Tests for Exponentiality based on characterizations

Emanuele Taufer
Notation

• $X_1, \ldots, X_n$ a random sample from a continuous distribution $F$
• $X_{(1)} \leq \cdots \leq X_{(n)}$ the order statistics
• The hypothesis of interest

$$H_0: F(x) = 1 - e^{-x/\theta}, \quad \theta > 0$$

• $X$ is $E(\theta)$
Classical approaches

• Kolmogorov-Smirnov

\[ KS_n = \sup_{x \geq 0} |F_n(x) - F_0(x)| \]

• Cramér-von Mises

\[ \omega_n^2 = \int_0^\infty (F_n(x) - F_0(x))^2 dF_0(x) \]
Some other approaches

- Empirical Characteristic function of Laplace transform
- Entropy
- Integrated Empirical Distribution function
- Lack of Memory
- Constancy of Mean Residual life
- Spacings
Laplace Transform and Ch.F.

\[ \psi_n(t) = \frac{1}{n} \sum_{i=1}^{n} \exp(-tY_i), \quad Y_i = \frac{X_i}{\bar{X}} \]

Baringhaus and Henze (1991): CVM type using \((1 + t)\psi'(t) + \psi(t) = 0, \ t \in \mathbb{R}\).

Henze (1993): CVM type on \(\psi_n(t)\) and \(\psi(t)\)

Henze and Meintanis (2002a): introducing a weight function

Memoryless property

\[ F(x + y) = F(x)F(y), \quad x, y > 0 \]

- Angus (1982): KS and CVM approach \( x = y \)
- Ahmad and Alwasel (1999): approach based on V-statistics
  \[ F(rx) = [F(x)]^r \]
- Alwasel (2001)
Entropy Characterizations

Entropy of Exponential is max among densities with positive support and given mean. Compare parametric versus non-parametric estimator of entropy.

Integrated EDF

- Klar (2001)

\[
\int_{0}^{\infty} \left[ \Psi_n(t) - \Psi(t) \right]^2 \exp(-at) \, dt,
\]

\[
\Psi(t) = \int_{t}^{\infty} (1 - F(x; 1)) \, dx = \exp(-t)
\]

\[
\Psi_n(t) = \int_{t}^{\infty} (1 - F_n(x)) \, dx = \frac{1}{n} \sum_{j=1}^{n} \max(Y_j - t, 0)
\]
Mean residual life characterization

\[ E(X - t | X > t) = \frac{\int_t^\infty \bar{F}(x)dx}{\bar{F}(t)} \]

is constant at \( \theta = E(X), \ \forall t \) for exponential

Equivalently

\[ \int_t^\infty \bar{F}(x)dx = \bar{F}(x)\theta, \ \forall t \]

\[ E[\min(X, t)] = F(t)\theta, \ \forall t \]
Some previous work on MRL

• Hollander and Proschan (1975), Bergman and Klefsjö (1989), Bandyopadhyay and Basu (1990): against DMRL
• Koul (1978), Bhattacharjee and Sen (1995): against NBUE
• Klar (2000): against HNBUE
Test statistics

• Define the sample MRL after $X_{(k)}$ as

\[ \bar{X}_{>k} = \frac{1}{n - k + 1} \sum_{i=k+1}^{n+1} (X_{(i)} - X_{(k)}) \]

\[ = \frac{1}{n - k + 1} \sum_{i=k+1}^{n+1} (n - i + 2)(X_{(i)} - X_{(i-1)}) . \]

• Under the null hypothesis

\[ E(\bar{X}_{>k}) = E(\bar{X}) = \theta, \quad k = 1, \ldots, n. \]
• First Idea: reject for large values of

\[ T'_n = \max_{1 \leq k \leq n} \frac{|\bar{X} - \bar{X}_{>k}|}{\bar{X}} \]

