

On new goodness-of-fit tests for the Rayleigh distribution

SC Liebenberg

 **orcid.org 0000-0002-0106-3084**

Thesis accepted in fulfilment of the requirements for the degree
Doctor of Philosophy in Science with Statistics at the North-
West University

Promoter: JS Allison

Graduation October 2020
20396236

Abstract

The Rayleigh distribution has been observed in numerous processes across multiple research disciplines. It has therefore become increasingly important to test if a specific data set originated from this distribution. Goodness-of-fit tests developed specifically for the Rayleigh distribution have become more researched over the past 10 years and there is no consensus on which test performs best in certain situations. The primary aim of this thesis is to develop new goodness-of-fit tests for the Rayleigh distribution. We propose several novel tests based on different approaches, the first of which is a conditional expectation characterization. The second approach deals with a differential equation that has the Rayleigh density function as the unique solution and the third approach is based on the Mellin transform. The asymptotic theory of some of the newly proposed tests are developed and the finite-sample performance of the new tests are compared to that of existing tests in an extensive Monte Carlo simulation study. In the simulation study, the tests are compared against several alternative distributions that are commonly used as alternatives for the Rayleigh distribution, as well as against two mixture distributions to assess the power performance with respect to local alternatives. When the power estimates of the goodness-of-fit tests are considered, it is clear that the newly developed tests are very competitive and tend to outperform or match the competitor tests that are considered in this study.

Key words: *Asymptotics, Characterisation, Differential equation, Goodness-of-fit, Mellin transform, Rayleigh distribution*

Acknowledgements

“And everything, whatever you do in word or deed, do all in the name of the Lord Jesus, giving thanks to God and the Father by Him.”

Colossians 3:17

First and foremost, I offer my gratitude and praise to God, without whom none of this would be possible.

I would like to thank the following:

- Jennifer Liebenberg, my wife and best friend, for being my support and motivation.
- Prof. James Allison, my promoter, to whom I owe a debt of gratitude for his patience, insight and willingness to instruct.

Contents

Abstract	1
Acknowledgements	2
1 Introduction	9
1.1 Overview	9
1.2 Objectives	10
1.3 Thesis outline	11
2 Characterisations and existing goodness-of-fit tests for the Rayleigh distribution	13
2.1 Introduction	13
2.2 Properties of the Rayleigh distribution	13
2.3 Characterisations of the Rayleigh distribution	18
2.3.1 Characterisations based on conditional expectations	18
2.3.2 Characterisations based on order statistics	20
2.3.3 Characterisations based on record values	20
2.3.4 Characterisations based on failure rate	21
2.3.5 Characterisations based on Entropy	22
2.3.6 Characterisations based on the Laplace transform	23
2.4 Goodness-of-fit tests for the Rayleigh distribution	24

2.4.1	Classical tests based on the empirical distribution function . .	25
2.4.2	A test based on the empirical Laplace transform	26
2.4.3	Tests based on entropy	27
2.4.4	Tests based on the Phi-divergence measure	32
2.4.5	A test based on the empirical likelihood ratio	35
2.4.6	Tests adapted for the Rayleigh distribution	37
2.4.7	On the use of weight functions	39
3	Kernel density estimation	41
3.1	Introduction	41
3.2	Notation and methodology	42
3.3	Measures of fit	44
3.4	Kernel function choices	46
3.5	Bandwidth selection	47
3.6	Boundary-corrected kernel density estimation	49
3.6.1	Reflection and pseudo-data methods	50
3.6.2	Transformation methods	53
3.6.3	Boundary kernel methods	55
3.6.4	Numerical investigation	59
4	New goodness-of-fit tests for the Rayleigh distribution	63
4.1	Introduction	63
4.2	New tests based on a conditional expectation characterisation	63
4.2.1	Asymptotic theory	66
4.3	New test based on a differential equation approach	73
4.4	New test based on the Mellin transform	76
5	Simulations and results	79
5.1	Introduction	79

5.2	Simulation setting	79
5.3	Simulation results	84
5.3.1	Discussion on power estimates for alternative distributions . .	84
5.3.2	Discussion on local power estimates for the $HN - \text{Ral}(1)$ mixture	87
5.3.3	Discussion on local power estimates for the $\Gamma(1.5) - \text{Ral}(1)$ mixture	89
5.4	Real data application	89
6	Concluding remarks and future research	103
6.1	Introduction	103
6.2	Concluding remarks and future research	105
6.2.1	Estimating F with the plug-in estimator $F_{\hat{\theta}_n}$ in CM_n	105
6.2.2	The T_n test with derivative estimation	107
6.2.3	Power-divergence choice for the test of Zamanzade & Mahdizadeh (2017)	112
6.2.4	Two choices of $\Psi(\cdot)$ for the test of Torabi et al. (2016)	112
	Appendix A	114

List of Figures

2.1	Density function (left) and distribution function (right) of the Rayleigh distribution.	15
3.1	Construction of the kernel density estimate (Koekemoer 2004). . . .	44
3.2	Various shapes of kernel functions.	47
3.3	Kernel density estimate plots for the KDE and cutnorm boundary-corrected method.	60
3.4	Kernel density estimate plots for the reflect and simple boundary-corrected method.	61
3.5	MISE, variance and bias components for the reflect (red line), cutnorm (blue line) and simple (purple line) boundary-corrected methods. . .	62
5.1	Power estimates across tuning parameter value for the $CH(1)$ (left) and $DL(1)$ (right) distributions with $n = 20$	86
6.1	First derivative of Rayleigh distribution (blue line) and density estimate (red line).	109
A.1	Power estimates across tuning parameter value for the $EXP(1)$ (top left), $EV(1.5)$ (top right), $\Gamma(1.5)$ (bottom left) and $IG(0.5)$ (bottom right) distributions and $n = 20$	114

A.2	Power estimates across tuning parameter value for the $CH(1)$ (top left), $DL(1.5)$ (top right), $EXP(1)$ (middle left), $EV(1.5)$ (middle right), $\Gamma(1.5)$ (bottom left) and $IG(0.5)$ (bottom right) distributions and $n = 30$	115
-----	--	-----

List of Tables

3.1	List of kernels functions.	47
5.1	Probability density functions for choices of the alternative distributions.	82
5.2	Notation references for new and existing test statistics.	83
5.3	Calculated p -values for the average wind speed data.	90
5.4	Estimated powers for the alternative distributions given in Table 5.1 for the MLE and sample size $n = 20$	91
5.5	Estimated powers for the alternative distributions given in Table 5.1 for the MLE and sample size $n = 30$	94
5.6	Estimated local powers for $n = 20$ (top row) and $n = 30$ (bottom row) for the $HN - Ral(1)$ mixture distribution and MLE.	97
5.7	Estimated local powers for $n = 20$ (top row) and $n = 30$ (bottom row) for the $\Gamma(1.5) - Ral(1)$ mixture distribution and MLE.	100
A.1	Estimated powers for the alternative distributions given in Table 5.1 for the MME and sample size $n = 20$	116
A.2	Estimated powers for the alternative distributions given in Table 5.1 for the MME and sample size $n = 30$	119

Chapter 1

Introduction

1.1 Overview

Since the inception of the Rayleigh distribution in conjunction with an acoustic problem (Rayleigh 1880) it has been observed in numerous processes across multiple research disciplines. It naturally arises in a two dimensional setting when the resultant of two independently normally distributed vectors are considered (see Chapter 2). This arrangement often emerges as the probability mechanism in environmental occurrences such as short-term average wind speeds (see e.g., Morgan, Lackner, Vogel & Baise 2011, Celik 2004, Dörvlo 2002) which is an important facet in sustainable energy development, and observations in oceanic wave height (see e.g., Longuet-Higgins 1952, Soares & Carvalho 2003, Casas-Prat & Holthuijsen 2010). Moreover, in the fields of astronomy and astrophysics the Rayleigh distribution is used to model planet formation theories (see e.g., Lawless 2002).

Regarding planet formation, the Rayleigh distribution is one of three distributions commonly used to describe the distribution of mutual inclinations between the orbital planes of planets. The distribution of the mutual inclination between the orbital planes of planets gives insight into theories regarding the building-blocks of planetary systems, and more specifically, theories of planet formation and growth of

galaxies. The mutual inclination of a planet in a planetary system's orbits can also give indication to significant events that could have caused altered inclinations and eccentricities of planets. The interested reader is referred to Lissauer et al. (2011), Fang & Margot (2012), Figueira et al. (2012), Fabrycky et al. (2014) and Bovaird & Lineweaver (2017) for more details on this.

The Rayleigh distribution has also proved particularly useful in imaging techniques in the fields of physics and medicine. In this context, it is known that the noise deviation data associated with magnetic resonance images is Rayleigh distributed (see e.g., Rajan, Poot, Juntu & Sijbers 2010, Dekker & Sijbers 2014, Toa, Sim, Lim & Lim 2019). Furthermore, in order to model grey level behaviour in ultrasound imaging the Rayleigh distribution is needed (see e.g., Belaid & Boukerroui 2018, Gai, Zhang, Yang & Yu 2018, Sarti, Corsi, Mazzini & Lamberti 2005).

In all of the above, the primary concern is if a specific data set originated from a Rayleigh distribution. Consequently, the need to assess the correctness of fit to the data becomes increasingly important. Researchers, therefore, concern themselves with constructing goodness-of-fit tests to verify if the observed data did indeed realize from a Rayleigh distribution. Tests specifically for the Rayleigh distribution have only become more researched in the past 10 years and there is still no consensus on which test performs best in certain situations. In this, there is a need for investigation.

1.2 Objectives

The primary aim of this thesis is to develop new goodness-of-fit tests for the Rayleigh distribution and to perform a thorough Monte Carlo study wherein the performance of the new goodness-of-fit tests are compared and evaluated against existing goodness-of-fit tests for the Rayleigh distribution.

The main objectives of this thesis can be summarized as follows:

- Review the existing literature on goodness-of-fit tests for the Rayleigh distribution.
- Review the literature on specific statistical tools that are used in developing our new goodness-of-fit tests for the Rayleigh distribution.
- Develop and explore a new goodness-of-fit test based on a conditional expectation characterisation for the Rayleigh distribution.
- Derive the asymptotic theory surrounding the new goodness-of-fit test based on the conditional expectation characterisation.
- Investigate the Mellin transform and apply it in constructing a new goodness-of-fit test for the Rayleigh distribution.
- Investigate and develop a new goodness-of-fit test for the Rayleigh distribution based on a differential equation approach.
- Adapt existing tests to specifically test for the Rayleigh distribution.
- Evaluate the performance of the newly suggested goodness-of-fit tests and existing goodness-of-fit tests for the Rayleigh distribution.

1.3 Thesis outline

Chapter 2 starts by exploring general properties of the Rayleigh distribution and continues in presenting an overview of existing characterisations and goodness-of-fit tests of the Rayleigh distribution.

Chapter 3 deals with an important tool in the development of a new goodness-of-fit tests in the form of the kernel density estimate. Several aspects of kernel density

estimation, such as bandwidth selection and boundary-corrected kernels are also considered. These tools will be utilized in Chapter 4.

In Chapter 4 we introduce and develop several new goodness-of-fit tests for the Rayleigh distribution based on three different approaches. The first approach utilizes a conditional expectation characterisation that was given in Chapter 2. Subsequently, the asymptotic properties of this new test statistic is thoroughly explored. The second approach revolves around the differential equation that characterizes the Rayleigh distribution. The test statistic is developed with the use of the kernel density estimate techniques discussed in Chapter 3. The third approach uses the lesser utilized Mellin transform to develop a test statistic for the Rayleigh distribution.

Chapter 5 deals with the finite-sample performance of the new and existing goodness-of-fit tests. Power estimates are calculated and presented for a variety of alternative distributions. Local power estimates are also presented for two mixtures of distributions and a real data example is employed to assess the new and existing tests performance to a real-world application.

The thesis concludes in Chapter 6 with some final remarks and avenues for future research.

Chapter 2

Characterisations and existing goodness-of-fit tests for the Rayleigh distribution

2.1 Introduction

This chapter aims to present the existing characterisations and goodness-of-fit tests for the Rayleigh distribution. In Section 2.2, we start with an overview of the properties of the Rayleigh distribution. In Section 2.3 we present some characterisations of the Rayleigh distribution. Some existing goodness-of-fit tests as well as two tests that we specifically modified to test for the Rayleigh distribution are discussed in Section 2.4.

2.2 Properties of the Rayleigh distribution

Properties of the univariate Rayleigh distribution and its relationship to other distributions are discussed in, e.g., Siddiqui (1962) and Johnson, Kotz & Balakrishnan (1994). We will now relate a selection of these characteristics and properties that we

deem important for the remaining chapters.

Consider a vector $\mathbf{Z} = \{Z_1, Z_2 \dots Z_n\}$ of size n sampled independently from a normal distribution with mean 0 and variance θ^2 and let this vector denote a point in the n -dimensional Euclidean space. Now, consider the distance from the origin to the point $\mathbf{Z}$, written as $D_{\mathbf{Z}} = \sqrt{\sum_{i=1}^n Z_i^2}$. There exists a probability density function for $D_{\mathbf{Z}}$ given by

$$f(x, n, \theta) = \frac{2x^{n-1} \exp\left(\frac{-x^2}{2\theta^2}\right)}{(2\theta^2)^{n/2} \Gamma(n/2)}, \quad (2.1)$$

where $x > 0$, $\theta > 0$ and $\Gamma(\cdot)$ is the gamma function.

In the case where $n = 2$, i.e., the point $\{Z_1, Z_2\}$, the probability density function (pdf) in (2.1) reduces to the Rayleigh distribution with probability density function given by

$$g(x, \theta) = \frac{x}{\theta^2} \exp\left(\frac{-x^2}{2\theta^2}\right), \quad (2.2)$$

and cumulative distribution function (cdf),

$$G(x, \theta) = 1 - \exp\left(\frac{-x^2}{2\theta^2}\right). \quad (2.3)$$

In Figure 2.1 we present the pdf and cdf of the Rayleigh distribution for various parameter values.

Remark 2.1. *There exists an alternative parametrization for the Rayleigh distribution with density and distribution functions,*

$$\begin{aligned} \tilde{g}(x, \theta) &= \frac{2x}{\theta^2} \exp\left(\frac{-x^2}{\theta^2}\right), \\ \tilde{G}(x, \theta) &= 1 - \exp\left(\frac{-x^2}{\theta^2}\right). \end{aligned}$$

Many of the characterisations given in Section 2.3 are based on this parametrization,

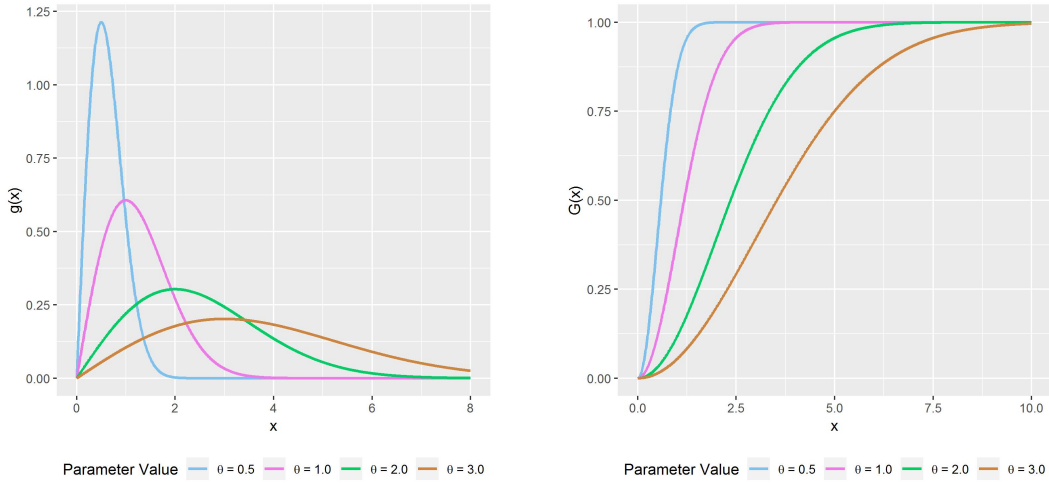


Figure 2.1: Density function (left) and distribution function (right) of the Rayleigh distribution.

whereas many of the goodness-of-fit tests in Section 2.4 are based on the parametrization in (2.2). We will therefore use the appropriate parametrization as needed and reference the density or distribution function when necessary. The properties of the Rayleigh distribution in the remainder of this section is applicable to the parametrization in (2.2).

The hazard (failure rate) function of the Rayleigh distribution is given by

$$h(x, \theta) = \frac{g(x, \theta)}{S(x, \theta)} = \frac{x}{\theta^2},$$

where $S(x, \theta) = 1 - G(x, \theta) = \exp(-x^2/2\theta^2)$ is the survival (or reliability) function. The Rayleigh distribution has a linearly increasing hazard rate that makes it an appropriate lifetime model for rapidly aging components (see e.g., Lawless 2002). This can be seen as follows: for small values of x the reliability of a component decreases with time more slowly than a component with a constant hazard rate and for large values of x the reverse is true (see Johnson et al. 1994).

The quantile function of the Rayleigh distribution is given by

$$Q(p, \theta) = G^{-1}(p, \theta) = \theta \{-2 \log(1 - p)\}^{1/2}, \quad 0 < p < 1,$$

from which the percentiles of the distribution can be obtained.

Next, suppose a random variable X follows a Rayleigh distribution with density function given in (2.2), then the raw moments can be written in the form

$$E[X^r] = \theta^r 2^{r/2} \Gamma\left[\frac{r}{2} + 1\right].$$

As an illustration, consider the first four raw moments given by

$$\begin{aligned} E[X] &= \theta \sqrt{\pi/2}, & E[X^2] &= 2\theta^2, \\ E[X^3] &= 3\theta^3 \sqrt{\pi/2}, & E[X^4] &= 8\theta^4, \end{aligned}$$

from which we have that $\text{Var}(X) = \theta^2(4 - \pi)/2$.

Other functions of the Rayleigh distribution which prove useful are the moment generating function,

$$M_X(t) = E[\exp(tX)] = 1 + \theta t e^{\theta^2 t^2/2} \sqrt{\frac{\pi}{2}} \left(\text{erf}\left(\frac{\theta t}{\sqrt{2}}\right) + 1 \right),$$

where $\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$ is the error function, and the characteristic function

$$\phi_X(t) = E[\exp(itX)] = 1 + \theta t e^{-\theta^2 t^2/2} \sqrt{\frac{\pi}{2}} \left(-\text{erfi}\left(\frac{\theta t}{\sqrt{2}}\right) + i \right),$$

where $\text{erfi}(x) = -i \text{erf}(ix)$ is the imaginary error function.

Most of the aforementioned expressions contain the population parameter θ , which

has to be estimated when inference is performed. In considering the estimation of the parameter θ from a random sample $X_1, X_2, \dots, X_n$, we have access to the maximum likelihood estimate,

$$\hat{\theta}_n^{ML} = \sqrt{(2n)^{-1} \sum_{j=1}^n X_j^2},$$

and the methods of moment estimate,

$$\hat{\theta}_n^{MM} = \sqrt{\frac{2}{\pi} n^{-1} \sum_{j=1}^n X_j}.$$

It can easily be shown that $\hat{\theta}_n^{MM}$ is an unbiased estimator of θ , while $\hat{\theta}_n^{ML}$ is biased. However, from large sample theory of maximum likelihood estimation we know that $\hat{\theta}_n^{ML}$ is asymptotically unbiased.

The Rayleigh distribution also has inherent connection with other distributions. It is well known that if a random variable X has a Rayleigh distribution with parameter θ , then X^2 is exponentially distributed with parameter $2\theta^2$. The Rayleigh distribution is also a special case of the Weibull distribution with shape parameter set to 2 and the chi-square distribution with 2 degrees of freedom. The Rice distribution is also closely related to the Rayleigh distribution which can be seen when the parameter v of the Rice distribution, the distance between a reference point and the center of the bivariate distribution, is set to zero.

The popularity of the Rayleigh distribution and its plethora of applications also sparked interest in generalizations or modification of this distribution. Balakrishnan & Kocherlakota (1985) studied the two parameter (double) Rayleigh distribution whereas Vodă (1976) considered the generalized Rayleigh distribution. Merovci (2013) developed a transmuted Rayleigh distribution by using a quadratic rank transmutation map, Roy (2004) utilized a general approach for discretization continuous life distributions to presented a discrete Rayleigh distribution. Tan & Beaulieu (1997)

constructed an infinite series representation of the bivariate Rayleigh distribution and Simon & Alouini (1998) built upon this to provide a single integral representation of the bivariate distribution. The generalized Rayleigh distribution was defined by Miller, Bernstein & Blumenson (1958) whereas Jensen (1970) discussed a generalization of the multivariate Rayleigh distribution. Each of the mentioned variations or extensions in turn yields their own set of applications.

2.3 Characterisations of the Rayleigh distribution

A characterisation of a probability distribution states certain properties that only hold true for that specific distribution. The concept of characterizing distributions has become more prevalent as distributional aspects increase in practical applications and as new distributions are implemented. In that sense, characterisations of probability distributions are valuable tools in the development of goodness-of-fit testing. More information on characterisation of a probability distribution can be found in Ahsanullah (2017) and Galambos & Kotz (1978).

In what follows we give characterisations specifically for the Rayleigh distribution.

2.3.1 Characterisations based on conditional expectations

The conditional expectation of a continuous random variable X given a random variable Y can be expressed as

$$E[X|Y = y] = \int x f_{X|Y}(X|Y = y) dx,$$

where $f_{X|Y}(X|Y = y) = f_{X,Y}(x, y)/f_Y(y)$. More information on conditional expectation and the important role that this concept plays in statistics can be found in Fristedt & Gray (2013).

Consider the following characterisations for the Rayleigh distribution utilizing conditional expectation.

Ahsanullah & Shakil (2013):

Theorem 2.1. *Let X be a nonnegative random variable with absolutely continuous distribution function $F(x)$ with $F(0) = 0$, $F(x) > 0$ for all $x > 0$ and finite $E[X^{2k}]$, for some fixed $k \geq 1$. Then X has a Rayleigh distribution with $F(x) = 1 - e^{-x^2/2\theta^2}$, $x > 0$, $\theta > 0$ if, and only if, $E[X^{2k}|X > t] = \sum_{i=0}^k 2^i \theta^{2i} k^{(i)} t^{2(k-i)}$, with $k^{(i)} = k(k-1)\dots(k-i+1)$ and $k^{(0)} = 1$.*

The above characterisation can be seen as a moment expression of order $2k$ conditioned on a tail event $\{X > t\}$. In this sense the parameter $k = 1, 2, \dots$, can be seen as a type of tuning parameter.

A second characterisation is given in the following theorem:

Theorem 2.2. *Let X be a nonnegative random variable with absolutely continuous distribution function $F(x)$ with $F(0) = 0$ and $F(x) > 0$ for all $x > 0$ and finite $E[X^{2k+1}]$, for some fixed $k \geq 1$. Then X has a Rayleigh distribution with $F(x) = 1 - e^{-x^2/2\theta^2}$, $x > 0$, $\theta > 0$ if, and only if,*

$$E[X^{2k-1}|X > t] = \sum_{j=0}^{k-1} \frac{(2k-1)!!}{(2k-1-2j)!!(1/\theta^2)^j} t^{2k-1-2j} + \frac{(2k-1)!!}{(1/\theta^2)^\ell} \sqrt{\pi/2\theta^2} (1 - \operatorname{erf}(t\sqrt{1/2\theta^2})) e^{t^2/\theta^2},$$

where $(2k-1)!! = 1, 3, \dots, (2k-1)$, $k \geq 1$, any nonnegative number ℓ and $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$ denotes the error function.

The above two theorems can be used to find the k^{th} -order moment for X based on the tail event $\{X > t\}$. Specifically, Theorem 2.1 gives access to the even conditional moments and Theorem 2.2 gives access to the odd conditional moments. For example, for $k = 1$ we derive from Theorem 2.2 that the first conditional moment is,

$$E[X|X > t] = t + \theta \sqrt{\frac{\pi}{2}} \left\{ 1 - \operatorname{erf}\left(\sqrt{\frac{1}{2\theta^2}} t\right) \right\} \exp(t^2\theta^2),$$

and the second conditional moment is obtained from Theorem 2.1 as,

$$E [X^2|X > t] = t^2 + 2\theta^2.$$

2.3.2 Characterisations based on order statistics

Consider random variables $X_1, X_2, \dots, X_n$ with associated order statistics $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$. Properties and applications of order statistic can be found in Arnold, Balakrishnan & Nagaraja (1992).

The following theorems characterizes the Rayleigh distribution exploiting order statistics.

Ahsanullah & Shakil (2013):

Theorem 2.3. *Let X be a nonnegative random variable with absolutely continuous distribution function $F(x)$ with $F(0) = 0$ and $F(x) > 0$ for all $x > 0$ and finite $E(X^2)$. Then X has the Rayleigh distribution with $F(x) = 1 - e^{-x^2/\theta^2}$, $x > 0$, $\theta > 0$ if, and only if,*

$$E [X_{(i)}^{2m} | X_{(i-1)} = t] = \sum_{j=0}^m \frac{m!}{(m-i)!} \left(\frac{\theta^2}{n-i+1} \right)^i t^{2(m-i)}, \quad (2.4)$$

for some $n \geq 1$, $m \geq 1$.

For the special case where $m = 1$ we have that (2.4) reduces to

$$E [X_{(i)}^2 | X_{(i-1)} = t] = t^2 + \frac{\theta^2}{n-i+1}.$$

2.3.3 Characterisations based on record values

Let $M_n = X_{(n)}$, where $X_{(n)}$ is the maximum value in a ordered set of values. Furthermore, if $M_1 \leq M_2 \leq \dots \leq M_n \leq M_{n+1} \leq \dots$ is a sequential list of maximums then the random variable X_j is the upper record value in the case

where $X_j > M_{n-1}$ with j the record time. Denote the record time as $U(n)$ where $U(n) = \min\{j | j > U(n-1), X_j > X_{U(n-1)}\}$ and $U(1) = 1$.

What follows are characterisations of the Rayleigh distribution based on these record values.

Ahsanullah & Shakil (2013):

Theorem 2.4. *Let X be a nonnegative random variable with absolutely continuous distribution function $F(x)$ with $F(0) = 0$ and $F(x) > 0$ for all $x > 0$. Assume $E[X_{U(n+1)}]$ is finite. Then X has a standard Rayleigh distribution with $F(x) = 1 - e^{-x^2}$, $x > 0$ if, and only if,*

$$E[X_{U(n+1)} | X_{U(n)} = t] = t + \sqrt{\frac{\pi}{2}} e^{t^2} - \sqrt{\frac{\pi}{2}} e^{t^2} \operatorname{erf}(\sqrt{2}t),$$

for some fixed $n \geq 1$ and $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$ denotes the error function.

It is further noted in Ahsanullah & Shakil (2013) that the conditional probability density function of the above conditional expectation is the same as that of $E(X | X > t)$. Therefore the characterisations based on $E[X_{U(n+1)}^{2m} | X_{U(n)} = t]$ and $E[X_{U(n+1)}^{2m-1} | X_{U(n)} = t]$ are equivalent to Theorem 2.1 and Theorem 2.2.

A second characterisation based on record values is given in the following theorem:

Theorem 2.5. *Let X be a nonnegative random variable with absolutely continuous distribution function $F(x)$ with $F(0) = 0$ and $F(x) > 0$ for all $x > 0$ and finite $E[X^2]$. Then X has a Rayleigh distribution with $F(x) = 1 - e^{-x^2/\theta^2}$, $x > 0$, $\theta > 0$ if, and only if, $X_{U(n)}^2$ is Erlang distributed, for some fixed $n \geq 1$.*

Some properties of the Erlang distribution can be found in Johnson et al. (1994).

2.3.4 Characterisations based on failure rate

Gupta & Bradley (2003) and Nanda (2010) provide several characterisations based

on the failure rate. A characterisation for the Rayleigh distribution can specifically be formulated in terms of the failure rate $h(x)$ and the expectation $E[h(X)/X]$ as follows.

Nanda (2010):

Theorem 2.6. *Denote the failure rate by $h(u) = \frac{f(u)}{S(u)}$ where $f(u)$ is the probability density function and $S(u) = 1 - F(u)$ is the survival function with associated distribution function $F(u)$. Furthermore, let X have finite second-order moment given by $\text{Var}(X) = \sigma^2$ with mean $E[X] = \mu$, so that the coefficient of variation of X is $c = \sigma/\mu$. Then for any nonnegative random variable X , it can be shown that*

$$E\left[\frac{h(X)}{X}\right] \geq \frac{2}{\mu^2(1+c^2)}.$$

The equality holds if, and only if, X has a Rayleigh distribution with probability density function $f(x) = x/\theta^2 \exp(-x^2/2\theta^2)$, $x \geq 0$, $\theta > 0$.

2.3.5 Characterisations based on Entropy

Define the differential entropy of a $\mathbb{R}^d$ -valued random variable as

$$H(f) = - \int f(x) \ln f(x) dx,$$

where $f(x)$ is the pdf of X .

The cumulative residual entropy (CRE) is the result of exchanging the density function in the definition of the well-known Shannon entropy by the complementary cumulative distribution function, $S(x) = P(X > x) = 1 - F(x)$. Rao, Chen, Vemuri & Wang (2004) established the new CRE as a nonnegative entropy measure of the form

$$\text{CRE}(X) = - \int_0^\infty S(x) \ln S(x) dx. \quad (2.5)$$

By using $S(x) = \exp(-x^2/2\theta^2)$ in (2.5) it can be shown that the CRE of the Rayleigh distribution is $\text{CRE}(X) = E[X/2] = \theta\sqrt{2\pi}/4$. This ultimately leads to the following characterisation of the Rayleigh distribution.

Baratpour & Khodadadi (2012):

Theorem 2.7. *The random variable X attains maximum CRE among all nonnegative, absolutely continuous random variables Y subject to $E[Y] = \nu$, $E[Y^3] = \omega$ and $\theta^2 = \frac{\omega}{3\nu}$ if, and only if, X has the Rayleigh distribution with parameter θ .*

2.3.6 Characterisations based on the Laplace transform

The Laplace transform of a random variable X with distribution function $F(x)$ is defined as

$$E[e^{-tx}] = \int_{-\infty}^{\infty} e^{-tx} dF(x),$$

where t is a real number. For more details on the Laplace transform refer to Schiff (1999).

The following characterisation for the Rayleigh distribution can be made based on the Laplace transform.

Meintanis & Iliopoulos (2003):

Theorem 2.8. *The Laplace transform*

$$L(t) = E[e^{-tX}] = 1 - \frac{\sqrt{\pi}}{2} t e^{t^2/4} \operatorname{erfc}\left(\frac{t}{2}\right),$$

of the standard Rayleigh distribution given by $F(x) = 1 - e^{-x^2}$, with the complement error function $\operatorname{erfc}(z) = 2/\sqrt{\pi} \int_z^{\infty} e^{-u^2} du$, is the unique solution to the differential equation $ty'(t) - [1 + (t^2/2)]y(t) + 1 = 0$ subject to $\lim_{t \rightarrow \infty} y(t) = 0$.

This characterizes the Rayleigh distribution since the distribution of a non-negative

random variable is uniquely determined by its corresponding Laplace transform.

2.4 Goodness-of-fit tests for the Rayleigh distribution

An important aspect of statistical inference is obtaining information about the form of the population from which the data realized. To evaluate the correctness of the fit of a certain distribution to data, a null hypothesis is designed wherein a statement is made about the probability function of the parent population. Hereafter tests of fit are employed to check the aforementioned hypothesis (Gibbons & Chakraborti 2014). Due to this fact, goodness-of-fit testing has a rich history in statistics. Classic tests of fit started with Pearson's chi-square test proposed in Pearson (1900). The Kolmogorov-Smirnov test was then independently established by Kolmogorov (1933) and Smirnov (1939) and became the standard for goodness-of-fit testing for an extended period. This test was adapted to test specifically for the normal distribution by Lilliefors (1967). The famous Cramér-von Mises test statistic found its origin in the work of Cramér (1928), Von Mises (1931) and Von Mises (1947). This test proved to be the leader in the field until the test of Anderson & Darling (1954), which improved on the former by introducing a variance stabilizing weight function. The aforementioned authors pioneered the field of goodness-of-fit testing and their work still influences this research area at present.

