



New classes of tests for the Weibull distribution using Stein's method in the presence of random right censoring

E. Bothma¹ · J. S. Allison¹ · I. J. H. Visagie¹ 

Received: 19 May 2021 / Accepted: 2 November 2021 / Published online: 9 January 2022

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

We develop two new classes of tests for the Weibull distribution based on Stein's method. The proposed tests are applied in the full sample case as well as in the presence of random right censoring. We investigate the finite sample performance of the new tests using a comprehensive Monte Carlo study. In both the absence and presence of censoring, it is found that the newly proposed classes of tests outperform competing tests against the majority of the distributions considered. In the cases where censoring is present we consider various censoring distributions. Some remarks on the asymptotic properties of the proposed tests are included. We present another result of independent interest; a test initially proposed for use with full samples is amended to allow for testing for the Weibull distribution in the presence of censoring. The techniques developed in the paper are illustrated using two practical examples.

Keywords Goodness-of-fit testing · Hypothesis testing · Random right censoring · Warp-speed bootstrap · Weibull distribution

1 Introduction

The Weibull distribution is often used in survival analysis as well as reliability theory, see e.g., Kalbfleisch and Prentice (2011). This flexible distribution is a popular model which allows for constant, increasing and decreasing hazard rates. The Weibull distribution is also frequently applied in various engineering fields, including electrical and industrial engineering to represent, for example, manufacturing times, see Jiang and Murthy (2011). As a result of its wide range of practical uses, a number of goodness-of-fit tests have been developed for the Weibull distribution; see e.g. Mann et al. (1973), Tiku and Singh (1981), Liao and Shimokawa (1999), Cabaña and Quiroz (2005) as well as Krit (2014).

✉ I. J. H. Visagie
jaco.visagie@nwu.ac.za

¹ Subject Group Statistics, North-West University, Potchefstroom, South Africa

The papers listed above deal with testing for the Weibull distribution in the full sample case; i.e., where all lifetimes are observed. However, random right censoring often occurs in the fields mentioned above. For example, we may study the duration that antibodies remain detectable in a patient's blood after receiving a specific type of Covid-19 vaccine, i.e. the duration of the protection that the vaccine affords the recipient. When gathering the relevant data, we will likely not be able to measure this duration in all of the patients. For example, some may leave the study by emigrating to a different country while still having detectable antibodies. In this case, the exact time of interest is not observed. This situation is referred to as random right censoring, see e.g., Cox and Oakes (1984).

In the presence of censoring, testing the hypothesis that the distribution of the lifetimes is Weibull is complicated by the fact that an incomplete sample is observed. Balakrishnan et al. (2015) suggests a way to perform the required goodness-of-fit tests by transforming the censored sample to a complete sample. Another approach is to modify the test statistics used in the full sample case to account for the presence of censoring. Although fewer in number, tests for the Weibull distribution in the presence of random censoring are available in the literature. For example, Koziol and Green (1976) and Kim (2017) propose modified versions of the Cramér-von Mises test and the test proposed in Liao and Shimokawa (1999), respectively, for use with censored data.

Throughout this paper we are primarily interested in the situation where censoring is present; the results relating to the full sample case are treated as special cases obtained when all lifetimes are observed. Before proceeding some notation is introduced. Let X_1, \dots, X_n be independent and identically distributed (i.i.d.) lifetime variables with continuous distribution function F and let C_1, \dots, C_n be i.i.d. censoring variables with distribution function H , independent of X_1, \dots, X_n . We assume non-informative censoring throughout. Let

$$T_j = \min(X_j, C_j) \quad \text{and} \quad \delta_j = \begin{cases} 1, & \text{if } X_j \leq C_j, \\ 0, & \text{if } X_j > C_j. \end{cases}$$

Note that in the full sample case $T_j = X_j$ and $\delta_j = 1$ for $j = 1, \dots, n$.

Based on the observed pairs (T_j, δ_j) , $j = 1, \dots, n$ we wish to test the composite hypothesis

$$H_0 : X \sim \text{Weibull}(\lambda, \theta), \quad (1)$$

for some unknown $\lambda > 0$ and $\theta > 0$. Here $X \sim \text{Weibull}(\lambda, \theta)$ refers to a Weibull distributed random variable with distribution function

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

This hypothesis is to be tested against general alternatives. We will make use of maximum likelihood estimation to estimate λ and θ . The log-likelihood of the Weibull distribution is

$$\mathcal{L}(\theta, \lambda | X_1, \dots, X_n) = d \log(\theta) - d\theta \log(\lambda) + (\theta - 1) \sum_{j=1}^n \delta_j \log(X_j) - \lambda^{-\theta} \sum_{j=1}^n X_j^\theta,$$

where $d = \sum_{j=1}^n \delta_j$. In the full sample case $d = n$. No closed form formulae for the maximum likelihood estimates $\hat{\lambda}$ and $\hat{\theta}$ exist, meaning that numerical optimisation techniques are required to arrive at parameter estimates.

