

Article

Testing for the Rayleigh Distribution: A New Test with Comparisons to Tests for Exponentiality Based on Transformed Data

Gerrit Lodewicus Grobler, Elzanie Bothma *  and James Samuel Allison

School of Mathematical and Statistical Sciences, Faculty of Natural and Agricultural Sciences, North-West University, Potchefstroom 2531, South Africa; gerrit.grobler@nwu.ac.za (G.L.G.); james.allison@nwu.ac.za (J.S.A.)

* Correspondence: elzanie.bothma@nwu.ac.za

Abstract: We propose a new goodness-of-fit test for the Rayleigh distribution which is based on a distributional fixed-point property of the Stein characterization. The limiting null distribution of the test is derived and the consistency against fixed alternatives is also shown. The results of a finite-sample comparison is presented, where we compare the power performance of the new test to a variety of other tests. In addition to existing tests for the Rayleigh distribution we also exploit the link between the exponential and Rayleigh distributions. This allows us to include some powerful tests developed specifically for the exponential distribution in the comparison. It is found that the new test outperforms competing tests for many of the alternative distributions. Interestingly, the highest estimated power, against all alternative distributions considered, is obtained by one of the tests specifically developed for the Rayleigh distribution and not by any of the exponentiality tests based on the transformed data. The use of the new test is illustrated on a real-world COVID-19 data set.

Keywords: asymptotics; goodness-of-fit; Monte Carlo simulation; Rayleigh distribution; Stein characterization

MSC: 62F03; 62F05



Citation: Grobler, G.L.; Bothma, E.; Allison, J.S. Testing for the Rayleigh Distribution: A New Test with Comparisons to Tests for Exponentiality Based on Transformed Data. *Mathematics* **2022**, *10*, 1316. <https://doi.org/10.3390/math10081316>

Academic Editors: Alexandru Agapie, Denis Enachescu, Vlad Stefan Barbu and Bogdan Iftimie

Received: 15 March 2022

Accepted: 7 April 2022

Published: 15 April 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In 1880 an acoustics problem gave rise to a distribution that nowadays plays a prominent role in research fields such as reliability theory, life testing and survival analysis (see, e.g., [1]). The Rayleigh distribution was introduced by [2], while undertaking a study regarding the resultant of a great number of sound waves with differing phases. Refs. [3,4] demonstrated the importance of the Rayleigh distribution in communication engineering and electro-vacuum devices, respectively. Ref. [5] found that the Rayleigh distribution has clinical applications, specifically estimating the noise variance of Magnetic Resonance Images (MRI). Ref. [6] discusses this phenomenon and proposed that this estimation can be done by fitting the density function of the Rayleigh distribution to the partial histogram of the MRI. Ref. [7] improved this estimation with the use of background segmentation, by fitting the density function of the Rayleigh distribution to the histogram of the segmented background in order to estimate the noise variance. The estimation of the noise forms a crucial part in efficiently denoising the MRI as well as in the quality assessment of these images. The Rayleigh distribution has also become a popular model in survival analysis and reliability theory, see, e.g., [8,9].

For any of the above-mentioned applications to be relevant, it is crucial to test the hypothesis that the observed data are indeed realisations from a Rayleigh distribution. Since the square of a Rayleigh distributed variable is exponentially distributed, goodness-of-fit tests designed for the exponential distribution can be used to test for the Rayleigh

distribution—a fact that we investigate further in Section 4. However, even though the applications of the Rayleigh distribution increased significantly over the past few decades, literature on tests specifically developed for the Rayleigh distribution is relatively scarce. Some of these include a test proposed by [10] based on the empirical Laplace transform, a test based on entropy suggested by [11] as well as [12] and an empirical likelihood based test by [13]. It has become a common approach to use distributional characterizations to propose goodness-of-fit testing procedures, see, e.g., Ref. [14] and the references therein. In this paper, we propose a new test for the Rayleigh distribution based on a modification of Stein’s characterization discussed by [15].

The standard Stein characterization (see [16]) of the normal distribution states that Z is standard normal if, and only if,

$$\mathbb{E}[g'(Z) - Zg(Z)] = 0 \tag{1}$$

is true for all absolute continuous functions g for which the expectation exists. Some applications, such as goodness-of-fit tests based on (1), are rather complicated, since the results depend on the choice of g . Instead of using this relationship, Ref. [17] characterised the standard normal distribution based on the zero bias distribution. A real valued random variable X^* is said to have a X zero-bias distribution if

$$\mathbb{E}[g'(X^*)] = \mathbb{E}[Xg(X)]$$

holds for all absolutely continuous functions g for which the expectation exists. If $\mathbb{E}X = 0$ and $\text{Var}(X) = 1$, the X zero-bias distribution exists, is unique and has distribution function

$$F(t) = \mathbb{E}[X(X - t)\mathbb{1}\{X \leq t\}], t \in \mathbb{R}.$$

Using this distribution function, ref. [17] showed that Z is standard normal if, and only if, the distribution function of Z is given by $F(t)$. Ref. [15] generalised this method to a wide range of continuous distributions by generalising Stein’s characterization. They showed that if X has support $[0, \infty)$, then it has distribution F if, and only if, the distribution function of X is given by

$$F(t) = \mathbb{E}\left[-\frac{f'(X)}{f(X)} \min\{X, t\}\right], t \in (0, \infty), \tag{2}$$

where f is the density of X . The result in (2) is true under some regularity conditions on f , which will be discussed in Section 2. The characterization in (2) will be used to develop a new goodness-of-fit test specifically for the Rayleigh distribution.

Before proceeding some notation is introduced. Let X_1, \dots, X_n be independent and identically distributed (i.i.d.) continuous realisations of a positive random variable X with unknown distribution function F and density f . If X follows a Rayleigh distribution with density function

$$f(x) = \frac{x}{\theta^2} e^{-\frac{x^2}{2\theta^2}}, x \geq 0, \theta > 0,$$

it will be denoted by $X \sim \text{Ral}(\theta)$. The composite goodness-of-fit hypothesis to be tested is

$$H_0 : \text{the distribution of } X \text{ is } \text{Ral}(\theta), \tag{3}$$

for some $\theta > 0$, against general alternatives.

The remainder of the article is organised as follows: In Section 2, the new test statistic is introduced. Section 3 contains the basic theoretical results pertaining to the asymptotic behaviour of the test. The results of a Monte Carlo study, where the power performance of the newly proposed test is compared to some existing tests, is given in Section 4. The competing tests also include five powerful tests for exponentiality based on transformed data.

The paper concludes in Section 5 with an application of the test to a real-world COVID-19 data set and some concluding remarks.

2. Test Statistic

For the characterization in (2) to be true the following regularity conditions, see [15], should hold:

- (I) f is continuously differentiable on $[0, \infty)$;
- (II) $f(x) > 0$ for every $x \in [0, \infty)$;
- (III) for $\kappa_f(x) = \left| \frac{f'(x) \min\{F(x), 1-F(x)\}}{f^2(x)} \right|$ we have $\sup_{x \in [0, \infty)} \kappa_f(x) < \infty$;
- (IV) $\int_0^\infty (1 + |x|) |f'(x)| dx < \infty$;
- (V) $\lim_{x \rightarrow 0} \frac{F(x)}{f(x)} = 0$;
- (VI) $\lim_{x \rightarrow \infty} \frac{1-F(x)}{f(x)} = 0$.