In establishing notation for this section, let $X_1, X_2, \dots, X_n$ be independent and identically distributed continuous realizations of a positive random variable X . If X follows a Rayleigh distribution with density function given in (2.2), it will be denoted by $X \sim \text{Ral}(\theta)$. It easily follows that $X/\theta \sim \text{Ral}(1)$. The composite goodness-of-fit hypothesis to be tested is

$$H_0 : \text{the distribution of } X \text{ is } \text{Ral}(\theta), \quad (2.6)$$

for some $\theta > 0$, against general alternatives. The majority of test statistics that we will consider are based on the scaled values $Y_j = X_j/\hat{\theta}_n$, where $\hat{\theta}_n$ is a consistent estimator for θ (either the $\hat{\theta}_n^{ML}$ or $\hat{\theta}_n^{MM}$). The use of scaled values is motivated from the invariance property of the Rayleigh distribution with respect to scale transformations. Furthermore, denote by $X_{(j)}$ and $Y_{(j)}$ the order statistics, i.e., $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$ and $Y_{(1)} < Y_{(2)} < \cdots < Y_{(n)}$.

In the following sections we present some existing goodness-of-fit tests for the Rayleigh distribution.

2.4.1 Classical tests based on the empirical distribution function

The empirical cumulative distribution function (ecdf) is given by

$$G_n(x) = \frac{1}{n} \sum_{j=1}^n I(Y_j \leq x).$$

where $I(\cdot)$ is the indicator function.

There are various tests based on the deviation of the ecdf and the cumulative distribution function specified under the null hypothesis. One such test is the Kolmogorov-Smirnov test which relies upon the maximum deviation between $G_n(x)$ and the hypothesized distribution $G_0(x)$. The Kolmogorov-Smirnov test statistic has a closed-form,

$$D_n = \max(D_n^+, D_n^-),$$

where $D_n^+ = \max_{1 \leq j \leq n} [j/n - G_0(Y_{(j)})]$ and $D_n^- = \max_{1 \leq j \leq n} [G_0(Y_{(j)}) - (j-1)/n]$, with $G_0(Y_{(j)}) = 1 - \exp(-Y_{(j)}^2/2)$. A test that utilizes the L_2 -norm and aforementioned deviation is the Cramér-von Mises test with closed-form,

$$W_n = \sum_{j=1}^n \left[G_0(Y_{(j)}) - \frac{2j-1}{2n} \right]^2 + \frac{1}{12n}.$$

The Anderson-Darling test is similar to that of the Cramér-von Mises test but with an incorporated weight function that gives it the closed-form,

$$A_n = -n - \frac{1}{n} \sum_{j=1}^n 2j - 1 [\log G_0(Y_{(j)}) + \log\{1 - G_0(Y_{(n-j+1)})\}] .$$

The Watson test incorporates the Cramér-von Mises test and is given by the closed form,

$$V_n = W_n - n(\bar{G}_0 - 1/2)^2,$$

where $\bar{G}_0 = 1/n \sum_{j=1}^n G_0(Y_{(j)})$.

Refer to Watson (1962), Gibbons & Chakraborti (2014) and D'Agostino (1986) for a more thorough treatment of these classical tests.

2.4.2 A test based on the empirical Laplace transform

Meintanis & Iliopoulos (2003):

The test statistic is based on the characterisation in Theorem 2.8 and utilizes the differential equation $ty'(t) - [1 + (t^2/2)]y(t) + 1 = 0$ to set up the test

$$MI_{n,\varphi} = n \int_0^\infty D_n^2(t)w(t)dt,$$

where $D_n(t) = tL'_n(t) - [1 + (t^2/2)]L_n(t) + 1$ and $w(t) = \exp(-\varphi t)$ with $\varphi > 0$ being a chosen tuning parameter (see Section 2.4.7 for a discussion on the use of weight functions). Furthermore, recall the Laplace transform $L(t)$ defined in Theorem 2.8 and note that $L_n(t)$,

$$L_n(t) = \frac{1}{n} \sum_{j=1}^n \exp(-tX_j),$$

is the empirical counterpart. The test statistic is calculated on the scaled values $Y_j = X_j/\hat{\theta}_n$ and rejects for large values of $MI_{n,\varphi}$. A closed-form expression of $MI_{n,\varphi}$

is given by

$$\begin{aligned}
MI_{n,\varphi} &= \frac{n}{\varphi} + \frac{1}{n} \sum_{j,k=1}^n \left[\frac{1}{(Y_j + Y_k + \varphi)} + \frac{Y_j + Y_k}{(Y_j + Y_k + \varphi)^2} + \frac{2Y_j Y_k + 2}{(Y_j + Y_k + \varphi)^3} \right] \\
&+ \frac{1}{n} \sum_{j,k=1}^n \left[\frac{3(Y_j + Y_k)}{(Y_j + Y_k + \varphi)^4} + \frac{6}{(Y_j + Y_k + \varphi)^5} \right] \\
&- 2 \sum_{j=1}^n \left[\frac{1}{(Y_j + \varphi)} + \frac{Y_j}{(Y_j + \varphi)^2} + \frac{1}{(Y_j + \varphi)^3} \right],
\end{aligned}$$

Meintanis & Iliopoulos (2003) derived the null distribution and proved the consistency of the test and further provided insightful theoretical properties of the test statistic when the MLE and MME are used. These specific cases are studied in a more general setting where it is shown that the test statistics are closed at the boundary $\varphi = \infty$ with the use of a limit statistic. The relation to the first nonzero component of Neyman's smooth test of the limit statistic obtained through the MME is also provided.

In addition to a power study, the proposed tests are calculated on two real data examples. From the power study, Meintanis & Iliopoulos (2003) ultimately concluded that the Laplace transform based test with $\varphi = 2$ leads to highly competitive results against existing tests that were considered in their study.

2.4.3 Tests based on entropy

Baratpour & Khodadadi (2012):

Baratpour & Khodadadi (2012) showed that the CRE for the Rayleigh distribution is $\text{CRE}(X) = \frac{\theta\sqrt{2\pi}}{4}$ (see Section 2.3.5 for a short derivation). The authors defined a new measure of distance between two distributions based on CRE and named it the cumulative Kullback-Leibler (CKL) divergence. If F and G are distributions of two

nonnegative random variables X_1 and X_2 then the CKL is given by

$$\text{CKL}(F, G) = \int_0^\infty \bar{F}(x) \log \frac{\bar{F}(x)}{\bar{G}(x)} dx - \{E(X_1) - E(X_2)\},$$

where $\bar{F}(x) = 1 - F(x)$ and $\bar{G}(x) = 1 - G(x)$ are the survival functions of X_1 and X_2 .

By using the characterisation given in Theorem 2.7 for the Rayleigh distribution and utilizing a discrimination information statistic based on the CKL, a test statistic of the form

$$CK_n = \frac{1}{\bar{X}} \left(\sum_{i=1}^{n-1} \binom{n-i}{n} \left(\log \left(\frac{n-i}{n} \right) \right) (X_{(i+1)} - X_{(i)}) + \sqrt{\frac{\pi}{2}} \sqrt{\frac{\sum_{i=1}^n X_i^3}{3 \sum_{i=1}^n X_i}} \right),$$

can be constructed, where $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$.

Rao et al. (2004) proved the consistency of the CRE and Baratpour & Khodadadi (2012) extends the proof to the test statistic CK_n . Power estimates are given for a nominal significance level of 5% and 1% and the test is applied to a real data example. It was concluded by the authors that the CK_n test appears to be more powerful than the classical tests which were the only other tests considered in the study.

Alizadeh Noughabi, Alizadeh Noughabi & Behabadi (2012):

This test is based on an estimator of the well-known Kullback–Leibler divergence function given by

$$KL(g||g_0) = \int_{-\infty}^{\infty} g(x) \log \left(\frac{g(x)}{g_0(x)} \right) dx,$$

where $g_0(x)$ is the density under the null hypothesis, i.e., the Rayleigh density function and $g(x)$ is the density function of a random variable X . The estimator is

formed by first noting that $KL(g||g_0)$ reduces to

$$KL(g||g_0) = -H(g) - \int_0^\infty g(x) \log\{g_0(x)\} dx,$$

where $H(g) = E[-\log g(X)]$ is the entropy of a random variable X . The sample estimate is then given by

$$KL_{n,m} = -H_{n,m} + 2\log(\hat{\theta}_n) - \frac{1}{n} \sum_{i=1}^n \log(X_i) + 1,$$

where $H_{n,m} = (1/n) \sum_{i=1}^n \log \{(n/2m)(X_{(i+m)} - X_{(i-m)})\}$ is the sample-entropy estimator introduced by Vasicek (1976) and m is a window width restricted to $m \leq n/2$. The choice of $\hat{\theta}_n$ is restricted to the maximum likelihood estimate in $KL_{n,m}$.

The consistency and standard normal asymptotic distribution under the null hypothesis was proven in the same paper by Alizadeh Noughabi et al. (2012). It is also noted that when the sample size is sufficiently large the critical values can be obtained from the limiting normal distribution of $KL_{n,m}$ derived by Song (2002). Critical values in the paper are, however, obtained and provided through Monte Carlo simulations for a sample size of $n = 10$ and $n = 20$ at a significance level of 5% and 1%. Ultimately, results of a power study and the implementation of the test on a real data example are included. In the limited power study of the paper, it was stated that the $KL_{n,m}$ test performed better than the other tests for a uniform alternative, however, for the majority of other alternatives, the Anderson–Darling test (defined in Section 2.4.1) had the greatest power.

Jahanshahi, Habibi Rad & Fakoor (2016):

This entropy-based statistic utilizes the Hellinger distance,

$$D_{g,g_0} = \frac{1}{2} \int_0^\infty \left\{ \sqrt{g(x)} - \sqrt{g_0(x)} \right\}^2 dx, \quad (2.7)$$

instead of the traditionally used Kullback–Leibler divergence which experiences difficulties when the probability density function is zero. It is trivial to see that the Hellinger distance evaluates the deviation of a density $g(x)$ from the hypothesized density $g_0(x)$, which in this case is the Rayleigh density, and the equality only holds when $g(x) = g_0(x)$. By setting the distribution function $G(x) = p$, (2.7) can be rewritten as

$$D_{g,g_0} = \frac{1}{2} \int_0^1 \left(\sqrt{\left(\frac{d}{dp} G^{-1}(p) \right)^{-1}} - \sqrt{\frac{G^{-1}(p) \exp(-(G^{-1}(p))^2/2\theta^2)}{\theta^2}} \right)^2 \frac{d}{dp} G^{-1}(p) dp.$$

Using the approximation $(d/dp \ G^{-1}(p))^{-1} \cong \left\{ \frac{n}{2m} (X_{(i+m)} - X_{(i-m)}) \right\}^{-1}$ leads to the following test statistic

$$DH_{n,m} = \frac{1}{2n} \sum_{i=1}^n \frac{\left[\sqrt{\left(\frac{n}{2m} (X_{(i+m)} - X_{(i-m)}) \right)^{-1}} - \sqrt{\left(X_{(i)} e^{(-X_{(i)}^2/2\hat{\theta}_n^2)} / \hat{\theta}_n^2 \right)} \right]^2}{\left\{ \frac{n}{2m} (X_{(i+m)} - X_{(i-m)}) \right\}^{-1}},$$

where $X_{(i)} = X_{(1)}$ for $i < 1$, $X_{(i)} = X_{(n)}$ for $i > n$ and m is a window width subject to $m \leq n/2$. Jahanshahi et al. (2016) provides a proof for the consistency of the test, and also includes a power study and two real data examples. It was concluded by the authors that the test $DH_{n,m}$ performed as good or better than the considered competitor tests for increasing, decreasing and nonmonotone hazard rates. It is also stated that it was difficult to choose a best performing test for increasing and nonmonotone hazard rate distributions. In the real data examples all the tests did not reject the null hypothesis of Rayleigh distributed data.

Ahrari, Baratpour, Habibirad & Fakoor (2019)

The quantile function of a cumulative distribution function $F(x)$ is given by

$$Q(u) = F^{-1}(u) = \inf\{x : F(x) \geq u\}, \ 0 \leq u \leq 1.$$

With this as starting point, Ahrari et al. (2019) proposed three new distance measures between the quantile functions of two distributions P and Q . The measures are stated to bear resemblance to the Kulback-Leibler divergence measures and Tsallis generalized entropy measure (see, Tsallis 1998).

Conforming to the notation of Ahrari et al. (2019), let Q_1 and Q_2 be the respective quantile functions of two nonnegative random variables X and Y . The three new distance measure are given by

$$\begin{aligned} D_{KL_1}(Q_1\|Q_2) &= \int_0^1 Q_1(x) \log \frac{Q_1(x)}{Q_2(x)} dx - \int_0^1 Q_1(x) dx \log \frac{\int_0^1 Q_1(x) dx}{\int_0^1 Q_2(x) dx}, \\ D_{KL_2}(Q_1\|Q_2) &= \int_0^1 Q_1(x) \log \frac{Q_1(x)}{Q_2(x)} dx - \int_0^1 Q_1(x) dx + \int_0^1 Q_2(x) dx, \\ D_T(Q_1\|Q_2) &= \frac{1}{(\alpha - 1)} \left\{ \int_0^1 Q_1^\alpha(x) Q_2^{1-\alpha}(x) dx - \alpha \int_0^1 Q_1(x) dx \right. \\ &\quad \left. - (1 - \alpha) \int_0^1 Q_2(x) dx \right\}, \end{aligned}$$

with $0 < \alpha < 1$. The authors' prove the divergence measures to be larger or equal to zero if, and only if, $Q_1 = Q_2$. The test statistics based on the aforementioned divergence measures are then

$$\begin{aligned} Q_{KL_1} &= \frac{D_{Q_{KL_1}}(Q_n\|Q_0(\cdot; \hat{\theta}_n))}{\bar{X}_n} \\ &= \frac{1}{n} \sum_{i=1}^n \frac{X_i}{\bar{X}_n} \log \frac{X_i}{\bar{X}_n} - \frac{1}{2} \sum_{i=1}^n \frac{X_{(i)}}{\bar{X}_n} \int_{\frac{i-1}{n}}^{\frac{i}{n}} \log(-2 \log(1-x)) dx + \log \left(\sqrt{\frac{\pi}{2}} \right), \\ Q_{KL_2} &= \frac{D_{Q_{KL_2}}(Q_n\|Q_0(\cdot; \hat{\theta}_n))}{\bar{X}_n} \\ &= \frac{1}{n} \sum_{i=1}^n \frac{X_i}{\bar{X}_n} \log X_i \hat{\theta}_n - 1 + \frac{\hat{\theta}_n}{\bar{X}_n} \sqrt{\frac{\pi}{2}} - \frac{1}{2} \sum_{i=1}^n \frac{X_{(i)}}{\bar{X}_n} \int_{\frac{i-1}{n}}^{\frac{i}{n}} \log(-2 \log(1-x)) dx, \\ Q_T &= D_T(Q_n\|Q_0(\cdot; \hat{\theta}_n)) \end{aligned}$$

$$= \frac{1}{\alpha - 1} \left\{ \sum_{i=1}^n \left(\frac{X_{(i)}^\alpha \hat{\theta}_n^{(1-\alpha)}}{\bar{X}_n} \int_{\frac{i-1}{n}}^{\frac{i}{n}} (-2 \log(1-x))^{\frac{1}{2}(1-\alpha)} dx \right) - \alpha \right\} + \frac{\hat{\theta}_n}{\bar{X}_n} \sqrt{\frac{\pi}{2}},$$

where $\hat{\theta}_n$ is the MLE of θ , $Q_n(t) = X_{(r)}$, $\frac{r-1}{n} < t < \frac{r}{n}$, is the r -th order statistic and the empirical counterpart of Q_1 . Furthermore, $Q_0(x, \theta) = \theta \{-2 \log(1-p)\}^{1/2}$ is the quantile function of the Rayleigh distribution.

The authors prove the test to be consistent but do not provide the asymptotic null distributions for the test. A power study by the authors concluded that the test Q_T performed better than the two other proposed tests and had higher power than the competitor tests for most of the alternatives considered.

2.4.4 Tests based on the Phi-divergence measure

Zamanzade & Mahdizadeh (2017):

Several test statistics can be based on the Phi-divergence measure

$$D_\phi(P_1 || P_2) = \int_{\Omega} \phi \left(\frac{dP_1}{dP_2} \right) dP_2,$$

where P_1 and P_2 are probability measures on the measurable space Ω and $\phi(\cdot)$ is a convex function such that $\phi(1) = 0$ and second derivative $\phi''(1) > 0$ (see, Pardo 2018). Consider the probability density functions $g(x)$, the hypothesized density function $g_0(x)$, and let $D_n(g_0 || \hat{g}_h)$ be a sample estimate of $D_\theta(g_0 || g)$ written in the form

$$D_n(g_0 || \hat{g}_h) = \frac{1}{n} \sum_{i=1}^n \phi \left(\frac{g_0(x_i)}{\hat{g}_h(x_i)} \right),$$

where $\hat{g}_h(x) = (nh)^{-1} \sum_{i=1}^n k((x_i - x)/h)$, k is a kernel function and h is a suitably chosen bandwidth (more details on kernel functions and bandwidth estimates are given in Chapter 3).

New test statistics can now be constructed by choosing appropriate functions for $\phi(\cdot)$. Zamanzade & Mahdizadeh (2017) specifically studied the following selection of functions and resultant test statistics:

- $\phi(t) = -\log(t)$ resulting in the Kullback-Leibler distance with test statistic

$$PKL_n = \frac{1}{n} \sum_{i=1}^n \log \left(\frac{\hat{g}_h(x)}{g_0(x)} \right).$$

- $\phi(t) = \frac{1}{2}(1 - \sqrt{t})^2$ resulting in the Hellinger distance with test statistic

$$PH_n = \frac{1}{2n} \sum_{i=1}^n \left(1 - \left(\frac{g_0(x)}{\hat{g}_h(x)} \right)^{1/2} \right)^2.$$

- $\phi(t) = (t - 1) \log(t)$ resulting in the Jeffreys distance with test statistic

$$PJ_n = \frac{1}{n} \sum_{i=1}^n \left(\frac{g_0(x)}{\hat{g}_h(x)} - 1 \right) \log \left(\frac{g_0(x)}{\hat{g}_h(x)} \right).$$

- $\phi(t) = |t - 1|$ resulting in the Total Variation distance with test statistic

$$PTV_n = \frac{1}{n} \sum_{i=1}^n \left| \frac{g_0(x)}{\hat{g}_h(x)} - 1 \right|.$$

- $\phi(t) = \frac{1}{2}(1 - t)^2$ resulting in the Chi-square distance with test statistic

$$PC_n = \frac{1}{2n} \sum_{i=1}^n \left(1 - \frac{g_0(x)}{\hat{g}_h(x)} \right)^2.$$

A power study was performed in Zamanzade & Mahdizadeh (2017) for sample sizes $n \in \{10, 20, 50\}$ and the results of the tests on two real data sets were included. The authors found that among the proposed tests, the Jeffreys and Hellinger distance tests

performed the best. However, it was stated that the Total Variation and Kullback-Leibler were superior in certain instances. All the proposed tests also did not reject the null hypothesis for the data examples that were known to be Rayleigh distributed.

Torabi, Montazeri & Grané (2016):

Torabi et al. (2016) suggested a new proximity measure which was inspired by the Phi-divergence approach. This measure is used to develop a test for the location-scale family of distribution and specifically implemented to test for the Gaussian distribution. The discrepancy measure between the hypothesized null distribution F_0 (in this case the normal distribution with unknown mean, μ , and variance, σ^2) and the unknown distribution F of the data, is defined as

$$D(F_0||F) = \int_{-\infty}^{\infty} \Psi \left(\frac{1 + F_0(x)}{1 + F(x)} \right) dF(x),$$

where $\Psi(\cdot) : (0, \infty) \rightarrow \mathbb{R}^+$ is continuous decreasing on $(0, 1)$ and increasing on $(1, \infty)$ with $\Psi(1) = 0$. Now, estimating F by the ecdf F_n , leads to the easily calculable test statistic

$$H_n = n^{-1} \sum_{i=1}^n \Psi \left(\frac{1 + F_0(Z_{(i)})}{1 + i/n} \right),$$

where $Z_{(i)} = (X_{(i)} - \hat{\mu})/\hat{\sigma}$ is the scaled observations for the location-scale families with consistent estimators $\hat{\mu}$ and $\hat{\sigma}$. Torabi et al. (2016) discussed possible options for the function $\Psi(\cdot)$ and suggest choosing $\Psi(x) = ((x - 1)/(x + 1))^2$ as it lead to the highest powers for testing normality in their simulation study.

The authors showed the test to be invariant under location-scale transformations and proved the test to be consistent.

We adapt the test of Torabi et al. (2016) specifically for the Rayleigh distribution.

In testing for the Rayleigh distribution the test statistic maintains the form

$$C_n = n^{-1} \sum_{i=1}^n \Psi \left(\frac{1 + F_0(Y_{(i)})}{1 + i/n} \right),$$

with the only change being the scaled observations are now $Y_{(i)} = X_{(i)}/\hat{\theta}_n$, and $F_0(Y_{(i)}) = 1 - \exp(-Y_{(i)}^2/2)$.

2.4.5 A test based on the empirical likelihood ratio

Safavinejad, Jomhoori & Alizadeh Noughabi (2015):

When testing for the Rayleigh distribution, the likelihood ratio tests statistic takes the form

$$R = \frac{\prod_{i=1}^n f_{H_1}(X_i)}{\prod_{i=1}^n f_{H_0}(X_i)} = \frac{\prod_{i=1}^n f_{H_1}(X_i)}{(\prod_{i=1}^n X_i/\theta^{2n}) \exp(-\sum_{i=1}^n X_i/2\theta^2)},$$

where f_{H_1} is the density under the alternative hypothesis and f_{H_0} is the density under H_0 . It is well known (from the Neyman-Pearson lemma) that if both f_{H_0} and f_{H_1} are fully specified, then the likelihood ratio statistic is the most powerful test. However, f_{H_1} is completely unknown and the parameter θ in f_{H_0} is also unknown, and hence requires estimation. To this end, a density-based empirical likelihood technique is employed by Safavinejad et al. (2015) to estimate $\prod_{i=1}^n f_{H_1}(X_i)$.

Given $X_1, X_2, \dots, X_n$ iid from a random sample Vexler & Gurevich (2010) and Safavinejad et al. (2015) states the empirical likelihood function to be $L_p = \prod_{i=1}^n p_i$ where p_i , $i = 1, 2, \dots, n$ are components that maximize the function L_p . The density based likelihood function under H_1 is then

$$L_f = \prod_{i=1}^n f(X_{(i)}) := \prod_{i=1}^n f_i.$$

The approach rests on finding values for f_i that maximizes L_f subject to empirical constraints dependent on H_1 that are exemplified in Vexler & Gurevich (2010) and Safavinejad et al. (2015). The authors conclude that using a Lagrange multiplier method to maximize $\log(f_i)$ yields a usable expression to estimate $f_{H_1}(X_i)$ in the form

$$f_j = \frac{2m}{n(X_{(j+m)} - X_{(j-m)})},$$

where $X_{(j)} = X_{(1)}$, if $j \leq n$ and $X_{(j)} = X_{(n)}$, if $j \geq 1$ and m is a window width. It is evident that this estimator has close ties to the sample-entropy estimator proposed by Vasicek (1976). The test statistic can now be implemented by estimating R with

$$\hat{R}_{n,m} = \frac{\prod_{i=1}^n \{2m/n (X_{(i+m)} - X_{(i-m)})\}}{\left(\prod_{i=1}^n X_i / \hat{\theta}_n^{2n}\right) \exp\left(-\sum_{i=1}^n X_i^2 / 2\hat{\theta}_n^2\right)},$$

where $\hat{\theta}_n$ is the MLE of the parameter θ . Noting the test statistic is dependent on the parameter m , Safavinejad et al. (2015) adopted the modification suggested by Vexler & Gurevich (2010) to consider choices of m in the range $(1, n^{1-\delta})$, $0 < \delta < 1$, which then leads to the tests statistic

$$\hat{R}_n = \frac{\min_{1 \leq m < n^{1-\delta}} \prod_{i=1}^n \{2m/n (X_{(i+m)} - X_{(i-m)})\}}{\left(\prod_{i=1}^n X_i / \hat{\theta}_n^{2n}\right) \exp\left(-\sum_{i=1}^n X_i^2 / 2\hat{\theta}_n^2\right)}.$$

A proof of the asymptotic consistency of the test is provided in the paper by Safavinejad et al. (2015). A power study is conducted, critical values are provided for a range of significance levels and a real data example is also included. The authors conclude that the proposed test $\hat{R}_{n,m}$ is outperformed by the test of Meintanis & Iliopoulos (2003) for almost all of the alternatives considered.

2.4.6 Tests adapted for the Rayleigh distribution

Meintanis (2009):

A goodness-of-fit test has been suggested which employs transformation to uniformity. The test by Meintanis (2009) states that for a suitable transformation $F_\theta(x) = U_\theta(x)$, where $F_\theta(x)$ is the hypothesized distribution under the null hypothesis with unknown parameter θ , a test statistic can be constructed between the empirical characteristic function (ECF) of $U_j = U_\theta(X_j)$ and the characteristic function (CF) of the standard uniform distribution, $\phi_U(t)$. More formally, the test statistic can be written as

$$M_n = \int_{-\infty}^{\infty} |\phi_n(t) - \phi_U(t)|^2 w(t) dt, \quad (2.8)$$

where $w(t)$ is a suitable chosen weight function, $\phi_n(t) = n^{-1} \sum_{i=1}^n \exp(itU_j)$ is the characteristic function of the (unknown) U_j and $\phi_U(t) = t^{-1} \{\sin t + i(1 - \cos t)\}$ is the characteristic function of a uniform random variable on $(0, 1)$. If we now estimate θ by $\hat{\theta}_n$ (in the case of testing for the Rayleigh distribution) we obtain

$$\hat{U}_j = F_{\hat{\theta}_n}(X_j) = 1 - \exp\left(\frac{-X_j}{2\hat{\theta}_n^2}\right).$$

Meintanis (2009) shows that different closed forms for (2.8) can be obtained for different choices of the weight function. Specifically, by choosing $w(t) = \exp(-\varphi|t|)$ with tuning parameter φ , and with U_j replaced by $\hat{U}_j$, M_n becomes

$$\begin{aligned} M1_{n,\varphi} &= \frac{1}{n} \sum_{j=1}^n \sum_{k=1}^n \frac{2\varphi}{\hat{U}_{jk}^2 + \varphi^2} + 2n \left[2 \tan^{-1} \left(\frac{1}{\varphi} \right) - \varphi \log \left(1 + \frac{1}{\varphi^2} \right) \right] \\ &- 4 \sum_{j=1}^n \left[\tan^{-1} \left(\frac{\hat{U}_j}{\varphi} \right) + \tan^{-1} \left(\frac{1 - \hat{U}_j}{\varphi} \right) \right], \end{aligned} \quad (2.9)$$

where $\widehat{U}_{jk}^2 = \widehat{U}_j - \widehat{U}_k$.

Instead of using the probability integral transforms as is the case in (2.9), we adapt the approach for the Rayleigh distribution by considering the following transformation for exponentiality given by Alzaid & Al-Osh (1991):

Theorem 2.9. *Let X_1 and X_2 be two independent observations from a distribution F , then $\frac{X_1}{X_1+X_2}$ is distributed standard uniform $U(0,1)$ if, and only if, F is exponential.*

The transformation in Theorem 2.9 holds true for the Rayleigh distribution by noting that, if $X \sim \text{Ral}(\theta)$, then $X^2/2\theta^2$ follows a standard exponential distribution. This result is now formally stated in Corollary 2.1.

Corollary 2.1. *Let X_1 and X_2 be two independent observations from a distribution G , then $\frac{X_1^2}{X_1^2+X_2^2}$ follows a standard uniform distribution $U(0,1)$ if, and only if, G is the Rayleigh distribution with parameter θ (i.e., $G(x) = 1 - \exp(-x^2/2\theta^2)$).*

Proof. If $X_1 \sim \text{Ral}(\theta)$, then $X_1^2/2\theta^2$ follows a standard exponential distribution. The same holds for X_2 . Now, from Theorem 2.9 we thus have that

$$\frac{\frac{X_1^2}{2\theta^2}}{\frac{X_1^2}{2\theta^2} + \frac{X_2^2}{2\theta^2}} = \frac{X_1^2}{X_1^2 + X_2^2},$$

follows a standard uniform distribution if, and only if, X_1^2 and X_2^2 are exponentially distributed, or then if, and only if, X_1 and X_2 follows a Rayleigh distribution. $\square$

Now, let $\widehat{Z}_{ij} = X_{(i)}^2/(X_{(i)}^2 + X_{(j)}^2)$, $i, j = 1, \dots, n$, $i \neq j$, then the test statistic in (2.8), which is now based on this new transformation to uniformity, becomes

$$M2_{n,\varphi} = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n \sum_{k=1}^n \sum_{\substack{l=1 \\ k \neq l}}^n \frac{2\varphi}{(\widehat{Z}_{ij} - \widehat{Z}_{kl})^2 + \varphi^2}$$

$$\begin{aligned}
& + 2n \left[2 \tan^{-1} \left(\frac{1}{\varphi} \right) - \varphi \log \left(1 + \frac{1}{\varphi^2} \right) \right] \\
& - 4 \sum_{i=1}^n \sum_{j=1}^n \left[\tan^{-1} \left(\frac{\widehat{Z}_{ij}}{\varphi} \right) + \tan^{-1} \left(\frac{1 - \widehat{Z}_{ij}}{\varphi} \right) \right], i, j, k, l = 1, 2, \dots n.
\end{aligned}$$

2.4.7 On the use of weight functions

Weight functions should be accepted as a standard in modern goodness-of-fit testing and in fact, has been used in tests as early as Anderson & Darling (1954). Bickel & Rosenblatt (1973) incorporated a weight function in a L2-distance based test which employs the celebrated kernel density estimator. Tenreiro (2007) studied a version of the Bickel-Rosenblatt test where the weight function is chosen based on the kernel used. Epps (2005) noted a relationship between the test utilizing the empirical characteristic function with a specific choice of weight function and the Anderson-Darling test. Meintanis, Swanepoel & Allison (2014) took the empirical characteristic function based test and introduced a weighted version, where the choice of weight function is reduced to only choosing the value of a tuning parameter.

Aside from being a necessary inclusion in some cases for the convergence of the integral and stability of the test, it is a well-studied fact that the weight function provides a degree of control with regards to the test statistic (see e.g., Baringhaus, Gürtler & Henze 2000, Baringhaus, Ebner & Henze 2017). For the popular choices $w(t) = \exp(-\varphi|t|)$ and $w(t) = \exp(-\varphi t^2)$, choosing a large value for the tuning parameter φ results in a rapidly decaying function. This allows lower order moments to feature more prominently in the test statistic while choosing a small value for the tuning parameter results in a slower decaying function allowing for higher order moments to enter into the test statistic (see e.g., Meintanis 2010). The effectiveness of the weight function ultimately depends on the moment structure of the underlying distribution. Aside from above choices, the weight function can be any function that adheres to certain conditions (depending on the form of the test) and is often chosen

for computational simplicity rather than optimality.

Chapter 3

Kernel density estimation

3.1 Introduction

Probability density estimation has become a powerful tool in the arsenal of non-parametric statistics. Once the unknown density function, which generated data in a real-world event, is estimated it opens the door to probabilities, functions and further subsequent calculations. An intuitive approach to estimating the probability density function is the histogram. The construction of the histogram relies on the choice of an origin x_0 and a bin-width b . The choice of bin-width plays a major role in the estimation process, as it defines the size of the neighbourhood over which frequencies are drawn. The histogram provides an adequate estimation procedure but suffers from some limitations. For example, the discontinuity of histograms makes it difficult to obtain derivatives of the estimates when required (Silverman 1986), and the placement of the bin edges is a sensitive issue (Wand & Jones 1995). These disadvantages motivated more advanced nonparametric density estimation techniques such as kernel density estimation (KDE), variable window-width estimators, series estimators and penalized likelihood estimators. This chapter will focus on kernel density estimation and the interested reader is referred to Pagan & Ullah (1999) for other nonparametric density estimation techniques.