Since the Weibull distribution has a shape parameter the first step for many goodness-of-fit tests for the Weibull distribution is to transform the data. If $X \sim \text{Weibull}(\lambda, \theta)$ then a frequently used transformation is $\log(X)$, which results in a random variable that is type I extreme value distributed with parameters $\log(\lambda)$ and $1/\theta$. The resulting transformed random variable is part of a location scale family, which is a desirable result when performing goodness-of-fit testing. We therefore have that if $X \sim \text{Weibull}(\lambda, \theta)$, then $X^{(t)} = \theta(\log(X) - \log(\lambda))$ follows a standard type I extreme value distribution with distribution function

$$G(x) = 1 - e^{-e^x}, \quad -\infty < x < \infty.$$

We denote a random variable with this distribution function by $EV(0, 1)$. As a result the hypothesis in (1) holds, if, and only if, $X^{(t)} \sim EV(0, 1)$. All of the test statistics considered make use of the transformed observed values

$$Y_j = \hat{\theta} \left[\log(T_j) - \log(\hat{\lambda}) \right], \quad (2)$$

with $\hat{\lambda}$ and $\hat{\theta}$ the maximum likelihood estimates of the Weibull distribution. Let

$$X_j^{(t)} = \hat{\theta} \left[\log(X_j) - \log(\hat{\lambda}) \right].$$

If X_1, \dots, X_n are realised from a $\text{Weibull}(\lambda, \theta)$ distribution, then $X_1^{(t)}, \dots, X_n^{(t)}$ will approximately follow an $EV(0, 1)$ distribution see e.g. Kotz and Nadarajah (2000). The resulting random variables are no longer independent. However, several properties of the classical testing procedures remain unaffected when performing this type of transformation; the interested reader is referred to Baringhaus and Henze (1991) as well as Gupta and Richards (1997) and the references therein for more details. The tests employed below are based on discrepancy measures between the calculated values of Y_1, \dots, Y_n and the standard type I extreme value distribution. The order statistics of Y_1, \dots, Y_n are denoted by $Y_{(1)} < \dots < Y_{(n)}$, while $\delta_{(j)}$ represents the indicator variable corresponding to $Y_{(j)}$.

The remainder of the paper is structured as follows. In Sect. 2, we propose two new classes of tests for the Weibull distribution for both the full sample and censored case. We also modify the test proposed in Krit (2014) to accommodate random right censoring. In the presence of censoring the null distribution of all the test statistics considered depend on the unknown censoring distribution, we therefore propose a parametric bootstrap procedure in Sect. 2 in order to compute critical values. Section 3 presents the results of a Monte Carlo study where the empirical powers of the newly

proposed classes of tests as well as the newly modified test are compared to those of existing tests. The paper concludes in Sect. 4 with two practical applications; one concerning the survival times of patients diagnosed with a certain type of leukemia (no censoring is present in these data) and the other relates to observed leukemia remission times (in the presence of censoring). Some avenues for future research are also discussed.

2 Proposed test statistics

Our newly proposed classes of tests are based on the following theorem, which characterises the standard type I extreme value distribution.

Theorem 1 *Let W be a random variable with absolutely continuous density and assume that $E[e^W] < \infty$. In this case $W \sim EV(0, 1)$, if, and only if,*

$$E \left[(it + 1 - e^W) e^{itW} \right] = 0, \quad \forall t \in R,$$

with $i = \sqrt{-1}$.

Proof The 'if' part of the theorem can easily be shown using direct calculation. The 'only if' part is shown below.

Assume $E[(it + 1 - e^W) e^{itW}] = 0$. From Fourier analysis we have that the Fourier transform of $f'(w)$ is $e^{itw} f'(w) = -itE[e^{itW}]$. This implies that

$$\begin{aligned} 0 &= E \left[(it + 1 - e^W) e^{itW} \right] \\ &= \int_{-\infty}^{\infty} (it + 1 - e^w) e^{itw} f(w) dw \\ &= \int_{-\infty}^{\infty} [-f'(w) + (1 - e^w) f(w)] e^{itw} dw, \end{aligned}$$

which is the Fourier transform of $h(w) := -f'(w) + (1 - e^w) f(w)$. The fact that the Fourier transform of h is 0 implies that $h(w) = 0$ for all $w \in R$. Thus, f satisfies the following differential equation $f'(w) - (1 - e^w) f(w) = 0$. Using separation of variables we have

$$\frac{f'(w)}{f(w)} = 1 - e^w \implies \log(f(w)) = w - e^w + c \implies f(w) = e^{w - e^w},$$

where the last step follows from the fact that f must be a density function and hence integrate to 1. This completes the proof. \square

Let $w(t)$ be a non-negative, symmetric weight function. From Theorem 1, we have that

$$\begin{aligned}\eta &= \int_{-\infty}^{\infty} E \left[(it + 1 - e^Y) e^{itY} \right] w(t) dt \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (it + 1 - e^y) e^{ity} dG(y) w(t) dt\end{aligned}\quad (3)$$

equals 0 if $Y \sim EV(0, 1)$. Note that the inclusion of the weight function, w , above is required to ensure that η is finite. Clearly η will be unknown because G is unknown. However, G can be estimated by the Kaplan-Meier estimator, G_n , of the distribution function given by

$$1 - G_n(t) = \begin{cases} 1, & t \leq Y_{(1)} \\ \prod_{j=1}^{k-1} \left(\frac{n-j}{n-j+1} \right)^{\delta_{(j)}}, & Y_{(k-1)} < t \leq Y_{(k)}, \quad k = 2, \dots, n. \\ \prod_{j=1}^n \left(\frac{n-j}{n-j+1} \right)^{\delta_{(j)}}, & t > Y_{(n)}. \end{cases}$$

More details about this estimator can be found in Kaplan and Meier (1958), Efron (1967) as well as Breslow and Crowley (1974). In the full sample case this estimator reduces to the standard empirical distribution function, $G_n(X_{(j)}) = j/n$.