It can easily be seen that conditions (I), (II), (V) and (VI) hold for the Rayleigh distribution. For $X \sim \text{Ral}(\theta)$, κ_f in condition (III) becomes

$$\kappa_f(x) = \frac{\theta^2 e^{x^2/2\theta^2}}{x^2} \min\{F(x), 1 - F(x)\} \left| 1 - \frac{x^2}{\theta^2} \right|$$

and for $x^2 > \theta^2$; $\left| 1 - \frac{x^2}{\theta^2} \right| = \frac{x^2}{\theta^2} - 1$. For x large enough we have that $1 - F(x) < F(x)$; thus,

$$\lim_{x \rightarrow \infty} \kappa_f(x) = \theta^2 \lim_{x \rightarrow \infty} \left(\frac{\exp(x^2/2\theta^2)}{x^2} \right) (1 - F(x)) \left(\frac{x^2}{\theta^2} - 1 \right) = 1$$

Because x is sufficiently small, we have that $F(x) < 1 - F(x)$; thus, we have

$$\lim_{x \rightarrow 0} \kappa_f(x) = \theta^2 \lim_{x \rightarrow 0} \exp(-x^2/2\theta^2) \left(1 - \frac{x^2}{\theta^2} \right) \frac{(1 - e^{x^2/2\theta^2})}{x^2} = \frac{1}{2\theta^2} \theta^2 = \frac{1}{2}.$$

Since $\kappa_f(x)$ is continuous with limits 1 and $\frac{1}{2}$ as x tends to infinity and zero, respectively, it implies that $\sup_{x \in [0, \infty)} \kappa_f(x) < \infty$.

The integral in condition (IV) can be written as follows in terms of expectations:

$$\int_0^\infty (1 + |x|) \left(1 - \frac{x^2}{\theta^2} \right) \left(\frac{1}{\theta^2} \right) e^{-x^2/2\theta^2} dx = \mathbb{E} \left(\{1 + X\} \left\{ 1 - \frac{X^2}{\theta^2} \right\} \right),$$

where X is Rayleigh distributed. The finite moments of the Rayleigh distribution exist, i.e., $\mathbb{E}(X^k) < \infty$, $k \in \mathbb{N}$. Therefore,

$$\int_0^\infty (1 + |x|) |f'(x)| dx < \infty.$$

In Proposition 1 below, the characterization in (2) is re-stated specifically for the Rayleigh distribution.

Proposition 1. Let $X : \Omega \rightarrow (0, \infty)$ be a random variable with distribution function F and density function f that satisfies conditions (I)–(VI) and $\mathbb{E}[X] < \infty$. Then $X \sim \text{Ral}(\theta)$ if, and only if,

$$\mathbb{E} \left[\left(\frac{X}{\theta^2} - \frac{1}{X} \right) \min\{X, t\} \right] - F(t) = 0, t > 0.$$

Note that $X \sim \text{Ral}(\theta)$ if, and only if, $Y = \frac{X}{\theta} \sim \text{Ral}(1)$. This follows from the invariance property of the Rayleigh distribution with respect to scale transformations. This implies that $Y \sim \text{Ral}(1)$ if, and only if, for all $t > 0$

$$\psi(t) = T^Y(t) - F^Y(t) = 0, \tag{4}$$

where $T^Y(t) = \mathbb{E}[(Y - 1/Y) \min(Y, t)]$ and F^Y is the distribution function of Y . Our newly proposed test is motivated by (4). Since $\psi(t)$ will be unknown, we estimate it by its empirical counterpart,

$$\hat{\psi}_n(t) = T_n^Y(t) - F_n^Y(t),$$

where $T_n^Y(t) = \frac{1}{n} \sum_{j=1}^n (Y_j - 1/Y_j) \min(Y_j, t)$, $F_n^Y(t) = \frac{1}{n} \sum_{j=1}^n I(Y_j \leq t)$ and $Y_j = X_j/\hat{\theta}_n$, with $\hat{\theta}_n = \sqrt{(2n)^{-1} \sum_{j=1}^n X_j^2}$ the maximum likelihood estimator for θ .

We propose the following weighted L^2 -distance between $\hat{\psi}(t)$ and 0 to test the hypothesis in (3):

$$R_{n,a} = n \int_0^\infty \hat{\psi}_n^2(t) w_a(t) dt, \tag{5}$$

where $w_a(t)$ is a positive, continuous weight function depending on a positive tuning parameter a . The test rejects for large values of $R_{n,a}$. Throughout the paper we use $w_a(t) = e^{-at}$ as the weight function, which results in the following easily calculable form of the test statistic:

$$\begin{aligned} R_{n,a} = & \frac{1}{n} \sum_{j=1}^n \left(-\frac{1}{a} e^{-aY(j)} \left[\left\{ Y(j) - \frac{1}{Y(j)} \right\}^2 \left\{ \frac{2}{a} Y(j) + \frac{2}{a^2} \right\} + 2Y(j)^2 - 3 \right] + \frac{2}{a^3} \left[Y(j)^2 - 2 + \frac{1}{Y(j)^2} \right] \right) \\ & + \frac{2}{n} \sum_{1 \leq j < k \leq n} \left(\left\{ Y(j) - \frac{1}{Y(j)} \right\} \left\{ Y(k) - \frac{1}{Y(k)} \right\} \left(-\frac{1}{a} e^{-aY(j)} \left\{ \frac{1}{a} Y(j) + \frac{2}{a^2} \right\} + \frac{2}{a^3} - \frac{Y(j)}{a^2} e^{-aY(k)} \right) \right. \\ & \quad \left. + \left\{ Y(j) - \frac{1}{Y(j)} \right\} \left\{ -\frac{Y(j)}{a} e^{-aY(k)} \right\} \right. \\ & \quad \left. + \left\{ Y(k) - \frac{1}{Y(k)} \right\} \left\{ \frac{1}{a^2} e^{-aY(k)} - \frac{1}{a} e^{-aY(j)} \left(Y(j) + \frac{1}{a} \right) \right\} + \frac{1}{a} e^{-aY(k)} \right), \end{aligned}$$

where $Y_{(1)} < Y_{(2)} < \dots < Y_{(n)}$ denotes the order statistics of Y_1, \dots, Y_n .

Remark 1. The most commonly used choices for the weight function $w_a(\cdot)$ are $w_a(t) = e^{-a|t|}$ and $w_a(t) = e^{-at^2}$ (see, e.g., [18,19]). Due to the positive support of the Rayleigh distribution, we use $w_a(t) = e^{-a|t|} = e^{-at}$, $t \geq 0$. This choice does not only provide a close form expression for the test statistic, but also competitive powers which are reported in the Monte Carlo simulation study (see Section 4).

3. Asymptotics

In this section, we will first show that, under the null hypothesis, $R_{n,a}$ converges in distribution to a norm of a Gaussian element of the Hilbert space $\mathcal{H} = L^2((0, \infty), \mathcal{B})$ of measurable, square integrable functions. The norm $\|\cdot\|_{\mathcal{H}}$ that will be used is defined in terms of a random element G_n of \mathcal{H} , $n \in \mathbb{N}$, by

$$\|G_n\|_{\mathcal{H}} = \left(\int_0^\infty \{G_n(t)\}^2 e^{-at} dt \right)^{\frac{1}{2}}.$$

We will also show that the newly proposed test is consistent.

First note that by substituting t with $\frac{s}{\hat{\theta}_n}$ and Y_j with $\frac{X_j}{\hat{\theta}_n}$ in (5) the test statistic $R_{n,a}$ can be rewritten as

$$R_{n,a} = \frac{1}{\hat{\theta}_n} \int_0^\infty \left(\sqrt{n} \left\{ \hat{T}_n^X(s) - F_n^X(s) \right\} \right)^2 e^{-as/\hat{\theta}_n} ds, \tag{6}$$

where

$$\hat{T}_n^X(s) = \frac{1}{n\hat{\theta}_n^2} \sum_{j=1}^n \left(X_j - \frac{\hat{\theta}_n^2}{X_j} \right) \min\{X_j, s\} \tag{7}$$

is a continuous function.

To obtain our two main results, we use the following Lemma, in which the notation $G_n \approx H_n$ is used when $\|G_n - H_n\|_{\mathcal{H}}^2 = o_{\mathbb{P}}(1)$, where $o_{\mathbb{P}}(1)$ denotes a sequence of random variables that converge to zero in probability. We also assume, w.l.o.g., that $\theta = 1$.