The aim is to present a brief overview of the concepts and techniques associated with this approach, without in-depth nor detailed concerns of the technical and mathematical aspects of the various concepts presented. Kernel density estimation and aspects associated with kernel density estimation are used in existing tests that are discussed in Section 2.4.4 and will also be used in developing a new goodness-of-fit test for the Rayleigh distribution (see Section 4.3). Both the existing tests and the newly developed test will be implemented in the simulation study in chapter 5. The layout of this chapter follows: In Section 3.2 the methodology and notation used throughout the chapter will be introduced. Section 3.3 contains various discrepancy measures that will be used in the motivation of several key concepts. The error criteria of interest are the mean square error (MSE) and the mean integrated square error (MISE). In Section 3.4 a brief overview of kernel functions are given. Section 3.5 deals with bandwidth selection. Section 3.6 will handle the discussion and implementation of a few selected boundary-corrected techniques and the effect of boundary-bias in certain situations.

This chapter relies on the content of Silverman (1986), Wand & Jones (1995) and Liebenberg (2014). Where applicable concepts and equations are with reference to these authors' work with notation adopted from the relevant text and publications.

3.2 Notation and methodology

Let $X_1, X_2, \dots, X_n$ be independent and identically distributed (iid) continuous realizations of a random variable X with unknown probability density function $f(x)$. The intuitive approach of the Parzen-Rosenblatt kernel density estimate, can be thought of as the average of some weight functions (W_i) where the weight function depends on the distance of observations X_i from x together with the sample size, n

and bandwidth, h :

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n W_i(x, h).$$

To improve on the idea of the histogram, the nature of the weight function should be such that values far from x should produce small values. Therefore, choosing the weight as a function that gradually applies less weight to points further away from the center of the neighbourhood leads to the formulation of the KDE by Rosenblatt (1956) and Parzen (1962) as

$$\hat{f}(x; h) = \frac{1}{nh} \sum_{i=1}^n k\left(\frac{x - X_i}{h}\right), \quad (3.1)$$

where h is the bandwidth, $k(\cdot)$ is the kernel function and n is the sample size. The kernel, k , is chosen to be a symmetric unimodal probability density that satisfies $\int k(z)dz = 1$, $\int zk(z)dz = 0$ and $\int z^2k(z)dz > 0$. More formally, the value of the kernel estimate at the point x is simply the average of the n kernel ordinates at that point. Combining contributions from each data point means that for the regions where there are many observations, the kernel estimates should assume large values. The opposite should occur in areas where there are few observations. The interested reader is referred to Wand & Jones (1995) for a more thorough explanation of kernel density estimation. The procedure is illustrated in Figure 3.1 in an attempt to estimate the density function of a sample obtained from the standard normal distribution $\phi(x)$. In the illustration, the kernel function was chosen as the normal density function.

We proceed by introducing the following notation that will be used throughout the chapter:

- A shorthand method of writing (3.1) is $\hat{f}(x; h) = \frac{1}{n} \sum_{i=1}^n k_h(x - X_i)$, where

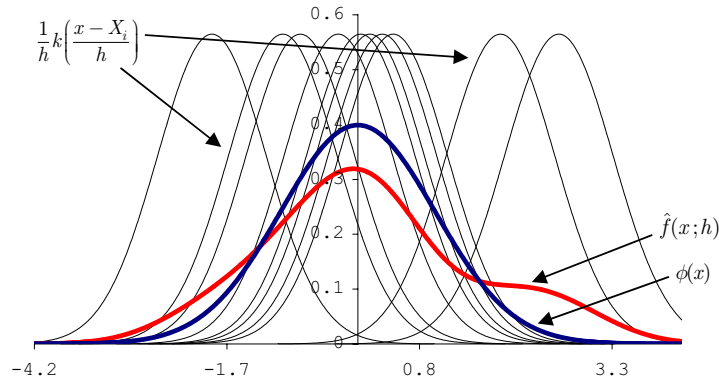


Figure 3.1: Construction of the kernel density estimate (Koekemoer 2004).

$$k_h(\cdot) = \frac{1}{h} k\left(\frac{\cdot}{h}\right).$$

- The convolution of two functions, say f and g , is defined by $(f * g)(x) = \int f(x - y)g(y)dy$.
- The expected value of $\hat{f}(x; h)$ is $E\hat{f}(x; h) = \int k_h(x - y)f(y)dy$.

3.3 Measures of fit

To proceed, several measures of fit for kernel density estimation are evaluated. These measures are generally used to assess how well a kernel estimate fits the data. The measures of fit will be used to discuss several properties and extensions of kernel density estimation such as bandwidth selection and boundary-corrected kernel density estimation. We will start with a well-known local distance measure in the form of the mean square error (MSE) and relate this to more general error criteria in the form of the integrated square error (ISE) which we will ultimately extend to the mean integrated square error (MISE).

The MSE is a measure of fit that measures discrepancy at a single point x on the real line. A well known advantage of the MSE is that it can be split into a variance

and bias component which accounts for a more detailed analysis of the performance of the kernel density estimator. The MSE can be presented as

$$\begin{aligned} MSE \left[\hat{f}(x; h) \right] &= Var \left[\hat{f}(x; h) \right] + \left\{ E \hat{f}(x; h) - f(x) \right\}^2 \\ &= Var \left[\hat{f}(x; h) \right] + \left\{ Bias \left[\hat{f}(x; h) \right] \right\}^2, \end{aligned} \quad (3.2)$$

where $Var \left[\hat{f}(x; h) \right] = n^{-1}(K_h^2 * f)(x) - n^{-1}(K_h * f)^2(x)$ and $Bias \left[\hat{f}(x; h) \right] = (K_h * f)(x) - f(x)$. We can rewrite (3.2) in the following form by writing out the convolution and simplifying the terms:

$$\begin{aligned} MSE \left[\hat{f}(x; h) \right] &= n^{-1} \left[\int K_h^2(x-y) f(y) dy \right] - n^{-1} \left[\int K_h(x-y) f(y) dy \right]^2 \\ &\quad + \left\{ \left[\int K_h(x-y) f(y) dy \right] - f(x) \right\}^2. \end{aligned}$$

Due to the fact that $\hat{f}(x; h)$ is an estimator of a function $f(x)$, the MSE is not an ideal measure of fit since it is evaluated at a single point x . A more appropriate measure is the integrated L_2 distance between $\hat{f}(x; h)$ and $f(x)$ defined as

$$ISE \left[\hat{f}(\cdot; h) \right] = \int \left\{ \hat{f}(x; h) - f(x) \right\}^2 dx.$$

In words, this measure takes the square distance between $\hat{f}(x; h)$ and $f(x)$ and evaluates it over all values of x , not just a single value. A more global measure of fit, the MISE, can be obtained by extending the ISE to incorporate multiple data sets from the density $f(x)$:

$$\begin{aligned} MISE \left[\hat{f}(\cdot; h) \right] &= E \left\{ ISE \left[\hat{f}(\cdot; h) \right] \right\} \\ &= (n)^{-1} \int K_h^2(x) dx + (1 - n^{-1}) \int (K_h * f)^2(x) dx \\ &\quad - 2 \int (K_h * f)^2(x) f(x) dx + \int f(x)^2 dx. \end{aligned} \quad (3.3)$$

The fact that (3.3) depends on the bandwidth h in a complicated way can be remedied by deriving an asymptotic approximation for the MISE, i.e.,

$$AMISE \left[\hat{f}(x; h) \right] = (nh)^{-1} R(k) + \frac{1}{4} h^4 \mu_2(k)^2 R(f''). \quad (3.4)$$

where $\mu_2(k) = \int z^2 k(z) dz$, $R(k) = \int k(z)^2 dz$ and $R(f'') = \int f''(z)^2 dz$ with $f''(\cdot)$ denoting the second derivative of $f(\cdot)$. The above large sample approximation requires several assumptions about the density function $f(x)$, the bandwidth h , and the kernel k , which are listed below.

- The second derivative of f is continuous, square integrable and ultimately monotone, i.e., monotone over both $(-\infty, -M)$ and (M, ∞) for some $M > 0$.
- The bandwidth h is a non-random sequence of positive numbers.
- The kernel k is a bounded density function with finite fourth moment and is symmetric around zero.

3.4 Kernel function choices

There are many choices for the kernel function, some of which are given in Table 3.1 and Figure 3.2. Kernel functions are generally subject to $\int k(x) dx = 1$, $\int x k(x) dx = 0$, $\int x^2 k(x) dx < \infty$, and $k(x) \geq 0$ for all x , which are necessary properties for the use in kernel density estimation. Wand & Jones (1995) discusses the choice of kernel functions in terms of efficiencies and canonical kernels. It is noted that the efficiencies of what can be considered suboptimal kernels are still very high and they differ only slightly from each other. It can therefore be concluded that one loses very little in performance if suboptimal kernel functions are used. The most popular kernels that are generally used are the Epanechnikov kernel (which can be considered the optimal kernel in terms of efficiencies) and the normal density function. Since the choice of kernel function can be left to the popular choices, we turn our attention to the choice

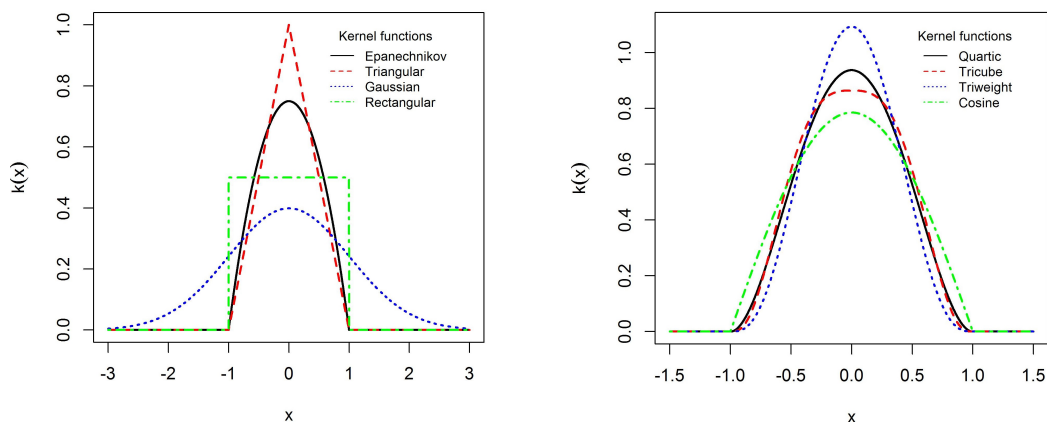


Figure 3.2: Various shapes of kernel functions.

of bandwidth which, in contrast to the kernel function choice, has a large effect on the kernel density estimate.

Table 3.1: List of kernels functions.

Name of kernel function	$k(u)$
Normal	$\frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$
Rectangular	$\frac{1}{2}$
Triangular	$1 - u $
Epanechnikov	$\frac{3}{4}(1 - u^2)$
Quartic	$\frac{15}{16}(1 - u^2)^2$
Triweight	$\frac{35}{32}(1 - u^2)^3$
Tricube	$\frac{70}{81}(1 - u ^3)^3$
Cosine	$\frac{\pi}{4} \cos\left(\frac{\pi}{2}u\right)$

3.5 Bandwidth selection

The choice of bandwidth plays a critical role in kernel density estimation. Choosing the bandwidth too large may obscure important features of the underlying density being estimated and choosing the bandwidth too small may result in an overly variable estimate (this occurrence is similarly observed with binwidth selection in

histograms). In the following section we will specifically explore the rule-of-thumb bandwidth selector.

Silverman (1986) declared that if the AMISE in (3.4) is minimized, it is possible to find an optimal value for the bandwidth h , which can be written in the form

$$h_{opt} = \mu_2(K)^{-2/5} \left\{ \int K(t)^2 dt \right\}^{1/5} \left\{ \int f''(x)^2 dx \right\}^{-1/5} n^{-1/5}. \quad (3.5)$$

The formula for the optimal bandwidth is useful in many ways, but it is still dependent on the unknown density being estimated through $\int f''(x)^2$. From the above expression note that $h \rightarrow 0$ as $n \rightarrow \infty$ at a very slow rate. Furthermore, small values of h will be appropriate for more rapid fluctuating densities. An optimal bandwidth can be obtained by using expression (3.5) and choosing h with reference to some standard family of densities (Silverman 1986, p.40). This can be done by assigning a value to the term $\int f''(x)^2 dx$ in (3.5) from a standard family of distributions. By substituting $\int f''(x)^2 dx$ with the standard Gaussian distribution in (3.5), the rule-of-thumb bandwidth of Silverman (1986) is obtained:

$$\begin{aligned} h_{rot} &= (4\pi)^{-1/10} \left(\frac{3}{8} \pi^{-1/2} \right)^{-1/5} \sigma n^{-1/5} \\ &= 1.06 \sigma n^{-1/5}. \end{aligned} \quad (3.6)$$

Note that the parameter σ still needs to be obtained and a quick and easy way of accomplishing this would be to estimate σ from the data. Silverman (1986) suggests estimating σ using a robust estimate of spread, namely

$$\hat{\sigma}^* = \min(S_{n-1}^2; IQR/1.34),$$

where $S_{n-1}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ and $IQR = Q_3 - Q_1$ with Q_1 being the first quartile and Q_3 the third quartile. It is accepted that h_{rot} will cope moderately well

with bimodal densities while still adequately handling unimodal densities. Wand & Jones (1995) also note that the use of $\hat{\sigma}^*$ guards against outliers if f has heavy tails. The ease of implementation and the desirable properties of the rule-of-thumb bandwidth, with the robust estimate of spread, makes it an overwhelmingly popular choice when performing kernel density estimation. The properties combined with the good performance of the kernel density estimate when this bandwidth choice is used, also motivates our use of it in the simulation study in Chapter 5. Theoretical properties of this bandwidth can be found in Wand & Jones (1995) and Silverman (1986). The interested reader is also referred to Hall & Marron (1987), Park & Marron (1990), Sheather & Jones (1991) and Heidenreich, Schindler & Sperlich (2013) for other plug-in bandwidth choices.

3.6 Boundary-corrected kernel density estimation

It is a well known fact that kernel density estimation experiences difficulties at the boundaries of the density being estimated (see, Wand & Jones 1995, Karunamuni & Alberts 2005, Marron & Ruppert 1994). This occurrence is described as boundary-bias and it is due to the kernel density estimate giving weight to the area outside of the data range where there are no data values. This can be explained by noting that the support of $\hat{f}(x; h)$ is the range $[X_{(1)} - h; X_{(n)} + h]$ where h is the chosen bandwidth. The kernel extends into these regions, $[X_{(1)} - h; X_{(1)}]$ and $[X_{(n)}, X_{(n)} + h]$, devoid of data and consequently penalizes the estimation in the boundary area containing data. For ease of presentation and without loss of generality, consider the data to span the range $[0, \infty)$ i.e., the densities of interest are bounded on the interval $[0, \infty)$. The boundary region for this support is then $[0, h)$ and the interior lies on (h, ∞) . Furthermore, to investigate the behavior of the kernel density estimate at the boundary, consider a sequence $x_n = \alpha h$ where $\alpha \in [0, 1)$. This sequence converges to the boundary such that it is always within h from the boundary. Now, taking

$E[\widehat{f}(x; h)] = \int_{-1}^{\alpha} k(t)f(x - ht)dt$ and by performing a Taylor expansion it is possible to show that

$$\begin{aligned} E[\widehat{f}(x; h)] &= f(x) \int_{-1}^{\alpha} k(t)dt - hf'(x) \int_{-1}^{\alpha} tk(t)dt \\ &\quad + \frac{h^2}{2} f''(x) \int_{-1}^{\alpha} t^2 k(t)dt + o(h^2). \end{aligned} \quad (3.7)$$

The boundary-bias problem is clear with closer inspection of (3.7) and noting that $\int_{-1}^{\alpha} k(t)dt$ no longer equates to 1. This fact leads to the conclusion that the kernel density estimate is not consistent at the boundaries of the support and further inspection will exhibit that the order of bias at the boundary stands at $O(h)$. Quantifiable, the expected value of the kernel density estimate asymptotically only reaches $(1/2)f(0)$ in the boundary area (see, Marron & Ruppert 1994).

Many authors worked on fixing this drawback and there exists a vast literature on the subject. In this section we are selectively going to explore only a few key approaches that can broadly be group into the following three categories:

- Reflection and pseudo-data methods.
- Transformation methods.
- Boundary kernel methods.

3.6.1 Reflection and pseudo-data methods

Since the lack of data outside of the data range affects the kernel density estimation, a first approach involves adding data in this area. Various authors, including Boneva, Kendall & Stefanov (1971) and Schuster (1985) pursued this approach, the details of which are given in the following sections.

Reflection method of Boneva et al. (1971)

The reflection method was introduced and evaluated by Boneva et al. (1971) and rest on the idea of adding the negative counterpart of the data $\{-X_1, -X_2, \dots, -X_n\}$, to the sample and in effect creating a new sample $\{-X_1, -X_2, \dots, -X_n, X_1, X_2, \dots, X_n\}$. To accommodate this technique the KDE expression in (3.1) is modified to

$$\hat{f}_R(x; h) = \begin{cases} \frac{1}{nh} \sum_{i=1}^n \left\{ k\left(\frac{x-X_i}{h}\right) + k\left(\frac{x+X_i}{h}\right) \right\} & , \ x \geq 0 \ , \\ 0 & , \ x < 0 \ . \end{cases}$$

Mirror-image kernel density estimator of Schuster (1985)

Schuster (1985) presented a ‘mirror-image’ modification to $\hat{f}(x)$ which has support $[c, \infty)$. Modifications of this method as well as the *mirror-image* concept can be studied from Schuster (1985). We only state the results of the paper below.

In the case where $X \geq c$, with c some known constant, a new sample can be constructed from $\{X_1, X_2, \dots, X_n\}$ with $Y_i = S_i(X_i - c)$, $i = 1, 2, \dots, n$ where $\{S_1, S_2, \dots, S_n\}$ are iid random variables assuming values

$$S_i = \begin{cases} +1, & \text{with probability } p \ , \\ -1, & \text{with probability } 1 - p \end{cases}$$

with $p = 0.5$.

The estimator for $f(x)$, presented by Schuster (1985), is then

$$\tilde{f}_S(x) = \begin{cases} g_n(x - c) + g_n^-(c - x), & \text{if } x \geq c \\ 0, & \text{if } x < c \end{cases} \ ,$$

where $g_n(x; h)$ is the KDE and $g_n^-(x)$ is the KDE with kernels $k^-(x) = k(-x)$.

However, Schuster (1985) further noted that if the kernels are symmetric, $\tilde{f}_S(x)$ does not depend on the randomization introduced by $\{S_1, S_2, \dots, S_n\}$. The density kernel estimator can therefore be written as

$$\hat{f}_S(x; h) = (nh)^{-1} \sum_{i=1}^n \left\{ K\left(\frac{x - X_i}{h}\right) + K\left(\frac{x - 2c + X_i}{h}\right) \right\}.$$

In the case where c and d are known constants and $c < d$, this estimator can be extended to

$$\begin{aligned} \hat{f}_S(x; h) &= \hat{f}_n^*(2c - x) + \hat{f}_n^*(x) + \hat{f}_n^*(2d - x) \\ &= (nh)^{-1} \sum_{i=1}^n K\left(\frac{x - 2c + X_i}{h}\right) + (nh)^{-1} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \\ &\quad + (nh)^{-1} \sum_{i=1}^n K\left(\frac{x + 2d - X_i}{h}\right), \end{aligned}$$

for $x \in [c, d]$ and zero elsewhere. For this density estimate kernels are limited to bounded, symmetrical kernels with support $[-1, 1]$, and bandwidth $0 < h < (d - c)/2$.

Pseudo-data kernel density estimator of Cowling & Hall (1996)

The pseudo-data method is similar to the reflection method in that it adds data outside the boundary. This is achieved by first transforming the data and then placing the transformed values on the outside of the data range. The data is obtained with the so-called pseudo-data generator derived from the relationship $E[X_i] = G(i/n) + O(n^{-1+\delta})$, $\forall \delta$ and $G = F^{-1}$. Cowling & Hall (1996) estimates an extension of G on $(-\infty, 0]$ by fitting a zero-intercept r^{th} -degree polynomial to the points $(A_1 i/n; X_{A_1 i}), (A_2 i/n; X_{A_2 i}), \dots, (A_s i/n; X_{A_s i})$ where $A_1, A_2, \dots, A_s$ are integer values and $s \geq r$. The pseudo-data generator then takes the form

$$X_{(-i)} = \sum_{j=1}^s a_j X_{(A_j i)}, \quad 1 \leq i \leq n/\max(A_i),$$

where $\{a_1, a_2, \dots, a_s\} \in \mathbb{R}$, such that $\sum_{i=1}^s a_i A_i^j = (-1)^j$, $1 \leq j \leq r$. Based on a numerical study, Cowling & Hall (1996) ultimately suggested a three-point rule based on the pseudo-data generator:

$$X_{(-i)} = -5X_{(i/3)} - 4X_{(2i/3)} + 10/3X_{(i)} ,$$

where $X_{(t)}$ linearly interpolates among $0, X_{(1)}, X_{(2)}, \dots, X_{(n)}$. The KDE is then applied in the modified form:

$$\hat{f}_{CH}(x; h) = \frac{1}{nh} \left\{ \sum_{i=1}^n k\left(\frac{x - X_i}{h}\right) + \sum_{i=1}^m k\left(\frac{x + X_{(-i)}}{h}\right) \right\} , \quad m \leq n .$$

It is important to note that the integer m in the above estimator is of a larger order than that of nh but of smaller order than n . It is concluded in a study by Cowling & Hall (1996) that the pseudo-data method exhibited a lower MSE than the reflection estimator.

3.6.2 Transformation methods

The transformation methods rests on the idea of applying a one-to-one transformation function to the data. Take, without loss of generality, the transformation to be $g : [0, \infty) \rightarrow [0, \infty)$. The KDE is then applied to the transformed data set, $g(X_1, X_2, \dots, X_n)$. Note that the kernel density estimate is calculated on the transformed data which will result in an estimate for the probability density function of the transformed data. This can potentially be problematic in terms of interpretation.

Transformation method of Marron & Ruppert (1994)

Marron & Ruppert (1994) describe their method as a three-step process. In the first step a choice is made on the transformation function g . The function is chosen from a parametric family and this is done to ensure that the density of $g(X_1, X_2, \dots, X_n)$

has a first derivative that is approximately equal to zero at the boundaries of the support. The second step involves applying the mirror-image kernel density estimator of Schuster (1985) to the transformed data. Finally, the estimator is converted to a density estimate on the original data.

On the topic of choosing a transformation g , Marron & Ruppert (1994) suggest the transformation to be restricted to a smooth, monotonically increasing function, $Y_i = g(X_i)$. Furthermore, let f_Y be the density of the transformed data Y_i . The approach of Marron & Ruppert (1994) is to choose the function g such that f_Y is flat at the boundaries. This results in f_Y being a reflected kernel estimator. The authors suggest a method to obtain g in the form of a quartic polynomial and provide an involved algorithm to ultimately achieve bias reduction with their three-step process.

Transformed-reflection kernel density estimate of Karunamuni & Alberts (2005)

Karunamuni & Alberts (2005) combined the transformation and reflection based estimators and suggested a generalized kernel density estimator which takes the form

$$\hat{f}_{KA}(x; h) = \frac{1}{nh} \sum_{i=1}^n \left\{ k\left(\frac{x + g_1(X_i)}{h}\right) + k\left(\frac{x - g_2(X_i)}{h}\right) \right\},$$

where g_1 and g_2 are transformation functions and are chosen such that the bias of the estimator is of order $O(h^2)$, $\forall x \geq 0$. The proposed density estimator is a generalized estimator in the sense that it reduces to the standard reflection estimator of Boneva et al. (1971) defined in section (3.6.1) when $g_1 = g_2 = t$. On the topic of transformations (g_1, g_2) , Karunamuni & Alberts (2005) states that when the transformation

pair is chosen to be $g = g_1 = g_2$ it must satisfy

$$g''(0) = \frac{2f'(0) \int_c^1 (t-c)k(t)dt}{f(0) \left(c + 2 \int_c^1 (t-c)k(t)dt \right)},$$

where c is a point estimate in the boundary region $[0, h)$ of the support $[0, \infty)$. This approach causes the transformation g to be locally adaptive at the boundary and reduces to the KDE in the interior. If g_1 and g_2 are chosen to be different transformations Karunamuni & Alberts (2005) suggests possible choices to satisfy $g_1''(0) = 2d$ and $g_2''(0) = 0$ where $d = f''(0)/f(0)$. The authors specifically study the density estimator of the form

$$\hat{f}_n^*(x, h) = \frac{1}{nh} \sum_{i=1}^n \left\{ k\left(\frac{x + \hat{g}_c}{h}\right) + k\left(\frac{x - \hat{g}_c}{h}\right) \right\},$$

where $\hat{g}_c(u) = u + (1/2)\hat{d}k'_c u^2 + \lambda_0(\hat{d}k'_c)^2 u^3$, $k'_c = 2 \int_c^1 (t-c)k(t)dt / (c + 2 \int_c^1 (t-c)k(t)dt)$ and λ_0 is an integer such that $12\lambda_0 > 1$. Furthermore, $\hat{d}$ is an algorithmic estimated pilot bandwidth. It is concluded by the authors that their density estimator possesses desirable attributes such as the estimators non-negative traits, but also state that among the competitor choices in the study none dominates outright in performance.

3.6.3 Boundary kernel methods

Boundary kernel methods differ from reflection and transformation methods in that the kernels adapt to the shape of the density in the boundary regions.

Cut-normalize method of Gasser & Müller (1979)

Recall the expression of $E[\hat{f}(x; h)]$ given in (3.7):

$$E[\hat{f}(x; h)] = f(x) \int_{-1}^{\alpha} k(t)dt - hf'(x) \int_{-1}^{\alpha} tk(t)dt$$

$$+\frac{h^2}{2}f''(x)\int_{-1}^{\alpha}t^2k(t)dt+o(h^2) .$$

Gasser & Müller (1979) suggested that an intuitive approach to fixing the boundary-bias problem would be to divide the kernel density estimate by $\int_{-1}^{\alpha} k(t)dt$ in order to renormalize each kernel. This ensures that each kernel integrates to 1 and that the estimator is consistent at the boundary. This consistency can be verified by writing the kernel density estimate as

$$\begin{aligned}\hat{f}_{GM}(x;h) &= \frac{\hat{f}(x;h)}{\int_{-1}^{\alpha} k(t)dt} \\ &= (nh)^{-1} \left(\int_{-1}^{\alpha} k(t)dt \right)^{-1} \sum_{i=1}^n \left(\frac{x - X_i}{h} \right) ,\end{aligned}$$

and obtaining the expected value of $\hat{f}_{GM}$:

$$E \left[\hat{f}_{GM}(x;h) \right] = f(x) - hf'(x) \frac{\int_{-1}^{\alpha} tk(t)dt}{\int_{-1}^{\alpha} k(t)dt} + O(h) .$$

Note, the boundary-bias of the estimator is that of order $O(h)$ and while it is an improvement it still has a drawback when considering the bias of order $O(h^2)$ in the interior region.

Simple boundary-correction method of Jones (1993)

An amelioration on the cut-renormalization method of Gasser & Müller (1979) is to attain a bias of order $O(h^2)$ at the boundary in order to match the order of bias in the interior. An approach that gained some traction is to use a kernel function that differs for each value in the sequence $x_n = \alpha h$, e.g., a kernel generated by the function

$$v(z; \alpha) = \frac{\int_{-1}^{\alpha} z^2 k(z) dz - \int_{-1}^{\alpha} z k(z) dz}{\int_{-1}^{\alpha} k(z) dz \int_{-1}^{\alpha} z^2 k(z) dz - \int_{-1}^{\alpha} z k(z) dz} k(z) I(-1 < z < \alpha) .$$

Jones (1993) adopted the variable kernel approach and formulated a method where generalized jackknifing is employed to seek the linear combination of

$$v(z; \alpha) = \frac{\left(\int_{-1}^{\alpha} x\ell(x)dx\right)k(x) - \left(\int_{-1}^{\alpha} xk(x)dx\right)\ell(x)}{\left(\int_{-1}^{\alpha} x\ell(x)dx\right)\left(\int_{-1}^{\alpha} k(x)dx\right) - \left(\int_{-1}^{\alpha} x\ell(x)dx\right)\left(\int_{-1}^{\alpha} \ell(x)dx\right)}, \quad (3.8)$$

that satisfies the $O(h^2)$ bias at the boundary. Note the use of two functions $k(\cdot)$ and $\ell(\cdot)$ in the above expression and consider the notion of how to choose the appropriate functions. Jones (1993) provides several choices on the matter and Jones & Foster (1996) motions to choose the function $\ell(\cdot)$ such that it is a function of $k(\cdot)$ as the resultant boundary-correction would then be derived entirely from the interior kernel function. To investigate one possible choice of the above mentioned kernel functions we conform to the notation used in Jones (1993). The formulation of the variable kernel approach with the use of generalized jackknifing related in (3.8) under the adopted notation states to seek a linear combination of

$$\widehat{f}_J(x) = \alpha_x \bar{f}(x) + \beta_x \tilde{f}(x). \quad (3.9)$$

In this let $\bar{f}(x) = n^{-1} \sum_{i=1}^n k_h(x - X_i) / \int_{-1}^p k(u)du$ and $n^{-1} \sum_{i=1}^n \ell_h(x - X_i) / \int_{-1}^p \ell(u)du$ with $p = x/h$. Furthermore,

$$\begin{aligned} \alpha_x &= c_1(p)a_0(p) / \{c_1(p)a_0(p) - a_1(p)c_0(p)\}, \\ \beta_x &= -a_1(p)c_0(p) / \{c_1(p)a_0(p) - a_1(p)c_0(p)\}, \end{aligned}$$

with $c_r = \int_{-1}^p u^r \ell(u)du$ and $a_r = \int_{-1}^p u^r k(u)du$. One possible choice for $\ell(\cdot)$ is then $\ell(u) = uk(u)$ which results in a kernel function of the form $v(x) = (L_x + xM_x)k(x)$ with

$$L_x = a_2(p) / \{a_2(p)a_0(p) - a_1^2(p)\},$$

and

$$M_x = -a_1(p) / \{a_2(p)a_0(p) - a_1^2(p)\} .$$

The disadvantage with general jackknifing boundary-correction is that it can produce negative values near the boundary. Jones & Foster (1996) however provides a solution in the form of a negative-correction estimator:

$$\hat{f}_{JF}(x) = \bar{f}(x) \exp \left\{ \frac{\hat{f}_J(x)}{\bar{f}(x)} - 1 \right\} ,$$

where $\hat{f}_J(x)$ and $\bar{f}(x)$ is defined in (3.9). The estimator $\hat{f}_{JF}(x)$ will be nonnegative since k and $\bar{f}(x)$ are nonnegative and the remaining terms reside in the exponent.

Adaptive Beta boundary-correction method:

Gamma and Beta kernel estimators were proposed in Chen (1999) and Chen (2000) and are based on the gamma and beta density function which varies the amount of smoothing by naturally varying the kernel shape without explicitly changing the value of the bandwidth. These estimators are claimed to be free of boundary bias, non-negative and achieve the optimal rate of convergence for the mean integrated square error.

Zhang (2010) investigated this claim and suggests that the gamma kernel estimator is the reflection estimator with double-exponential kernel at $x = 0$, which makes it boundary-bias free only if the condition $f'(0) = 0$ holds.

Zhang & Karunamuni (2010) later also considered the beta kernel estimates and suggested that the kernels only exhibit the desired properties of being nonnegative and show superior boundary performance when certain conditions are adhered to. If these conditions are not present it was stated that the adaptive beta kernels are similar to the reflection estimator. The reflection kernel suffers when the density

being estimated does not have the shoulder property, i.e., $f'(0) = 0$, and this restriction carries over to the beta kernel estimator. Zhang & Karunamuni (2010) claims that the study by Chen (2000) was ultimately confounded by non-optimal bandwidth choices and by incomparable quantities.