Let Δ_j denote the size of the jump in $G_n(T_{(j)})$;

$$\Delta_j = G_n(T_{(j)}) - \lim_{t \uparrow T_{(j)}} G_n(t), \quad j = 1, \dots, n.$$

Simple calculable expressions for the Δ_j 's are

$$\begin{aligned}\Delta_1 &= \frac{\delta_{(1)}}{n}, \quad \Delta_n = \prod_{j=1}^{n-1} \left(\frac{n-j}{n-j+1} \right)^{\delta_{(j)}} \text{ and} \\ \Delta_j &= \prod_{k=1}^{j-1} \left(\frac{n-k}{n-k+1} \right)^{\delta_{(k)}} - \prod_{k=1}^j \left(\frac{n-k}{n-k+1} \right)^{\delta_{(k)}} \\ &= \frac{\delta_{(j)}}{n-j+1} \prod_{k=1}^{j-1} \left(\frac{n-k}{n-k+1} \right)^{\delta_{(k)}}, \quad j = 2, \dots, n-1.\end{aligned}$$

In the full sample case $\Delta_j = 1/n$, $j = 1, \dots, n$.

Estimating G by G_n in (3), we propose the test statistic

$$S_{n,a} = n \int_{-\infty}^{\infty} \left| \sum_{j=1}^n \Delta_j \left[ite^{itY_j} + (1 - e^{Y_j})e^{itY_j} \right] \right|^2 w_a(t) dt, \quad (4)$$

where $w_a(t)$ is a weight function containing a user-defined tuning parameter $a > 0$. The null hypothesis in (1) is rejected for large values of $S_{n,a}$.

Straightforward algebra shows that, if $w_a(t) = e^{-at^2}$, then the test statistic simplifies to

$$S_{n,a}^{(1)} = n \sqrt{\frac{\pi}{a}} \sum_{j=1}^n \sum_{k=1}^n \Delta_j \Delta_k e^{-(Y_j - Y_k)^2 / 4a} \left\{ -\frac{1}{4a^2} \left((Y_j - Y_k)^2 - 2a \right) + 2 \left(1 - e^{Y_j} \right) \left(\frac{1}{2a} \right) (Y_j - Y_k) + \left(1 - e^{Y_j} \right) \left(1 - e^{Y_k} \right) \right\},$$

and if $w_a(t) = e^{-a|t|}$ the test statistic has the following easily calculable form

$$S_{n,a}^{(2)} = n \sum_{j=1}^n \sum_{k=1}^n \Delta_j \Delta_k \left\{ \frac{-4a \left(3(Y_j - Y_k)^2 - a^2 \right)}{\left((Y_j - Y_k)^2 + a^2 \right)^3} + \frac{8a(Y_j - Y_k) \left(1 - e^{Y_j} \right)}{\left((Y_j - Y_k)^2 + a^2 \right)^2} + \frac{2a \left(1 - e^{Y_j} \right) \left(1 - e^{Y_k} \right)}{\left(Y_j - Y_k \right)^2 + a^2} \right\}.$$

New goodness-of-fit tests containing a tuning parameter are often accompanied by a recommended value of this parameter; this choice is typically based on the finite sample power performance of the test. Another approach which may be used is to choose the value of the tuning parameter data-dependently; see e.g., Allison and Santana (2015). In this paper, we opt to use the values recommended in the literature for tests containing a tuning parameter.

The weight functions specified above correspond to scaled Gaussian and Laplace kernels. These weight functions are popular choices found in the goodness-of-fit literature; see e.g., Meintanis and Iliopoulos (2003), Allison et al. (2017), Betsch and Ebner (2018), Betsch and Ebner (2019) as well as Henze and Visagie (2020). The popularity of these weight functions is, at least in part, due to the fact that their inclusion typically results in simple calculable forms for L^2 -type statistics which do not require numerical integration. As an alternative to the weight functions used, one may employ a symmetric uniform kernel as a weight function; see e.g., Fernández et al. (2008). However, in the mentioned paper, the authors found that the computational time required for the test statistic obtained using the symmetric uniform kernel was substantial. This, coupled with the simple computational forms obtained using the weight functions defined above and the favourable power performance discussed in Sect. 3, motivated us to restrict our attention to the scaled Gaussian and Laplace kernels.

Although we do not derive the asymptotic results related to the proposed classes of test statistics we include some remarks in this regard. $S_{n,a}$ is a characteristic function based weighted L^2 -type statistic. The asymptotic properties of this class of statistics, in the complete sample case, are studied in detail in Feuerverger and Mureika (1977), while more recent references include Baringhaus and Henze (1988), Klar and Meintanis (2005) as well as Baringhaus et al. (2017). A convenient setting for the derivation of the asymptotic properties of these tests is the separable Hilbert space of square integrable functions. Typically, the asymptotic null distribution of $S_{n,a}$ corresponds to that of $\int_{-\infty}^{\infty} |Z(t)|^2 w_a(t) dt =: S_a$, where $Z(\cdot)$ is a zero-mean Gaussian process. The

distribution of S_a is the same as that of $\sum_{j=1}^{\infty} \lambda_j U_j$, where U_j are i.i.d. chi-squared random variables with parameter 1 and where λ_j are eigenvalues of an integral operator (see e.g., Allison et al. (2021)). These tests are consistent against a large class of fixed alternative distributions. Additionally, these tests are frequently consistent against contiguous alternatives converging to the null at a rate of $n^{-1/2}$.