Lemma 1. *Suppose X, X_1, X_2, \dots are i.i.d. random variables with distribution function F^X and $\mathbb{E}[X^4] < \infty$. Let $\hat{T}_n^X(s)$ be defined as in (7), then*

$$\hat{T}_n^X(s) = \frac{1}{\hat{\theta}_n^2} \left\{ T_n^X(s) + (1 - \hat{\theta}_n^2) r_n^X(s) \right\},$$

where

$$r_n^X(s) = \frac{1}{n} \sum_{j=1}^n \frac{1}{X_j} \min\{X_j, s\},$$

and

$$T_n^X(s) = \frac{1}{n} \sum_{j=1}^n \left(X_j - \frac{1}{X_j} \right) \min(X_j, s).$$

We also have that

$$\sqrt{n} \hat{T}_n^X(s) \approx \frac{\sqrt{n}}{\hat{\theta}_n^2} \left\{ T_n^X(s) + (1 - \hat{\theta}_n^2) r^X(s) \right\},$$

where

$$r^X(s) = \mathbb{E} \left[\frac{1}{X} \min\{X, s\} \right].$$

Proof. The first result follows immediately by rewriting $\hat{T}_n^X(s)$ in (7) as

$$\hat{T}_n^X(s) = \frac{1}{n\hat{\theta}_n^2} \sum_{j=1}^n \left[\left(X_j - \frac{1}{X_j} \right) \min\{X_j, s\} + \left(\frac{1}{X_j} - \frac{\hat{\theta}_n^2}{X_j} \right) \min\{X_j, s\} \right].$$

To show the second result we notice that

$$\sqrt{n} \left\{ \hat{T}_n^X(s) - \frac{1}{\hat{\theta}_n^2} \left[T_n^X(s) + (1 - \hat{\theta}_n^2) r^X(s) \right] \right\} = \frac{\sqrt{n}(1 - \hat{\theta}_n^2)}{\hat{\theta}_n^2} \left\{ r_n^X(s) - r^X(s) \right\}.$$

Applying a weak form of the law of large numbers in separable Hilbert spaces, we have that $r_n^X(s) = r^X(s) + o_{\mathbb{P}}(1)$ and by the continuous mapping theorem $\|r_n^X - r^X\|_{\mathcal{H}}^2 = o_{\mathbb{P}}(1)$. Since $\hat{\theta}_n^2$ is the maximum likelihood estimator of θ^2 , we have that $\sqrt{n}(1 - \hat{\theta}_n^2) = \mathcal{O}_{\mathbb{P}}(1)$, where $\mathcal{O}_{\mathbb{P}}(1)$ denotes a sequence of random variables that is bounded in probability. The result then follows from Slutsky’s theorem. \square

Theorem 1. *Let X, X_1, X_2, \dots be i.i.d. standard Rayleigh random variables. There exists a centred Gaussian element \mathcal{W} of \mathcal{H} such that $R_{n,a} \xrightarrow{\mathcal{D}} \|\mathcal{W}\|_{\mathcal{H}}^2$, where the covariance kernel of \mathcal{W} is given by*

$$\begin{aligned}
 K(s, t) &= \text{Cov}[W_j(s), W_j(t)] \\
 &= F^X(s \wedge t) + (s \wedge t)[I_3(s \wedge t, s \vee t) - 2I_1(s \wedge t, s \vee t) + I_{-1}(s \wedge t, s \vee t)] \\
 &\quad + st[2F^X(s \vee t) - 2 + I_2(s \vee t, \infty) + I_{-2}(s \vee t, \infty)] + I_4(0, s \wedge t) - 2I_2(0, s \wedge t) \\
 &\quad + r^X(s)r^X(t) + 2F^X(s)F^X(t) \\
 &\quad + \left\{ -\frac{1}{2}I_4(0, s) + \frac{3}{2}I_2(0, s) + s \left[-\frac{1}{2}I_3(s, \infty) + \frac{3}{2}I_1(s, \infty) + I_{-1}(s, \infty) \right] \right\} r^X(t) \\
 &\quad + \left\{ -\frac{1}{2}I_4(0, t) + \frac{3}{2}I_2(0, t) + t \left[-\frac{1}{2}I_3(t, \infty) + \frac{3}{2}I_1(t, \infty) + I_{-1}(t, \infty) \right] \right\} r^X(s) \\
 &\quad - \frac{1}{2} \{ I_4(0, s) - I_2(0, s) + s[I_3(s, \infty) + I_1(s, \infty)] \} F_X(t) \\
 &\quad - \frac{1}{2} \{ I_4(0, t) - I_2(0, t) + t[I_3(t, \infty) + I_1(t, \infty)] \} F_X(s),
 \end{aligned}$$

where

$$I_k(a, b) = \mathbb{E}[X_j^k \mathbb{1}(a \leq X_j \leq b)]$$

and

$$I_k(a, \infty) = \lim_{b \rightarrow \infty} I_k(a, b).$$

Proof. First note that

$$\sqrt{n} \{ \hat{T}_n^X(s) - F_n^X(s) \} \approx \frac{\sqrt{n}}{\hat{\theta}_n^2} \left\{ T_n^X(s) + \left(1 - \hat{\theta}_n^2 \right) r^X(s) - \hat{\theta}_n^2 F^X(s) \right\},$$

since $\|F^X - F_n^X\|_{\mathcal{H}}^2 = o_{\mathbb{P}}(1)$. We can therefore write

$$\sqrt{n} \{ \hat{T}_n^X(s) - F_n^X(s) \} \approx \frac{1}{\sqrt{n\hat{\theta}_n^2}} \sum_{j=1}^n W_j(s),$$

where

$$W_j(s) = \left(X_j - \frac{1}{X_j} \right) \min\{X_j, s\} + \left(1 - \frac{1}{2}X_j^2 \right) r^X(s) - \frac{1}{2}X_j^2 F^X(s).$$

We note that W_1, \dots, W_n are i.i.d. random variables with $E(W_1) = 0$ and $E\|W_1\|_{\mathcal{H}}^2 < \infty$. Therefore, by the central limit theorem for separable Hilbert spaces (see [20]) there exists a centred Gaussian element $\mathcal{W} \in \mathcal{H}$ with

$$\frac{1}{\sqrt{n}} \sum_{j=1}^n W_j(\cdot) \xrightarrow{\mathcal{D}} \mathcal{W}(\cdot).$$

From this we have that $\sqrt{n} \{ \hat{T}_n^X(s) - F_n^X(s) \} = \mathcal{O}_{\mathbb{P}}(1)$. Therefore, since $\hat{\theta}_n = 1 + o_{\mathbb{P}}(1)$ and by Holder’s inequality we have that

$$\begin{aligned}
 &\left| \int_0^\infty \left(\sqrt{n} \{ \hat{T}_n^X(s) - F_n^X(s) \} \right)^2 e^{-as/\hat{\theta}_n} ds - \int_0^\infty \left(\sqrt{n} \{ \hat{T}_n^X(s) - F_n^X(s) \} \right)^2 e^{-as} ds \right| \\
 &\leq \sup_{s>0} \left| e^{-as\left(\frac{1}{\hat{\theta}_n} - 1\right)} - 1 \right| \|\sqrt{n} \{ \hat{T}_n^X - F_n^X \}\|_{\mathcal{H}}^2 = o_{\mathbb{P}}(1).
 \end{aligned}$$

Therefore,

$$R_{n,a} = \|\sqrt{n} \{ \hat{T}_n^X - F_n^X \}\|_{\mathcal{H}}^2 + o_{\mathbb{P}}(1). \tag{8}$$

The final result then follows from Slutsky’s theorem. \square

Remark 2. A closed form expression for the covariance kernel for the limiting centred Gaussian distribution does not exist. However, for non-negative even numbers of k closed form formulas for functions $I_k(a, b)$ exist by using the following recursive formulas

$$I_0(a, b) = F^X(b) - F^X(a)$$

$$I_k(a, b) = a^k e^{-1/2a^2} - b^k e^{-1/2b^2} + kI_{k-2}(a, b).$$

Now that we have shown that, under the null hypothesis, $R_{n,a}$ converges in distribution to a norm of a Gaussian element of the Hilbert space \mathcal{H} , we can continue to show that the newly proposed test is consistent. Therefore, we will show that $\frac{R_{n,a}}{n} = \Delta + o_{\mathbb{P}}(1)$, where $\Delta = \|T^X - F^X\|_{\mathcal{H}}^2$ with the properties that $\Delta = 0$ under the null hypothesis and $\Delta > 0$ under fixed alternatives. This is as a result of the characterization of the Rayleigh distribution in Proposition 1.