3.6.4 Numerical investigation

In what follows we conduct a small and by no means comprehensive numerical study of a few selected boundary-corrected methods. The aim of this investigation is to comprehend and visualize the boundary-bias problem. Similarly, this aspires to visualize how the chosen boundary-corrected methods adjusts and compensates for this flaw.

In this investigation, 10 000 random samples were generated from a $Ral(1)$ -distribution. The kernel density estimate and the boundary-corrected methods were calculated on the 10 000 samples. The boundary-corrected kernels under investigation are as follows, the reflection method ($\hat{f}_S$) of Schuster (1985), the cutnorm procedure ($\hat{f}_{GM}$) of Gasser & Müller (1979) and the simple boundary kernel ($\hat{f}_J$) of Jones & Foster (1996). These methods were chosen for the ease-of-use and accessibility to researchers offered by the statistical software R, within the package *evmix* (see, Hu & Scarrott 2018). The bandwidth was chosen with the rule-of-thumb method h_{rot} in (3.6). The calculations were repeated for a sample size of $n = 20$ and $n = 200$. The sample sizes were chosen to inspect the behavior in a small and large sample situation. Figures 3.3 and 3.4 exhibit the plots for the KDE and boundary-corrected methods for the first 100 generated samples. The MISE, bias and variance components are displayed in Figure 3.5. The left column of plots in all figures correspond to the sample size $n = 20$ and the right column correspond to the sample size $n = 200$. The boundary-bias problem caused by the kernel spilling out of the range of data is clearly visible in the left boundary of the KDE plots. Notice the problem is persistent for both the $n = 20$

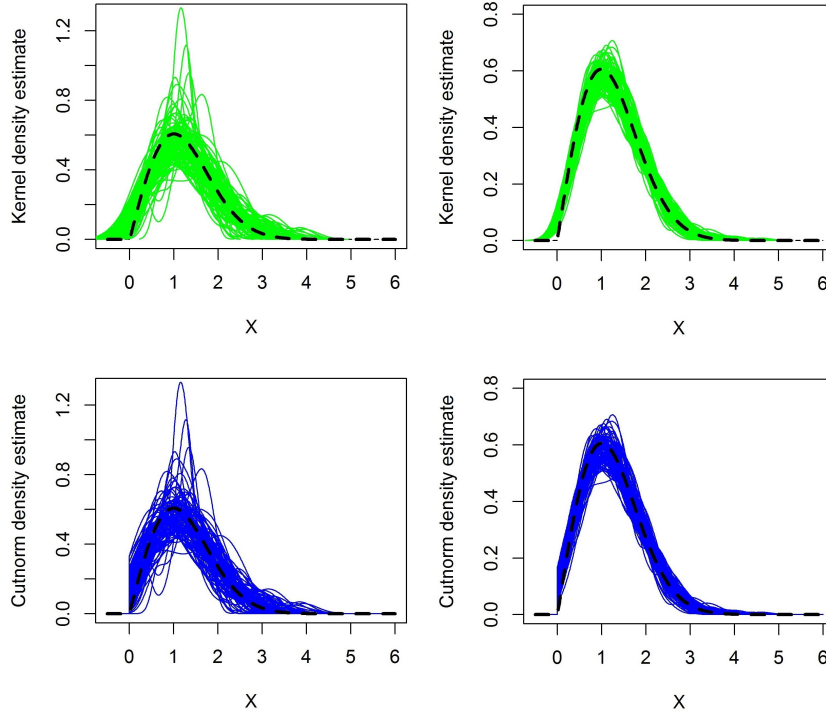


Figure 3.3: Kernel density estimate plots for the KDE and cutnorm boundary-corrected method.

and $n = 200$ cases even though the variance seems to be decreased for the $n = 200$ case. The variability and the chance of peculiar density estimates as a function of odd samples are also clearly seen in the small sample case. If we turn our attention to the cutnorm procedure in the bottom row of Figure 3.3 we clearly see the correction this procedure offers. The extent of the boundary-bias is severely decreased on the left boundary and disappears below zero. As mentioned before, the cutnorm procedure remedies the inconsistency of the KDE in the boundary area. The variability persists in the interior and this is because the cutnorm procedure is equivalent to the KDE in the interior. Figure 3.4 exhibits the performance of the reflect and simple boundary corrected density estimates. Shifting our attention to the reflect method in the top row we see that it also attempts to remedy the boundary-bias problem albeit to a slightly worse extent than that of the cutnorm method. This however,

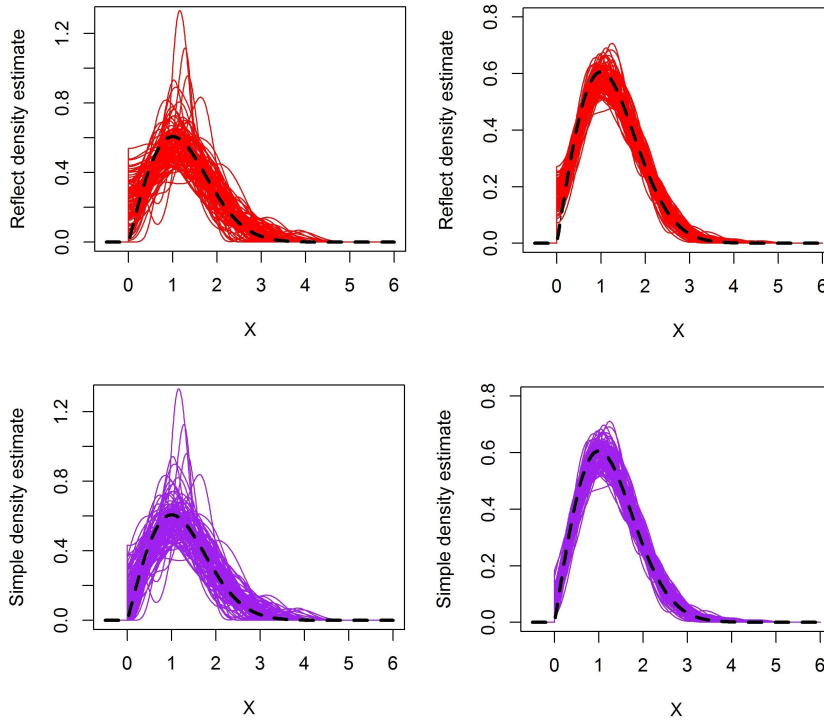


Figure 3.4: Kernel density estimate plots for the reflect and simple boundary-corrected method.

is not surprising as the reflect method was specifically designed for the cases where $f'(0) = 0$ or $f'(a) = 0$. The $Ral(1)$ -density does not adhere to this setup in that it produces $f'(0) = 1/\theta^2$.

The simple boundary corrected procedure on the other hand seems to also perform adequately. For a better overall view we turn our attention to the MISE error criterion in Figure 3.5. The variance component for both the sample sizes seems relatively equivalent and there is not much difference in the bias component between the $\hat{f}_{GM}$ and $\hat{f}_J$ estimates for $n = 20$. The bias component for these two methods seem to differ more substantially for the $n = 200$ sample size but on closer inspection it is only a moderate improvement for the $\hat{f}_J$ method. The overall MISE echoes the same conclusion that there is not much to choose between the two methods consid-

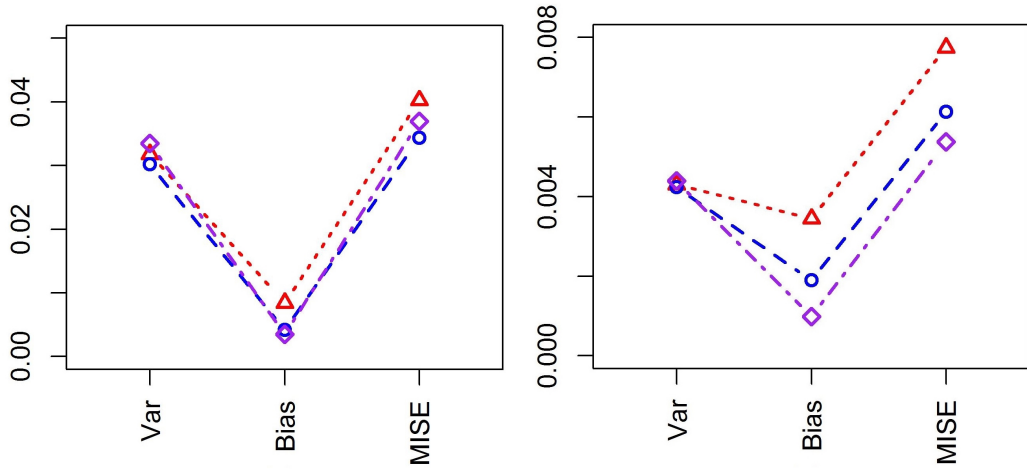


Figure 3.5: MISE, variance and bias components for the reflect (red line), cutnorm (blue line) and simple (purple line) boundary-corrected methods.

ering the sample size and density considerations. The choice of boundary-corrected kernel in this case can be left to other considerations such as ease of implementation and computational speed. At this point we note that the simple boundary-corrected method requires renormalization and negative-correction which increases computation time to a fair degree over that of the cutnorm procedure. Since both procedures perform well, we choose to use the cutnorm procedure for the simulation study in Chapter 5 with the differential equation based test in Section 4.3 as it yields much faster computation time.

Chapter 4

New goodness-of-fit tests for the Rayleigh distribution

4.1 Introduction

In this chapter we introduce and develop new goodness-of-fit tests for the Rayleigh distribution based on several different approaches. Section 4.2 introduces tests for the Rayleigh distribution based on a characterisation of this distribution. In Section 4.3 we propose a test for the Rayleigh distribution based on a differential equation and in Section 4.4 we introduce a test based on the Mellin transform.

4.2 New tests based on a conditional expectation characterisation

Consider the characterisation of the Rayleigh distribution by Ahsanullah & Shakil (2013) in Section 2.3 given here in terms of the parametrization in (2.3).

Characterisation. *Let X be a nonnegative random variable with absolutely continuous cumulative distribution function F satisfying $F(0) = 0$, $F(x) > 0$ for all $x > 0$,*

and with finite $E(X^{2k})$, for some fixed $k \geq 1$. Then X has a Rayleigh distribution with parameter θ (that is $F(x) = 1 - e^{-x^2/2\theta^2}$, $x \geq 0$), if, and only if, for all $t > 0$,

$$E(X^{2k}|X > t) = \sum_{l=0}^k 2^l \theta^{2l} k^{(l)} t^{2(k-l)},$$

where $k^{(l)} = k(k-1) \dots (k-l+1)$ and $k^{(0)} = 1$.

Now, from this Characterisation, and with $S(t) = P(X > t) = 1 - F(t)$, we have the following corollary.

Corollary 4.1. *Let X be a nonnegative random variable with absolutely continuous cumulative distribution function F satisfying $F(0) = 0$, $F(x) > 0$ for all $x > 0$, and with finite $E(X^{2k})$, for some fixed $k \geq 1$. Then X has a Rayleigh distribution with parameter θ , if, and only if, for all $t > 0$,*

$$E[X^{2k}I(X > t)] = S(t)\nu_{k,\theta}(t),$$

where $\nu_{k,\theta}(t) := \sum_{l=0}^k 2^l \theta^{2l} k^{(l)} t^{2(k-l)}$, with $k^{(l)} = k(k-1) \dots (k-l+1)$ and $k^{(0)} = 1$.

Proof. From straightforward computations, one can show that for all $t > 0$,

$$E(X^{2k}|X > t) = \frac{1}{P(X > t)} E[X^{2k}I(X > t)] = \frac{1}{S(t)} E[X^{2k}I(X > t)].$$

From the Characterisation, it follows that X has a Rayleigh distribution with parameter θ if, and only if, for all $t > 0$,

$$\frac{1}{S(t)} E[X^{2k}I(X > t)] = \nu_{k,\theta}(t),$$

equivalently, if and only if, for all $t > 0$, $E[X^{2k}I(X > t)] = S(t)\nu_{k,\theta}(t)$. $\square$

Based on Corollary 4.1, a random variable X admitting moments of order $2k$ has a

Rayleigh distribution with parameter θ if, and only if, for all $t > 0$

$$E[X^{2k}I(X > t)] - S(t)\nu_{k,\theta}(t) = E\{I(X > t)[X^{2k} - \nu_{k,\theta}(t)]\} = 0,$$

or equivalently,

$$\frac{E[X^{2k}I(X > t)]}{S(t)} - \nu_{k,\theta}(t) = 0.$$

It is then clear that such a random variable has a Rayleigh distribution with parameter θ if, and only if, either of the functions $t \mapsto E\{I(X > t)[X^{2k} - \nu_{k,\theta}(t)]\}$ or $t \mapsto E[X^{2k}I(X > t)]/S(t) - \nu_{k,\theta}(t)$ are zero. Since these functions are unknown, their discrepancies to 0 can be measured by Kolmogorov-Smirnov (KS) or Cramér-von Mises (CM) distances applied to their suitable normalized empirical versions defined respectively for all $t > 0$, by

$$\begin{aligned}\mathcal{T}_n(t) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n I(X_i > t)[X_i^{2k} - \nu_{k,\hat{\theta}_n}(t)] \quad \text{or} \\ \tilde{\mathcal{T}}_n(t) &= \frac{\sum_{i=1}^n I(X_i > t)X_i^{2k}}{\sum_{i=1}^n I(X_i > t)} - \nu_{k,\hat{\theta}_n}(t).\end{aligned}$$

This would lead to the Kolmogorov-Smirnov (KS_n and $\widetilde{KS}_n$) and Cramér-von Mises (CM_n and $\widetilde{CM}_n$) test statistics

$$KS_n = \sup_{t \in \Theta} |\mathcal{T}_n(t)| \quad \text{and} \quad CM_n = \int_{\Theta} \mathcal{T}_n^2(t) w(t) d\tilde{F}_n(t),$$

or

$$\widetilde{KS}_n = \sup_{t \in \Theta} |\tilde{\mathcal{T}}_n(t)| \quad \text{and} \quad \widetilde{CM}_n = n \int_{\Theta} \tilde{\mathcal{T}}_n^2(t) w(t) d\tilde{F}_n(t),$$

where Θ is a compact subset of $(0, \infty)$, w is a suitable positive weight function satisfying some integrability conditions (see Section 4.2.1), and $\tilde{F}_n$ is any consistent estimator of F . One can take $\tilde{F}_n$ as the empirical distribution function $F_n(t) = (1/n) \sum_{i=1}^n I(X_i \leq t)$, or the plug-in estimator $\hat{F}_{\hat{\theta}_n}$ of $F = F_{\theta}$, obtained by substi-

tuting $\widehat{\theta}_n$ for θ into F_θ .

The tests KS_n and $\widetilde{KS}_n$ on the one hand and CM_n and $\widetilde{CM}_n$ on the other hand are based on equivalent characterisations. Therefore, they are equivalent. However, the fractional form of $\widetilde{\mathcal{T}}_n$ makes it difficult to study the asymptotic behavior of $\widetilde{KS}_n$ and $\widetilde{CM}_n$. In what follows, we just study the asymptotic properties of KS_n and CM_n through the study of the asymptotic behavior of $\mathcal{T}_n$.

4.2.1 Asymptotic theory

We make the following assumption regarding the estimator $\widehat{\theta}_n$ for θ :

(A) The estimator $\widehat{\theta}_n$ satisfies the Bahadur representation

$$\sqrt{n}(\widehat{\theta}_n^2 - \theta^2) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_\theta(X_i) + o_P(1),$$

where ψ_θ is some continuous function such that $E[\psi_\theta(X_1)] = 0$, $E[\psi_\theta^{2+\delta}(X_1)] < \infty$, for some $\delta > 0$, and $o_P(1)$ stands for a sequence of random variables tending to 0 in probability as n tends to infinity.

Remark 4.1. *The maximum likelihood estimator of θ satisfies the above property with $\psi_\theta(x) = (x^2 - 2\theta^2)/2$ and $o_P(1) = 0$. The moment estimator of θ , which is obtained from the empirical counterpart of the first-order moment, also satisfies (A) with $\psi_\theta(x) = 2\theta(2x - \theta\sqrt{\pi})/\sqrt{\pi}$.*

Now, define for all $t > 0$, the following process

$$\mathcal{S}_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^n G_i(t),$$

where, for any $i = 1, \dots, n$,

$$G_i(t) = I(X_i > t) \left[X_i^{2k} - \nu_{k,\theta}(t) \right] - \psi_\theta(X_i) \omega_{k,\theta}(t), \quad (4.1)$$

with

$$\omega_{k,\theta}(t) := \sum_{l=1}^k 2^{l+1} l \theta^{2(l-1)} k^{(l)} t^{2(k-l)} S(t),$$

for all $t > 0$.

Let η be a positive measure defined by $d\eta(t) = w(t)dF(t)$. Let $\mathcal{L}_2$ be the collection of functions g defined on Θ such that $\int_{\Theta} g^2(s)d\eta(s) < \infty$. It is well known that $\mathcal{L}_2$ endowed with the inner product $\langle g_1, g_2 \rangle = \int_{\Theta} g_1(s)g_2(s)d\eta(s)$ and the norm $\|g\| = \langle g, g \rangle^{\frac{1}{2}}$ is a separable Hilbert space. One can note that the G_i 's and the processes $\mathcal{S}_n$ and $\mathcal{T}_n$ are random elements of $\mathcal{L}_2$.

Now, define on $(0, \infty) \times (0, \infty)$ and on $(0, \infty)$ respectively, the following real-valued functions:

$$\mathcal{K}(x, t) = I(x > t) \left[x^{2k} - \nu_{k,\theta}(t) \right] - \psi_{\theta}(x) \omega_{k,\theta}(t) \quad \text{and}$$

$$K(t) = \int_0^{\infty} \mathcal{K}(x, t) dF(x).$$

Theorem 4.1. *Assume that $E(X^{4k+2\delta}) < \infty$, for some fixed $k \geq 1$. Then*

(i) *as n tends to infinity,*

$$\sup_{t \in \Theta} |\mathcal{T}_n(t) - \mathcal{S}_n(t)| \longrightarrow 0,$$

in probability.

(ii) *as n tends to infinity,*

$$\mathcal{T}_n - \mathcal{S}_n \longrightarrow 0,$$

in probability in $\mathcal{L}_2$.

Proof. Let $\Theta \subset (0, \infty)$ be compact. For Part (i), one can write for all $t \in \Theta$,

$$\mathcal{T}_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^n I(X_i > t) \left[X_i^{2k} - \nu_{k,\theta}(t) \right]$$

$$\begin{aligned}
& -\frac{1}{\sqrt{n}} \sum_{i=1}^n [I(X_i > t) - S(t)] \left[\nu_{k, \hat{\theta}_n}(t) - \nu_{k, \theta}(t) \right] \\
& - S(t) \sqrt{n} \left[\nu_{k, \hat{\theta}_n}(t) - \nu_{k, \theta}(t) \right] \\
& := \mathcal{T}_{1n}(t) - \mathcal{T}_{2n}(t) - \mathcal{T}_{3n}(t).
\end{aligned}$$

Now, with the convention that $x^b = 1$ for all $x > 0$ and $b \leq 0$, observe that for all $t \in \Theta$,

$$\begin{aligned}
\sqrt{n} \left[\nu_{k, \hat{\theta}_n}(t) - \nu_{k, \theta}(t) \right] &= \sqrt{n} \sum_{l=1}^k \left(\hat{\theta}_n^{2l} - \theta^{2l} \right) 2^l k^{(l)} t^{2(k-l)} \\
&= \sqrt{n} \left(\hat{\theta}_n^2 - \theta^2 \right) \sum_{l=1}^k \left(\hat{\theta}_n^{2(l-1)} + \hat{\theta}_n^{2(l-2)} \theta^2 + \dots + \hat{\theta}_n^2 \theta^{2(l-2)} + \theta^{2(l-1)} \right) 2^l k^{(l)} t^{2(k-l)}.
\end{aligned}$$

From assumption (A), one has asymptotically

$$\sqrt{n} \left[\nu_{k, \hat{\theta}_n}(t) - \nu_{k, \theta}(t) \right] = \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_{\theta}(X_i) \sum_{l=1}^k 2^{l+1} l \theta^{2(l-1)} k^{(l)} t^{2(k-l)} + o_P(1).$$

Thus, asymptotically, for all $t \in \Theta$,

$$\mathcal{T}_{3n}(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_{\theta}(X_i) \sum_{l=1}^k 2^{l+1} l \theta^{2(l-1)} k^{(l)} t^{2(k-l)} S(t) + o_P(1)$$

and

$$\begin{aligned}
\mathcal{T}_{2n}(t) &= \frac{1}{n} \sum_{i=1}^n [I(X_i > t) - S(t)] \sqrt{n} \left[\nu_{k, \hat{\theta}_n}(t) - \nu_{k, \theta}(t) \right] \\
&= \frac{1}{n} \sum_{i=1}^n [I(X_i \leq t) - F(t)] \sqrt{n} \left(\hat{\theta}_n^2 - \theta^2 \right) \sum_{l=1}^k 2^{l+1} l \theta^{2(l-1)} k^{(l)} t^{2(k-l)}.
\end{aligned}$$

It is easy to see that

$$\sup_{t \in \Theta} |\mathcal{T}_{2n}(t)| \leq Ct \cdot \sup_{t \in \Theta} \left| \frac{1}{n} \sum_{i=1}^n [I(X_i \leq t) - F(t)] \right| \times \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_\theta(X_i) \right|,$$

for some constant $C > 0$.

By the classical central limit theorem the sequence of random variables $\sum_{i=1}^n \psi_\theta(X_i)/\sqrt{n}$ converges in distribution to a zero-mean Gaussian random variable with variance $E[\psi_\theta^2(X_1)]$. Also, by the Glivenko-Cantelli theorem, as n tends to infinity, in probability,

$$\sup_{t \in \Theta} \left| \frac{1}{n} \sum_{i=1}^n [I(X_i \leq t) - F(t)] \right| \longrightarrow 0.$$

Thus $\sup_{t \in \Theta} |\mathcal{T}_{2n}(t)|$ tends in probability to 0, as n tends to infinity.

This clearly implies that for any compact subset Θ of $(0, \infty)$, as n tends to infinity, in probability,

$$\sup_{t \in \Theta} |\mathcal{T}_n(t) - \mathcal{S}_n(t)| \longrightarrow 0.$$

For the proof of the second part of the theorem, one can just observe that it follows from the above convergence that, in probability, $\mathcal{T}_n(t) - \mathcal{S}_n(t)$ is uniformly bounded. Then by the dominated convergence theorem, it converges in probability to 0 in $\mathcal{L}_2$. $\square$

Theorem 4.2. Assume that $E(X^{4k+2\delta}) < \infty$, for some fixed $k \geq 1$.

(i) If $K(t) = 0$, for all $t > 0$, then $\mathcal{T}_n(t)$ converges weakly in $\mathcal{L}_2$ to a zero-mean Gaussian process $\mathcal{T}(t)$ with covariance kernel Γ defined for all $s, t > 0$ by

$$\begin{aligned} \Gamma(s, t) = & E \left\{ I(X_1 > s \vee t) \left[X_1^{2k} - \nu_{k, \theta}(t) \right] \left[X_1^{2k} - \nu_{k, \theta}(s) \right] \right\} \\ & - E \left\{ I(X_1 > s) \psi_\theta(X_1) \left[X_1^{2k} - \nu_{k, \theta}(s) \right] \right\} \omega_{k, \theta}(t) \\ & - E \left\{ I(X_1 > t) \psi_\theta(X_1) \left[X_1^{2k} - \nu_{k, \theta}(t) \right] \right\} \omega_{k, \theta}(s) \\ & + E \left[\psi_\theta^2(X_1) \right] \omega_{k, \theta}(s) \omega_{k, \theta}(t), \end{aligned} \quad (4.2)$$

where $x \vee y$ denotes the largest number between x and y .

(ii) if $K(t) \neq 0$, on some $\Theta_0 \subset (0, \infty)$ satisfying $\eta(\Theta_0) > 0$, then $n^{-1/2}\mathcal{T}_n \rightarrow K$ in probability in $\mathcal{L}_2$.

Proof. We just have to check the conditions of the application of the central limit theorem for Hilbert spaces to the G_i 's. These are iid elements of $\mathcal{L}_2$, as the X_i 's are iid random variables. For all $G \in \mathcal{L}_2$,

$$E(\langle G_1, G \rangle) = \int_0^\infty \int_\Theta \mathcal{K}(x, t) G(t) d\eta(t) dF(x) = \int_\Theta K(t) G(t) d\eta(t).$$

To handle Part (i) of the theorem, it is easy to see that if $K(t) = 0$ for all $t \in (0, \infty)$ then $E(\langle G_1, G \rangle) = 0$.

Now,

$$E(\|G_1\|^2) = E \left[\int_\Theta G_1^2(t) d\eta(t) \right] = \int_0^\infty \int_\Theta \mathcal{K}^2(x, t) d\eta(t) dF(x),$$

which is finite by our assumptions. The central limit theorem for Hilbert spaces then applies to $\mathcal{S}_n$. Hence, it converges to a zero-mean Gaussian element of $\mathcal{L}_2$, $\mathcal{T}$ with covariance kernel

$$E[\mathcal{T}(s)\mathcal{T}(t)] = E[G_1(s)G_1(t)] = \Gamma(s, t).$$

For establishing the second part of the theorem, observe that if there exists some $\Theta_0 \subset (0, \infty)$ such that $\eta(\Theta_0) > 0$ on which $K(t) \neq 0$, then we can write

$$\mathcal{S}_n(t) = \mathcal{U}_n(t) + \sqrt{n}K(t),$$

where

$$\mathcal{U}_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^n [G_i(t) - K(t)],$$

with $G_i(t)$, $i = 1, \dots, n$, given in (4.1). From this, we have

$$\begin{aligned}
\|\mathcal{S}_n\|^2 &= \|\mathcal{U}_n\|^2 + 2\sqrt{n}\langle \mathcal{U}_n, K \rangle + n\|K\|^2 \\
&= n \left(\frac{1}{n} \|\mathcal{U}_n\|^2 + \frac{2}{\sqrt{n}} \langle \mathcal{U}_n, K \rangle + \|K\|^2 \right).
\end{aligned}$$

Note that the random functions $G_i(t) - K(t)$ are zero-mean iid elements of $\mathcal{L}_2$. Thus, as in the proof of Part (i), by elementary arguments, one can see that $\|\mathcal{U}_n\|^2$ converges in distribution to $\|\mathcal{T}\|^2$. By the classical central limit theorem, $\langle \mathcal{U}_n, K \rangle$ converges in distribution to $\langle \mathcal{T}, K \rangle$. With all these arguments, it is easily seen that the first and second terms in brackets in the last equality converge in distribution to 0 and the second part of Theorem 4.2 follows. $\square$

From the above theorem, we will study the asymptotic behavior of CM_n . In what follows, we indicate how to approximate the tails of the limiting distribution. Consider the positive continuous function w introduced earlier. We assume that it is continuous and satisfies

$$w(t) \geq 0, \quad t \geq 0, \quad 0 < \int_{\Theta} t^{4k+2\delta} w(t) dF(t) = \int_{\Theta} t^{4k+2\delta} d\eta(t) < \infty.$$

Corollary 4.2. *Assume (A) holds, then*

- (i) *under H_0 , the random variable CM_n converges in distribution to the random variable*

$$CM = \int_{\Theta} \mathcal{T}^2(t) d\eta(t),$$

where $\mathcal{T}$ is the Gaussian process defined in Theorem 4.1.

- (ii) *under H_1 , the sequence of random variables CM_n diverges in probability to infinity.*

Proof. This corollary is a direct consequence of Theorems 4.1 and 4.2. $\square$

Corollary 4.2 has a number of consequences. Its second part implies that the power of the test based on CM_n against any fixed alternative tends to one. This can be

seen from the fact that, under H_1 ,

$$\frac{CM_n}{n} \xrightarrow{P} \int \mathcal{K}^2(t)w(t)dF(t) := I > 0.$$

Now, because under H_1 , $I > 0$, we have that $CM_n \xrightarrow{P} \infty$ as $n \rightarrow \infty$. It then follows that

$$P_{H_1}(CM_n \geq c_n(\alpha)) \rightarrow 1 \text{ as } n \rightarrow \infty,$$

where $c_n(\alpha)$ is the critical value obtained from the null distribution of CM_n .

Its first part allows for an approximation of its null distribution. Indeed, it is easy to check that the kernel $\Gamma(s, t)$ defined in (4.2) is a Mercer kernel. It results that, the integral operator ∇_Γ defined on $\mathcal{L}_2$ by

$$\nabla_\Gamma g(s) = \int_{\Theta} \Gamma(s, t)g(t)d\eta(t), \quad s > 0$$

admits eigenvalues $\zeta_1, \zeta_2, \dots$ sorted so that $\zeta_1 \geq \zeta_2 \geq \dots \geq 0$, and eigenfunctions $h_1, h_2, \dots$ which form an orthonormal basis for $\mathcal{L}_2$.

Corollary 4.3. *Under the conditions of Corollary 4.2, if $K(t) = 0$ for all $t > 0$, then CM_n has asymptotically the same distribution as $\sum_{j \geq 1} \zeta_j \chi_j^2$, where ζ_j and χ_j^2 , $j \geq 1$, are respectively the eigenvalues of ∇_Γ and iid random variables following a chi-squared distribution with one degree of freedom, respectively.*

Proof. From our assumptions, the Gaussian process $\mathcal{T}$ defined in Theorem 4.1 is a random element of $\mathcal{L}_2$. It has the following Karhunen-Loève representation

$$\mathcal{T}(t) = \sum_{j=1}^{\infty} N_j h_j(t), \quad t > 0,$$

where for all $j \geq 1$, $N_j = \langle \mathcal{T}, h_j \rangle$ are independent zero-mean Gaussian random

variables with variances ζ_j . It follows from this that $||\mathcal{T}||^2 = \sum_{j=1}^{\infty} N_j^2$, since $E(N_j^2) = \zeta_j \geq 0$, $N_j = 0$ in probability whenever $\zeta_j = 0$. For positive ζ_j 's, one can observe that $Z_j = N_j/\sqrt{\zeta_j}$, $j \geq 1$, are iid standard Gaussian random variables. Thus,

$$CM = \int_{\Theta} \mathcal{T}^2(t) d\eta(t) = ||\mathcal{T}||^2 = \sum_{j=1}^{\infty} \zeta_j Z_j^2.$$

The result then follows from Corollary 4.2. $\square$

The distribution of the random variable $\sum_{j=1}^{\infty} \zeta_j \chi_j^2$ can be approximated by that of $\sum_{j=1}^J \zeta_j \chi_j^2$, for an arbitrary large J . However, the ζ_j 's are unknown and must be estimated. They can be estimated here by considering the eigenvalues of the operator $\nabla_{\hat{\Gamma}_n}(s, t)$, where $\hat{\Gamma}_n(s, t)$ is any consistent estimator of $\Gamma(s, t)$. Possible estimators are the $\hat{\zeta}_j$'s solutions of the Fredholm integral equations

$$\nabla_{\hat{\Gamma}_n} \hat{f}_j = \hat{\zeta}_j \hat{f}_j, \quad j \geq 1.$$

Remark 4.2. *A natural estimator of $\Gamma(s, t)$ can be obtained by taking the empirical counterpart in the expression given in (4.2), in which θ is replaced by $\hat{\theta}_n$. For the computation of the cumulative distribution function of $\sum_{j=1}^J \hat{\zeta}_j \chi_j^2$ the reader can refer to Ngatchou-Wandji & Harel (2013). However, in our Monte Carlo study, we make use of the parametric bootstrap in order to obtain approximate critical values. This is a standard approach in these type of goodness-of-fit tests.*

4.3 New test based on a differential equation approach

Our second approach for constructing a goodness-of-fit test for the hypothesis in (2.6) is based on the differential equation

$$\theta^2 x g'(x) + g(x)(x^2 - \theta^2) = 0,$$

subject to $g(1) = (1/\theta^2) \exp(-1/2\theta^2)$, where $\theta > 0$ is a constant. As shown below, we have that the Rayleigh density in (2.2) is the unique solution to this differential equation.