In the case of random censoring, very little asymptotic results are available in the literature for test statistics of this type. Very recently, advances have been made in this regard; see e.g., Cuparić and Milošević (2021), in which a test for exponentiality is considered based on so-called inverse probability censoring weights, where the authors derive the asymptotic properties of the given test. In addition Fernández and Rivera (2020) studied Kaplan-Meier U - and V -statistics in order to derive some asymptotic results relating to the lifetime distribution. Some of these results may be helpful in deriving the asymptotic properties of the tests proposed in this paper in future research.

The null distribution of each of the test statistics considered depends on the unknown censoring distribution, even in the case of a simple hypothesis, see D'Agostino and Stephens (1986). Since we will not assume any known form of the censoring distribution, we propose the following parametric bootstrap algorithm to estimate the critical values of the tests.

1. Based on the pairs (T_j, δ_j) , $j = 1, \dots, n$ estimate θ and λ by $\hat{\theta}$ and $\hat{\lambda}$, respectively, using maximum likelihood estimation.
2. Transform T_j to Y_j using the transformation in (2) for $j = 1, \dots, n$.
3. Calculate the test statistic, say $W_n := W(Y_1, \dots, Y_n; \delta_1, \dots, \delta_n)$.
4. Obtain a parametric bootstrap sample X_1^*, \dots, X_n^* by sampling from a Weibull distribution with parameters $\hat{\theta}$ and $\hat{\lambda}$.
5. Obtain a non-parametric bootstrap sample, C_1^*, \dots, C_n^* , by sampling from the Kaplan-Meier estimate of the distribution of $C_j^{(t)} = \hat{\theta} [\log(C_j) - \log(\hat{\lambda})]$.
6. Set

$$T_j^* = \min(X_j^*, C_j^*) \text{ and } \delta_j^* = \begin{cases} 1, & \text{if } X_j^* \leq C_j^* \\ 0, & \text{if } X_j^* > C_j^*. \end{cases}$$

7. Calculate $\hat{\theta}^*$ and $\hat{\lambda}^*$ based on (T_j^*, δ_j^*) , $j = 1, \dots, n$.
8. Obtain $Y_j^* = \hat{\theta}^* [\log(T_j^*) - \log(\hat{\lambda}^*)]$, $j = 1, \dots, n$.
9. Based on the pairs (Y_j^*, δ_j^*) , $j = 1, \dots, n$, calculate the value of the test statistic, say $W_n^* := W(Y_1^*, \dots, Y_n^*; \delta_1^*, \dots, \delta_n^*)$.
10. Repeat steps 4-9 B times to obtain W_1^*, \dots, W_B^* . Obtain the order statistics, $W_{(1)}^* \leq \dots \leq W_{(B)}^*$. The estimated critical value is then $\hat{c}_n(\alpha) = W_{[B(1-\alpha)]}^*$ where $[c]$ denotes the floor of c .

The algorithm provided above is quite general and can easily be amended in order to test for any lifetime distribution in the presence of random censoring. In the absence of censoring there is no need to implement this algorithm; in this case, the critical values can be obtained via Monte Carlo simulation by sampling from any Weibull distribution and effecting the transformation discussed above.

3 Numerical results

In this section, we compare the power performances of the newly proposed tests to those of existing tests via a Monte Carlo simulation study. The existing tests used include the classical Kolmogorov-Smirnov (KS_n) and Cramér-von Mises (CM_n) tests. These tests have been modified for use with censored data, see Koziol and Green (1976). The test introduced in Liao and Shimokawa (1999) is considered in the case of full samples. A modification making this test suitable for use with censored data is proposed in Kim (2017); we denote the test statistic by LS_n in both the full sample and censored cases. The calculable forms of the test statistics mentioned above are

$$KS_n = \max \left[\max_{1 \leq j \leq n} \left\{ G_n(Y_{(j)}) - \left(1 - e^{-Y_j} \right) \right\}, \right. \\ \left. \max_{1 \leq j \leq n} \left\{ \left(1 - e^{-Y_j} \right) - G_n^-(Y_{(j)}) \right\} \right],$$

$$CM_n = \frac{n}{3} + n \sum_{j=1}^{d+1} \left\{ G_n \left(X_{j-1}^{(t)} \right) \left(X_j^{(t)} - X_{j-1}^{(t)} \right) \right. \\ \left. \times \left[G_n \left(X_{j-1}^{(t)} \right) - \left(X_j^{(t)} + X_{j-1}^{(t)} \right) \right] \right\},$$

$$LS_n = \frac{1}{\sqrt{n}} \sum_{j=1}^n \left(\frac{\max \left[j/n - G_n(Y_j), G_n(Y_j) - (j-1)/n \right]}{\sqrt{G_n(Y_j) \left[1 - G_n(Y_j) \right]}} \right).$$

Krit (2014) proposes a test for the Weibull distribution in the full sample case. This test compares the empirical Laplace transform of the random variables resulting from the transformation in (2) to the Laplace transform of an $EV(0, 1)$ random variable; $\psi(t) = \Gamma(1-t)$ for $t < 1$. Let ψ_n be the empirical Laplace transform of the transformed observations, obtained using the Kaplan-Meier estimate of the distribution function;

$$\psi_n(t) = \int_{-\infty}^{\infty} e^{-tx} dG_n(x) = \sum_{j=1}^n \Delta_j e^{-tY_j}. \quad (5)$$

The resulting test statistic is

$$KR_n^* = n \int_I [\psi_n(t) - \Gamma(1-t)]^2 w_a(t) dt, \quad (6)$$

where $w_a(t) = e^{at-e^{at}}$ is a weight function, a a user-specified tuning parameter and I some interval. Based on numerical considerations, Krit (2014) suggests that $I = (-1, 0]$ should be used. The quantity in (6) can be approximated by a Riemann sum;

$$K R_n = n \sum_{k=-m}^{-1} \left[\sum_{j=1}^n \Delta_j e^{-Y_j k/m} - \Gamma(1 - k/m) \right]^2 e^{ak/m - e^{ak/m}},$$

where m is the number of points at which the integrand is evaluated. In the numerical results shown below, we use $a = -5$ and $m = 100$ as recommended in Krit (2014). In the full sample case, G_n , in (5), is taken to be the empirical distribution function. Upon setting $\Delta_j = 1/n$ we obtain the test statistic in Krit (2014).