Theorem 2. Suppose X, X_1, X_2, \dots are i.i.d. random variables with distribution function F^X and $\mathbb{E}[X^2] < \infty$. As $n \rightarrow \infty$, we have

$$\frac{R_{n,a}}{n} = \|T^X - F^X\|_{\mathcal{H}}^2 + o_{\mathbb{P}}(1).$$

Proof. From (8) we have that

$$\frac{R_{n,a}}{n} = \|\hat{T}_n^X - F_n^X\|_{\mathcal{H}}^2 + o_{\mathbb{P}}(1).$$

To prove the theorem we need to show that

$$\|\hat{T}_n^X - F_n^X\|_{\mathcal{H}}^2 = \|T^X - F^X\|_{\mathcal{H}}^2 + o_{\mathbb{P}}(1).$$

By a weak form of the law of large numbers for separable Hilbert spaces we have that $T_n^X(s) = T^X(s) + o_{\mathbb{P}}(1)$ and $F_n^X(s) = F^X(s) + o_{\mathbb{P}}(1)$. Moreover, from Lemma 1 we have that $\hat{T}_n^X(s) = T_n^X(s) + o_{\mathbb{P}}(1)$ and hence $\hat{T}_n^X(s) = T^X(s) + o_{\mathbb{P}}(1)$. We also have that

$$\hat{T}_n^X(s) - F_n^X(s) = \left(\hat{T}_n^X(s) - T^X(s)\right) + \left(T^X(s) - F^X(s)\right) + \left(F^X(s) - F_n^X(s)\right),$$

and by the continuous mapping theorem the result follows. \square

4. Simulation Study

In this section, Monte Carlo simulations are used to compare the finite sample performance of the newly proposed test to the following existing goodness-of-fit tests for the Rayleigh distribution:

- The traditional Kolmogorov–Smirnov (KS_n), Cramér–von Mises (CM_n) and Anderson–Darling (AD_n) tests;
- A test based on the empirical Laplace transform proposed by [10], $EL_{n,a}$;
- A test based on the cumulative residual entropy proposed by [11], CR_n , and
- A test based on an estimator of the Kullback–Leibler divergence proposed by [12], $KL_{n,a}$.

The estimated powers of $R_{n,a}$, $EL_{n,a}$ and $KL_{n,a}$ are functions of a tuning parameter, a . For $R_{n,a}$ we report the results for $a = 1$ and $a = 5$, for $EL_{n,a}$ a is 1 and 5 and for $KL_{n,a}$ results are reported for $a = 3$ and $a = 4$. The motivation for these choices of a will be discussed in Section 4.2.

In addition to the existing tests, we also compare the performance of the new test to the following five powerful tests for exponentiality (see, e.g., the overview papers by [21] as well as [22] for a discussion on a variety of tests for exponentiality);

- The modified Kolmogorov–Smirnov (\widetilde{KS}_n) and Cramér-von Mises (\widetilde{CM}_n) tests based on the mean residual life proposed by [23];
- Two tests based on the empirical Laplace transform; one proposed by [24], $BH_{n,a}$, and the other one by [25], $HM_{n,a}$, and
- A test based on the empirical characteristic function proposed by [26], EP_n .

Here, we test for the Rayleigh distribution by testing for exponentiality of the transformed data (using the well known property that the square of a Rayleigh distributed random variable follows an exponential distribution). The estimated powers of $BH_{n,a}$ and $HM_{n,a}$ are functions of a tuning parameter, a . For both $BH_{n,a}$ and $HM_{n,a}$ we report the results for $a = 0.75, 1$ and 1.25 .

4.1. Simulation Setting

A significance level of 5% is used throughout. Critical values of all the tests are obtained using 50,000 independent Monte Carlo replications drawn from a standard Rayleigh distribution (all the test statistics are invariant with respect to scale transformations). Power estimates are calculated and reported for sample sizes $n = 20$ and $n = 30$ using 10,000 independent Monte Carlo replications obtained from various alternative distributions. These include some ‘local’ alternatives as well as those given in Table 1. These alternative distributions were chosen since they are frequently used alternatives for the Rayleigh distribution, which has an increasing hazard rate. The hazard rates of the considered alternative distributions include constant hazard rates (CHR), increasing hazard rates (IHR), decreasing hazard rates (DHR) and non-monotone hazard rates (NMHR). These alternatives all have support in \mathbb{R}^+ and are used in many other empirical studies for goodness-of-fit tests of lifetime distributions (see, e.g., [10,21,27]). In Table 1, all scale parameters are set to one due to the scale transformation $Y_j = X_j/\hat{\theta}_n, j = 1, \dots, n$. All simulations and calculations are done in Ref. [28]. The tables are produced using the *Stargazer* package, see [29].

Table 1. Probability density functions of the alternative distributions considered in the Monte Carlo study.

Alternative	f(x)	Notation
Gamma	$\frac{1}{\Gamma(\theta)} x^{\theta-1} \exp(-x)$	$\Gamma(\theta)$
Weibull	$\theta x^{\theta-1} \exp(-x^\theta)$	$W(\theta)$
Power	$\frac{1}{\theta} x^{(1-\theta)/\theta}, 0 < x < 1$	$PW(\theta)$
Linear Failure Rate	$(1 + \theta x) \exp\left(-x - \frac{\theta x^2}{2}\right)$	$LFR(\theta)$
Lognormal	$\exp\left\{-\frac{1}{2}\left(\frac{\log(x)}{\theta}\right)^2\right\} \left\{\theta x \sqrt{2\pi}\right\}^{-1}$	$LN(\theta)$
Inverse Gaussian	$\left(\frac{\theta}{2\pi x^3}\right)^{1/2} \exp\left\{-\frac{\theta(x-1)^2}{2x}\right\}$	$IG(\theta)$
Gompertz	$\exp(-\theta x) \exp\left\{-\left(\frac{1}{\theta}\right)(\exp(\theta x) - 1)\right\}$	$GO(\theta)$
Exponential	$\theta \exp(-\theta x)$	$Exp(\theta)$
Extreme value	$\frac{1}{\theta} \exp\left(x + \frac{1-\exp(x)}{\theta}\right)$	$EV(\theta)$
Exponential geometric	$(1 - \theta) \exp(-x)(1 - \theta \exp(-x))^{-2}$	$EG(\theta)$

We first consider some local power estimates. Here, we consider a mixture distribution, which is obtained by sampling with probability p from a standard exponential distribution ($Exp(1)$) and with probability $(1 - p)$ from a $Ral(1)$ distribution. The value $p = 0$ corresponds to the standard Rayleigh distribution, whereas increasing values of p implies a larger deviation from the null distribution. These estimated powers are given in Table 2 and the estimated powers for the exponentiality tests based on the transformed data are given in Table 3. The estimated powers for sample sizes 20 and 30 against every alternative distribution in Table 1 are given in Tables 4 and 5, respectively. The estimated powers, obtained using the tests for exponentiality based on the transformed data, for sample sizes 20 and 30 are given in Tables 6 and 7, respectively. The entries in these tables are the

percentages of 10,000 independent Monte Carlo samples that resulted in the rejection of the null hypothesis (rounded to the nearest integer). For the reader’s convenience, the highest estimated power for each alternative distribution among the existing tests, as well as the tests for exponentiality based on the square of the data, are displayed separately in bold in each of their respective tables. The last column of Tables 2, 4 and 5 contain the highest estimated powers from the corresponding exponentiality tests based on the transformed data (i.e., the highest powers obtained from Tables 3, 6 and 7 are also reported in the last column of Tables 2, 4 and 5); this will make comparison easier.