• Note: no convergence, use instead

\[ T_n = \max_{1 \leq k \leq n-[n\gamma]} \frac{|\bar{X} - \bar{X}_{>k}|}{\bar{X}}, \quad \gamma \in (0, 1). \]

\( \gamma \) is a trimming parameter
Some asymptotics

\[ \bar{X}_k \xrightarrow{\text{a.s.}} \theta \quad 1 \leq k \leq n - \lfloor n^\gamma \rfloor \quad T_n \xrightarrow{\text{a.s.}} 0. \]

**Theorem 1**  
Let \( m(t) < \infty \) and \( F \) be a continuous d.f. with mean \( \theta \). Then, as \( n \to \infty \),

\[
\max_{1 \leq k \leq n - \lfloor n^\gamma \rfloor} |\bar{X} - \bar{X}_k| \xrightarrow{\text{a.s.}} \sup_{0 \leq t < \infty} |\theta - m(t)|.
\]
Next, exploiting equality in distribution

\[ \frac{\bar{X} - \bar{X}_{>k}}{\bar{X}} = \frac{i/(n + 1) - U_{(i)}}{i/(n + 1)} \]

for \( k = 1, \ldots, n, i = n - k + 1 \).

\( U_{(k)} \) is the kth order statistics from a uniform random sample

**Theorem 2**  Let \( \gamma \in (0, 1) \), then under the null hypothesis of exponentiality

\[ n^{\gamma/2} T_n \xrightarrow{D} \sup_{0 \leq t \leq 1} |W(t)| \]

where \( W(t) \) is a Wiener process.
# Power simulations

## Table 2. FR and MRL classification of the distributions used in simulations, $\theta > 0$.

<table>
<thead>
<tr>
<th>Distribution</th>
<th>IFR (DMRL)</th>
<th>DFR (IMRL)</th>
<th>DIFR (IDMRL)</th>
<th>IDFRL (DIMRL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weibull ($\theta$)</td>
<td>$\theta &gt; 1$</td>
<td>$\theta &lt; 1$</td>
<td>$-$</td>
<td>$-$</td>
</tr>
<tr>
<td>Power ($\theta$)</td>
<td>$\theta \leq 1$</td>
<td>$-$</td>
<td>$\theta &gt; 1$</td>
<td>$-$</td>
</tr>
<tr>
<td>Lomax ($\theta$)</td>
<td>$-$</td>
<td>$\theta &gt; 0$</td>
<td>$-$</td>
<td>$-$</td>
</tr>
<tr>
<td>Dhillon ($\theta$)</td>
<td>$\theta \geq 1$</td>
<td>$-$</td>
<td>$\theta &lt; 1$</td>
<td>$-$</td>
</tr>
<tr>
<td>Log-logistic ($\theta$)</td>
<td>$-$</td>
<td>$\theta \leq 1$</td>
<td>$-$</td>
<td>$\theta &gt; 1$</td>
</tr>
<tr>
<td>Compound Rayleigh ($\theta$)</td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
<td>$\theta &gt; 0$</td>
</tr>
</tbody>
</table>

Figure 1. Power values, in percentage, of $T_n$ (dashed), $KS_n$ (solid) and $L_n$ (dotted). Distributions are Weibull (1.2), x (cross); Power (0.5), o (circle).
Figure 2. Power values, in percentage, of $T_n$ (dashed), $KS_n$ (solid) and $L_n$ (dotted). Distributions are Lomax (0.5), x (cross); Weibull (0.8), o (circle).
Figure 3. Power values, in percentage, of $T_n$ (dashed), $KS_n$ (solid) and $L_n$ (dotted). Distributions are Power (2), x (cross); Dhillon (0.5), o (circle).
Figure 4. Power values, in percentage, of $T_n$ (dashed), $KS_n$ (solid) and $L_n$ (dotted). Distributions are Log-logistic (3), $x$ (cross); Compound Rayleigh (1), $o$ (circle).
Tests based on spacings

- Normalized spacings from $E(\theta)$ are $E(\theta)$.