For each of the tests considered above, the null hypothesis in (1) is rejected for large values of the test statistics.

3.1 Simulation setting

In the numerical results presented below, we use a nominal significance level of 10% throughout. Empirical powers are presented for sample sizes $n = 50$ and $n = 100$. The empirical powers for complete and censored samples are reported; censoring proportions of 10% and 20% are included. For each lifetime distribution considered, we report the powers obtained using three different censoring distributions. The first censoring distribution used is the exponential distribution, the parameter of which is chosen so as to obtain the specified level of censoring. The second censoring distribution used is the uniform distribution with support $(0, m)$; again, m is chosen such that the required censoring level is achieved. The final censoring distribution used is the Koziol-Green model proposed in Koziol and Green (1976). Denote the survival function of a given lifetime distribution by S . In this case, the Koziol-Green censoring distribution, indexed by β , has survival function $S^\beta(t)$. It can be shown that the censoring proportion is $\beta/(1 + \beta)$. We chose the value of β so as to ensure that the required level of censoring is achieved. The alternative lifetime distributions considered are listed in Table 1. Note that each of the alternatives considered have the same support as that of the Weibull distribution with the exception of the beta distribution which is restricted to the unit interval $(0, 1)$ (Tables 10, 11, 12 and 13).

The obtained empirical powers are presented in Tables 2, 3, 4, 5, 6, 7, 8 and 9. These tables report the percentages of 50 000 independent Monte Carlo samples that

Table 1 Density functions of the alternative distributions

Alternative	Density	Notation	Support
Weibull	$\theta x^{\theta-1} \exp(-x^\theta)$	$W(\theta)$	$(0, \infty)$
Gamma	$(\Gamma(\theta))^{-1} x^{\theta-1} \exp(-x)$	$\Gamma(\theta)$	$(0, \infty)$
Lognormal	$(\theta x \sqrt{2\pi})^{-1} \exp\left(-\log^2(x) (2\theta^2)^{-1}\right)$	$LN(\theta)$	$(0, \infty)$
Chi square	$(2^{\theta/2} \Gamma(\theta/2))^{-1} x^{\theta/2-1} \exp(-x/2)$	$\chi^2(\theta)$	$(0, \infty)$
Beta	$x^{\alpha-1} (1-x)^{\theta-1} \Gamma(\alpha+\theta) (\Gamma(\alpha)\Gamma(\theta))^{-1}$	$\beta(\alpha, \theta)$	$(0, 1)$
Lindley	$\frac{\theta^2}{\theta+1} (1+x) \exp(-\theta x)$	$Lind(\theta)$	$(0, \infty)$

Table 2 Estimated powers for the full sample case where $n = 50$

F	KS_n	CM_n	LS_n	KR_n	$S_{n,1}^{(1)}$	$S_{n,2}^{(1)}$	$S_{n,5}^{(1)}$	$S_{n,1}^{(2)}$	$S_{n,2}^{(2)}$	$S_{n,5}^{(2)}$
$W(0.5)$	10	10	10	10	10	10	10	10	10	10
$W(1)$	10	10	10	10	10	10	10	10	10	10
$W(1.5)$	10	11	10	11	11	10	10	11	11	10
$W(2)$	10	10	9	10	10	10	10	10	10	10
$\Gamma(2)$	13	15	17	14	17	17	16	14	16	17
$\Gamma(3)$	17	20	25	19	24	25	24	18	22	25
$LN(0.5)$	50	62	70	63	72	76	76	51	68	76
$LN(1)$	50	61	70	63	73	77	77	51	69	77
$\chi^2(8)$	21	24	31	23	30	32	31	20	28	31
$\chi^2(10)$	23	28	35	27	34	36	35	23	31	36
$\beta(1, 1)$	69	81	92	90	85	87	88	80	86	89
$\beta(0.5, 1)$	70	81	93	91	86	88	88	80	86	89
$Lind(0.5)$	11	11	12	13	11	13	14	10	11	13
$Lind(2)$	10	10	11	11	11	11	11	10	10	11

Table 3 Heatmap of the estimated powers for the full sample case where $n = 50$

F	KS_n	CM_n	LS_n	KR_n	$S_{n,1}^{(1)}$	$S_{n,2}^{(1)}$	$S_{n,5}^{(1)}$	$S_{n,1}^{(2)}$	$S_{n,2}^{(2)}$	$S_{n,5}^{(2)}$
$W(0.5)$										
$W(1)$										
$W(1.5)$										
$W(2)$										
$\Gamma(2)$										
$\Gamma(3)$										
$LN(0.5)$										
$LN(1)$										
$\chi^2(8)$										
$\chi^2(10)$										
$\beta(1, 1)$										
$\beta(0.5, 1)$										
$Lind(0.5)$										
$Lind(2)$										

lead to the rejection of the null hypothesis, rounded to the nearest integer. For ease of comparison, the highest power in each line is printed in bold. Tables 2 and 4 contain the results relating to full samples. In order to ease visual comparison of the results obtained, we include so called "heatmaps", see Döring and Cramer (2019), of these results in Tables 3 and 5. For each test considered, Tables 6, 7, 8 and 9 show three empirical powers against each lifetime distribution, corresponding to the three different censoring distributions used. In each case, the results for the exponential, uniform and Koziol-Green models are shown in the first, second and third lines, respectively.