Table 2. Estimated local powers for the mixture of the Rayleigh and exponential distributions for various choices of the mixture parameter, p .

p	n	KS_n	CM_n	AD_n	$EL_{n,1}$	$EL_{n,5}$	CR_n	$KL_{n,3}$	$KL_{n,4}$	$R_{n,1}$	$R_{n,5}$	Exp
0	20	5	5	5	5	5	5	5	5	5	5	5
	30	5	4	5	4	5	4	5	5	6	5	5
0.05	20	6	6	7	8	7	7	6	5	7	9	7
	30	6	7	8	9	8	9	6	6	8	9	8
0.1	20	8	8	12	15	10	10	6	6	10	13	10
	30	8	9	14	16	13	13	8	7	13	16	12
0.15	20	9	11	17	20	15	14	8	8	14	20	14
	30	12	14	21	25	19	16	12	11	18	23	17
0.2	20	12	14	22	27	19	17	11	11	19	27	18
	30	15	17	27	34	25	21	14	15	25	34	23
0.25	20	15	18	29	34	25	21	14	14	24	34	25
	30	21	24	36	44	34	27	19	20	32	44	18
0.3	20	19	22	35	42	30	24	16	15	30	42	30
	30	26	30	44	54	41	32	25	27	41	53	42
0.35	20	22	27	42	50	37	30	21	22	37	50	36
	30	32	37	54	62	49	38	31	32	50	61	49
0.4	20	28	32	48	56	43	35	26	27	42	56	43
	30	41	45	62	70	58	44	39	38	57	70	57
0.45	20	34	39	56	64	52	41	32	32	49	62	51
	30	46	51	69	77	65	52	46	46	64	76	66
0.5	20	39	43	61	69	57	45	36	36	55	69	57
	30	51	59	76	82	72	57	52	53	71	83	72

Table 3. Estimated local powers for the mixture of the Rayleigh and exponential distributions, using transformed data, for various choices of the mixture parameter, p .

p	n	\widetilde{KS}_n	\widetilde{CV}_n	EP_n	$BH_{n,0.75}$	$BH_{n,1}$	$BH_{n,1.25}$	$HM_{n,0.75}$	$HM_{n,1}$	$HM_{n,1.25}$
0	20	5	5	5	5	5	5	5	5	5
	30	5	5	5	5	5	5	5	5	5
0.05	20	6	6	7	7	7	7	7	6	7
	30	6	7	7	8	7	8	7	7	7
0.1	20	7	9	10	10	10	10	10	9	9
	30	8	10	11	11	12	10	12	12	11
0.15	20	8	12	12	14	13	13	14	13	13
	30	10	15	15	17	16	17	17	16	16
0.2	20	10	15	16	18	17	17	18	18	17
	30	14	19	20	23	22	22	23	22	22

Table 3. Cont.

p	n	\widetilde{KS}_n	\widetilde{CV}_n	EP_n	$BH_{n,0.75}$	$BH_{n,1}$	$BH_{n,1.25}$	$HM_{n,0.75}$	$HM_{n,1}$	$HM_{n,1.25}$
0.25	20	14	18	20	23	22	22	25	23	22
	30	18	25	27	31	30	29	34	31	31
0.3	20	15	23	24	28	28	26	30	28	27
	30	23	31	34	40	37	36	42	39	38
0.35	20	19	29	30	35	33	33	36	34	34
	30	27	38	40	46	44	44	49	46	45
0.4	20	23	35	36	41	39	38	43	41	40
	30	36	46	49	55	54	52	57	55	55
0.45	20	29	41	43	49	46	46	51	48	48
	30	42	53	55	63	61	59	66	63	62
0.5	20	32	46	47	53	52	51	57	54	53
	30	47	60	62	70	67	67	72	70	69

Table 4. Estimated powers for general alternatives for the Rayleigh distribution for sample size $n = 20$.

F	KS_n	CM_n	AD_n	$EL_{n,1}$	$EL_{n,5}$	CR_n	$KL_{n,3}$	$KL_{n,4}$	$R_{n,1}$	$R_{n,5}$	Exp
CHR											
$Exp(1)$	86	90	96	97	94	89	84	85	94	97	95
IHR											
$\Gamma(1.5)$	57	63	73	75	73	64	44	44	72	77	72
$\Gamma(2)$	32	38	44	43	46	41	19	18	44	44	44
$W(1.2)$	64	69	80	84	79	68	54	55	79	85	79
$W(1.4)$	37	42	53	58	53	43	26	25	52	60	52
$PW(1)$	16	20	40	42	13	21	45	46	18	41	21
$LFR(2)$	38	45	63	70	56	42	36	36	56	69	59
$LFR(4)$	26	29	47	56	41	29	25	25	40	55	42
$EV(0.5)$	56	61	79	84	74	58	54	55	73	84	76
$EV(1.5)$	22	24	46	56	33	20	26	28	35	56	38
$GO(0.5)$	56	62	79	84	74	57	54	54	75	84	76
$GO(1.5)$	22	25	44	54	33	20	27	29	33	55	38
DHR											
$\Gamma(0.4)$	100	100	100	100	100	100	100	100	100	100	100
$\Gamma(0.7)$	97	98	100	100	99	97	98	98	99	100	99
$W(0.8)$	97	98	100	100	100	98	98	98	100	100	100
$EG(0.2)$	94	96	99	100	99	95	95	95	98	96	98
$EG(0.5)$	97	98	99	88	100	97	98	98	92	71	99
$EG(0.8)$	86	90	95	55	100	99	100	99	75	60	94
NMHR											
$PW(2)$	87	89	98	99	93	83	96	96	95	99	95
$PW(3)$	99	99	100	100	100	99	100	100	99	100	100
$LN(0.8)$	67	71	72	66	75	74	51	50	75	68	75
$LN(1)$	90	92	94	93	95	92	82	82	94	94	94
$LN(1.5)$	100	100	100	100	100	100	99	99	100	100	100
$IG(0.5)$	97	98	98	98	98	97	94	93	99	98	98
$IG(1.5)$	56	61	60	48	62	64	42	42	64	52	63

Table 5. Estimated powers for general alternatives for the Rayleigh distribution for sample size $n = 30$.

F	KS_n	CM_n	AD_n	$EL_{n,1}$	$EL_{n,5}$	CR_n	$KL_{n,3}$	$KL_{n,4}$	$R_{n,1}$	$R_{n,5}$	Exp
CHR											
$Exp(1)$	96	98	99	100	99	97	96	96	99	100	99
IHR											
$\Gamma(1.5)$	76	81	87	89	88	79	63	65	88	90	87
$\Gamma(2)$	47	53	57	56	59	54	28	28	60	57	59
$W(1.2)$	81	86	92	94	92	83	74	76	91	95	92
$W(1.4)$	53	58	69	72	69	56	37	39	68	74	68
$PW(1)$	22	28	55	51	14	37	66	68	24	50	27
$LFR(2)$	53	59	76	83	71	53	51	53	72	84	74
$LFR(4)$	36	43	61	69	55	37	37	37	54	68	56
$EV(0.5)$	74	80	90	94	88	72	71	73	88	94	89
$EV(1.5)$	32	35	58	68	42	23	39	41	46	69	51
$GO(0.5)$	74	78	91	94	88	72	73	74	88	94	89
$GO(1.5)$	33	37	58	68	44	25	41	42	47	68	51
DHR											
$\Gamma(0.4)$	100	100	100	100	100	100	100	100	100	100	100
$\Gamma(0.7)$	100	100	100	100	100	100	100	100	100	100	100
$W(0.8)$	100	100	100	100	100	100	100	100	100	100	100
$EG(0.2)$	99	99	100	100	100	99	100	99	99	96	100
$EG(0.5)$	100	100	100	87	100	100	100	100	93	68	100
$EG(0.8)$	96	98	99	52	100	100	99	99	75	59	99
NMHR											
$PW(2)$	97	98	100	100	98	94	100	100	99	100	99
$PW(3)$	100	100	100	100	100	100	100	100	99	100	100
$LN(0.8)$	83	86	87	81	88	87	70	71	89	82	88
$LN(1)$	97	98	99	99	99	98	94	95	99	99	99
$LN(1.5)$	100	100	100	100	100	100	100	100	100	100	100
$IG(0.5)$	100	100	100	100	100	100	99	99	100	100	100
$IG(1.5)$	73	77	77	62	78	79	62	62	80	67	80

Table 6. Estimated powers for general alternatives for the exponential distribution for sample size $n = 20$.