\[ Y_i = (n - i + 1)(X_{(i)} - X_{(i-1)}) \quad i = 1, \ldots, n \]

- Relationship with $U(0,1)$ spacings

\[ U_{(k)} = \sum_{i=1}^{k} \frac{Y_i}{(\bar{X} n)}, \quad k = 1, \ldots, n - 1 \]
• Tests of spacings are typically of the form

\[ \sum_{i=1}^{n} g_n(D_i) \]

\[ D_i = U(i) - U(i-1), \quad i = 1, 2, \ldots, n + 1, \]

\[ U(0) \equiv 0, \quad U(n+1) \equiv 1. \]

Pike (1965),
Sethuraman and Jammalamadaka (1970, 1975),
A different spacings test

- $F_n(t)$ and $G_n(t)$ the EDF of $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ respectively

$$T_{1,n} = \sqrt{\frac{n}{2}} \sup_{0 \leq t < \infty} |F_n(t) - G_n(t)|$$

$$T_{2,n} = \frac{n}{2\bar{X}} \int (F_n(t) - G_n(t))^2 e^{-t/\bar{X}} \, dt,$$
Some asymptotics

- $F_n(t)$ and $G_n(t)$ are not independent

\[
F_n(t) = \frac{1}{n} \sum_{i=1}^{n} I\{X_i \leq t\} = \frac{1}{n} \sum_{i=1}^{n} I\{X_{(i)} \leq t\} = \frac{1}{n} \sum_{i=1}^{n} I\{\Sigma_{j=1}^{i} D_{j} \leq t\}
\]

\[
G_n(t) = \frac{1}{n} \sum_{i=1}^{n} I\{Y_i \leq t\} = \frac{1}{n} \sum_{i=1}^{n} I\{(n-i+1)D_{i} \leq t\}.
\]
For $0 \leq s \leq 1$, we define the two processes

\[ \alpha_n(s) \equiv \sqrt{n}(F_n(F_E^{-1}(s)) - s) \]

\[ \beta_n(s) \equiv \sqrt{n}(G_n(F_E^{-1}(s)) - s), \]

where $F_E$ indicates the distribution function of an $E(\theta)$

Then

\[ \sqrt{n}(F_n(t) - G_n(t)) = [\alpha_n(F_E(t)) - \beta_n(F_E(t))] \]

$T_{1,n}$ and $T_{2,n}$ are distribution-free under the null
Consistency

under exponentiality

\[ \sup_{0 \leq t < \infty} |F_n(t) - F_E(t)| \xrightarrow{\text{a.s.}} 0 \quad \text{and} \quad \sup_{0 \leq t < \infty} |G_n(t) - F_E(t)| \xrightarrow{\text{a.s.}} 0, \]

under the alternative \( F_A \), following the results in Pyke (1965),

\[ F_n(t) - G_n(t) \xrightarrow{\text{a.s.}} F_A(t) - 1 + \int_0^\infty f_A(y) \exp\{-th_A(y)\} \, dy \]

where \( f_A(t) \) denotes the density and

\[ h_A(t) = \frac{f_A(t)}{1 - F_A(t)}, \]
Asymptotic Null Distribution

For $0 \leq s, u \leq 1$, let

$$G^\alpha_K(s) = \int_0^1 [K(u, s) - sK(u, 1)] \, dF^{-1}(u)$$

$$G^\beta_K(s) = K(s, 1)$$

where $K(\cdot, \cdot)$ denotes a Kiefer process.