In order to reduce the computational cost associated with the numerical powers a warp-speed bootstrap procedure, see Giacomini et al. (2013), is employed. This methodology has been employed by a number of authors in the literature to compare

Table 4 Estimated powers for the full sample case where $n = 100$

F	KS_n	CM_n	LS_n	KR_n	$S_{n,1}^{(1)}$	$S_{n,2}^{(1)}$	$S_{n,5}^{(1)}$	$S_{n,1}^{(2)}$	$S_{n,2}^{(2)}$	$S_{n,5}^{(2)}$
$W(0.5)$	10	10	10	10	10	10	10	9	10	10
$W(1)$	10	10	10	10	10	10	10	10	10	10
$W(1.5)$	10	10	10	10	10	10	10	10	10	10
$W(2)$	10	9	10	10	10	10	10	9	10	10
$\Gamma(2)$	17	19	26	19	24	26	27	17	22	26
$\Gamma(3)$	26	31	41	31	39	43	45	24	35	44
$LN(0.5)$	77	88	94	90	95	97	98	78	93	97
$LN(1)$	78	88	93	90	95	97	98	78	93	97
$\chi^2(8)$	32	39	51	41	49	54	56	31	45	55
$\chi^2(10)$	36	44	58	46	56	62	64	34	51	63
$\beta(1, 1)$	94	98	100	100	99	99	99	99	99	100
$\beta(0.5, 1)$	94	98	100	100	99	99	99	99	99	99
$Lind(0.5)$	12	12	13	13	12	14	16	11	12	15
$Lind(2)$	11	11	11	11	11	11	12	10	11	12

Table 5 Heatmap of the estimated powers for the full sample case where $n = 100$

F	KS_n	CM_n	LS_n	KR_n	$S_{n,1}^{(1)}$	$S_{n,2}^{(1)}$	$S_{n,5}^{(1)}$	$S_{n,1}^{(2)}$	$S_{n,2}^{(2)}$	$S_{n,5}^{(2)}$
$W(0.5)$										
$W(1)$										
$W(1.5)$										
$W(2)$										
$\Gamma(2)$										
$\Gamma(3)$										
$LN(0.5)$										
$LN(1)$										
$\chi^2(8)$										
$\chi^2(10)$										
$\beta(1, 1)$										
$\beta(0.5, 1)$										
$Lind(0.5)$										
$Lind(2)$										

Monte Carlo performances; see e.g., Meintanis et al. (2018), Allison et al. (2019) as well as Mijburgh and Visagie (2020). The bootstrap algorithm in Sect. 2 is implemented to calculate the critical values used to obtain the results in Tables 6, 7, 8 and 9.

For $S_{n,a}^{(1)}$ and $S_{n,a}^{(2)}$, we include numerical powers in the cases where a is set to 1, 2 and 5 in the full sample case and set to 0.75, 1 and 2 in the presence of censoring. The difference between these two sets of choices is due to the fact that the newly proposed tests exhibits higher powers in the presence of censoring if slightly smaller values of a are used. All calculations are performed in R, see [44]. The *LindleyR* package is used to generate samples from censored distributions, see Mazucheli et al. (2016). Parameter estimation is performed using the *parmsurvfit* package, see Jacobson et al. (2018), while the tables are produced using the *Stargazer* package, see Hlavac (2018).

Table 6 Estimated powers for 10% censoring for a sample size of $n = 50$ with three different censoring distributions

F	KS_n	CM_n	LS_n	KR_n	$S_{n,0.75}^{(1)}$	$S_{n,1}^{(1)}$	$S_{n,2}^{(1)}$	$S_{n,0.75}^{(2)}$	$S_{n,1}^{(2)}$	$S_{n,2}^{(2)}$
$W(0.5)$	9	9	9	7	8	8	8	8	8	8
	9	9	9	4	9	9	9	9	8	9
	10	9	10	9	8	8	9	9	9	8
$W(1)$	9	9	9	9	8	8	10	9	9	8
	9	9	9	8	8	8	10	9	9	8
	9	9	9	9	8	8	9	9	9	8
$W(1.5)$	10	10	10	9	9	9	9	9	9	9
	10	10	10	8	9	9	9	10	10	9
	10	10	10	9	8	8	10	9	9	8
$W(2)$	9	10	10	8	9	9	9	9	9	9
	10	10	10	9	9	9	10	10	9	9
	9	10	10	9	8	9	9	9	9	8
$\Gamma(2)$	13	14	13	21	15	15	15	12	13	14
	13	14	13	21	15	15	15	12	13	14
	12	14	12	21	14	14	15	12	12	14
$\Gamma(3)$	17	19	18	29	21	22	23	15	17	21
	16	19	17	28	20	21	21	14	16	20
	16	19	17	29	20	20	20	14	16	19
$LN(0.5)$	46	56	57	64	64	67	69	36	46	63
	45	56	57	63	64	66	68	36	46	62
	43	54	56	64	61	63	63	35	44	59
$LN(1)$	43	54	53	56	60	62	58	34	43	58
	42	53	49	48	58	59	53	31	41	56
	44	54	56	64	61	63	63	35	44	59
$\chi^2(8)$	19	23	22	34	26	27	27	16	20	25

3.2 Simulation results

First, we consider the results associated with the full sample case, given in Tables 2 and 4, together with the heatmaps shown in Tables 3 and 5. All of the tests considered attain the nominal size for both sample sizes used. The tests associated with the highest powers are $S_{n,2}^{(1)}$ and $S_{n,5}^{(1)}$, although LS_n and $S_{n,5}^{(2)}$ also performs well. When analysing complete samples, we recommend using $S_{n,2}^{(1)}$ or $S_{n,5}^{(1)}$.