Solving the differential equation:

$$\theta^2 x g'(x) + g(x)(x^2 - \theta^2) = 0$$

$$\int \frac{g'(x)}{g(x)} dx = - \int \frac{x^2 - \theta^2}{\theta^2 x} dx$$

$$\log |g(x)| = -\frac{x^2}{2\theta^2} + \log x + C$$

$$g(x) = e^{-\frac{x^2}{2\theta^2} + \log x + C}$$

But $g(1) = 1/\theta^2 \exp(-1/2\theta^2)$, hence

$$\begin{aligned} \frac{1}{\theta^2} e^{-\frac{1}{2\theta^2}} &= e^{-\frac{1}{2\theta^2}} e^C \\ e^C &= \frac{1}{\theta^2} \end{aligned}$$

Thus we have that $g(x) = (x/\theta^2) \exp(-x^2/2\theta^2)$, $x \geq 0$.

Note that $e^C = 1/\theta^2$ is also the only value of C that ensures that $g(x)$ is a valid density function for $x \geq 0$.

Now, if $X \sim \text{Ral}(\theta)$, then it follows that

$$E[\theta^2 X g'(X) + g(X)(X^2 - \theta^2)] = 0,$$

and if $Y = X/\theta \sim \text{Ral}(1)$, then

$$E[Y h'(X) + h(X)(Y^2 - 1)] = 0, \tag{4.3}$$

where $h(y) = y \exp(-y^2/2)$ (i.e., the standard Rayleigh density function) and $h'(y) = (1 - y^2) \exp(-y^2/2)$ is the first derivative of $h(y)$. On the basis of (4.3) we suggest the following test statistic,

$$\begin{aligned} T_n &= n \int_0^\infty \left[y h'_0(y) + \hat{h}_{\hat{\theta}}(y)(y^2 - 1) \right]^2 w(y) d\tilde{F}(y), \\ &= n \int_0^\infty \left[(y - y^3) e^{-y^2/2} + \hat{h}_{\hat{\theta}}(y)(y^2 - 1) \right]^2 w(y) d\tilde{F}(y), \end{aligned} \quad (4.4)$$

where $h'_0(y)$ is the standard Rayleigh density function, $w(y)$ is a suitable nonnegative weight function, $\tilde{F}(y) = P(X/\hat{\theta} \leq y)$ and $\hat{h}_{\hat{\theta}}(y) = (nh)^{-1} \sum_{j=1}^n k \left\{ (y - X_j/\hat{\theta})/h \right\}$ is a kernel density estimator based on the data $X_j/\hat{\theta}$, with $\hat{\theta}$ an estimator for θ . Furthermore, $k(\cdot)$ is a suitable chosen kernel function and h is the bandwidth (See Chapter 3 for details on kernel and bandwidth choices). Clearly the test rejects for large values of the test statistic.

Estimating $\tilde{F}(y)$ by its empirical counterpart $\tilde{F}_n(y) = (1/n) \sum_{i=1}^n I(X_i/\hat{\theta} \leq y)$ and choosing $w(y) = \exp(-\varphi y)$, we obtain the following expression for the test statistic,

$$T_{n,\varphi} = \sum_{j=1}^n \left[(Y_j - Y_j^3) e^{\frac{-Y_j^2}{2}} + \hat{h}_{\hat{\theta}}(Y_j)(Y_j^2 - 1) \right]^2 \exp(-\varphi Y_j).$$

Remark 4.3. *It is well known that the kernel density estimator suffers from boundary bias near the endpoints of the support of the density being estimated. In order to rectify this shortcoming, we suggest using a boundary-corrected density estimator as a modification to the test statistic in (4.4). The modified test statistic using boundary-corrected density estimator will be denoted as $Tb_{n,\varphi}$ (see Chapter 3 for more details on boundary-corrected density estimators).*

4.4 New test based on the Mellin transform

In this section we propose a goodness-of-fit test that utilizes the Mellin transform. This integral transform was traditionally used in applied mathematics, but was adopted by the field of mathematical statistics in the study of quotients of independent random variables and the distribution of products (see, Epstein 1948). The Mellin transform of a function $h(x)$ is defined as

$$\mathcal{M}(h, s) = \int_0^\infty x^{s-1} h(x) dx, \quad (4.5)$$

and generally holds for a complex value s (see Hahn & Poularikas 2000). However, in the spirit of Meintanis (2008a) we take $t = s - 1 > 0$ to be real valued. This allows the formulation in (4.5) to be identified as the moment process of X and can be written formally as

$$M_X(t) = E[X^t] = \int_0^\infty x^t dF(x). \quad (4.6)$$

Goodness-of-fit tests based on the conceptualization in (4.6) have been developed for the Lognormal distribution in Meintanis (2008a) and for generalized exponential laws in Meintanis (2008b). Besides aforementioned papers, the Mellin transform has been sparsely explored in the context of goodness-of-fit testing. In contrast, the Laplace transform has enjoyed more attention in the literature and possesses an inherent connection to the Mellin transform. If $L_X(t) = E[e^{-tX}]$ then the Mellin transform can be obtained by $M_X(t) = L_Y(t) = E[e^{-tY}]$ with $Y = -\log(X)$. Therefore, we note that the test statistic proposed in this section holds a certain connection to the test in Meintanis & Iliopoulos (2003). The newly defined test is derived from the transformed variable through the Mellin transform and has the advantage of not containing an error function after transformation.

A starting point for developing a goodness-of-fit test with the Mellin transform is to consider the popular Kolmogorov-Smirnov and Cramér-von Mises distance measures,

$$\mathcal{S}(t) = \int_0^\infty [M(t) - M_n(t)]^2 dP(t) \quad \text{and} \quad \tilde{\mathcal{S}}(t) = \sup_{t>0} |M(t) - M_n(t)|,$$

where $M_n(t)$ is the empirical counterpart of (4.6) and P is any assigned measure. In this, we depart from the classical discrepancy measures in favour of the differential equation

$$D(t, \theta) := 2M'(t) - \left\{ \log(2) - 2\log(\theta) + \psi\left(1 + \frac{t}{2}\right) \right\} M(t) = 0, \quad (4.7)$$

with digamma function $\psi(\cdot)$ and boundary condition $M(0) = 1$. A goodness-of-fit test can be constructed by exploiting the fact that the Mellin transform $M(t) = 2^{t/2}\theta^t\Gamma(1 + t/2)$ of the Rayleigh distribution given in (2.3) is the unique solution to the differential equation given in (4.7). Now we can estimate (4.7) by its empirical counterpart

$$D_n(t, \hat{\theta}_n) = 2M'_n(t) - \left\{ \log(2) - 2\log(\hat{\theta}_n) + \psi\left(1 + \frac{t}{2}\right) \right\} M_n(t), \quad (4.8)$$

where $M_n(t) = n^{-1} \sum_{j=1}^n X_j^t$ and $M'_n(t) = n^{-1} \sum_{j=1}^n X_j^t \log(X_j)$ and $\hat{\theta}_n$ is a consistent estimator of θ . Ultimately, Cramér-von Mises and Kolmogorov-Smirnov type test statistics can be formulated based on (4.8) to test the proximity to zero of the differential equation. Based on (4.8) consider

$$LA_n = n \int_0^\infty D_n^2(t, \hat{\theta}_n) w(t) d\hat{F}_n(t) \quad \text{and} \quad \tilde{T}_n = \sup_{t>0} |D_n(t, \hat{\theta}_n)|,$$

where $w(t)$ is an appropriate nonnegative weight function and $\hat{F}_n(t)$ is any consistent estimator of F . Once again, we can take $\hat{F}_n$ as the empirical distribution function $F_n(t) = (1/n) \sum_{i=1}^n I(X_i \leq t)$, or the plug-in estimator $F_{\hat{\theta}_n}$ of $F = F_\theta$, obtained by

substituting $\widehat{\theta}_n$ for θ into F_θ . For implementation of this test statistic we choose the weight function $w(t) = \exp(-\varphi t^2)$ with tuning parameter $\varphi > 0$ and $\widehat{F}_n(t)$ as the empirical distribution function. By using scaled values $Y_j = X_j/\widehat{\theta}_n$, we can obtain desirable calculable forms

$$LA_{n,\varphi} = \sum_{k=1}^n \left[2 \sum_{j=1}^n Y_j^{Y_k} \log(Y_j) - \left\{ \log(2) + \psi \left(1 + \frac{Y_k}{2} \right) \right\} \sum_{j=1}^n Y_j^{Y_k} \right]^2 \exp(-\varphi Y_k^2),$$

$$\widetilde{LA}_n = \max_{1 \leq k \leq n} \left| 2 \sum_{j=1}^n Y_j^{Y_k} \log(Y_j) - \left\{ \log(2) + \psi \left(1 + \frac{Y_k}{2} \right) \right\} \sum_{j=1}^n Y_j^{Y_k} \right|.$$

The two test statistics would be close to zero only if the observations realised from the Rayleigh distribution and consequently will reject the hypothesis in (2.6) for large values. We will only consider the Cramér-von Mises type test as it demonstrated greater performance in power.

Now, consider the consistency of the test LA_n . First note from (4.9) we have

$$\frac{LA_n}{n} = \int_0^\infty D_n^2(t, \widehat{\theta}_n) w(t) d\widehat{F}_n(t).$$

From the Law of Large Numbers we have that $M_n(t) \xrightarrow{P} M(t)$ and $M'_n(t) \xrightarrow{P} M'(t)$. Consequently, it holds that $D_n(t, \widehat{\theta}_n) \xrightarrow{a.s.} D(t, \theta)$ and therefore by Fatou's Lemma and the Glivenko-Cantelli Theorem we have that

$$\liminf_{n \rightarrow \infty} \frac{LA_n}{n} \xrightarrow{P} \mathcal{I} := \int_0^\infty D^2(t, \theta) w(t) dF(t).$$

Note that $\mathcal{I} = 0$ if, and only if, H_0 is true and $\mathcal{I} > 0$ otherwise. This shows that the test LA_n is consistent against fixed alternative distributions.

Chapter 5

Simulations and results

5.1 Introduction

In this chapter, we compare the finite-sample performances of the newly proposed tests to those of the existing tests. Section 5.2 establishes the simulation setting and in Section 5.3 we provide the power estimates against general alternatives and the local power estimates for two distribution mixtures. The discussion of the power estimates precedes the tables which can be found at the end of the chapter. In Section 5.4, the new and existing tests are applied to a real data example.

5.2 Simulation setting

Critical values were calculated for a significance level of 5% with the use of 50 000 independent Monte Carlo replicates. Note that each of the test statistics in Table 5.2 rejects the null hypothesis in (2.6) for large values. That is, the null hypothesis is rejected when the test statistic is larger than some value c (obtained using Monte Carlo simulations) that is dependent on the sample size n . A number of tests rely on a tuning parameter $\varphi > 0$ or window-width $m > 0$ that has to be chosen. In this study, the power estimates were obtained over a grid of values

$\varphi \in \{0.00, 0.25, 0.50, 1.00, 1.50, 2.00, 3.00, 5.00, 7.50, 10.00\}$ and $m \in \{1, 2, \dots, n/2\}$. Only parameters that exhibited the highest power estimates are reported in Table 5.4 and Table 5.5. For comparability, an attempt was made to keep the tuning parameter and window-width values as similar as possible across the tests and consistently the same over tables. Therefore, power estimates for tests that require tuning parameters or a window-width are reported for

- $KL_{n,m}$ and $DH_{n,m}$ with $m \in \{3, 4, 6\}$,
- $MI_{n,\varphi}$ with $\varphi \in \{0.5, 2, 5\}$,
- $T_{n,\varphi}$, $Tb_{n,\varphi}$, $M1_{n,\varphi}$, $M2_{n,\varphi}$, $CM_{n,\varphi}$ and $LA_{n,\varphi}$ with $\varphi \in \{1, 2, 5\}$.

The $CM_{n,\varphi}$ and KS_n tests also require a choice of the tuning parameter k . Observed power estimates differ only slightly over different choices of k and therefore only results for $k = 1$ are reported. Furthermore, as mentioned in Section 4.3, the $Tb_{n,\varphi}$ test utilizes a boundary-bias corrected kernel. In this study, the cut-normalized boundary corrected kernel of Gasser & Müller (1979) is used for performance and ease of implementation (see Section 3.6.4 for further motivation). The tests of Torabi et al. (2016), adapted for the Rayleigh distribution, are implemented for functions $\psi(x) = [(x-1)/(x+1)]^2$ in $C1_n$ and $\psi(x) = x \log(x) - x + 1$ in $C2_n$.

For ease of reference, Table 5.1 lists the parametrization for the various alternative distributions used in the study. These distributions are commonly used as alternatives for the Rayleigh distribution and include distributions with constant hazard rates (CHR), decreasing hazard rates (DHR), increasing hazard rates (IHR) and non-monotone hazard rates (NMHR). Table 5.2 provides a reference to the test statistics that were discussed in Chapter 2.

Power estimates are obtained with 10 000 independent Monte Carlo replications for each of the tests listed in Table 5.2 and displayed in Table 5.4 for sample sizes $n = 20$ and in Table 5.5 for sample sizes $n = 30$. The power estimates in the afore-

mentioned tables are obtained using the MLE. Tables for the MME can be found in Appendix A. Similar results were obtained using the two estimation techniques.

Results for local power estimates are, in turn, displayed in Tables 5.6 and 5.7 and are obtained as follows: A sample of size n is generated as individual values from either a $Ral(1)$ -distribution with probability p , or a chosen alternative distribution with probability $1 - p$. The parameter p determines the level of contamination, i.e., $p = 0$ indicates a sample purely from the $Ral(1)$ -distribution and $p > 0$ indicates the introduction of values from the chosen alternative distribution. Table 5.6 shows power estimates for the case where the Rayleigh distribution is contaminated by the half-normal distribution while Table 5.7 shows power estimates in the case where the $\Gamma(1.5)$ distribution is used. The local power estimates are also obtained using the MLE. All calculations and simulations were performed using R software (R Core Team 2013).

Table 5.1: Probability density functions for choices of the alternative distributions.

Alternative	$f(x)$	Notation
Gamma	$\frac{1}{\Gamma(\theta)} x^{\theta-1} e^{-x}$	$\Gamma(\theta)$
Weibull	$\theta x^{\theta-1} \exp(-x^\theta)$	$W(\theta)$
Power	$\frac{1}{\theta} x^{(1-\theta)/\theta}, 0 < x < 1$	$PW(\theta)$
Dhillon	$\frac{\theta+1}{x+1} \exp\left\{-\left(\log(x+1)\right)^{\theta+1}\right\} (\log(x+1))^\theta$	$DL(\theta)$
Chen	$2\theta x^{\theta-1} \exp\left\{x^\theta + 2(1 - \exp(x^\theta))\right\}$	$CH(\theta)$
Linear failure rate	$(1+\theta x) \exp(-x - \theta x^2/2)$	$LFR(\theta)$
Extreme value	$\frac{1}{\theta} \exp\left(x + \frac{1-e^x}{\theta}\right)$	$EV(\theta)$
Lognormal	$\exp\left\{-\frac{1}{2}(\log(x)/\theta)^2\right\} / \left\{\theta x \sqrt{2\pi}\right\}$	$LN(\theta)$
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	$\theta \exp(-\theta x)$	$EXP(\theta)$
Beta exponential	$\theta e^{-x} (1 - e^{-x})^{\theta-1}$	$BEX(\theta)$
Exponential logarithmic	$\frac{1}{-\ln \theta} \frac{(1-\theta)e^{-x}}{1 - (1-\theta)e^{-x}}$	$EL(\theta)$
Exponential Nadarajah Haghighi (1)	$\frac{\theta(1+x)^{-0.5} e^{1-(1+x)^{0.5}}}{2[1 - e^{1-(1+x)^{0.5}}]^{1-\theta}}$	$ENH1(\theta)$
Exponential geometric	$\frac{(1-\theta)e^{-x}}{(1 - \theta e^{-x})^2}$	$EG(\theta)$
Beta	$\frac{\Gamma(\theta_1 + \theta_2)}{\Gamma(\theta_1)\Gamma(\theta_2)} x^{\theta_1-1} (1-x)^{\theta_2-1}$	$B(\theta_1, \theta_2)$
Half normal	$\left(\frac{2}{\pi}\right)^{1/2} \exp\left(-\frac{x^2}{2}\right)$	HN

5.3 Simulation results

Below we discuss the power estimates obtained for the alternative distributions and the local powers obtained for the mixture distributions.

5.3.1 Discussion on power estimates for alternative distributions

Note that Table 5.4 (results for $n = 20$) and Table 5.5 (results for $n = 30$) respectively span over three pages to accommodate all columns. For ease of comparison, the highest power for each alternative is printed in bold.

If we observe the classical tests (D_n , V_n , W_n , A_n) in isolation, it is clear that the Anderson-Darling test (A_n) outperforms all the other classical tests. The Cramér-von Mises (W_n) test is overall the second best performer in terms of power estimates with the exception of the $IG(0.5)$ case, where the W_n test slightly outperforms the A_n test. This trend is observed in both the $n = 20$ and $n = 30$ cases.

Turning our attention to the entropy tests with window-width m ($KL_{n,m}$, $DH_{n,m}$), the $DH_{n,m}$ test performs better than the $KL_{n,m}$ test for both sample sizes considered. Interestingly, $KL_{n,m}$ outperforms its counterpart for the $PW(1)$ and $n = 30$ case. Considering entropy tests together ($KL_{n,m}$, $DH_{n,m}$, CK_n), we observe that CK_n is outperformed by its counterparts.

Next we turn our attention to tests based on the Hellinger distance ($C1_n$, $C2_n$, CK_n , PKL_n , PH_n , PJ_n , PTV_n , PC_n). Here we observe that PC_n either matches or outperforms the PH_n , PJ_n , PTV_n tests. The aforementioned tests generally seem to outperform $C1_n$, $C2_n$, CK_n in most instances. For $n = 20$, the $C1_n$ however outperforms its competitors against the $\Gamma(2)$, $DL(1.5)$, $IG(1.5)$ and $LN(0.8)$ distributions. For $n = 30$ it tends to rather match than outperform the other Hellinger distance

tests for the specific distributions mentioned. The PKL_n provides lower powers than the other tests based on the Hellinger distance.

For the tests with a tuning parameter ($MI_{n,\varphi}$, $M1_{n,\varphi}$, $M2_{n,\varphi}$, $T_{n,\varphi}$, $Tb_{n,\varphi}$, $CM_{n,\varphi}$, $LA_{n,\varphi}$), we find that $LA_{n,\varphi}$ outperforms or matches its counterparts in the majority of the cases considered. The boundary-corrected kernel test proves to be a strong contender with power estimates close to that of the $LA_{n,\varphi}$ test. The $MI_{n,\varphi}$ and $CM_{n,\varphi}$ seem to be closely matched and exhibit competitive results when compared to other tuning parameter tests. $M1_{n,\varphi}$, $M2_{n,\varphi}$ lacks power when compared to the competitor tests. In general, the tests with tuning parameters provide powers superior to those associated with the majority of competing tests considered in this study.

The newly developed tests ($T_{n,\varphi}$, $Tb_{n,\varphi}$, $CM_{n,\varphi}$, KS_n , $LA_{n,\varphi}$) perform very well when compared to the other tests in the study. When compared to each other in isolation, we find that $LA_{n,\varphi}$ is superior and KS_n shows the lowest power performance. The lower power estimates of KS_n when compared to its counterparts is not surprising since Cramér-von Mises type tests traditionally perform better than Kolmogorov-Smirnov type tests in most situations. Figure 5.1 illustrates the effect and stability of the power estimates across the parameter value φ for the newly developed tests (see Appendix A for further illustration). We find that $CM_{n,\varphi}$ is the most stable across the grid of power values $\varphi \in \{0.00, 0.25, 0.50, 1.00, 1.50, 2.00, 3.00, 5.00, 7.50, 10.00\}$, closely followed by the $LA_{n,\varphi}$ test. The result indicating that for any value $\varphi \geq 1$ the $CM_{n,\varphi}$ and $LA_{n,\varphi}$ tests have relatively high power. If we turn our attention to $T_{n,\varphi}$ and $Tb_{n,\varphi}$ we find that the bias-corrected kernel test ($Tb_{n,\varphi}$) is less sensitive than $T_{n,\varphi}$ to choices of the φ -value. This fact is possibly supplemented by the nature of the weight functions used. When choosing a large value for the weight functions $w(t) = \exp(-\varphi|t|)$ and $w(t) = \exp(-\varphi t^2)$ the test becomes more sensitive to deviations in the tail of the distributions. This, in turn, causes the boundary

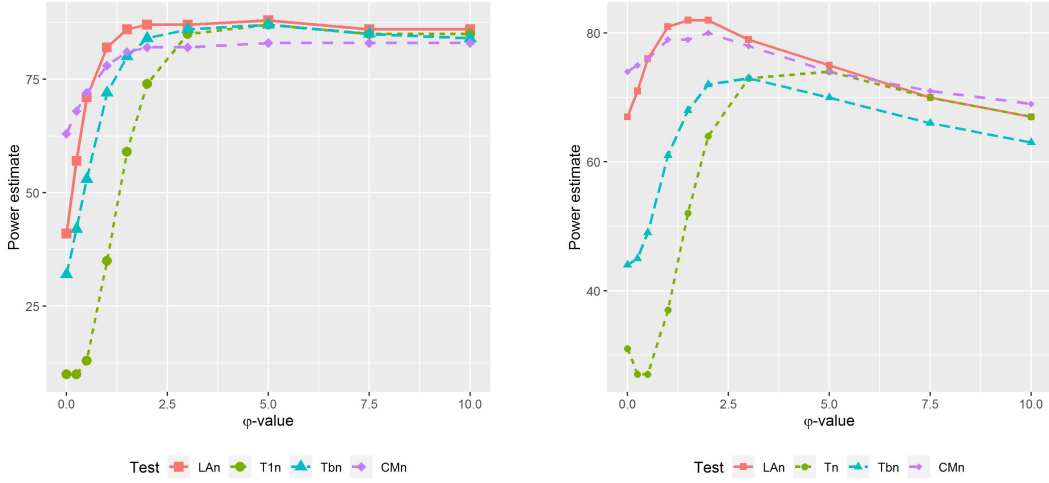


Figure 5.1: Power estimates across tuning parameter value for the $CH(1)$ (left) and $DL(1)$ (right) distributions with $n = 20$.

effects around zero for the tests $T_{n,\phi}$ and $Tb_{n,\phi}$ to have less influence on the power estimates. This trend can be observed both when $n = 20$ and $n = 30$.

Focusing on the tests adapted for the Rayleigh distribution ($M1_{n,\phi}$, $M2_{n,\phi}$, $C1_n$, $C2_n$), we find that the transformation test $M1_{n,\phi}$ by Meintanis (2009) tends to either match or outperform the $M2_{n,\phi}$, $C1_n$ and $C2_n$ tests. The tuning parameter for $M1_{n,\phi}$ and $M2_{n,\phi}$ provides an edge in performance over the non-tuning parameter counterparts. Furthermore, $M2_{n,\phi}$ only slightly falls short of the performance of $M1_{n,\phi}$. However, $M2_{n,\phi}$ proves the better performer by some margin for the $G(2)$, $W(3)$, $DL(1)$, $IG(0.5)$, $IG(1.5)$ and $LN(0.8)$ distributions.

We find that there is very little difference between the powers of the tests considered against the alternatives with constant and decreasing hazard rates. The tests have extremely high powers against these alternatives, this could be attributed to the fact that the Rayleigh distribution has an increasing hazard rate. For distributions with increasing hazard rates, we see that the $LA_{n,\phi}$ and $Tb_{n,\phi}$ tests outperform their com-

petitors in terms of power estimates for $n = 20$ and $n = 30$. However, the $DH_{n,m}$ test performs better than $LA_{n,\varphi}$ and $Tb_{n,\varphi}$ against the distributions $\Gamma(2)$, $GO(0.5)$ and $n = 20$. For $n = 30$ we find that $MI_{n,\varphi}$ performs the best against $CH(1.5)$ and $GO(0.5)$. The $CM_{n,\varphi}$ and $MI_{n,\varphi}$ tests both prove to be strong contenders for the best performing test against distributions with increasing hazard rates as they exhibit comparable powers in this case. For distributions with non-monotone hazard rates, we find that $LA_{n,\varphi}$ exhibits the highest powers in the majority of cases considered, closely followed by the $DH_{n,m}$. The $DH_{n,m}$ test outperforms the $LA_{n,\varphi}$ test against the $IG(0.5)$ and $IG(1.5)$ distributions for the sample size $n = 20$, however for sample size $n = 30$ it only outperforms the $LA_{n,\varphi}$ test against the $IG(0.5)$ distribution.

Overall, $LA_{n,\varphi}$ and $Tb_{n,\varphi}$ performs the best among all tests considered. These tests either match or outperform all remaining competitors. The tuning parameter choices $\varphi = 2$ and $\varphi = 5$ for the tests $LA_{n,\varphi}$, $Tb_{n,\varphi}$ and $CM_{n,\varphi}$ result in the highest powers in most cases. The tuning parameter value $\varphi = 0.5$ mostly results in the highest power for the $MI_{n,\varphi}$ test. For the $DH_{n,m}$ test, the window-width exhibiting the highest powers is $m = 6$. In general, $LA_{n,\varphi}$, $Tb_{n,\varphi}$, $CM_{n,\varphi}$, $DH_{n,m}$ and $MI_{n,\varphi}$ prove to be the superior tests.

5.3.2 Discussion on local power estimates for the $HN - \text{Ral}(1)$ mixture

Table 5.6 displays the local power estimates of various tests against a mixed $HN - \text{Ral}(1)$ distribution. The top row shows power estimates against samples of size $n = 20$, while the bottom row shows power estimates against samples of size $n = 30$.

From Table 5.6 it is clear that all of the tests attain the nominal significance level of 5% for $n = 20$ and $n = 30$ under the null hypothesis (corresponding to the case

where $p = 0$). For alternatives close to the Rayleigh distribution ($p = 0.05$), and sample size $n = 20$, we find that the $MI_{n,\varphi}$ test tends to best identify the deviation from the null hypothesis. As we move further away from the Rayleigh distribution ($p = 0.10$ to $p = 0.30$), $MI_{n,\varphi}$ still performs the best. For moderate departure from the Rayleigh distribution ($p = 0.35$ to $p = 0.50$) the $Tb_{n,\varphi}$ test outperforms $MI_{n,\varphi}$ for $n = 20$. For $n = 30$, however, the $MI_{n,\varphi}$ test tends to match $Tb_{n,\varphi}$ in most cases.

The classical tests have the weakest power performance in all cases for $n = 20$ and $n = 30$ with A_n performing the best among the classical tests. For $p = 0.05$ and $p = 0.10$ there is not much to choose between the tests based on entropy in terms of power. The $LA_{n,\varphi}$ test is a strong contender for the cases $p = 0.30$ to $p = 0.50$ and matches the highest power estimate for $p = 0.35$ and $n = 20$. Overall, $T_{n,\varphi}$ and $MI_{n,\varphi}$ shows superior power performance compared to the competing tests for all values of p .

The test $T_{n,\varphi}$ has difficulty detecting departures from the Rayleigh distribution across all values of p for $n = 20$ and also performs poorly for $n = 30$ for the tuning parameter choice, $\varphi = 1$. This seems to be largely isolated to the choice $\varphi = 1$ as it shows improved performance for choices $\varphi = 2$ and $\varphi = 5$. The test with boundary-corrected kernel ($Tb_{n,\varphi}$) does not exhibit this problem. This can be attributed to the boundary-corrected kernel partially remedying the boundary effect of the kernel density estimate around zero for small values of the tuning parameter. The powers of the $CM_{n,\varphi}$ and $LA_{n,\varphi}$ tests once again proves relatively insensitive to the choice of φ .

5.3.3 Discussion on local power estimates for the $\Gamma(1.5) - \text{Ral}(1)$ mixture

Table 5.7 displays the local power estimates for a mixture distribution comprising of the $\Gamma(1.5)$ and $\text{Ral}(1)$ distributions. As before, the top row shows the results for $n = 20$ and the bottom row shows the results for $n = 30$. The $DH_{n,m}$ test with $m = 6$ attains the highest power in the case where $p = 0.05$ and $n = 20$. For $n = 30$ we find that the $LA_{n,\varphi}$ test performs the best. For the cases $p = 0.10$ to $p = 0.20$ ($n = 20$ and $n = 30$), the $LA_{n,\varphi}$ test with $\varphi = 1$ outperforms all other tests. As we move further from the Rayleigh distribution ($p = 0.25$ to 0.50), it is evident that $LA_{n,\varphi}$ still performs the best, however, for $p = 0.25$ ($n = 20$) and $p = 0.45$ ($n = 20$) the $LA_{n,\varphi}$ test is matched by $DH_{n,m}$ test with $m = 6$. The adapted test $M2_{n,\varphi}$ consistently performs better than the $M1_{n,\varphi}$ test for this mixture distribution.

As with the $HN - \text{Ral}(1)$ mixture, we find that $T_{n,\varphi}$ with $\varphi = 1$ lacks in power, but this is remedied in the case of the boundary-corrected kernel test ($Tb_{n,\varphi}$). Furthermore, the tests $Tb_{n,\varphi}$ and $CM_{n,\varphi}$ prove to be serious contenders for the best performing test.

5.4 Real data application

The following data set is analyzed in order to demonstrate the use of the newly proposed tests as well as the existing tests in a real-world setting.

A popular data set associated with the Rayleigh distribution is the average wind speed observations that appear in Best, Rayner & Thas (2010). The aforementioned authors fitted a Rayleigh distribution to the data and it is therefore often used to assess goodness-of-fit for the Rayleigh distribution (see, Alizadeh Noughabi et al. 2012, Jahanshahi et al. 2016). The data set consists of 30 average wind speed

observations (km/h) recorded in 2007 in a suburb of Sydney, Australia. The observations are as follows:

2.7, 3.2, 2.1, 4.8, 7.6, 4.7, 4.2, 4.0, 2.9, 2.9,
 4.6, 4.8, 4.3, 4.6, 3.7, 2.4, 4.9, 4.0, 7.7, 10.0,
 5.2, 2.6, 4.2, 3.6, 2.5, 3.3, 3.1, 3.7, 2.8, 4.0.

The newly proposed and existing tests were calculated using the data after the known location parameter ($\mu = 1.5$) was subtracted. Thereafter the p -values were calculated with the use of Monte Carlo simulations. For the test statistics with a tuning parameter or window-width a compromise choice was made, i.e., the tuning parameter or window-width was chosen for which the power performance was satisfactory in the power study. The hypothesis to be tested is that the data are Rayleigh distributed and the p -values are given in Table 5.3. It is clear that all of the tests, with the exception of $DH_{n,m}$, CK_n and KS_n were not significant at a 5% level. The majority of the tests therefore do not reject the null hypothesis that the data realized from a Rayleigh distribution. These findings are in line with those found in Best et al. (2010).

Table 5.3: Calculated p -values for the average wind speed data.

Test statistic	D_n	V_n	W_n	A_n	$KL_{n,6}$	$DH_{n,6}$	$MI_{n,5}$
p-value	0.086	0.151	0.100	0.093	0.557	0.002	0.667
Test statistic	$T_{n,5}$	$Tb_{n,5}$	$M1_{n,2}$	$M2_{n,2}$	$C1_n$	$C2_n$	CK_n
p-value	0.617	0.667	0.989	0.053	0.104	0.111	0.005
Test statistic	PKL_n	PH_n	PJ_n	PTV_n	PC_n	$CM_{n,5}$	KS_n
p-value	0.074	0.157	0.118	0.901	0.618	0.477	0.015
Test statistic	$LA_{n,5}$	R_n					
p-value	0.293	0.246					

Table 5.4: Estimated powers for the alternative distributions given in Table 5.1 for the MLE and sample size $n = 20$.