$(G^\alpha_K(s), G^\beta_K(s))$ is Gaussian

$$\text{cov}(G^\alpha_K(s), G^\beta_K(u)) = 0, \quad \text{(Barbe, 1994).}$$
THEOREM 1  Under exponentiality one can construct the sequences \( \alpha_n(s) \) and \( \beta_n(s) \) and a sequence of Kiefer processes \( K_n(\cdot, \cdot) \) on the same probability space, such that

\[
\sup_{0 \leq s \leq 1} |\alpha_n(s) - G_{K_n}^{\alpha}(s)| = O_p(n^{-1/2} \phi_n \log^2 n)
\]

\[
\sup_{0 \leq s \leq 1} |\beta_n(s) - G_{K_n}^{\beta}(s)| = O_p(n^{-1/2} \log^2 n).
\]

where

\[
\phi_n = \left| F^{-1} \left( \frac{1}{n} \right) \right| \vee \left| F^{-1} \left( 1 - \frac{1}{n} \right) \right|
\]
**Corollary** Under exponentiality the process $\sqrt{(1/2)}[\alpha_n(s) - \beta_n(s)]$
converges weakly to a Brownian Bridge $B(s)$, $0 \leq s \leq 1$.

**Theorem 2** Under the null hypothesis of exponentiality,

$$T_{1,n} \xrightarrow{\mathcal{D}} \sup_{0 < s < 1} |B(s)|,$$

$$T_{2,n} \xrightarrow{\mathcal{D}} \int_0^1 |B(s)|^2 \, ds,$$

where $B(s)$, $0 \leq s \leq 1$ denotes a Brownian Bridge.
Approximate Bahadur efficiency
denote,

\[ F_A(t) - 1 + \int_0^\infty f_A(y) \exp\{-th_A(y)\} \, dy = H(t, \theta) \]

under some regularity conditions on \( F(t, \theta) \) and \( f(t, \theta) \)

\[ H(t, \theta) = \left[ F'(t, 0) - te^{-t} \int_0^\infty F'(y, 0) \, dy \right] \theta + O(\theta^2). \]
Let $c_{T_1}(\theta)$ and $c_{T_2}(\theta)$ denote the approximate Bahadur slopes of $T_{1,n}$ and $T_{2,n}$, respectively.

$$c_{T_1}(\theta) = \left[ \sup_{0 \leq t < \infty} |H(t, \theta)| \right]^2$$
$$c_{T_2}(\theta) = \frac{\pi^2}{4} \left[ \int_0^\infty H(t, \theta)^2 e^{-t} \, dt \right].$$

<table>
<thead>
<tr>
<th></th>
<th>$e^B_{T_1,KS}$</th>
<th>$e^B_{T_2,\omega^2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear Failure</td>
<td>0.18</td>
<td>0.25</td>
</tr>
<tr>
<td>Makeham</td>
<td>0.07</td>
<td>0.08</td>
</tr>
<tr>
<td>Weibull</td>
<td>0.45</td>
<td>0.44</td>
</tr>
</tbody>
</table>
**Power comparisons**

**TABLE I  Goodness of Fit Tests Studied.**

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition/characterization</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_{S_n}$</td>
<td>Two sided Kolmogorov–Smirnov test with estimated mean</td>
<td>Durbin (1975)</td>
</tr>
<tr>
<td>$\omega_n^2$</td>
<td>Cramer–von Mises statistic with estimated mean</td>
<td>Darling (1957)</td>
</tr>
<tr>
<td>$A_{1,n}$</td>
<td>KS type based on loss of memory functional equation</td>
<td>Angus (1982)</td>
</tr>
<tr>
<td>$A_{2,n}$</td>
<td>CVM type based on loss of memory functional equation</td>
<td>Angus (1982)</td>
</tr>
<tr>
<td>$H_{1,n}$</td>
<td>KS type based on mean residual life function</td>
<td>Baringhaus and Henze (2000)</td>
</tr>
<tr>
<td>$H_{2,n}$</td>
<td>Cramer–von Mises type based on mean residual life</td>
<td>Baringhaus and Henze (2000)</td>
</tr>
<tr>
<td>$V_{m,n}$</td>
<td>Based on Vasicek’s (1976) entropy estimator</td>
<td>Ebrahimi et al. (1992)</td>
</tr>
<tr>
<td>$G_n$</td>
<td>Based on Gini’s index</td>
<td>Gail and Gastwirth (1978)</td>
</tr>
</tbody>
</table>
TABLE III  Monte Carlo Power Estimates Based on 10,000 Samples of Size 20.