We now turn our attention to the powers achieved in the presence of censoring. The size of the tests are maintained closely for all sample sizes for censoring proportions of 10% and 20%, with the single exception of $KR_{m,a}$ in the case of small sample sizes. As expected, the powers generally increase with sample size and decrease marginally as the censoring proportion increases.

Table 6 continued

F	KS_n	CM_n	LS_n	KR_n	$S_{n,0.75}^{(1)}$	$S_{n,1}^{(1)}$	$S_{n,2}^{(1)}$	$S_{n,0.75}^{(2)}$	$S_{n,1}^{(2)}$	$S_{n,2}^{(2)}$
$\chi^2(10)$	19	22	21	33	26	26	27	16	19	25
	19	22	21	34	24	25	24	16	18	23
	21	25	25	38	29	31	32	17	21	28
	21	25	25	38	29	31	32	18	21	28
	21	25	24	37	27	28	27	18	20	26
$\beta(1, 1)$	64	75	83	1	78	80	82	68	73	80
	63	75	83	1	78	79	81	68	72	80
	61	70	79	0	55	50	31	62	63	47
$\beta(0.5, 1)$	61	72	79	0	74	75	76	66	70	76
	60	72	79	0	73	74	74	65	69	74
	61	69	79	0	54	49	28	62	62	46
$Lind(0.5)$	10	11	12	8	10	10	11	9	9	9
	10	11	12	8	10	10	12	9	9	10
	10	10	12	8	9	10	11	9	9	9
$Lind(2)$	10	10	10	7	9	9	10	9	9	8
	10	10	10	6	9	9	10	9	9	9
	10	10	11	7	9	9	9	9	9	8

Table 7 Estimated powers for 20% censoring for a sample size of $n = 50$ with three different censoring distributions

F	KS_n	CM_n	LS_n	KR_n	$S_{n,0.75}^{(1)}$	$S_{n,1}^{(1)}$	$S_{n,2}^{(1)}$	$S_{n,0.75}^{(2)}$	$S_{n,1}^{(2)}$	$S_{n,2}^{(2)}$
$W(0.5)$	7	6	8	6	7	8	7	7	7	7
	8	6	8	6	8	8	8	8	8	8
	9	9	9	6	7	7	8	9	8	7
$W(1)$	8	8	9	6	7	7	7	8	8	7
	8	8	8	7	8	8	8	8	8	8
	8	8	9	6	7	7	7	8	8	7
$W(1.5)$	9	9	10	10	8	8	9	9	8	8
	9	9	9	9	8	8	9	9	8	8
	8	8	9	6	7	7	7	9	8	7
$W(2)$	9	9	10	8	8	8	9	9	9	8
	9	9	9	10	8	8	9	9	9	8
	8	8	9	7	7	7	8	8	8	7
$\Gamma(2)$	12	13	12	12	13	13	13	11	12	12
	11	12	11	11	12	13	14	10	11	13
	11	12	11	9	11	11	12	11	11	11
$\Gamma(3)$	15	17	16	12	19	19	19	14	15	18
	15	17	16	15	18	18	18	13	14	18
	13	16	15	10	15	15	14	13	14	15

Table 7 continued

F	KS_n	CM_n	LS_n	KR_n	$S_{n,0.75}^{(1)}$	$S_{n,1}^{(1)}$	$S_{n,2}^{(1)}$	$S_{n,0.75}^{(2)}$	$S_{n,1}^{(2)}$	$S_{n,2}^{(2)}$
$LN(0.5)$	40	50	52	24	57	59	58	32	41	55
	40	50	50	20	56	57	53	31	40	53
	35	45	48	16	49	49	42	29	37	46
$LN(1)$	34	45	44	17	45	45	37	27	34	42
	32	42	34	21	37	37	33	21	28	35
	35	45	48	16	50	50	42	29	36	46
$\chi^2(8)$	17	20	20	12	23	24	23	15	18	22
	17	20	19	13	22	23	22	15	17	22
	15	18	18	12	19	19	17	14	16	18
$\chi^2(10)$	20	23	23	14	26	27	27	16	19	26
	19	22	22	14	26	27	26	16	19	25
	17	21	20	12	21	21	19	16	18	20
$\beta(1, 1)$	55	67	74	34	70	71	73	59	64	71
	55	66	72	32	68	69	69	58	63	68
	51	51	64	0	19	9	0	49	44	6
$\beta(0.5, 1)$	50	59	64	19	56	54	47	52	55	53
	48	54	61	3	39	32	12	48	48	29
	51	51	64	0	19	9	0	48	43	6
$Lind(0.5)$	9	10	12	9	9	9	9	9	8	8
	9	9	11	6	9	9	8	8	8	8
	9	9	12	6	8	8	8	9	8	8
$Lind(2)$	8	8	10	6	7	7	7	8	8	7
	8	8	9	6	8	8	7	8	8	7
	9	9	10	6	7	7	7	8	8	7