CHR	D_n	V_n	W_n	A_n	$KL_{n,3}$	$KL_{n,4}$	$KL_{n,6}$	$DH_{n,3}$	$DH_{n,4}$	$DH_{n,6}$	$MI_{n,5}$	$MI_{n,2}$	$MI_{n,5}$	$T_{n,1}$	$T_{n,2}$	$T_{n,5}$
<i>Exp(1)</i>	86	77	89	95	84	84	84	94	95	96	96	97	95	73	94	97
DHR																
<i>BEX(0.7)</i>	97	94	98	100	97	98	97	99	100	100	100	100	99	94	99	100
<i>BEX(0.9)</i>	91	84	93	98	90	90	90	97	97	98	98	98	97	81	97	99
<i>EG(0.2)</i>	90	83	93	97	89	89	89	96	97	98	97	98	97	81	96	98
<i>EG(0.5)</i>	96	92	97	99	95	96	95	99	99	99	99	99	99	91	99	99
<i>EG(0.8)</i>	99	98	99	100	99	99	99	100	100	100	100	100	100	98	100	100
<i>EL(0.2)</i>	97	94	98	99	97	97	97	99	99	100	99	100	99	94	99	100
<i>EL(0.5)</i>	92	86	95	98	92	92	91	97	98	99	98	99	98	85	97	99
<i>EL(0.8)</i>	88	80	92	96	87	87	87	95	96	97	97	98	96	77	95	98
<i>W(0.8)</i>	97	95	98	100	98	98	98	99	100	100	100	100	100	95	99	100
IHR																
<i>CH(1)</i>	55	47	61	78	53	55	55	73	77	80	84	83	76	35	74	87
<i>CH(1.5)</i>	7	7	7	12	7	8	7	10	11	11	21	13	9	4	10	22
<i>EV(1.5)</i>	23	22	25	45	27	29	30	39	43	46	63	51	38	11	43	63
$\Gamma(1.5)$	57	45	63	73	44	44	42	71	74	77	70	76	74	30	64	76
$\Gamma(2)$	32	23	36	43	19	18	15	46	49	52	33	44	45	10	30	44
<i>GO(0.5)</i>	16	17	17	36	22	23	25	30	32	35	56	39	26	8	34	56
<i>GO(1.5)</i>	48	40	52	71	47	48	48	67	70	73	80	77	68	27	67	83
<i>LF(2)</i>	37	31	42	61	35	36	37	56	61	65	71	67	58	19	56	74
<i>LF(4)</i>	25	21	29	46	24	25	25	42	47	50	59	53	44	11	41	61
<i>PW(1)</i>	15	26	19	39	44	47	48	41	43	39	55	34	17	25	43	53
<i>W(1.4)</i>	36	27	41	53	25	25	24	52	56	59	55	58	53	14	44	62
NMHR																
<i>B(0.5)</i>	87	87	89	98	96	97	97	98	98	98	99	98	95	87	98	99
<i>DL(1)</i>	64	51	69	74	49	48	45	75	77	80	61	75	76	37	64	74
<i>DL(1.5)</i>	26	17	29	31	13	13	10	38	40	43	16	29	34	8	19	27
<i>ENH1(2)</i>	91	85	93	96	86	86	85	96	96	97	94	97	96	79	93	97
<i>HN(1)</i>	47	38	52	71	45	46	46	65	69	73	78	75	67	26	65	81
<i>IG(0.5)</i>	36	26	40	39	27	25	20	50	52	53	6	28	38	14	24	23
<i>IG(1.5)</i>	92	87	94	95	86	86	84	95	96	97	87	95	96	78	91	94
<i>LN(0.8)</i>	66	54	70	72	52	51	45	76	79	79	47	68	74	40	58	65
<i>PW(2)</i>	87	87	88	98	96	97	97	97	98	98	99	99	95	87	98	99

CHR	$Tb_{n,1}$	$Tb_{n,2}$	$Tb_{n,5}$	$M1_{n,1}$	$M1_{n,2}$	$M1_{n,5}$	$M2_{n,1}$	$M2_{n,2}$	$M2_{n,5}$	$C1_n$	$C2_n$	CK_n	PKL_n	PH_n	PJ_n	PTV_n
DHR																
<i>Exp(1)</i>	93	97	97	93	94	94	91	92	91	95	94	88	90	97	97	96
<i>BEX(0.7)</i>	99	100	100	99	99	99	98	98	98	99	99	97	99	100	100	100
<i>BEX(0.9)</i>	96	98	98	96	96	96	94	94	94	97	97	92	94	99	99	98
<i>EG(0.2)</i>	95	98	98	95	96	96	94	94	94	97	96	92	93	98	98	98
<i>EG(0.5)</i>	98	99	99	98	98	98	98	98	98	99	99	96	97	99	99	99
<i>EG(0.8)</i>	100	100	100	99	100	100	99	100	99	100	100	99	99	100	100	100
<i>EL(0.2)</i>	99	100	100	99	99	99	99	99	98	99	99	98	98	100	100	100
<i>EL(0.5)</i>	97	99	99	97	97	97	96	96	96	98	97	94	95	99	99	98
<i>EL(0.8)</i>	94	98	98	95	95	95	93	93	93	96	96	90	91	98	98	97
<i>W(0.8)</i>	99	100	100	99	99	99	98	99	98	99	99	98	99	100	100	100
IHR																
<i>CH(1)</i>	72	84	87	78	78	78	64	64	62	75	74	58	64	85	85	83
<i>CH(1.5)</i>	10	18	22	13	13	13	7	6	6	11	11	5	8	19	18	19
<i>EV(1.5)</i>	41	59	64	48	49	49	26	24	23	39	37	20	36	59	59	58
$\Gamma(1.5)$	61	75	75	63	64	64	67	68	68	75	74	64	54	79	79	74
$\Gamma(2)$	27	41	41	30	30	29	41	42	43	48	47	40	24	49	49	41
<i>GO(0.5)</i>	31	49	56	36	38	39	17	15	14	29	27	12	30	51	50	49
<i>GO(1.5)</i>	65	80	83	72	73	73	55	54	53	68	66	48	58	80	80	78
<i>LF(2)</i>	53	70	74	60	61	61	46	46	45	58	56	41	47	72	72	69
<i>LF(4)</i>	39	57	61	46	47	47	32	32	32	44	42	29	35	60	59	56
<i>PW(1)</i>	42	50	53	33	34	35	13	8	7	25	24	21	39	51	50	55
<i>W(1.4)</i>	41	58	60	47	48	47	46	46	46	55	54	43	34	63	62	58
NMHR																
<i>B(0.5)</i>	98	99	99	98	98	98	88	84	81	95	95	84	97	99	99	99
<i>DL(1)</i>	61	72	70	55	56	55	73	74	74	78	77	72	56	78	78	70
<i>DL(1.5)</i>	16	25	24	14	14	14	32	34	35	38	36	36	18	35	36	24
<i>ENHI(2)</i>	92	96	96	90	91	91	94	95	95	96	96	93	89	97	97	95
<i>HN(1)</i>	62	78	81	69	70	70	54	55	53	68	66	50	57	80	80	77
<i>IG(0.5)</i>	17	21	14	4	4	4	41	43	44	46	45	47	22	35	36	16
<i>IG(1.5)</i>	89	93	92	81	82	81	95	95	95	96	96	94	86	95	95	93
<i>LN(0.8)</i>	55	62	58	37	37	38	74	74	75	77	76	73	53	72	72	60
<i>PW(2)</i>	98	99	99	98	98	98	88	84	81	95	94	84	97	99	99	99

	PC_n	$CM_{n,1}$	$CM_{n,2}$	$CM_{n,5}$	KS_n	$LA_{n,1}$	$LA_{n,2}$	$LA_{n,5}$	R_n
CHR									
<i>Exp(1)</i>	97	95	97	96	88	97	98	98	84
DHR									
<i>BEX(0.7)</i>	100	99	100	100	97	100	100	100	98
<i>BEX(0.9)</i>	99	97	98	98	92	98	99	99	90
<i>EG(0.2)</i>	98	97	98	98	92	98	99	98	89
<i>EG(0.5)</i>	99	99	99	99	97	99	100	99	95
<i>EG(0.8)</i>	100	100	100	100	99	100	100	100	99
<i>EL(0.2)</i>	100	99	100	100	98	100	100	100	97
<i>EL(0.5)</i>	99	98	99	98	94	99	99	99	92
<i>EL(0.8)</i>	98	96	98	97	91	97	98	98	87
<i>W(0.8)</i>	100	99	100	100	98	100	100	100	98
IHR									
<i>CH(1)</i>	85	78	82	83	59	82	87	88	54
<i>CH(1.5)</i>	20	12	16	18	7	11	17	21	8
<i>EV(1.5)</i>	59	42	50	55	23	48	59	64	28
$\Gamma(1.5)$	78	76	78	76	63	79	82	77	44
$\Gamma(2)$	45	49	51	45	38	53	54	46	19
<i>GO(0.5)</i>	51	28	37	44	15	36	47	55	22
<i>GO(1.5)</i>	80	69	74	76	50	76	82	83	48
<i>LF(2)</i>	71	59	65	66	41	67	74	74	36
<i>LF(4)</i>	58	44	50	53	28	54	61	62	25
<i>PW(1)</i>	57	23	32	42	14	27	41	52	45
<i>W(1.4)</i>	61	55	59	57	41	61	66	62	26
NMHR									
<i>B(0.5)</i>	99	96	97	98	86	98	99	99	96
<i>DL(1)</i>	75	79	80	74	70	81	82	75	49
<i>DL(1.5)</i>	29	37	37	29	32	43	40	28	14
<i>ENHI(2)</i>	96	97	97	97	93	97	98	97	86
<i>HN(1)</i>	79	67	73	75	50	75	81	82	45
<i>IG(0.5)</i>	25	41	39	25	42	50	42	21	27
<i>IG(1.5)</i>	95	96	96	95	94	96	97	95	86
<i>LN(0.8)</i>	67	77	75	68	71	80	77	66	52
<i>PW(2)</i>	99	95	97	98	85	98	99	99	96

Table 5.5: Estimated powers for the alternative distributions given in Table 5.1 for the MLE and sample size $n = 30$.

	D_n	V_n	W_n	A_n	$KL_{n,3}$	$KL_{n,4}$	$KL_{n,6}$	$DH_{n,3}$	$DH_{n,4}$	$DH_{n,6}$	$MI_{n,5}$	$MI_{n,2}$	$MI_{n,5}$	$T_{n,1}$	$T_{n,2}$	$T_{n,5}$
CHR																
<i>EXP(1)</i>	96	92	98	99	96	96	96	98	99	99	99	100	99	93	99	100
DHR																
<i>BEX(0.7)</i>	100	99	100	100	100	100	100	100	100	100	100	100	100	100	100	100
<i>BEX(0.9)</i>	91	84	93	98	90	90	90	97	97	98	98	98	97	81	97	99
<i>EG(0.2)</i>	98	95	99	100	98	98	98	99	100	100	100	100	100	96	99	100
<i>EG(0.5)</i>	99	99	100	100	99	99	99	100	100	100	100	100	100	99	100	100
<i>EG(0.8)</i>	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
<i>EL(0.2)</i>	100	99	100	100	100	100	100	100	100	100	100	100	100	100	100	100
<i>EL(0.5)</i>	99	97	99	100	99	99	99	100	100	100	100	100	100	97	100	100
<i>EL(0.8)</i>	97	94	98	99	97	97	97	99	99	99	99	100	99	95	99	100
<i>W(0.8)</i>	100	99	100	100	100	100	100	100	100	100	100	100	100	100	100	100
IHR																
<i>CH(1)</i>	74	65	79	90	74	75	74	85	87	90	94	93	89	61	89	95
<i>CH(1.5)</i>	8	9	8	14	10	10	10	12	12	13	28	15	10	6	15	27
<i>EV(1.5)</i>	32	31	35	58	41	42	42	49	52	55	77	63	50	24	60	77
$\Gamma(1.5)$	75	64	81	88	64	65	63	82	86	89	84	89	88	53	82	90
$\Gamma(2)$	46	34	52	57	29	29	26	55	59	64	45	57	60	18	45	57
<i>GO(0.5)</i>	21	23	24	46	34	34	35	37	40	42	69	50	35	17	49	67
<i>GO(1.5)</i>	65	56	71	86	66	67	67	78	81	85	92	89	83	52	84	92
<i>LF(2)</i>	53	44	58	76	53	53	52	68	72	76	85	80	74	38	74	86
<i>LF(4)</i>	36	30	42	60	37	37	37	52	56	61	74	66	57	22	58	75
<i>PW(1)</i>	21	39	27	53	67	70	72	57	58	57	69	41	20	53	63	66
<i>W(1.4)</i>	52	41	59	69	39	40	37	62	66	71	70	72	70	28	61	76
NMHR																
<i>B(0.5)</i>	97	97	97	100	100	100	100	100	100	100	100	100	99	98	100	100
<i>DH(1)</i>	81	70	85	88	69	69	66	85	88	91	78	88	89	60	82	86
<i>DH(1.5)</i>	36	24	40	42	22	22	19	45	50	54	22	39	46	12	28	35
<i>ENH(2)</i>	98	96	99	99	96	96	96	99	99	99	99	99	99	95	99	99
<i>HN(1)</i>	64	55	70	84	64	65	64	77	81	84	90	88	83	49	83	92
<i>IG(0.5)</i>	51	39	55	54	42	42	37	61	65	67	9	36	51	23	35	31
<i>IG(1.5)</i>	99	97	99	99	96	97	96	99	99	99	96	99	99	95	98	99
<i>LN(0.8)</i>	82	73	86	87	71	71	68	86	88	90	62	83	88	60	78	80
<i>PW(2)</i>	97	97	97	100	99	100	100	100	100	100	100	100	99	98	100	100

CHR	$Tb_{n,1}$	$Tb_{n,2}$	$Tb_{n,5}$	$M1_{n,1}$	$M1_{n,2}$	$M1_{n,5}$	$M2_{n,1}$	$M2_{n,2}$	$M2_{n,5}$	$C1_n$	$C2_n$	CK_n	PKL_n	PH_n	PJ_n	PTV_n
CHR																
<i>EXP(1)</i>	99	100	100	99	99	99	98	98	98	99	99	96	98	100	100	99
DHR																
<i>BEX(0.7)</i>	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
<i>BEX(0.9)</i>	96	98	98	96	96	96	94	94	94	97	97	92	94	99	99	98
<i>EG(0.2)</i>	99	100	100	99	99	99	99	99	99	100	99	98	98	100	100	100
<i>EG(0.5)</i>	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
<i>EG(0.8)</i>	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
<i>EL(0.2)</i>	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
<i>EL(0.5)</i>	100	100	100	100	100	100	99	99	99	100	100	99	99	100	100	100
<i>EL(0.8)</i>	99	100	100	99	99	99	99	99	98	99	99	97	98	100	100	100
<i>W(0.8)</i>	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
IHR																
<i>CH(1)</i>	88	94	95	91	90	91	81	80	78	88	87	72	80	94	94	93
<i>CH(1.5)</i>	14	23	27	15	16	17	7	6	6	13	12	5	9	22	22	23
<i>EV(1.5)</i>	56	73	77	61	62	64	36	33	31	51	49	23	47	73	72	70
$\Gamma(1.5)$	81	89	88	80	80	81	83	84	83	89	88	79	72	92	92	87
$\Gamma(2)$	42	55	53	41	41	42	56	57	57	62	61	54	34	63	63	52
<i>GO(0.5)</i>	45	62	67	48	49	52	23	20	18	37	35	14	39	63	62	60
<i>GO(1.5)</i>	83	91	92	86	86	87	72	71	70	83	82	61	73	92	91	90
<i>LF(2)</i>	71	84	86	76	76	77	62	62	61	73	71	52	62	85	85	81
<i>LF(4)</i>	55	71	75	60	61	62	45	44	44	56	54	36	47	73	73	67
<i>PW(1)</i>	66	66	66	41	43	46	17	9	7	33	32	35	55	69	68	77
<i>W(1.4)</i>	59	73	75	63	63	65	62	61	62	71	69	56	48	77	76	70
NMHR																
<i>B(0.5)</i>	100	100	100	100	100	100	97	95	92	99	99	94	100	100	100	100
<i>DH(1)</i>	79	86	84	72	71	73	87	88	88	90	90	85	72	90	90	83
<i>DH(1.5)</i>	25	34	30	19	19	19	45	46	48	48	47	47	25	45	46	29
<i>ENH1(2)</i>	99	99	99	98	98	98	99	99	99	99	99	98	97	100	100	99
<i>HN(1)</i>	81	90	92	85	84	85	72	71	70	82	81	63	72	91	90	88
<i>IG(0.5)</i>	27	29	18	6	5	5	57	57	58	60	59	60	35	49	50	22
<i>IG(1.5)</i>	98	99	98	94	93	94	99	99	99	99	99	99	96	99	99	98
<i>LN(0.8)</i>	73	79	74	52	52	53	88	88	89	89	89	86	70	86	87	74
<i>PW(2)</i>	100	100	100	100	100	100	97	95	92	99	99	94	100	100	100	100

	PC_n	$CM_{n,1}$	$CM_{n,2}$	$CM_{n,5}$	KS_n	$LA_{n,1}$	$LA_{n,2}$	$LA_{n,5}$	R_n
CHR									
<i>EXP(1)</i>	100	99	99	99	97	99	100	100	96
DHR									
<i>BEX(0.7)</i>	100	100	100	100	100	100	100	100	100
<i>BEX(0.9)</i>	99	97	98	98	92	98	99	99	90
<i>EG(0.2)</i>	100	100	100	100	98	100	100	100	98
<i>EG(0.5)</i>	100	100	100	100	100	100	100	100	99
<i>EG(0.8)</i>	100	100	100	100	100	100	100	100	100
<i>EL(0.2)</i>	100	100	100	100	100	100	100	100	100
<i>EL(0.5)</i>	100	100	100	100	99	100	100	100	98
<i>EL(0.8)</i>	100	99	100	100	98	100	100	100	97
<i>W(0.8)</i>	100	100	100	100	100	100	100	100	100
IHR									
<i>CH(1)</i>	94	89	92	93	76	92	95	95	75
<i>CH(1.5)</i>	23	13	17	23	7	12	20	26	10
<i>EV(1.5)</i>	72	52	61	68	31	58	70	76	41
$\Gamma(1.5)$	90	88	90	89	80	91	93	91	65
$\Gamma(2)$	57	61	62	57	52	68	68	59	29
<i>GO(0.5)</i>	63	37	47	58	20	42	58	67	33
<i>GO(1.5)</i>	91	83	87	89	66	87	92	93	67
<i>LF(2)</i>	84	74	78	81	55	80	86	87	53
<i>LF(4)</i>	71	57	63	68	39	65	73	75	36
<i>PW(1)</i>	76	29	41	55	18	27	49	64	69
<i>W(1.4)</i>	74	70	73	73	58	75	79	77	40
NMHR									
<i>B(0.5)</i>	100	99	100	100	96	100	100	100	100
<i>DH(1)</i>	87	90	90	87	85	92	92	88	68
<i>DH(1.5)</i>	36	47	46	37	43	56	53	37	21
<i>ENH1(2)</i>	99	100	100	99	99	100	100	100	97
<i>HN(1)</i>	90	82	86	88	66	87	92	92	64
<i>IG(0.5)</i>	36	56	50	35	57	65	57	29	42
<i>IG(1.5)</i>	99	99	99	99	99	99	100	99	97
<i>LN(0.8)</i>	81	89	88	82	86	91	90	81	71
<i>PW(2)</i>	100	99	100	100	96	100	100	100	100

Table 5.6: Estimated local powers for $n = 20$ (top row) and $n = 30$ (bottom row) for the $HN - Ral(1)$ mixture distribution and MLE.

D_n	V_n	W_n	A_n	$KL_{n,3}$	$KL_{n,4}$	$KL_{n,6}$	$DH_{n,3}$	$DH_{n,4}$	$DH_{n,6}$	$MI_{n,5}$	$MI_{n,2}$	$MI_{n,5}$	$T_{n,1}$	$T_{n,2}$	$T_{n,5}$
$p = 0$															
$n = 20$	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
$n = 30$	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
$p = 0.05$															
$n = 20$	5	5	5	5	5	5	7	7	7	10	7	5	5	6	10
$n = 30$	5	5	5	5	5	5	7	7	7	12	7	6	4	7	10
$p = 0.10$															
$n = 20$	6	5	6	9	5	6	8	9	10	16	9	7	5	8	14
$n = 30$	6	6	6	9	6	6	9	9	10	19	10	8	4	9	16
$p = 0.15$															
$n = 20$	6	6	7	11	6	6	10	11	13	21	13	9	4	9	20
$n = 30$	7	7	12	7	7	7	11	12	14	25	14	11	4	12	24
$p = 0.20$															
$n = 20$	7	6	8	14	7	8	13	14	16	27	17	12	4	13	26
$n = 30$	8	8	9	16	9	9	15	16	18	33	19	14	5	17	31
$p = 0.25$															
$n = 20$	8	7	9	17	9	9	15	18	20	32	20	15	4	15	31
$n = 30$	10	10	11	21	11	12	18	20	22	39	25	18	5	21	38
$p = 0.30$															
$n = 20$	10	9	11	21	10	11	19	21	24	37	25	19	4	18	36
$n = 30$	13	12	15	27	15	15	23	25	28	46	31	23	6	26	44
$p = 0.35$															
$n = 20$	11	10	13	25	12	12	22	25	27	41	29	23	5	22	42
$n = 30$	16	14	18	32	19	18	28	29	33	52	37	29	8	32	51
$p = 0.40$															
$n = 20$	14	12	16	29	14	14	27	29	32	46	34	26	5	26	47
$n = 30$	19	17	22	38	21	22	32	34	39	58	44	34	10	39	58
$p = 0.45$															
$n = 20$	16	14	19	34	17	18	31	33	36	50	38	31	6	30	51
$n = 30$	23	20	26	44	25	27	36	40	45	63	51	40	13	45	63
$p = 0.50$															
$n = 20$	19	17	22	39	20	20	34	37	41	54	44	34	8	34	55
$n = 30$	27	23	31	50	29	30	42	45	50	68	56	45	15	50	67

$Tb_{n,1}$	$Tb_{n,2}$	$Tb_{n,5}$	$M1_{n,1}$	$M1_{n,2}$	$M1_{n,5}$	$M2_{n,1}$	$M2_{n,2}$	$M2_{n,5}$	$C1_n$	$C2_n$	CK_n	PKL_n	PH_n	PJ_n	PTV_n
$p = 0$															
$n = 20$	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
$n = 30$	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
$p = 0.05$															
$n = 20$	6	8	10	7	6	6	5	5	6	6	5	6	9	9	8
$n = 30$	6	9	10	6	6	7	5	6	6	6	5	7	10	10	9
$p = 0.10$															
$n = 20$	8	12	14	8	8	6	6	6	8	8	7	8	14	14	12
$n = 30$	8	14	17	8	9	6	6	6	9	8	7	9	17	17	13
$p = 0.15$															
$n = 20$	10	16	21	11	11	11	8	7	10	10	8	10	20	20	17
$n = 30$	11	20	24	12	13	14	8	8	11	11	7	12	23	23	18
$p = 0.20$															
$n = 20$	11	21	26	14	15	15	8	9	12	12	8	13	24	24	21
$n = 30$	15	27	31	16	17	18	10	10	15	14	9	15	29	29	24
$p = 0.25$															
$n = 20$	13	26	31	17	18	19	11	10	15	14	9	15	29	29	26
$n = 30$	18	33	38	21	22	24	12	12	18	17	11	18	36	36	30
$p = 0.30$															
$n = 20$	17	31	37	21	22	23	12	12	19	18	12	18	34	34	30
$n = 30$	24	40	45	28	28	29	16	15	24	23	13	23	43	42	36
$p = 0.35$															
$n = 20$	21	35	42	25	26	27	15	15	23	21	14	20	38	38	35
$n = 30$	29	46	52	33	34	35	19	19	28	27	16	26	49	49	43
$p = 0.40$															
$n = 20$	23	41	47	29	31	31	17	17	27	25	16	24	43	43	40
$n = 30$	34	53	58	39	41	42	23	23	34	32	19	31	55	54	48
$p = 0.45$															
$n = 20$	27	45	51	33	35	36	20	20	30	29	19	26	48	47	44
$n = 30$	40	59	64	45	47	48	28	27	40	38	22	35	60	59	54
$p = 0.50$															
$n = 20$	31	50	56	39	39	40	24	23	35	33	22	30	54	53	49
$n = 30$	45	65	68	51	52	53	33	32	45	43	26	39	66	65	59

	PC_n	$CM_{n,1}$	$CM_{n,2}$	$CM_{n,5}$	KS_n	$LA_{n,1}$	$LA_{n,2}$	$LA_{n,5}$	R_n
$p = 0$									
$n = 20$	5	5	5	5	5	5	5	5	5
$n = 30$	5	5	5	5	5	5	5	5	5
$p = 0.05$									
$n = 20$	9	7	7	7	5	8	9	9	5
$n = 30$	9	7	8	8	6	8	9	10	5
$p = 0.10$									
$n = 20$	13	9	10	10	6	10	12	13	5
$n = 30$	14	9	11	13	6	11	15	16	6
$p = 0.15$									
$n = 20$	18	11	13	15	7	14	18	20	6
$n = 30$	20	12	15	18	8	15	20	23	7
$p = 0.20$									
$n = 20$	22	14	17	19	8	18	22	25	8
$n = 30$	26	16	20	24	9	21	27	31	9
$p = 0.25$									
$n = 20$	27	17	21	24	9	22	28	31	8
$n = 30$	32	19	24	29	11	25	33	37	11
$p = 0.30$									
$n = 20$	32	21	25	28	11	25	32	36	11
$n = 30$	39	25	31	35	14	31	40	44	15
$p = 0.35$									
$n = 20$	37	25	29	34	13	31	37	42	13
$n = 30$	46	30	36	42	17	37	46	51	18
$p = 0.40$									
$n = 20$	41	28	34	38	16	34	43	46	14
$n = 30$	51	35	43	49	20	43	53	57	21
$p = 0.45$									
$n = 20$	46	32	38	42	19	39	47	50	17
$n = 30$	57	57	58	57	24	49	60	63	26
$p = 0.50$									
$n = 20$	51	37	43	46	21	44	52	55	20
$n = 30$	63	46	55	60	29	46	55	60	30

Table 5.7: Estimated local powers for $n = 20$ (top row) and $n = 30$ (bottom row) for the $\Gamma(1.5) - \text{Ral}(1)$ mixture distribution and MLE.

D_n	V_n	W_n	A_n	$KL_{n,3}$	$KL_{n,4}$	$KL_{n,6}$	$DH_{n,3}$	$DH_{n,4}$	$DH_{n,6}$	$MI_{n,5}$	$MI_{n,2}$	$MI_{n,5}$	$T_{n,1}$	$T_{n,2}$	$T_{n,5}$
$p = 0$															
$n = 20$	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
$n = 30$	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
$p = 0.05$															
$n = 20$	7	6	7	8	6	5	9	10	11	8	8	9	5	6	9
$n = 30$	8	6	8	9	7	6	11	12	12	8	8	9	5	7	8
$p = 0.10$															
$n = 20$	9	7	10	11	7	6	14	15	14	10	10	11	5	8	12
$n = 30$	10	7	12	13	8	7	17	17	18	11	11	14	5	8	12
$p = 0.15$															
$n = 20$	11	8	12	14	7	7	18	19	20	13	14	15	6	10	15
$n = 30$	14	9	16	17	10	9	22	24	25	15	16	19	5	12	17
$p = 0.20$															
$n = 20$	13	10	15	18	9	8	22	24	24	16	17	19	6	12	19
$n = 30$	17	11	20	22	12	11	28	29	31	20	21	24	6	15	22
$p = 0.25$															
$n = 20$	16	11	18	21	10	9	27	28	30	19	21	12	6	14	23
$n = 30$	21	14	24	27	14	13	32	34	37	24	27	30	7	19	28
$p = 0.30$															
$n = 20$	19	12	22	25	12	10	31	31	34	23	26	27	7	17	28
$n = 30$	25	16	29	32	17	16	37	40	43	28	33	36	9	23	34
$p = 0.35$															
$n = 20$	22	15	25	29	13	12	35	36	38	26	30	31	8	20	31
$n = 30$	29	18	33	37	20	18	42	44	48	33	39	42	11	28	39
$p = 0.40$															
$n = 20$	24	16	28	32	15	14	38	40	42	29	33	36	9	23	37
$n = 30$	32	21	37	43	23	20	46	49	53	38	44	47	13	33	44
$p = 0.45$															
$n = 20$	27	18	31	36	16	16	41	44	48	33	38	40	11	26	40
$n = 30$	36	24	42	48	26	24	51	53	57	43	49	52	15	37	50
$p = 0.50$															
$n = 20$	29	20	34	40	19	15	45	48	50	36	42	44	12	30	45
$n = 30$	41	28	46	52	29	28	54	57	61	47	54	57	19	42	55

$Tb_{n,1}$	$Tb_{n,2}$	$Tb_{n,5}$	$M1_{n,1}$	$M1_{n,2}$	$M1_{n,5}$	$M2_{n,1}$	$M2_{n,2}$	$M2_{n,5}$	$C1_n$	$C2_n$	CK_n	PKL_n	PH_n	PJ_n	PTV_n
$p = 0$															
$n = 20$	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
$n = 30$	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
$p = 0.05$															
$n = 20$	6	7	6	6	6	8	8	9	8	8	10	7	10	10	7
$n = 30$	6	8	6	5	6	9	9	10	9	9	12	9	12	12	7
$p = 0.10$															
$n = 20$	8	10	11	7	7	11	11	11	12	12	14	10	15	15	9
$n = 30$	7	11	12	7	7	12	13	13	14	14	18	13	17	18	9
$p = 0.15$															
$n = 20$	9	13	14	8	8	14	15	15	16	15	18	12	19	19	12
$n = 30$	10	15	16	9	9	17	18	19	19	18	24	17	23	24	12
$p = 0.20$															
$n = 20$	11	16	18	10	10	16	18	19	19	19	22	14	24	24	15
$n = 30$	12	20	21	11	11	22	23	24	24	23	29	20	30	31	16
$p = 0.25$															
$n = 20$	13	20	22	12	12	20	22	23	23	22	25	17	28	28	18
$n = 30$	16	26	26	14	15	16	27	30	29	28	34	23	35	36	20
$p = 0.30$															
$n = 20$	15	24	26	14	15	23	26	26	28	27	29	19	32	33	22
$n = 30$	21	32	31	18	19	32	34	35	35	34	38	26	41	42	24
$p = 0.35$															
$n = 20$	18	28	29	17	17	28	29	30	32	31	32	22	36	37	25
$n = 30$	25	38	37	22	23	36	40	40	40	39	43	30	47	47	29
$p = 0.40$															
$n = 20$	21	32	35	20	20	31	32	35	36	34	36	24	41	42	30
$n = 30$	28	42	42	25	27	41	44	45	45	44	47	34	52	53	34
$p = 0.45$															
$n = 20$	25	36	38	23	23	35	36	38	40	38	39	27	45	45	34
$n = 30$	33	47	48	31	32	45	48	50	51	49	50	37	57	57	40
$p = 0.50$															
$n = 20$	28	40	43	27	28	38	39	41	44	42	42	29	49	50	38
$n = 30$	39	52	52	36	37	51	53	54	55	53	54	41	62	62	45

	PC_n	$CM_{n,1}$	$CM_{n,2}$	$CM_{n,5}$	KS_n	$LA_{n,1}$	$LA_{n,2}$	$LA_{n,5}$	R_n
$p = 0$									
$n = 20$	5	5	5	5	5	5	5	5	5
$n = 30$	5	5	5	5	5	5	5	5	5
$p = 0.05$									
$n = 20$	8	9	8	8	8	10	10	8	5
$n = 30$	8	9	9	8	9	13	12	9	6
$p = 0.10$									
$n = 20$	11	12	11	11	11	16	14	12	6
$n = 30$	11	13	13	12	13	20	17	13	7
$p = 0.15$									
$n = 20$	14	16	16	14	14	21	19	15	8
$n = 30$	15	18	18	16	18	27	24	18	9
$p = 0.20$									
$n = 20$	18	19	19	17	17	26	24	19	8
$n = 30$	20	23	23	21	22	34	31	23	11
$p = 0.25$									
$n = 20$	21	23	23	21	20	30	29	23	10
$n = 30$	25	29	29	27	27	40	38	29	13
$p = 0.30$									
$n = 20$	25	27	28	25	24	35	34	28	11
$n = 30$	30	35	35	32	31	46	44	35	15
$p = 0.35$									
$n = 20$	30	32	33	28	27	39	38	31	13
$n = 30$	35	40	41	38	36	51	50	40	18
$p = 0.40$									
$n = 20$	34	36	36	33	30	44	43	37	15
$n = 30$	41	45	46	43	40	56	56	46	21
$p = 0.45$									
$n = 20$	38	41	40	36	33	35	32	34	17
$n = 30$	47	50	51	48	44	61	60	52	24
$p = 0.50$									
$n = 20$	42	45	45	42	37	53	52	45	19
$n = 30$	52	55	56	53	48	66	66	57	28

Chapter 6

Concluding remarks and future research

6.1 Introduction

In this thesis we aimed to develop new goodness-of-fit tests for the Rayleigh distribution and furthermore to conduct Monte Carlo investigation into the performance of the new and existing goodness-of-fit tests. Several goodness-of-fit tests for the Rayleigh distribution were developed in Chapter 4 and the finite-sample performance of these tests were explored alongside the existing tests which were discussed in Chapter 2. The main objectives and outcomes of the objectives can be summarized as follows:

- Review the existing literature on goodness-of-fit tests for the Rayleigh distribution.