<table>
<thead>
<tr>
<th>Distribution</th>
<th>$T_{1,n}$</th>
<th>KS$_n$</th>
<th>$A_{1,n}$</th>
<th>$H_{1,n}$</th>
<th>$V_{4,n}$</th>
<th>G$_n$</th>
<th>$T_{2,n}$</th>
<th>$\omega_{n}^2$</th>
<th>A$_{2,n}$</th>
<th>H$_{2,n}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pareto(2.2)</td>
<td>87</td>
<td>93</td>
<td>99</td>
<td>83</td>
<td>99</td>
<td>44</td>
<td>67</td>
<td>91</td>
<td>97</td>
<td>74</td>
</tr>
<tr>
<td>Pareto(2.5)</td>
<td>97</td>
<td>98</td>
<td>99</td>
<td>95</td>
<td>99</td>
<td>53</td>
<td>91</td>
<td>98</td>
<td>99</td>
<td>86</td>
</tr>
<tr>
<td>Weibull(0.8)</td>
<td>04</td>
<td>17</td>
<td>02</td>
<td>13</td>
<td>04</td>
<td>24</td>
<td>04</td>
<td>20</td>
<td>04</td>
<td>21</td>
</tr>
<tr>
<td>Weibull(1.4)</td>
<td>27</td>
<td>29</td>
<td>25</td>
<td>36</td>
<td>37</td>
<td>36</td>
<td>26</td>
<td>35</td>
<td>29</td>
<td>37</td>
</tr>
<tr>
<td>Lognormal(0.6)</td>
<td>73</td>
<td>84</td>
<td>81</td>
<td>84</td>
<td>90</td>
<td>80</td>
<td>74</td>
<td>89</td>
<td>84</td>
<td>84</td>
</tr>
<tr>
<td>Lognormal(0.8)</td>
<td>28</td>
<td>30</td>
<td>42</td>
<td>29</td>
<td>42</td>
<td>24</td>
<td>26</td>
<td>34</td>
<td>38</td>
<td>27</td>
</tr>
<tr>
<td>Shifted exp(0.2)</td>
<td>28</td>
<td>25</td>
<td>42</td>
<td>23</td>
<td>59</td>
<td>22</td>
<td>25</td>
<td>28</td>
<td>34</td>
<td>23</td>
</tr>
<tr>
<td>LFR(0.5)</td>
<td>12</td>
<td>10</td>
<td>09</td>
<td>14</td>
<td>13</td>
<td>10</td>
<td>12</td>
<td>11</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>LFR(3.0)</td>
<td>30</td>
<td>30</td>
<td>23</td>
<td>39</td>
<td>36</td>
<td>38</td>
<td>30</td>
<td>37</td>
<td>29</td>
<td>38</td>
</tr>
<tr>
<td>Dhillon(0.7)</td>
<td>06</td>
<td>06</td>
<td>03</td>
<td>06</td>
<td>05</td>
<td>05</td>
<td>06</td>
<td>06</td>
<td>05</td>
<td>05</td>
</tr>
<tr>
<td>Dhillon(0.9)</td>
<td>17</td>
<td>15</td>
<td>10</td>
<td>21</td>
<td>20</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>14</td>
<td>18</td>
</tr>
</tbody>
</table>

Note: Significance level $\alpha = 0.05$. 
Test based on Gini’s index

• If $g(x) = x$, the spacings test is equivalent to

$$G = \sum_{j,k=1}^{n} \frac{|X_j - X_k|}{(2n(n - 1)\bar{X})}$$

Gail and Gastwirth (1978).
References


• Taufer, Emanuele (2009). Wilcoxon-signed rank test for long memory sequences. Communications in Statistic, Theory Methods 38 no. 16-17, 3240–3248.