (percentage of rejection rounded to the nearest integer) obtained from 10,000 samples of size $n = 50$. For simplicity we write $Z_a$ for our asymptotic test. Comparison of the figures in Table 2 indicate that the moment–based test is quite efficient in discriminating between the exponential distribution and some standard alternatives. In addition, and although the power of the new test varies considerably with the Laplace parameter $\alpha$, it may be seen that a compromise value, say $a = 1.0$ *or* $a = 1.5$, produces a test that outperforms the classical tests based on the EDF under most sampling situations, and often by a wide margin.

TABLE 2

*Percentage of rejection observed at nominal level 1% (left entry), 5% (middle entry) and 10% (right entry), with sample size n = 50*

| ↓ Model<br>test → | $Z_{0.5}$ | $Z_{1.0}$ | $Z_{1.5}$ | $Z_{2.0}$ | $Z_{4.0}$ | $Z_{5.0}$ | KS* | CM* | AD* |
|---|---|---|---|---|---|---|---|---|---|
| G(1.0) | 1 6 12 | 1 6 12 | 1 6 12 | 1 6 12 | 1 6 12 | 1 5 12 | 1 5 10 | 1 5 10 | 1 5 10 |
| G(1.6) | 50 80 89 | 54 81 90 | 53 80 89 | 51 79 89 | 43 75 86 | 40 72 85 | 22 47 60 | 29 56 69 | 29 58 71 |
| G(1.8) | 73 93 97 | 77 94 98 | 76 93 98 | 74 93 97 | 65 89 96 | 62 88 95 | 39 67 79 | 50 77 86 | 52 79 88 |
| W(1.3) | 36 66 80 | 44 72 84 | 46 74 85 | 46 75 86 | 42 74 86 | 39 72 85 | 18 42 56 | 26 52 65 | 24 52 65 |
| W(1.5) | 77 94 98 | 85 97 99 | 87 98 99 | 87 98 99 | 85 98 99 | 84 97 99 | 54 80 89 | 69 89 95 | 69 90 95 |
| LF(1.0) | 15 37 52 | 23 49 63 | 27 55 69 | 29 58 72 | 30 62 77 | 28 63 78 | 12 32 46 | 17 40 54 | 14 37 51 |
| LF(2.0) | 28 56 70 | 42 70 81 | 49 76 86 | 52 79 88 | 54 83 91 | 53 83 92 | 26 52 66 | 36 64 75 | 31 60 73 |
| LN(1.0) | 15 43 60 | 8 25 39 | 5 17 28 | 4 14 22 | 2 8 14 | 1 7 12 | 10 24 36 | 12 29 43 | 12 33 50 |
| LN(0.75) | 97 100 100 | 91 98 100 | 83 96 98 | 77 92 96 | 56 80 88 | 50 75 85 | 62 87 49 | 67 90 96 | 75 95 98 |
| HN | 18 42 57 | 28 55 69 | 31 61 75 | 36 65 75 | 37 70 83 | 31 71 84 | 16 38 52 | 22 47 62 | 19 44 59 |
| IG(1.0) | 71 93 98 | 49 78 88 | 36 65 78 | 28 55 69 | 14 36 50 | 12 31 45 | 22 53 71 | 22 55 73 | 30 70 87 |
| IG(1.5) | 99 100 100 | 95 99 100 | 90 98 99 | 84 96 98 | 64 87 93 | 58 83 91 | 76 95 98 | 75 95 98 | 84 98 100 |

*Department of Economics*
*National and Kapodistrian University of Athens*

SIMOS G. MEINTANIS

## REFERENCES

S. ASCHER, (1990). *A survey of tests for exponentiality*, "Communications in Statistics - Theory and Methods", 19, pp. 1811-1825.

J. BARTOSZEWICZ, M. SKOLIMOWSKA, (2008). *New classes of life distributions related to the Laplace transform*, "Communications in Statistics - Theory and Methods", to appear.

M. CARRASCO, J.P FLORENS, (2000). *Generalization of GMM to a continuum of moment conditions*, "Econometric Theory", 16, pp. 797-834.

M. CARRASCO, J.P. FLORENS, (2002). *Simulation–based method of moments and efficiency*, "Journal of Business & Economic Statistic", 20, pp. 482-492.

R. D'AGOSTINO, M.A. STEPHENS, (1986). *Goodness of Fit Techniques*, Marcel Dekker, New York.

N. HENZE, S.G. MEINTANIS, (2005). *Recent and classical tests for expionentiality: a partial review with comparisons*, "Metrika", 61, pp. 29-45.

R.H. RANDLES, (1982). *On the asymptotic normality of statistics with estimated parameters*, "Annals of Statistics", 10, pp. 462-474.

T.A. SEVERINI, (2005). *Elements of Distribution Theory*, Cambridge University Press, New York.

J.D. SPURRIER, (1984). *An overview of tests for exponentiality*, "Communications in Statistics - Theory and Methods", 13, pp. 1635-1654.

A.I. ZAYED, (1996). *Handbook of Function and Generalized Function Transformations.* CRC Press, New York.

## SUMMARY

*A consistent test for exponentiality based on the empirical moment process*

A test for exponentiality is proposed which is consistent within the newly defined class of LIFRA life distributions. The test may be viewed as a test for uniformity based on a continuum of moment conditions. The limit null distribution of the test statistic is derived, and the finite-sample properties of the proposed procedures are investigated via simulation.