F	\widetilde{KS}_n	\widetilde{CV}_n	EP_n	$BH_{n,0.75}$	$BH_{n,1}$	$BH_{n,1.25}$	$HM_{n,0.75}$	$HM_{n,1}$	$HM_{n,1.25}$
CHR									
$Exp(1)$	82	90	90	94	93	93	95	94	93
IHR									
$\Gamma(1.5)$	52	66	68	72	71	70	72	71	71
$\Gamma(2)$	29	40	43	43	43	43	43	43	44
$W(1.2)$	57	70	72	77	77	76	79	78	76
$W(1.4)$	31	44	46	52	50	50	52	52	51
$PW(1)$	12	11	7	17	14	11	21	17	14
$LFR(2)$	30	43	46	55	53	52	59	57	54
$LFR(4)$	20	30	31	40	36	36	42	39	39
$EV(0.5)$	48	61	62	72	71	68	76	72	71
$EV(1.5)$	15	21	22	34	30	27	38	35	33
$GO(0.5)$	47	62	63	72	71	68	76	72	71
$GO(1.5)$	14	22	23	33	29	28	38	34	31

Table 6. Cont.

<i>F</i>	\widetilde{KS}_n	\widetilde{CV}_n	EP_n	$BH_{n,0.75}$	$BH_{n,1}$	$BH_{n,1.25}$	$HM_{n,0.75}$	$HM_{n,1}$	$HM_{n,1.25}$
DHR									
$\Gamma(0.4)$	100	100	100	100	100	100	100	100	100
$\Gamma(0.7)$	95	98	98	99	99	99	99	99	99
$W(0.8)$	96	98	98	99	99	99	100	100	99
$EG(0.2)$	91	95	95	98	98	97	98	98	98
$EG(0.5)$	95	98	98	99	99	99	99	99	99
$EG(0.8)$	82	90	90	93	93	93	94	94	93
NMHR									
$PW(2)$	75	84	79	93	91	89	95	93	92
$PW(3)$	97	99	98	100	100	99	100	100	100
$LN(0.8)$	64	73	74	74	75	75	72	73	73
$LN(1)$	88	93	93	94	94	94	94	94	94
$LN(1.5)$	99	100	100	100	100	100	100	100	100
$IG(0.5)$	96	98	98	98	98	98	98	98	98
$IG(1.5)$	53	64	64	62	63	63	58	59	61

Table 7. Estimated powers for general alternatives for the exponential distribution for sample size $n = 30$.

<i>F</i>	\widetilde{KS}_n	\widetilde{CV}_n	EP_n	$BH_{n,0.75}$	$BH_{n,1}$	$BH_{n,1.25}$	$HM_{n,0.75}$	$HM_{n,1}$	$HM_{n,1.25}$
CHR									
$Exp(1)$	95	98	98	99	99	99	99	99	99
IHR									
$\Gamma(1.5)$	73	82	84	87	86	85	87	86	87
$\Gamma(2)$	43	57	58	57	59	59	58	57	58
$W(1.2)$	78	87	87	91	90	90	92	91	91
$W(1.4)$	47	60	62	67	66	65	68	68	66
$PW(1)$	17	15	6	21	18	14	27	22	18
$LFR(2)$	44	58	60	68	67	66	74	70	68
$LFR(4)$	31	43	44	53	51	48	56	54	51
$EV(0.5)$	67	79	80	87	86	84	89	87	87
$EV(1.5)$	22	30	29	45	41	37	51	46	42
$GO(0.5)$	67	78	79	87	85	84	89	88	86
$GO(1.5)$	22	31	29	46	42	39	51	47	44
DHR									
$\Gamma(0.4)$	100	100	100	100	100	100	100	100	100
$\Gamma(0.7)$	100	100	100	100	100	100	100	100	100
$W(0.8)$	100	100	100	100	100	100	100	100	100
$EG(0.2)$	99	99	99	100	100	100	100	100	100
$EG(0.5)$	100	100	100	100	100	100	100	100	100
$EG(0.8)$	95	98	98	99	99	99	99	99	99
NMHR									
$PW(2)$	93	96	91	99	98	97	99	99	98
$PW(3)$	100	100	100	100	100	100	100	100	100
$LN(0.8)$	82	88	89	88	88	88	86	87	88
$LN(1)$	97	98	99	99	99	99	99	99	99
$LN(1.5)$	100	100	100	100	100	100	100	100	100
$IG(0.5)$	99	100	100	100	100	100	100	100	100
$IG(1.5)$	72	80	80	78	78	80	73	75	76

4.2. Simulation Results

We will now present some general conclusions regarding the tabulated estimated powers of the different tests considered. Since the performance of the tests are affected by the type of hazard rate of the alternative distribution, we will discuss the overall performance as well as the performance when the results are grouped according to the type of hazard rate.

First, we will consider the estimated local powers, presented in Tables 2 and 3. We find that KS_n and CM_n exhibit poor power performance, displaying the lowest powers among the tests for the majority of the choices of the mixture probability, p . We note that $EL_{n,1}$ and $R_{n,5}$ are tied for the best test for the majority of mixture proportions. Figure 1 displays the local powers of AD , $EL_{n,1}$, CR and $R_{n,5}$ over the complete range of mixture probabilities. The superior performance of $EL_{n,1}$ and $R_{n,5}$, for this mixture distribution, is clear from this figure.

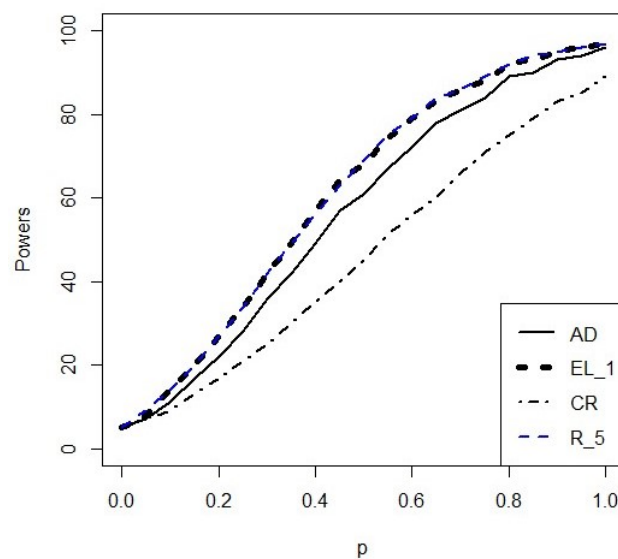


Figure 1. Local powers for some of the tests over the entire range of mixture probabilities of the Rayleigh exponential mixture distribution for $n = 20$.

For the transformed data, \widetilde{KS}_n exhibits the lowest powers overall and $HM_{n,0.75}$ has the highest overall powers for the majority of the alternatives considered.

We will now consider the performance of the tests, developed specifically for the Rayleigh distribution, in general against all of the general alternative distributions listed in Table 1. From both Tables 4 and 5 we see that, in general, the powers of KS_n and CR_n are lower for the majority of the alternatives considered and perform unfavourably in comparison to the other tests, for both sample sizes. On the other hand, $EL_{n,1}$ and $R_{n,5}$ perform quite well as we find that they outperform the other tests, having the highest estimated power for the majority of the alternatives considered. All tests considered perform quite well against the standard exponential distribution (which has a constant hazard rate) for both sample sizes.

Shifting our attention now to results associated with alternatives with increasing hazard rates, one finds, once again, that KS_n and CR_n have lower powers for both sample sizes considered. For most of the alternatives in this category $EL_{n,1}$ and $R_{n,5}$ have the highest power, only being outperformed, or equaled, for a handful of these alternatives by other tests.

Moving our attention to alternatives with a decreasing hazard, we see that all the tests considered perform very well and, since there are such minor differences in the power performance between all the tests, it is difficult to identify a single ‘best’ test for this set of

alternatives. However, for the smaller sample size, KS_n still attains powers that are slightly lower than the rest of the tests.