The existing literature were presented in Chapter 2. General properties of the Rayleigh distribution as well as characterisations and existing tests were discussed.

- Review the literature on specific statistical tools that are used in developing

our new goodness-of-fit tests for the Rayleigh distribution.

The kernel density estimate as well as aspects of kernel density estimation were reviewed in Chapter 3. The Mellin transform was discussed in Chapter 4, Section 4.4.

- Develop and explore a new goodness-of-fit test based on a conditional expectation characterisation for the Rayleigh distribution.
- Derive the asymptotic theory surrounding the new goodness-of-fit test based on a conditional expectation characterisation.

This test statistic and the asymptotic theory were developed in Chapter 4, Section 4.2.

- Investigate the Mellin transform and apply it in constructing a new goodness-of-fit test for the Rayleigh distribution.

The theory of the Mellin transform and the test statistic based on this transform takes place in Chapter 4, Section 4.4.

- Investigate and develop a new goodness-of-fit test for the Rayleigh distribution based on a differential equation approach.

The development of this test statistic can be found in Chapter 4 , Section 4.3.

- Adapt existing tests to specifically test for the Rayleigh distribution.

The tests statistics by Torabi et al. (2016) and Meintanis (2009) were adapted specifically for the Rayleigh distribution and can be found in Chapter 2.

- Evaluate the performance of the newly suggested goodness-of-fit tests and existing goodness-of-fit tests for the Rayleigh distribution.

The finite-sample performance of the test statistics and subsequent discussions can be found in Chapter 5.

6.2 Concluding remarks and future research

A major aim of the study was to develop new goodness-of-fit tests for the Rayleigh distribution and assess the performance of the newly developed and existing tests. In the Monte Carlo study it is found that the best performing tests are $DH_{n,m}$, $Tb_{n,\varphi}$, $LA_{n,\varphi}$, $CM_{n,\varphi}$. These tests performed superior in terms of power estimates against general alternatives and local power estimates. We find that three of the newly developed tests are among the set of identified best performing tests. The $LA_{n,\varphi}$ and $Tb_{n,\varphi}$ tests consistently attained or matched the highest power estimates, while the $CM_{n,\varphi}$ test proved to be the most stable across tuning parameters values. That is, the $CM_{n,\varphi}$ test achieved competitive power estimates for any $\varphi \geq 1$. Furthermore, the asymptotic theory for this test was thoroughly explored which also allows for obtaining the critical values from the limiting distribution (if one chooses to go this route). The adapted tests $C1_n$, $C2_n$ and $M2_{n,\varphi}$ lacked in performance to the best performing tests. For implementation of the newly developed tests, we advise choosing the tuning parameter as $\varphi = 2$ or $\varphi = 5$ as these choices exhibited good performance in most cases. For implementation of the $DH_{n,m}$ test a choice of $m = 6$ proved to perform well.

The remaining section collects a few avenues for future research.

6.2.1 Estimating F with the plug-in estimator $F_{\hat{\theta}_n}$ in CM_n

In section 4.2 we developed a test statistic based on a conditional expectation characterisation. The test statistic rests on the fact that a random variable X will be Rayleigh distributed if the function $t \mapsto E[X^{2k}I(X > t)]/S(t) - \nu_{k,\theta}(t)$ is zero. Based on this, a Cramér-von Mises type test statistic,

$$CM_n = \int_{\Theta} \mathcal{T}_n^2(t) w(t) d\tilde{F}_n(t), \quad (6.1)$$

was considered. Furthermore, it was noted that one can take $\tilde{F}_n$ as the empirical distribution function $F_n(t) = (1/n) \sum_{i=1}^n I(X_i \leq t)$, or the plug-in estimator $F_{\hat{\theta}_n}$ of $F = F_\theta$, obtained by substituting $\hat{\theta}_n$ for θ into F_θ . In the simulation study of Chapter 5, $\tilde{F}_n$ was taken as the empirical distribution function, however, when $\tilde{F}_n$ is taken as the plug-in estimator $F_{\hat{\theta}_n}$ we can consider (6.1) in the rewritten form

$$S_{n,k,\varphi} = n \int_0^\infty \mathcal{L}_n^2(t) w_\varphi(t) dt,$$

where $\mathcal{L}_n(t) = \frac{n^{-1} \sum_{j=1}^n Y_j^{2k} I(Y_j > t)}{\exp(-t^2/2)} - \sum_{j=0}^k 2^j k^{(j)} t^{2(k-j)}$, $w_\varphi(t) = t \exp(-\varphi t^2)$ and $Y_j = X_j/\hat{\theta}_n$ is scaled data. Note that when $\varphi = 1/2$, the weight $w_\varphi(t) = t \exp(-t^2/2)$ is the density of the standard Rayleigh distribution. With some algebra $S_{n,k,\varphi}$, results in the following closed form expression

$$S_{n,k,\varphi} := I_1 + I_2 + I_3,$$

where

$$I_1 = \begin{cases} \frac{1}{n} \sum_{j=1}^n \sum_{\ell=1}^n Y_j^{2k} Y_\ell^{2k} \left(\frac{\min(Y_j, Y_\ell)^2}{2} \right), & \text{if } \varphi = 1, \\ \frac{1}{n} \sum_{j=1}^n \sum_{\ell=1}^n Y_j^{2k} Y_\ell^{2k} \left(\frac{\exp[(1-\varphi) \min(Y_j, Y_\ell)^2] - 1}{2(1-\varphi)} \right), & \varphi > 0, \varphi \neq 1, \end{cases}$$

$$I_2 = \begin{cases} \sum_{j=1}^n \sum_{\ell=0}^k 2^\ell k^{(\ell)} \left(\frac{Y_j^{4k-2\ell+2}}{k-\ell+1} \right), & \text{if } \varphi = \frac{1}{2} \\ 2 \sum_{j=1}^n \sum_{\ell=0}^k 2^\ell k^{(\ell)} Y_j^{2k} \left(\frac{(\varphi-0.5)^{\ell-k} \gamma(k-\ell+1; (\varphi-0.5) Y_j^2)}{(2\varphi-1)} \right), & \text{if } \varphi > \frac{1}{2} \end{cases}$$

with $\gamma(\cdot)$ the incomplete gamma function, and

$$I_3 = n \sum_{j=0}^k \sum_{\ell=0}^k 2^j 2^\ell k^{(j)} k^{(\ell)} \left(\frac{\varphi^{j-2k+\ell-1}}{2} \right) \Gamma(-j+2k-\ell+1), \quad \varphi > 0,$$

with $\Gamma(\cdot)$ being the gamma function.

Power estimates and asymptotic properties are still to be investigated for this form

of the test statistic.

6.2.2 The T_n test with derivative estimation

In Section 4.3 a test statistic was developed based on the differential equation $\theta^2 x g'(x) + g(x)(x^2 - \theta^2) = 0$ and resulted in the form

$$T_n = n \int_0^\infty \left[y h'_0(y) + \hat{h}_{\hat{\theta}}(y)(y^2 - 1) \right]^2 w(y) d\hat{F}(y).$$

The rationale behind the test statistic was to determine how far $\hat{h}_{\hat{\theta}}(y)$ is from $h_0(y)$ by using a L_2 distance and the approach ultimately rested on taking $h'_0(y)$ as the derivative of the standard Rayleigh distribution and $\hat{h}_{\hat{\theta}}(y)$ as the KDE. Two variants to this test statistic can be constructed by considering the estimation of different features. The first being, estimating $h'_0(y)$ with a kernel derivative density estimate and using $\hat{h}_{\hat{\theta}}(y)$ as the estimate of the underlying density. Denote the first kernel derivative density estimate as $\hat{h}_{\hat{\theta}}^{(1)}(y)$ and consider the test statistic

$$Td_n = n \int_0^\infty \left[y \hat{h}_{\hat{\theta}}^{(1)}(y) + \hat{h}_{\hat{\theta}}(y)(y^2 - 1) \right]^2 w(y) d\hat{F}(y).$$

The second variant test considers the first kernel derivative density estimate as before, and rather than estimating the underlying density, uses the distribution under the null hypothesis denoted by $h_0(y)$. This test takes the form

$$Tr_n = n \int_0^\infty \left[y \hat{h}_{\hat{\theta}}^{(1)}(y) + h_0(y)(y^2 - 1) \right]^2 w(y) d\hat{F}(y).$$

For the implementation of these tests, the following selection of literature needs to be explored.

Kernel density derivative estimation

It is possible to use kernel density estimation to estimate the derivatives of densities. This is an important feature when the gradient of a density is to be investigated. The density derivative estimate is a innate extension of the formulation for the KDE in that it is the derivative of the KDE estimate:

$$\widehat{f}^{(r)}(x; h) = n^{-1}h^{-r-1} \sum_{i=1}^n k^{(r)}\left(\frac{x - X_i}{h}\right), \quad (6.2)$$

where $k^{(r)}(\cdot)$ is the r -th derivative of the kernel function. It logically follows that the derivative of the function in question has to exist and for this reason the Gaussian kernel is a natural choice.

Figure 6.1 illustrates a kernel density derivative estimate for the first derivative of the *Ral*(1)-distribution using a Gaussian kernel. We clearly see the deviation of the estimate on the left boundary. This can be ascribed to the boundary-bias inherent in kernel density estimation amplified to the density derivative estimate. As stated by Jones (1994) this deviation increases as the order of the derivative increases.

Maltz (1974) took a different approach to density estimation and showed that the density derivative can also be estimated by using differences of the empirical distribution function. Jones (1992) investigated this fact and provided a difference estimator in the form

$$\widetilde{f}^{(r)}(x; \ell) = (2\ell)^{-(r+1)} \sum_{j=0}^{r+1} \binom{r+1}{j} (-1)^j F_n(x + (r+1-2j)\ell),$$

contingent on that differences are taken between points symmetric about x and at increments 2ℓ . Furthermore, Jones (1992) states that taking $r = 0$ will give back an expression equivalent to the KDE with the kernel chosen to be the uniform

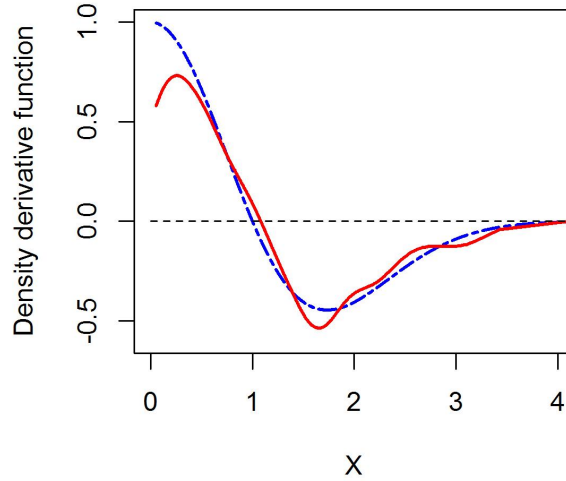


Figure 6.1: First derivative of Rayleigh distribution (blue line) and density estimate (red line).

kernel $U[-1, 1]$ with $h = \ell$. A further result by the author is that $\tilde{f}^{(r)}(x; \ell)$ is equivalent to $\hat{f}^{(r)}(x; h)$ in (6.2) with the kernel function being the density function of the distribution of the convolution $(r + 1 * U[-1, 1])$ random variable.

Bandwidth selection for kernel density derivative estimation

A rule-of-thumb bandwidth can be obtained by minimizing the AMISE and choosing the Gaussian kernel as the unknown density in the expression similar to the case in Section 3.5. An easy to calculate expression arises from the work of Chacón, Duong & Wand (2011)

$$h_{rot} = \{4/(3 + 2r)\}^{(1/(5+2r))} \hat{\sigma} n^{-(1/(5+2r))},$$

where r is once again the order of the derivative.

In the vain of automatic bandwidth selectors for density derivative estimation Härdle, Marron & Wand (1990) suggested a cross-validation method. Analogous to the motivation for cross-validation, the derivative density version starts by noting that an optimal bandwidth can be selected by minimizing

$$\int \hat{f}^{(r)}(x; h) dx - 2 \int \hat{f}^{(r)}(x; h) f(x) dx + \int f(x)^2 dx.$$

An estimator can now be written in the form

$$CV_r(h) = \int \hat{f}^{(r)}(x; h) dx - 2n^{-1}(-1)^r \sum_{i=1}^n \hat{f}_{-i}^{(2r)}(X_i; h),$$

where $\hat{f}_{-i}^{(2r)}$ is the density derivative version of the leave-one-out density estimator

$$\hat{f}_{-i}^{(2r)}(X_i; h) = (n-1)^{-1} \sum_{j \neq i} h^{-2r-1} k^{(2r)} \left\{ \frac{x - X_j}{h} \right\}.$$

Härdle et al. (1990) provides a more direct expression in the form:

$$\begin{aligned} CV_r(h) = & (-1)^r n^{-1} h^{-2r-1} \left[n^{-1} \sum_i \sum_j (k * k)^{(2r)} \left\{ \frac{X_i - X_j}{h} \right\} \right. \\ & \left. - 2(n-1)^{-1} \sum_i \sum_{\substack{j \\ i \neq j}} k^{(2r)} \left\{ \frac{X_i - X_j}{h} \right\} \right]. \end{aligned}$$

Boundary-corrections for kernel density derivative estimation

Recall the kernel density derivative estimate in (6.2) and consider the specific case where $r = 1$, i.e., the first derivative of the kernel density estimate,

$$\hat{f}^{(1)}(x; h) = n^{-1} h^{-2} \sum_{i=1}^n k' \left(\frac{x - X_i}{h} \right). \quad (6.3)$$

From Jones (1993) we find that the expectation on the boundary of $\widehat{f}^{(1)}(x; h)$ is

$$\begin{aligned} E \left[\widehat{f}^{(1)}(x; h) \right] &= h^{-1} a_0^{(1)}(p) f(x) - a_1(p) f'(x) + O(h) \\ &= h^{-1} \int_{-1}^p k'(x) dx f(x) - \int_{-1}^p x k'(x) dx f'(x) + O(h). \end{aligned}$$

On closer inspection of above expression notice that $f'(x)$ is amplified by the term $\int_{-1}^p x k'(x) dx$ and the leading term contains h^{-1} . To adjust the discrepancy and attain a bias of $O(h^2)$ the method of Jones (1993) discussed in Section 3.6.3 can be implemented once again. The generalized jackknife procedure to find a linear combination with desired properties applied to the density derivative problem results in the first derivative boundary kernel,

$$v(u, p) = h^{-1} \left\{ \frac{-A_0(p)k(x) + B_0(p)\ell(x) - C_0(p)m(x)}{a_0(p)A_1(p) - c_0(p)B_1(p) + d_0(p)C_1(p)} \right\}.$$

with appropriate functions $k(\cdot)$, $\ell(\cdot)$ and $m(\cdot)$. Furthermore, $d_s(p) = \int_{-1}^p u^s m(u) du$ and

$$\begin{aligned} A_i(p) &= c_i(p)d_2(p) - c_2(p)d_i(p), \\ B_i(p) &= a_i(p)d_2(p) - a_2(p)d_i(p), \\ C_i(p) &= a_i(p)c_2(p) - a_2(p)c_i(p). \end{aligned}$$

Choices of the functions $\ell(\cdot)$ and $m(\cdot)$ are once again taken to be a function of the kernel $k(\cdot)$ and different options can be found in Jones (1993).

The viability of test statistics Td_n and Tr_n needs to be investigated. Furthermore, the boundary-correction in these statistics will have to be investigated as there are noticeable problems near the boundary with kernel derivative estimation as seen in Figure 6.1.

6.2.3 Power-divergence choice for the test of Zamanzade & Mahdizadeh (2017)

Section 2.4.4 detailed several test statistics based on the phi-divergence measure,

$$D_\phi(P_1||P_2) = \int_{\Omega} \phi\left(\frac{dP_1}{dP_2}\right) dP_2,$$

and an appropriate function for $\phi(\cdot)$. In addition to the options for $\phi(\cdot)$ suggested by Zamanzade & Mahdizadeh (2017), one can consider a distance measure from the power-divergence family of Cressie & Read (1984) as a choice for $\phi(\cdot)$, i.e.,

$$\phi(t) = \frac{t^{\lambda+1} - t - \lambda(t-1)}{\lambda(\lambda+1)}, \quad \lambda \neq 0, -1. \quad (6.4)$$

The motivation for this ϕ -function is a more flexible test statistic in the form of

$$PPD_n = \frac{1}{n} \sum_{i=1}^n \left\{ \left(\frac{g_0(x)}{\hat{g}(x, h)} \right)^{\lambda+1} - \left(\frac{g_0(x)}{\hat{g}(x, h)} \right) - \lambda \left(\frac{g_0(x)}{\hat{g}(x, h)} - 1 \right) \right\} \{\lambda(\lambda+1)\}^{-1},$$

where $\{\lambda \neq 0, -1\}$ can be seen as a tuning parameter. This test statistic potentially makes it possible to emulate the powers of the tests of Zamanzade & Mahdizadeh (2017) with a single choice of the ϕ -function and varying λ values.

6.2.4 Two choices of $\Psi(\cdot)$ for the test of Torabi et al. (2016)

In Section 2.4.4 we explored a test for the normal distribution also based on the phi-divergence measure and ultimately adapted this test specifically for the Rayleigh distribution. The test is given by the form

$$D(F_0, F) = \int_{-\infty}^{\infty} \Psi\left(\frac{1 + F_0(x)}{1 + F(x)}\right) dF(x),$$

and is dependent on choosing a suitable function for $\Psi(\cdot)$. Aside from the functions considered by Torabi et al. (2016), two more functions as possible choices for $\Psi(\cdot)$

can be considered in the form of a measure from the power-divergence family of Cressie & Read (1984) given in (6.4) and also the Matusita divergence measure (see, Pardo 2018) expressed as

$$\Psi(t) = \left| 1 - t^\lambda \right|^{1/\lambda}, \quad 0 < \lambda < 1.$$

By utilizing one of the aforementioned choices a wide range of function shapes can be achieved by varying the associated λ parameter. The parameter in context of the test statistic can be thought of as a type of tuning parameter, allowing for a measure of flexibility.

Appendix A

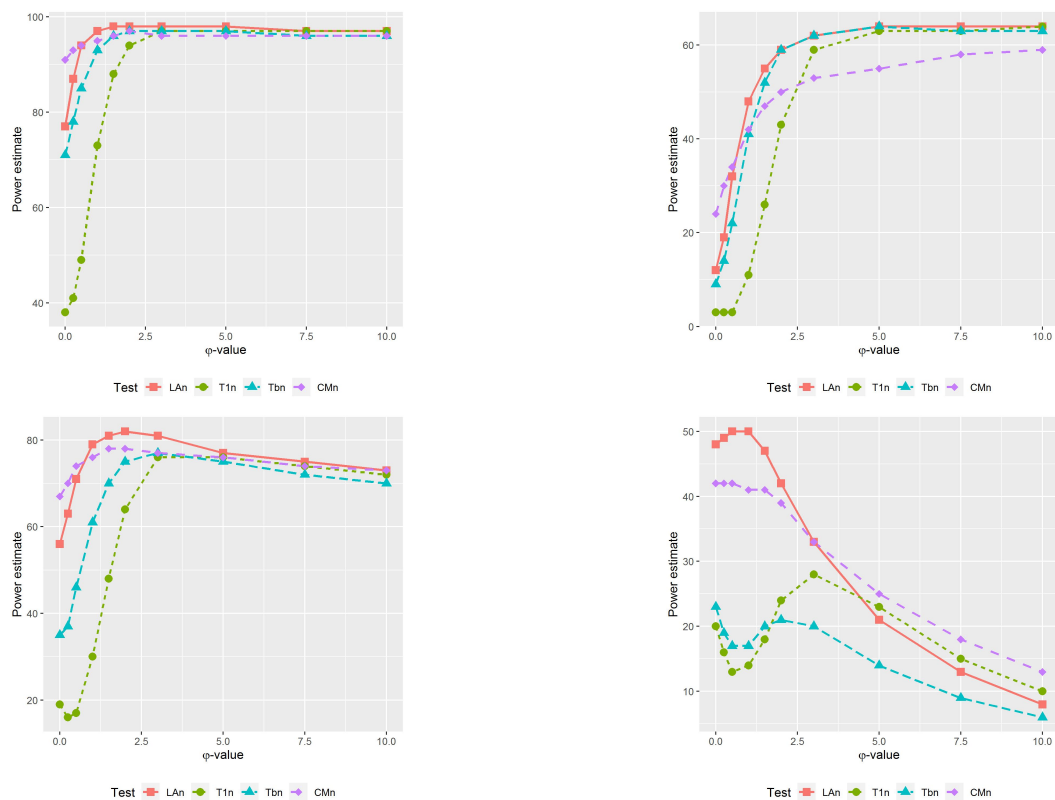


Figure A.1: Power estimates across tuning parameter value for the $EXP(1)$ (top left), $EV(1.5)$ (top right), $\Gamma(1.5)$ (bottom left) and $IG(0.5)$ (bottom right) distributions and $n = 20$.

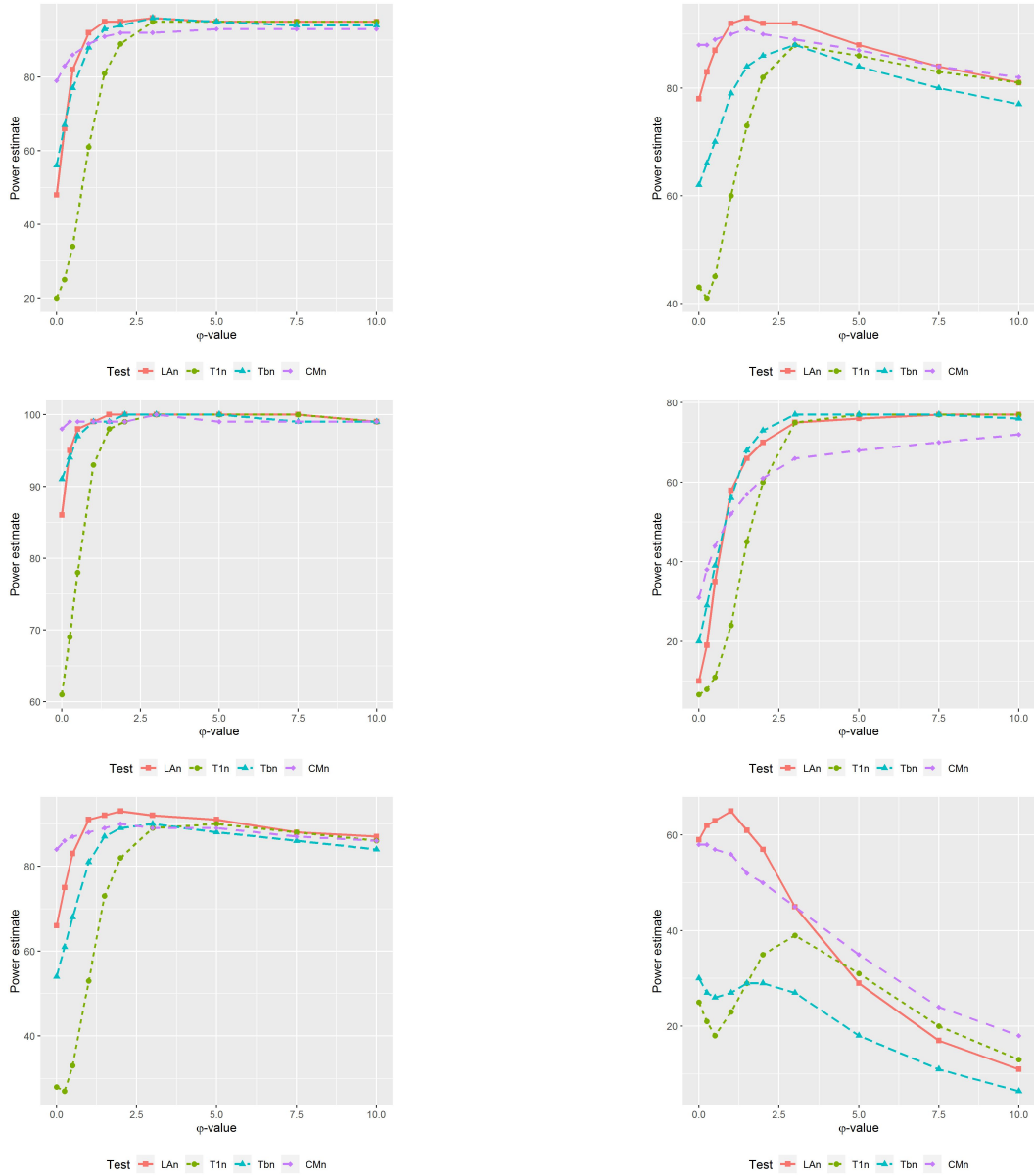


Figure A.2: Power estimates across tuning parameter value for the $CH(1)$ (top left), $DL(1.5)$ (top right), $EXP(1)$ (middle left), $EV(1.5)$ (middle right), $\Gamma(1.5)$ (bottom left) and $IG(0.5)$ (bottom right) distributions and $n = 30$.

Table A.1: Estimated powers for the alternative distributions given in Table 5.1 for the MME and sample size $n = 20$.

	D_n	V_n	W_n	A_n	$KL_{n,3}$	$KL_{n,4}$	$KL_{n,6}$	$DH_{n,3}$	$DH_{n,4}$	$DH_{n,6}$	$MI_{n,5}$	$MI_{n,2}$	$MI_{n,5}$	$T_{n,1}$	$T_{n,2}$	$T_{n,5}$
CHR																
<i>Exp(1)</i>	81	73	85	95	84	84	84	95	96	97	93	97	96	60	89	95
DHR																
<i>BEX(0.7)</i>	95	92	97	100	97	98	97	99	100	100	99	100	100	89	99	100
<i>BEX(0.9)</i>	87	81	90	97	90	90	90	97	97	98	97	98	98	70	94	97
<i>EG(0.2)</i>	86	79	90	97	89	89	89	97	97	98	96	98	98	88	92	97
<i>EG(0.5)</i>	94	89	96	99	95	96	95	99	99	99	98	99	99	97	97	99
<i>EG(0.8)</i>	98	96	99	100	99	99	99	100	100	100	100	100	100	95	99	100
<i>EL(0.2)</i>	96	92	97	99	97	97	97	99	99	100	99	100	100	88	98	99
<i>EL(0.5)</i>	89	83	92	98	92	92	91	98	98	99	97	99	99	74	94	98
<i>EL(0.8)</i>	84	76	88	96	87	87	87	96	97	98	95	98	97	65	91	96
<i>W(0.8)</i>	96	93	98	100	98	98	98	99	100	100	99	100	100	90	99	100
IHR																
<i>CH(1)</i>	52	48	58	78	53	55	55	76	78	81	81	84	81	33	68	84
<i>CH(1.5)</i>	8	8	9	14	7	8	7	11	12	12	21	17	13	6	11	22
<i>EV(1.5)</i>	25	26	29	49	27	29	30	42	46	47	62	56	17	18	42	63
$\Gamma(1.5)$	49	40	55	71	44	44	42	72	75	78	61	74	76	20	49	69
$\Gamma(2)$	25	18	29	39	19	18	15	46	49	52	26	40	46	6	16	34
<i>GO(0.5)</i>	18	21	22	40	22	24	24	33	36	36	55	46	15	14	35	56
<i>GO(1.5)</i>	44	43	51	72	47	48	48	68	72	75	77	79	75	28	62	80
<i>LF(2)</i>	35	32	40	61	35	36	37	58	63	66	67	69	65	19	50	71
<i>LF(4)</i>	24	23	28	47	24	25	25	45	49	52	57	55	50	13	37	59
<i>PW(1)</i>	24	30	30	46	44	47	48	43	44	39	56	43	27	42	49	57
<i>W(1.4)</i>	31	26	35	51	25	25	24	53	57	60	49	57	57	11	33	55
NMHR																
<i>B(0.5)</i>	89	90	92	99	96	97	97	98	98	99	99	99	98	91	99	100
<i>DL(1)</i>	54	40	59	70	49	48	45	75	78	80	48	71	77	20	44	61
<i>DL(1.5)</i>	19	12	21	27	13	13	10	37	40	42	10	23	33	4	8	17
<i>ENH1(2)</i>	61	77	90	95	86	86	85	96	96	97	88	96	97	77	86	93
<i>HN(1)</i>	44	41	50	71	45	46	46	67	71	75	75	77	73	25	60	78
<i>IG(0.5)</i>	29	17	31	33	27	25	20	48	51	52	1	17	34	7	8	8
<i>IG(1.5)</i>	88	79	91	94	86	86	84	95	96	97	74	93	96	55	78	86
<i>LN(0.8)</i>	56	40	60	67	52	51	45	75	78	80	27	60	73	20	35	46
<i>PW(2)</i>	88	90	92	98	96	97	97	98	98	98	99	99	97	91	98	99

CHR	$Tb_{n,1}$	$Tb_{n,2}$	$Tb_{n,5}$	$M1_{n,1}$	$M1_{n,2}$	$M1_{n,5}$	$M2_{n,1}$	$M2_{n,2}$	$M2_{n,5}$	$C1_n$	$C2_n$	CK_n	PKL_n	PH_n	PJ_n	PTV_n
DHR																
<i>Exp(1)</i>	90	95	95	93	94	94	91	93	91	93	93	88	93	97	97	96
<i>BEX(0.7)</i>	99	100	100	99	99	99	98	99	97	99	99	97	99	100	100	100
<i>BEX(0.9)</i>	94	97	97	96	96	96	94	95	93	96	96	92	96	99	99	98
<i>EG(0.2)</i>	98	97	97	95	96	96	94	95	94	96	96	92	95	98	98	98
<i>EG(0.5)</i>	98	99	99	98	98	98	98	98	97	98	98	96	98	100	100	99
<i>EG(0.8)</i>	99	100	100	99	100	100	99	100	99	100	100	99	100	100	100	100
<i>EL(0.2)</i>	95	99	99	99	99	99	99	99	98	99	99	98	99	100	100	100
<i>EL(0.5)</i>	95	98	97	97	97	97	95	96	95	97	97	94	97	99	99	98
<i>EL(0.8)</i>	92	96	96	95	95	95	93	94	92	95	94	90	94	98	98	97
<i>W(0.8)</i>	99	100	100	99	99	99	99	99	98	99	99	98	99	100	100	100
IHR																
<i>CH(1)</i>	71	82	84	78	78	78	68	69	62	73	72	58	71	86	86	84
<i>CH(1.5)</i>	13	19	22	13	13	13	10	9	7	13	12	5	9	18	17	20
<i>EV(1.5)</i>	45	59	63	48	49	49	35	32	24	41	40	20	41	58	57	58
$\Gamma(1.5)$	53	67	67	63	64	64	65	70	70	70	69	64	63	81	81	75
$\Gamma(2)$	21	31	32	30	30	29	38	44	47	42	40	40	32	52	53	43
<i>GO(0.5)</i>	38	51	55	37	38	39	26	22	16	32	31	13	34	48	47	48
<i>GO(1.5)</i>	65	77	80	72	73	72	60	61	54	66	65	48	64	81	81	80
<i>LF(2)</i>	54	67	70	60	61	61	49	51	46	55	54	41	54	73	72	70
<i>LF(4)</i>	41	55	59	46	47	47	35	37	34	41	40	29	41	60	59	57
<i>PW(1)</i>	53	55	56	33	34	35	30	21	11	33	33	21	40	45	44	49
<i>W(1.4)</i>	38	51	54	47	48	47	45	49	48	50	49	43	43	64	64	60
NMHR																
<i>B(0.5)</i>	98	99	100	98	98	99	94	93	80	96	96	84	98	99	99	99
<i>DL(1)</i>	47	59	57	55	56	55	69	76	76	72	71	72	65	80	80	70
<i>DL(1.5)</i>	10	14	15	14	14	14	28	35	40	31	29	36	24	39	39	25
<i>ENHI(2)</i>	87	92	92	90	91	91	94	96	95	95	95	93	92	97	97	95
<i>HN(1)</i>	62	75	77	69	70	70	58	60	54	66	64	50	64	81	80	78
<i>IG(0.5)</i>	7	6	3	4	4	4	37	45	51	39	38	47	30	40	42	15
<i>IG(1.5)</i>	79	85	81	81	82	81	94	96	96	95	95	94	91	96	96	91
<i>LN(0.8)</i>	37	43	38	37	37	38	71	76	78	71	70	73	62	76	76	58
<i>PW(2)</i>	98	99	99	98	98	98	95	93	80	96	96	84	98	99	99	99

	PC_n	$CM_{n,1}$	$CM_{n,2}$	$CM_{n,5}$	KS_n	$LA_{n,1}$	$LA_{n,2}$	$LA_{n,5}$	R_n
CHR									
<i>Exp(1)</i>	97	96	97	97	95	97	98	97	86
DHR									
<i>BEX(0.7)</i>	100	100	100	100	99	100	100	100	98
<i>BEX(0.9)</i>	99	98	98	99	97	98	99	99	92
<i>EG(0.2)</i>	98	98	98	98	97	98	99	98	91
<i>EG(0.5)</i>	99	99	100	100	99	99	100	99	96
<i>EG(0.8)</i>	100	100	100	100	100	100	100	100	99
<i>EL(0.2)</i>	100	100	100	100	99	100	100	100	98
<i>EL(0.5)</i>	99	99	99	99	98	98	99	99	93
<i>EL(0.8)</i>	98	97	98	98	96	97	98	98	89
<i>W(0.8)</i>	100	100	100	100	99	100	100	100	98
IHR									
<i>CH(1)</i>	86	81	83	85	75	81	87	87	57
<i>CH(1.5)</i>	20	12	14	16	10	11	17	21	8
<i>EV(1.5)</i>	60	45	50	54	38	46	58	64	29
$\Gamma(1.5)$	77	79	80	80	74	79	82	77	49
$\Gamma(2)$	46	52	52	51	47	53	54	45	22
<i>GO(0.5)</i>	50	34	38	42	27	34	47	54	23
<i>GO(1.5)</i>	81	74	77	79	67	74	81	83	50
<i>LF(2)</i>	72	65	68	70	58	66	74	74	39
<i>LF(4)</i>	59	51	54	56	43	52	60	62	27
<i>PW(1)</i>	51	24	28	34	19	23	40	52	45
<i>W(1.4)</i>	62	61	62	63	54	61	65	62	29
NMHR									
<i>B(0.5)</i>	99	96	98	99	95	98	99	100	96
<i>DL(1)</i>	73	80	81	80	77	81	82	74	54
<i>DL(1.5)</i>	28	38	39	35	37	43	41	27	16
<i>ENHI(2)</i>	96	97	98	98	96	97	98	97	88
<i>HN(1)</i>	80	73	76	78	68	74	81	81	49
<i>IG(0.5)</i>	22	41	40	33	44	50	44	20	29
<i>IG(1.5)</i>	94	97	97	97	96	96	97	94	88
<i>LN(0.8)</i>	64	78	77	75	76	80	78	65	55
<i>PW(2)</i>	99	96	98	98	94	98	99	99	96

Table A.2: Estimated powers for the alternative distributions given in Table 5.1 for the MME and sample size $n = 30$.