We now observe the results associated with alternatives with non-monotone hazard rates. The tests that generally perform well are AD_n , $EL_{n,1}$ and $R_{n,1}$. However, the test that exhibits the highest power for the majority of the alternatives, for both sample sizes, is $R_{n,5}$.

Finally, we consider the performance of the tests for exponentiality based on the transformed data. The tests with the lowest powers are \widetilde{KS}_n and \widetilde{CM}_n . $BH_{n,1}$ and $HM_{n,1.25}$ perform very well, exhibiting high powers for most of the alternatives considered, especially for alternatives with decreasing or non-monotone hazard rates. $HM_{n,0.75}$ displays the highest overall powers for the majority of the alternatives considered. However, the highest estimated power, against all alternative distributions considered, is obtained by one of the tests specifically developed for the Rayleigh distribution and not by any of the exponentiality tests based on the transformed data. Therefore, we recommend that the tests proposed specifically for the Rayleigh distribution is used when goodness-of-fit testing is performed for the Rayleigh distribution.

To conclude, we provide a brief demonstration of how the choice of the tuning parameter, a , influences the powers of the newly proposed test. In order to visualise the behaviour of the powers for different values of a , Figure 2 present the powers for $R_{n,a}$ over a grid of a values and six different alternative distributions. This figure is also used to motivate the choice of a values included in the study.

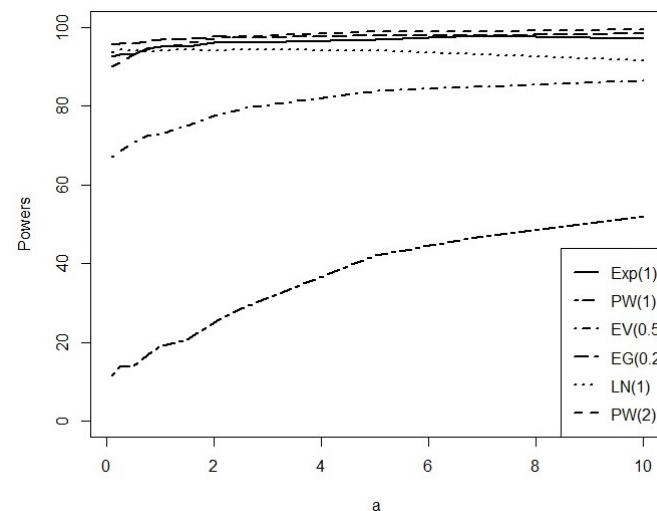


Figure 2. Estimated powers for $R_{100,a}$ for some alternatives appearing in Table 1.

The choice of $a = 1$ was made since it is the point where the powers for most of the alternative distributions start to stabilize and reach a plateau. The choice for $a = 5$ is due to the fact that it is the point where the powers for most of the alternative distributions reach their maximum value.

5. Practical Application

As noted in Section 1, the Rayleigh distribution found various applications in the fields of survival analysis and reliability theory. In this section we demonstrate the use of the tests specifically developed for the Rayleigh distribution by applying them to a real-world survival data set: the COVID-19 data set of Italy given in Table 8—for a discussion on the data set, see [30]. The data set displays the COVID-19 mortality rates recorded for 59 days in Italy from 27 February 2020 to 27 April 2020. Ref. [30] discussed and analysed the use of an extended three parameter Rayleigh distribution to model the data. They concluded that the newly extended Rayleigh distribution is a good fit to the data. We, however, will investigate the goodness-of-fit of the traditional one parameter Rayleigh distribution as

well as that of the exponential distribution. Figure 3 represents the probability plots of both the Rayleigh (grey dots) and exponential (black dots) distribution fitted to the data, where $\hat{\theta} = 6.583$ and $\hat{\lambda} = 0.123$ in the case of the exponential distribution.

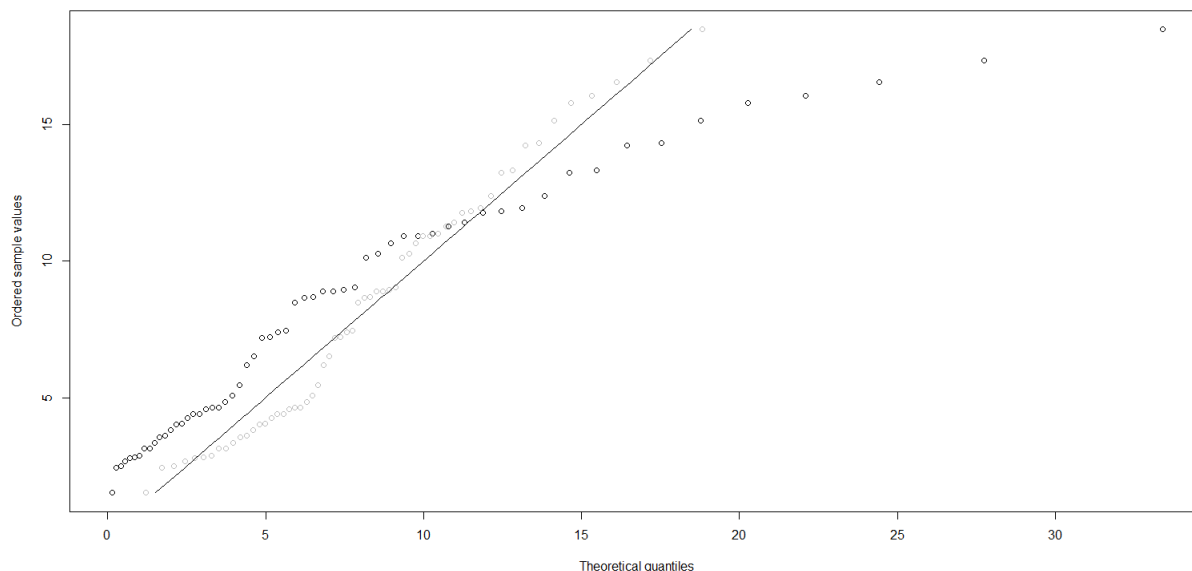


Figure 3. Probability plot of a fitted Rayleigh (grey dots) and exponential (black dots) distribution.

The probability plot suggests that the underlying distribution of the data might be the Rayleigh distribution instead of the exponential distribution.

Table 8. COVID-19 data set of Italy.

1.518	2.450	2.508	2.686	2.780	2.814	2.881	3.134	3.148	3.341
3.564	3.606	3.827	4.011	4.040	4.253	4.408	4.416	4.571	4.639
4.640	4.859	5.073	5.452	6.194	6.503	7.201	7.214	7.407	7.445
8.479	8.646	8.697	8.905	8.906	8.961	9.037	10.138	10.282	10.644
10.908	10.919	11.010	11.273	11.410	11.775	11.822	11.950	12.396	13.226
13.333	14.242	14.330	15.137	15.787	16.046	16.561	17.337	18.474	

Table 9 contains the estimated p-value (calculated based on 50,000 samples of size 59 simulated from the standard Rayleigh distribution) of each test for testing formally whether the data originated from a Rayleigh distribution.

Table 9. p-values for the COVID-19 data of Italy.

TestStatistic	KS_n	CM_n	AD_n	$EL_{n,1}$	$EL_{n,5}$	CR_n	$KL_{n,3}$	$KL_{n,4}$	$R_{n,1}$	$R_{n,5}$
p-value	0.05974	0.12038	0.12906	0.62786	0.5679	0.4104	0.09604	0.07	0.8903	0.91866

From these p-values it is clear that all the tests do not reject the null hypothesis in (3) at a 5% significance level and we can therefore conclude that the Rayleigh distribution is also a feasible option to model the data.

Having found that the Rayleigh distribution is a good fit to the observed data, one can now go about calculating quantiles, moments and other useful distributional properties by using the theoretical Rayleigh distribution with estimated parameter $\hat{\theta} = 6.583$. For example, by fitting this Rayleigh distribution we find that the mean mortality rate over the 59 days is 8.2506.

6. Conclusions and Future Research

In this article, a new goodness-of-fit test statistic specifically designed for the Rayleigh distribution was considered. The finite-sample performance of this newly suggested test was studied via the use of a Monte Carlo simulation. From the results, it is clear that this new test is not only feasible when testing goodness-of-fit for the Rayleigh distribution, it also outperforms or equals competitor tests for the majority of the alternative distributions considered. For practical implementation we suggest using the choice $a = 5$ for $R_{n,a}$. Alternatively, one can use a data-dependent choice of this tuning parameter suggested, e.g., in [31].

In analysing mortality or survival data (like the COVID-19 data set) one will, more often than not, deal with observations that are censored. For our newly proposed test to be applicable in these kinds of situations, it needs to be modified to accommodate censoring. Naturally, this modification will complicate some of the asymptotic derivations and might be an avenue for future research. Some work in this regard has been started by [32] as well as [33].

Author Contributions: The authors (G.L.G., E.B., J.S.A.) contributed equally. All authors have read and agreed to the published version of the manuscript

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The work of E. Bothma and J.S. Allison are based on research supported by the National Research Foundation (NRF). Any opinion, finding and conclusion or recommendation expressed in this material is that of the authors and the NRF does not accept any liability in this regard.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Johnson, N.L.; Kotz, S.; Balakrishnan, N. *Continuous Univariate Distributions*, 2nd ed.; Wiley: New York, NY, USA, 1994.
2. Rayleigh, F.R.S. Xii. On the resultant of a large number of vibrations of the same pitch and of arbitrary phase. *Lond. Edinb. Dublin Philosophical Mag. J. Sci.* **1880**, *10*, 73–78. [\[CrossRef\]](#)
3. Dyer, D.; Whisenand, C. Best linear unbiased estimator of the parameter of the Rayleigh distribution. *IEEE Trans. Reliab.* **1973**, *22*, 229–231. [\[CrossRef\]](#)
4. Polovko, A. *Fundamentals of Reliability Theory*; Academic Press: Cambridge, MA, USA, 1968.
5. Brummer, M.; Mersereau, R.; Eisner, R.; Lewine, R. Automatic detection of brain contours in MRI data sets. *IEEE Trans. Med. Imaging* **1993**, *12*, 153–166. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Sijbers, J.; Poot, D.; den Dekker, A.; Pintjens, W. Automatic estimation of the noise variance from the histogram of a magnetic resonance image. *Phys. Med. Biol.* **2007**, *52*, 1335–1348. [\[CrossRef\]](#)
7. Rajan, J.; Poot, D.; Juntu, J.; Sijbers, J. *Segmentation Based Noise Variance Estimation from Background MRI Data*; ICIAR 2010 Part I LNCS; Springer: Berlin/Heidelberg, Germany, 2010; Volume 6111, pp. 62–70.
8. Guemana, M.; Hafafa, A.; Mohamed, B.R. Reliability modeling using Rayleigh distribution: Industrial pump application. In Proceedings of the 19th European Conference on Mathematics for Industry ECMI2016, Santiago de Compostela, Spain, 13–17 June 2016.
9. Elviana, E.; Purwadi, J. *Parameters Estimation of Rayleigh Distribution in Survival Analysis on Type II Censored Data Using the Bayesian Method*; Journal of Physics: Conference Series; IOP Publishing: Bristol, UK, 2020; Volume 1503, p. 012004.
10. Meintanis, S.G.; Iliopoulos, G. Tests of fit for the Rayleigh distribution based on the empirical Laplace transform. *Ann. Inst. Stat. Math.* **2003**, *55*, 137–151. [\[CrossRef\]](#)
11. Baratpour, B.S.; Khodadadi, F. A cumulative residual entropy characterization of the Rayleigh distribution and related goodness-of-fit test. *J. Stat. Res. Iran* **2012**, *9*, 115–1294. [\[CrossRef\]](#)
12. Alizadeh Noughabi, R.; Alizadeh Noughabi, H.; Ebrahimi Moghaddam Behabadi, A. An entropy test for the Rayleigh distribution and power comparison. *J. Stat. Comput. Simul.* **2014**, *84*, 151–158. [\[CrossRef\]](#)
13. Safavinejad, M.; Jomhoori, S.; Alizadeh Noughabi, H. A density-based empirical likelihood ratio goodness-of-fit test for the Rayleigh distribution and power comparison. *J. Stat. Comput. Simul.* **2015**, *85*, 3322–3334. [\[CrossRef\]](#)
14. Nikitin, Y.Y. Tests based on characterizations, and their efficiencies: A survey. *arXiv* **2017**, arXiv:1707.01522.

15. Betsch, S.; Ebner, B. Characterizations of continuous univariate probability distributions with applications to goodness-of-fit testing. *arXiv* **2018**, arXiv:1810.06226.
16. Stein, C. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, CA, USA, 9–22 July 1971; Volume 2: Probability Theory; University of California Press: Berkeley, CA, USA, 1972; Volume 6, pp. 583–603.
17. Betsch, S.; Ebner, B. Testing normality via a distributional fixed point property in the Stein characterization. *TEST* **2018**, *29*, 105–138. [[CrossRef](#)]
18. Henze, N.; Meintanis, S.G.; Ebner, B. Goodness-of-fit tests for the gamma distribution based on the empirical Laplace transform. *Commun. -Stat.-Theory Methods* **2012**, *41*, 1543–1556. [[CrossRef](#)]
19. Allison, J.S.; Pretorius, C. A Monte Carlo evaluation of the performance of two new tests for symmetry. *Comput. Stat.* **2017**, *32*, 1323–1338. [[CrossRef](#)]
20. Ledoux, M.; Talagrand, M. *Probability in Banach Spaces: Isoperimetry and Processes*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2013.
21. Allison, J.S.; Santana, L.; Smit, N.; Visagie, I.J.H. An “apples-to-apples” comparison of various tests for exponentiality. *Comput. Stat.* **2017**, *32*, 1241–1283. [[CrossRef](#)]
22. Henze, N.; Meintanis, S.G. Recent and classical tests for exponentiality: A partial review with comparisons. *Metrika* **2005**, *61*, 29–45. [[CrossRef](#)]
23. Baringhaus, L.; Henze, N. Tests of fit for exponentiality based on a characterization via the mean residual life function. *Stat. Pap.* **2000**, *41*, 225–236. [[CrossRef](#)]
24. Baringhaus, L.; Henze, N. A class of consistent tests for exponentiality based on the empirical Laplace transform. *Ann. Inst. Stat. Math.* **1991**, *43*, 551–564. [[CrossRef](#)]
25. Henze, N.; Meintanis, S.G. Tests of fit for exponentiality based on the empirical Laplace transform. *Statistics* **2002**, *36*, 147–161. [[CrossRef](#)]
26. Epps, T.; Pulley, L. A test of exponentiality vs. monotone-hazard alternatives derived from the empirical characteristic function. *J. R. Stat. Soc. Ser. (Methodol.)* **1986**, *48*, 206–213. [[CrossRef](#)]
27. Allison, J.; Milošević, B.; Obradović, M.; Smuts, M. Distribution-free goodness-of-fit tests for the pareto distribution based on a characterization. *Comput. Stat.* **2022**, *37*, 403–418. [[CrossRef](#)]
28. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2019.
29. Hlavac, M. Stargazer: Well-formatted regression and summary statistics tables. *R Package Version* **2018**, *5*, 2.
30. Almongy, H.M.; Almetwally, E.M.; Aljohani, H.M.; Alghamdi, A.S.; Hafez, E. A new extended rayleigh distribution with applications of COVID-19 data. *Results Phys.* **2021**, *23*, 104012. [[CrossRef](#)]
31. Allison, J.; Santana, L. On a data-dependent choice of the tuning parameter appearing in certain goodness-of-fit tests. *J. Stat. Comput. Simul.* **2015**, *85*, 3276–3288. [[CrossRef](#)]
32. Bothma, E.; Allison, J.S.; Cockeran, M.; Visagie, I.J.H. Characteristic function and Laplace transform-based tests for exponentiality in the presence of random right censoring. *Stat* **2021**, *10*, e394. [[CrossRef](#)]
33. Cuparić, M.; Milošević, B. New characterization based exponentiality tests for randomly censored data. *arXiv* **2020**, arXiv:2011.07998.