	D_n	V_n	W_n	A_n	$KL_{n,3}$	$KL_{n,4}$	$KL_{n,6}$	$DH_{n,3}$	$DH_{n,4}$	$DH_{n,6}$	$MI_{n,5}$	$MI_{n,2}$	$MI_{n,5}$	$T_{n,1}$	$T_{n,2}$	$T_{n,5}$
CHR																
<i>EXP(1)</i>	94	90	96	99	96	96	96	99	99	99	100	100	100	86	98	99
DHR																
<i>BEX(0.7)</i>	99	99	100	100	100	100	100	100	100	100	100	100	100	99	100	100
<i>BEX(0.9)</i>	87	81	90	97	90	90	90	97	97	98	97	98	98	70	94	97
<i>EG(0.2)</i>	96	94	98	100	98	98	98	99	100	100	99	100	100	92	99	100
<i>EG(0.5)</i>	99	98	99	100	99	99	99	100	100	100	100	100	100	97	100	100
<i>EG(0.8)</i>	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
<i>EL(0.2)</i>	100	99	100	100	100	100	100	100	100	100	100	100	100	99	100	100
<i>EL(0.5)</i>	98	95	98	100	99	99	99	100	100	100	99	100	100	94	99	100
<i>EL(0.8)</i>	95	92	97	99	97	97	97	99	99	100	99	100	100	89	98	100
<i>W(0.8)</i>	100	99	100	100	100	100	100	100	100	100	100	100	100	99	100	100
IHR																
<i>CH(1)</i>	70	67	77	91	74	75	74	86	89	91	92	93	92	57	86	94
<i>CH(1.5)</i>	9	10	10	17	10	10	10	14	14	14	27	19	15	9	16	27
<i>EV(1.5)</i>	33	37	40	62	41	42	42	53	56	59	76	68	60	30	60	77
$\Gamma(1.5)$	67	58	73	86	64	65	63	82	86	89	76	88	90	38	71	84
$\Gamma(2)$	37	29	42	52	29	29	26	53	58	64	34	52	60	11	29	45
<i>GO(0.5)</i>	24	29	30	51	34	34	35	42	44	46	69	57	46	25	51	68
<i>GO(1.5)</i>	61	60	69	87	66	67	67	80	83	86	90	90	88	50	81	91
<i>LF(2)</i>	48	46	55	76	53	53	52	70	74	78	82	82	79	35	69	84
<i>LF(4)</i>	33	32	39	60	37	37	37	54	58	62	72	68	64	24	54	73
<i>PW(1)</i>	34	42	43	60	67	70	72	59	60	58	71	52	34	66	69	70
<i>W(1.4)</i>	43	38	50	67	39	40	37	63	67	72	63	71	74	21	51	69
NMHR																
<i>B(0.5)</i>	98	98	99	100	100	100	100	100	100	100	100	100	100	99	100	100
<i>DH(1)</i>	71	58	77	85	69	69	66	84	88	91	64	84	89	37	66	76
<i>DH(1.5)</i>	27	17	29	35	22	22	19	42	48	53	14	30	43	5	13	21
<i>ENH(2)</i>	96	92	98	99	96	96	96	99	99	99	96	99	100	86	97	99
<i>HN(1)</i>	60	58	67	84	64	65	64	79	82	85	88	89	87	47	79	90
<i>IG(0.5)</i>	42	26	44	46	42	42	37	57	63	66	1	22	45	9	14	10
<i>IG(1.5)</i>	97	93	98	99	96	97	96	99	99	99	89	99	99	83	94	96
<i>LN(0.8)</i>	74	59	78	83	71	71	68	85	88	90	37	76	87	34	56	61
<i>PW(2)</i>	97	98	99	100	99	100	100	100	100	100	100	100	100	99	100	100

CHR	$Tb_{n,1}$	$Tb_{n,2}$	$Tb_{n,5}$	$M1_{n,1}$	$M1_{n,2}$	$M1_{n,5}$	$M2_{n,1}$	$M2_{n,2}$	$M2_{n,5}$	$C1_n$	$C2_n$	CK_n	PKL_n	PH_n	PJ_n	PTV_n
DHR																
<i>EXP(1)</i>	98	99	99	99	99	99	98	98	98	99	98	96	98	100	100	99
<i>BEX(0.7)</i>	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
<i>BEX(0.9)</i>	94	97	97	96	96	96	94	95	93	96	96	92	96	99	99	98
<i>EG(0.2)</i>	99	100	100	99	99	99	99	99	99	99	99	98	99	100	100	100
<i>EG(0.5)</i>	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
<i>EG(0.8)</i>	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
<i>EL(0.2)</i>	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
<i>EL(0.5)</i>	99	100	100	100	100	100	99	99	99	100	99	99	99	100	100	100
<i>EL(0.8)</i>	99	99	99	99	99	99	98	99	98	99	99	97	99	100	100	100
<i>W(0.8)</i>	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
IHR																
<i>CH(1)</i>	87	93	93	91	90	91	82	83	76	87	86	72	85	95	94	94
<i>CH(1.5)</i>	18	25	27	15	16	17	12	9	8	15	14	5	10	21	20	24
<i>EV(1.5)</i>	61	73	76	61	62	64	46	42	30	52	51	23	52	73	72	72
$\Gamma(1.5)$	74	83	82	80	80	81	81	85	83	84	84	79	79	93	93	88
$\Gamma(2)$	32	44	43	41	41	42	50	57	60	54	53	54	43	66	66	54
<i>GO(0.5)</i>	51	64	68	48	49	52	35	29	18	40	39	14	43	62	60	61
<i>GO(1.5)</i>	83	90	91	86	86	87	76	75	67	81	80	61	79	92	92	91
<i>LF(2)</i>	70	82	84	76	76	77	64	65	59	70	68	52	69	86	86	83
<i>LF(4)</i>	55	70	73	60	61	62	48	48	43	53	51	36	53	74	73	70
<i>PW(1)</i>	74	71	68	41	43	46	40	29	13	44	44	35	56	62	61	68
<i>W(1.4)</i>	54	68	68	63	63	65	60	62	62	65	63	56	57	79	78	73
NMHR																
<i>B(0.5)</i>	100	100	100	100	100	100	99	99	91	99	99	94	100	100	100	100
<i>DH(1)</i>	67	76	73	72	71	73	84	88	88	85	84	85	79	92	92	83
<i>DH(1.5)</i>	16	21	19	19	19	19	39	47	52	38	37	47	32	49	50	30
<i>ENH1(2)</i>	97	98	98	98	98	98	99	99	99	99	99	98	98	100	100	99
<i>HN(1)</i>	80	89	89	85	84	85	74	75	68	79	78	63	78	91	91	89
<i>IG(0.5)</i>	12	10	3	6	5	5	52	60	65	51	50	60	42	52	55	18
<i>IG(1.5)</i>	94	96	93	94	93	94	99	99	99	99	99	99	98	99	99	98
<i>LN(0.8)</i>	56	61	52	52	52	53	84	89	90	84	84	86	77	89	89	71
<i>PW(2)</i>	100	100	100	100	100	100	99	98	91	99	99	94	100	100	100	100

	PC_n	$CM_{n,1}$	$CM_{n,2}$	$CM_{n,5}$	KS_n	$LA_{n,1}$	$LA_{n,2}$	$LA_{n,5}$	R_n
CHR									
EXP(1)	100	100	100	100	99	99	100	100	97
DHR									
BEX(0.7)	100	100	100	100	100	100	100	100	100
BEX(0.9)	99	98	98	99	97	98	99	99	92
EG(0.2)	100	100	100	100	100	100	100	100	98
EG(0.5)	100	100	100	100	100	100	100	100	100
EG(0.8)	100	100	100	100	100	100	100	100	100
EL(0.2)	100	100	100	100	100	100	100	100	100
EL(0.5)	100	100	100	100	100	100	100	100	99
EL(0.8)	100	100	100	100	99	100	100	100	97
W(0.8)	100	100	100	100	100	100	100	100	100
IHR									
CH(1)	95	92	93	94	88	92	95	95	78
CH(1.5)	24	14	17	20	11	11	19	26	10
EV(1.5)	74	57	62	67	49	56	69	76	42
$\Gamma(1.5)$	90	91	92	92	89	91	93	90	69
$\Gamma(2)$	58	65	66	65	61	68	68	59	34
GO(0.5)	63	42	47	54	34	40	56	67	34
GO(1.5)	92	87	89	91	82	86	92	93	70
LF(2)	85	79	81	84	72	79	85	87	56
LF(4)	72	63	67	70	55	64	73	75	39
PW(1)	68	28	34	42	23	24	47	64	69
W(1.4)	75	75	76	78	70	75	79	77	45
NMHR									
B(0.5)	100	99	100	100	99	100	100	100	100
DH(1)	87	92	92	91	90	92	93	87	72
DH(1.5)	36	50	50	47	48	57	54	36	24
ENH1(2)	99	100	100	100	99	100	100	100	97
HN(1)	90	86	88	90	81	87	91	92	67
IG(0.5)	30	55	52	43	58	65	58	27	44
IG(1.5)	99	100	99	100	99	99	99	99	98
LN(0.8)	79	90	90	88	89	92	91	80	74
PW(2)	100	100	100	100	99	100	100	100	100

Bibliography

- Ahrari, V., Baratpour, S., Habibirad, A. & Fakoor, V. (2019). Goodness of fit tests for Rayleigh distribution based on quantiles, *Communications in Statistics-Simulation and Computation* pp. 1–17.
- Ahsanullah, M. (2017). *Characterizations of Univariate Continuous Distributions*, Springer.
- Ahsanullah, M. & Shakil, M. (2013). Characterizations of Rayleigh distribution based on order statistics and record values, *Bulletin of the Malaysian Mathematical Sciences Society. Second Series* **3**.
- Alizadeh Noughabi, R., Alizadeh Noughabi, H. & Behabadi, A. E. M. (2012). An entropy test for the Rayleigh distribution and power comparison, *Journal of Statistical Computation and Simulation* **84**(1): 151–158.
- Alzaid, A. A. & Al-Osh, M. A. (1991). Characterization of probability distributions based on the relation, *Sankhyā: The Indian Journal of Statistics, Series B* pp. 188–190.
- Anderson, T. W. & Darling, D. A. (1954). A test of goodness of fit, *Journal of the American Statistical Association* **49**(268): 765–769.
- Arnold, B. C., Balakrishnan, N. & Nagaraja, H. N. (1992). *A First Course in Order Statistics*, Vol. 54, Society for Industrial and Applied Mathematics.
- Balakrishnan, N. & Kocherlakota, S. (1985). On the double Weibull distribution: order statistics and estimation, *Sankhyā: The Indian Journal of Statistics, Series B* pp. 161–178.

- Baratpour, B. S. & Khodadadi, F. (2012). A cumulative residual entropy characterization of the Rayleigh distribution and related goodness-of-fit test, *Journal of Statistical Research of Iran* **9**(2): 115–1294.
- Baringhaus, L., Ebner, B. & Henze, N. (2017). The limit distribution of weighted L2 goodness of fit statistics under fixed alternatives, with applications, *Annals of the Institute of Statistical Mathematics* **69**(5): 969–995.
- Baringhaus, L., Gürtler, N. & Henze, N. (2000). Theory & methods: weighted integral test statistics and components of smooth tests of fit, *Australian & New Zealand Journal of Statistics* **42**(2): 179–192.
- Belaid, A. & Boukerroui, D. (2018). Local maximum likelihood segmentation of echocardiographic images with Rayleigh distribution, *Signal, Image and Video Processing* **12**(6): 1087–1096.
- Best, D. J., Rayner, J. C. W. & Thas, O. (2010). Easily applied tests of fit for the Rayleigh distribution, *Sankhyā: The Indian Journal of Statistics, Series B* **72**(2): 254–263.
- Bickel, P. J. & Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates, *Annals of Statistics* **1**(6): 1071–1095.
- Boneva, L. I., Kendall, D. & Stefanov, I. (1971). Spline transformations: Three new diagnostic aids for the statistical data-analyst, *Journal of the Royal Statistical Society. Series B (Methodological)* **33**(1): 1–71.
- Bovaird, T. & Lineweaver, C. H. (2017). A flat inner disc model as an alternative to the Kepler dichotomy in the Q1-Q16 planet population, *Monthly Notices of the Royal Astronomical Society* **468**(2): 1493–1504.
- Casas-Prat, M. & Holthuijsen, L. H. (2010). Short-term statistics of waves observed in deep water, *Journal of Geophysical Research: Oceans* **115**(C9).

- Celik, A. N. (2004). A statistical analysis of wind power density based on the Weibull and Rayleigh models at the southern region of Turkey, *Renewable Energy* **29**(4): 593–604.
- Chacón, J. E., Duong, T. & Wand, M. P. (2011). Asymptotics for general multivariate kernel density derivative estimators, *Statistica Sinica* pp. 807–840.
- Chen, S. X. (1999). Probability density function estimation using Gamma kernels, *Annals of the Institute of Statistical Mathematics* **52**(3): 471–480.
- Chen, S. X. (2000). Beta kernel smoothers for regression curves, *Statistica Sinica* **10**(1): 73–91.
- Cowling, A. & Hall, P. (1996). On pseudodata methods for removing boundary effects in kernel density estimation, *Journal of the Royal Statistical Society: Series B (Methodological)* **58**(3): 551–563.
- Cramér, H. (1928). On the composition of elementary errors: First paper: Mathematical deductions, *Scandinavian Actuarial Journal* **1928**(1): 13–74.
- Cressie, N. & Read, T. R. C. (1984). Multinomial goodness-of-fit tests, *Journal of the Royal Statistical Society: Series B (Methodological)* **46**(3): 440–464.
- D’Agostino, R. B. (1986). *Goodness-of-Fit Techniques*, Vol. 68, CRC Press.
- Dekker, A. J. & Sijbers, J. (2014). Data distributions in magnetic resonance images: A review, *Physica Medica* **30**(7): 725–741.
- Dorvlo, A. S. S. (2002). Estimating wind speed distribution, *Energy Conversion and Management* **43**(17): 2311–2318.
- Epps, T. W. (2005). Tests for location-scale families based on the empirical characteristic function, *Metrika* **62**(1): 99–114.

- Epstein, B. (1948). Some applications of the Mellin transform in statistics, *The Annals of Mathematical Statistics* **19**(3): 370–379.
- Fabrycky, D. C., Lissauer, J. J., Ragozzine, D., Rowe, J. F., Steffen, J. H., Agol, E., Barclay, T., B., N., Borucki, W., Ciardi, D. R., Ford, E. B., Gautier, T. N., Geary, J. C., Holman, M. J., Jenkins, J. M., Li, J., Morehead, R. C., Morris, R. L., Shporer, A., Smith, J. C., Still, M. & Cleve, J. V. (2014). Architecture of Kepler’s multi-transiting systems. ii. new investigations with twice as many candidates, *The Astrophysical Journal* **790**(2): 146.
- Fang, J. & Margot, J.-L. (2012). Architecture of planetary systems based on Kepler data: Number of planets and coplanarity, *The Astrophysical Journal* **761**: 92.
- Figueira, P., Marmier, M., Boué, G., Lovis, C., Santos, N. C., Montalto, M., Udry, S., Pepe, F. & Mayor, M. (2012). Comparing HARPS and Kepler surveys: The alignment of multiple-planet systems., *Astronomy & Astrophysics / Astronomie et Astrophysique* **541**: 1–10.
- Fristedt, B. E. & Gray, L. F. (2013). *A Modern Approach To Probability Theory*, Springer Science & Business Media.
- Gai, S., Zhang, B., Yang, C. & Yu, L. (2018). Speckle noise reduction in medical ultrasound image using monogenic wavelet and Laplace mixture distribution, *Digital Signal Processing* **72**: 192–207.
- Galambos, J. & Kotz, S. (1978). Characterizations of Probability Distributions (Lecture Notes in Mathematics 675), *New York: Springer Verlag* .
- Gasser, T. & Müller, H.-G. (1979). Kernel estimation of regression functions, *Lecture Notes in Mathematics*, Springer Berlin Heidelberg, pp. 23–68.
- Gibbons, J. D. & Chakraborti, S. (2014). *Nonparametric Statistical Inference: Revised and Expanded*, CRC Press.

- Gupta, R. C. & Bradley, D. M. (2003). Representing the mean residual life in terms of the failure rate, *Mathematical and Computer Modelling* **37**(12-13): 1271–1280.
- Hahn, S. L. & Poularikas, A. D. (2000). The transforms and applications handbook, *CRC Press L*.
- Hall, P. & Marron, J. S. (1987). Estimation of integrated squared density derivatives, *Stat. Prob. Lett.* **6**(2): 109–115.
- Härdle, W., Marron, J. S. & Wand, M. P. (1990). Bandwidth choice for density derivatives, *Journal of the Royal Statistical Society: Series B (Methodological)* **52**(1): 223–232.
- Heidenreich, N.-B., Schindler, A. & Sperlich, S. (2013). Bandwidth selection for kernel density estimation: a review of fully automatic selectors, *AStA Advances in Statistical Analysis* **97**(4): 403–433.
- Hu, Y. & Scarrott, C. (2018). evmix: An R package for extreme value mixture modeling, threshold estimation and boundary corrected kernel density estimation, *Journal of Statistical Software* **84**(5): 1–27.
- Jahanshahi, S. M. A., Habibi Rad, A. & Fakoor, V. (2016). A goodness-of-fit test for Rayleigh distribution based on Hellinger distance, *Annals of Data Science* **3**(4): 401–411.
- Jensen, D. R. (1970). A generalization of the multivariate Rayleigh distribution, *Sankhyā: The Indian Journal of Statistics, Series A* pp. 193–208.
- Johnson, N., Kotz, S. & Balakrishnan, N. (1994). *Continuous Univariate Distributions*, Wiley.
- Jones, M. C. (1992). Differences and derivatives in kernel estimation, *Metrika* **39**(1): 335–340.

- Jones, M. C. (1993). Simple boundary correction for kernel density estimation, *Statistics and Computing* **3**(3): 135–146.
- Jones, M. C. (1994). On kernel density derivative estimation, *Communications in Statistics-Theory and Methods* **23**(8): 2133–2139.
- Jones, M. C. & Foster, P. J. (1996). A simple nonnegative boundary correction method for kernel density estimation, *Statistica Sinica* pp. 1005–1013.
- Karunamuni, R. J. & Alberts, T. (2005). On boundary correction in kernel density estimation, *Statistical Methodology* **2**(3): 191–212.
- Koekemoer, G. (2004). *A new method for transforming data to normality with application to density estimation*, PhD thesis, North-West University.
- Kolmogorov, A. (1933). Sulla determinazione empirica di una legge di distribuzione, *Giornale dell'Istituto Italiano degli Attuari* **4**: 83–91.
- Lawless, J. F. (2002). *Statistical Models and Methods for Lifetime Data*, 2 edn, New York: Wiley.
- Liebenberg, S. C. (2014). *Multiple Imputation in the Presence of a Detection Limit, with Applications: an Empirical Approach*, Master's thesis, North-West University.
- Lilliefors, H. W. (1967). On the Kolmogorov-Smirnov test for normality with mean and variance unknown, *Journal of the American Statistical Association* **62**(318): 399–402.
- Lissauer, J. J., Ragozzine, D., Fabrycky, D. C., Steffen, J. H., Ford, E. B., Jenkins, J. M., Shporer, A., Holman, M. J., Rowe, J. F., Quintana, E. V., Batalha, N. M., Borucki, W. J., Bryson, S. T., Caldwell, D. A., Carter, J. A., Ciardi, D., Dunham, E. W., Fortney, J. J., Gautier, T. N., Howel, S. B. I., Koch, D. G.,

- Latham, D. W., Marcy, G. W., Morehead, R. C. & Sasselov, D. (2011). Architecture and dynamics of Kepler's candidate multiple transiting planet systems, *The Astrophysical Journal Supplement Series* **197**(1): 8.
- Longuet-Higgins, M. S. (1952). On the statistical distribution of the height of sea waves, *Journal of Marine Research* **11**: 245–266.
- Maltz, C. (1974). Estimation of the k th derivative of a distribution function, *The Annals of Statistics* pp. 359–361.
- Marron, J. S. & Ruppert, D. (1994). Transformations to reduce boundary bias in kernel density estimation, *Journal of the Royal Statistical Society: Series B (Methodological)* **56**(4): 653–671.
- Meintanis, S. G. (2008a). A powerful method of assessing the fit of the lognormal distribution, *Communications in Statistics-Theory and Methods* **37**(12): 1948–1958.
- Meintanis, S. G. (2008b). Tests for generalized exponential laws based on the empirical Mellin transform, *Journal of Statistical Computation and Simulation* **78**(11): 1077–1085.
- Meintanis, S. G. (2009). Goodness-of-fit tests and minimum distance estimation via optimal transformation to uniformity, *Journal of Statistical Planning and Inference* **139**(2): 100–108.
- Meintanis, S. G. (2010). Testing skew normality via the moment generating function, *Mathematical Methods of Statistics* **19**(1): 64–72.
- Meintanis, S. G. & Iliopoulos, G. (2003). Tests of fit for the Rayleigh distribution based on the empirical Laplace transform, *Annals of the Institute of Statistical Mathematics* **55**(1): 137–151.

- Meintanis, S. G., Swanepoel, J. & Allison, J. (2014). The probability weighted characteristic function and goodness-of-fit testing, *Journal of Statistical Planning and Inference* **146**: 122–132.
- Merovci, F. (2013). Transmuted Rayleigh distribution, *Austrian Journal of Statistics* **42**(1): 21–31.
- Miller, K. S., Bernstein, R. I. & Blumenson, L. E. (1958). Generalized Rayleigh processes, *Quarterly of Applied Mathematics* **16**(2): 137–145.
- Morgan, E. C., Lackner, M., Vogel, R. M. & Baise, L. G. (2011). Probability distributions for offshore wind speeds, *Energy Conversion and Management* **52**(1): 15–26.
- Nanda, A. K. (2010). Characterization of distributions through failure rate and mean residual life functions, *Statistics & Probability Letters* **80**(9-10): 752–755.
- Ngatchou-Wandji, J. & Harel, M. (2013). A Cramér-von Mises test for symmetry of the error distribution in asymptotically stationary stochastic models, *Statistical Inference for Stochastic Processes* **16**(3): 207–236.
- Pagan, A. & Ullah, A. (1999). *Nonparametric Econometrics*, Cambridge University Press.
- Pardo, L. (2018). *Statistical Inference Based on Divergence Measures*, Chapman and Hall/CRC.
- Park, B. U. & Marron, J. S. (1990). Comparison of data-driven bandwidth selectors, *Journal of the American Statistical Association* **85**(409): 66–72.
- Parzen, E. (1962). On estimation of a probability density function and mode, *The Annals of Mathematical Statistics* **33**(3): 1065–1076.

- Pearson, K. (1900). X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **50**(302): 157–175.
- R Core Team (2013). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
URL: <http://www.R-project.org/>
- Rajan, J., Poot, D., Juntu, J. & Sijbers, J. (2010). Noise measurement from magnitude MRI using local estimates of variance and skewness, *Physics in Medicine & Biology* **55**(16): N441.
- Rao, M., Chen, Y., Vemuri, B. C. & Wang, F. (2004). Cumulative residual entropy: a new measure of information, *IEEE Transactions on Information Theory* **50**(6): 1220–1228.
- Rayleigh, F. R. S. (1880). Xii. On the resultant of a large number of vibrations of the same pitch and of arbitrary phase, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **10**(60): 73–78.
- Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition, *Proceedings of the National Academy of Sciences of the United States of America* **42**(1): 43.
- Roy, D. (2004). Discrete Rayleigh distribution, *IEEE Transactions on Reliability* **53**(2): 255–260.
- Safavinejad, M., Jomhoori, S. & Alizadeh Noughabi, H. (2015). A density-based empirical likelihood ratio goodness-of-fit test for the Rayleigh distribution

- and power comparison, *Journal of Statistical Computation and Simulation* **85**(16): 3322–3334.
- Sarti, A., Corsi, C., Mazzini, E. & Lamberti, C. (2005). Maximum likelihood segmentation of ultrasound images with Rayleigh distribution, *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* **52**(6): 947–960.
- Schiff, J. L. (1999). *The Laplace Transform. Undergraduate Texts in Mathematics*, Springer-Verlag, New York.
- Schuster, E. (1985). Incorporating support constraints into nonparametric estimators of densities, *Communications in Statistics - Theory and Methods* **14**: 1123–1136.
- Sheather, S. J. & Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation, *Journal of the Royal Statistical Society: Series B (Methodological)* **53**(3): 683–690.
- Siddiqui, M. M. (1962). Some problems connected with Rayleigh distributions, *Journal of Research of the National Bureau of Standards* **66**(2): 167–174.
- Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, New York.
- Simon, M. K. & Alouini, M. (1998). A simple single integral representation of the bivariate Rayleigh distribution, *IEEE Communications Letters* **2**(5): 128–130.
- Smirnov, N. (1939). Sur les écarts de la courbe de distribution empirique, *Matematicheskii Sbornik* **48**(1): 3–26.
- Soares, C. G. & Carvalho, A. N. (2003). Probability distributions of wave heights and periods in measured combined sea-states from the Portuguese coast, *Journal of Offshore Mechanics and Arctic Engineering* **125**(3): 198–204.

- Song, K.-S. (2002). Goodness-of-fit tests based on Kullback-Leibler discrimination information, *IEEE Transactions on Information Theory* **48**(5): 1103–1117.
- Tan, C. C. & Beaulieu, N. C. (1997). Infinite series representations of the bivariate Rayleigh and Nakagami-m distributions, *IEEE Transactions on Communications* **45**(10): 1159–1161.
- Tenreiro, C. (2007). On the finite sample behavior of fixed bandwidth Bickel–Rosenblatt test for univariate and multivariate uniformity, *Communications in Statistics—Simulation and Computation* **36**(4): 827–846.
- Toa, C.-K., Sim, K.-S., Lim, Z.-Y. & Lim, C.-P. (2019). Magnetic resonance imaging noise filtering using adaptive polynomial-fit non-local means., *Engineering Letters* **27**(3): 1733–1736.
- Torabi, H., Montazeri, N. H. & Grané, A. (2016). A test for normality based on the empirical distribution function, *SORT: Statistics and Operations Research Transactions* **40**(1): 0055–88.
- Tsallis, C. (1998). Generalized entropy-based criterion for consistent testing, *Physical Review E* **58**(2): 1442.
- Vasicek, O. (1976). A test for normality based on sample entropy, *Journal of the Royal Statistical Society. Series B (Methodological)* **38**(1): 54–59.
- Vexler, A. & Gurevich, G. (2010). Empirical likelihood ratios applied to goodness-of-fit tests based on sample entropy, *Computational Statistics & Data Analysis* **54**(2): 531–545.
- Vodă, V. G. (1976). Inferential procedures on a generalized Rayleigh variate. I, *Aplikace Matematiky* **21**(6): 395–412.

- Von Mises, R. (1931). *Vorlesungen Aus Dem Gebiete Der Angewandten Mathematik: Wahrscheinlichkeitsrechnung Und Ihre Anwendung In Der Statistik Und Theoretischen Physik*, F. Deuticke.
- Von Mises, R. V. (1947). On the asymptotic distribution of differentiable statistical functions, *The Annals of Mathematical Statistics* **18**(3): 309–348.
- Wand, M. P. & Jones, M. C. (1995). *Kernel Smoothing*, Chapman & Hall, London.
- Watson, G. S. (1962). Goodness-of-fit tests on a circle. II, *Biometrika* **49**(1/2): 57–63.
- Zamanzade, E. & Mahdizadeh, M. (2017). Goodness of fit tests for Rayleigh distribution based on Phi-divergence, *Revista Colombiana de Estadística* **40**(2): 279–290.
- Zhang, S. (2010). A note on the performance of the gamma kernel estimators at the boundary, *Statistics & Probability Letters* **80**(7-8): 548–557.
- Zhang, S. & Karunamuni, R. J. (2010). Boundary performance of the beta kernel estimators, *Journal of Nonparametric Statistics* **22**(1): 81–104.