Table 8 Estimated powers for 10% censoring for a sample size of $n = 100$ with three different censoring distributions

F	KS_n	CM_n	LS_n	KR_n	$S_{n,0.75}^{(1)}$	$S_{n,1}^{(1)}$	$S_{n,2}^{(1)}$	$S_{n,0.75}^{(2)}$	$S_{n,1}^{(2)}$	$S_{n,2}^{(2)}$
$W(0.5)$	10	10	9	6	7	7	7	8	8	7
	9	9	9	7	8	8	8	8	8	8
	10	10	10	10	9	8	9	10	9	8
$W(1)$	10	10	10	10	8	8	9	9	9	8
	10	10	10	11	8	8	10	9	9	8
	10	10	10	10	9	9	9	10	9	9
$W(1.5)$	10	10	10	9	9	9	9	10	10	9
	10	10	10	9	9	9	9	9	9	8
	10	10	10	11	8	8	9	9	9	8
$W(2)$	10	10	10	8	9	9	9	10	10	9

Table 8 continued

F	KS_n	CM_n	LS_n	KR_n	$S_{n,0.75}^{(1)}$	$S_{n,1}^{(1)}$	$S_{n,2}^{(1)}$	$S_{n,0.75}^{(2)}$	$S_{n,1}^{(2)}$	$S_{n,2}^{(2)}$
$\Gamma(2)$	10	10	10	9	9	9	10	10	10	9
	10	10	10	11	9	9	9	9	9	8
	17	19	19	18	21	22	23	14	16	20
	17	19	18	18	20	21	22	14	16	20
$\Gamma(3)$	17	18	18	15	19	20	21	14	16	19
	25	29	30	30	34	36	38	19	23	33
	24	29	29	29	33	35	37	19	23	32
	23	28	29	23	32	33	34	19	23	31
$LN(0.5)$	74	85	87	84	91	92	95	61	74	90
	72	84	87	83	90	92	94	61	74	90
	71	83	86	79	89	91	93	60	72	89
	71	83	85	42	88	90	90	58	71	87
$LN(1)$	70	82	83	26	87	88	87	54	68	85
	71	83	86	78	89	91	93	60	73	88
$\chi^2(8)$	30	36	38	38	43	45	49	22	28	41
	30	36	38	37	42	45	48	23	28	41
	29	34	36	28	39	42	43	22	27	38
	34	41	43	44	48	51	56	26	33	47
$\chi^2(10)$	34	41	44	44	49	52	56	26	33	48
	33	40	42	33	46	48	50	25	31	45
$\beta(1, 1)$	91	97	99	98	98	98	99	96	98	99
	91	97	99	98	98	98	98	96	97	99
	90	95	97	0	84	79	50	91	91	77
	90	96	98	96	97	98	98	95	97	98
$\beta(0.5, 1)$	90	96	98	95	97	98	98	95	97	98
	91	95	98	0	83	77	45	91	90	74
$Lind(0.5)$	11	11	13	10	11	11	13	10	10	10
	11	11	13	11	10	11	12	10	10	10
	11	11	13	11	10	11	12	10	10	10
	10	10	11	10	9	9	10	9	9	9
$Lind(2)$	10	10	11	12	9	9	10	9	9	8
	10	10	11	10	9	9	10	9	9	9

Comparing the results associated with a sample size of 50, we see that $S_{n,1}^{(1)}$, KR_n and LS_n generally tend to provide the highest powers. However, it should be noted that KR_n achieves very low power against certain alternatives; notably against the beta distributions considered when the censoring distribution is the Kozoil-Green model. When considering the empirical powers associated with samples of size 100, $S_{n,2}^{(1)}$ exhibits the highest powers, followed by $S_{n,1}^{(1)}$ and LS_n .

Table 9 Estimated powers for 20% censoring for a sample size of $n = 100$ with three different censoring distributions

F	KS_n	CM_n	LS_n	KR_n	$S_{n,0.75}^{(1)}$	$S_{n,1}^{(1)}$	$S_{n,2}^{(1)}$	$S_{n,0.75}^{(2)}$	$S_{n,1}^{(2)}$	$S_{n,2}^{(2)}$
$W(0.5)$	8	7	9	6	7	7	7	7	7	7
	9	7	9	6	8	8	8	7	7	8
	9	9	9	6	7	7	7	8	8	7
$W(1)$	9	9	9	6	7	7	7	8	8	7
	8	8	8	7	7	8	8	7	7	8
	9	9	9	6	7	7	7	8	8	7
$W(1.5)$	9	10	10	10	8	8	9	9	9	8
	9	10	9	12	9	9	10	9	9	9
	9	9	9	11	8	8	9	9	8	8
$W(2)$	10	10	10	8	9	9	9	9	9	8
	10	10	10	10	8	8	9	9	9	8
	9	9	9	6	7	7	7	9	8	7
$\Gamma(2)$	15	17	17	12	18	18	18	14	15	17
	15	17	15	12	15	16	16	12	13	15
	15	16	16	10	15	15	14	13	13	14
$\Gamma(3)$	23	27	28	17	30	32	32	18	22	29
	22	26	26	15	28	29	29	17	20	27
	20	24	25	12	23	24	21	17	19	22
$LN(0.5)$	67	80	83	57	86	88	90	55	68	85
	65	79	82	37	85	87	87	53	66	83
	61	76	80	18	81	82	79	51	63	78
$LN(1)$	61	75	77	20	76	76	67	47	59	72
	43	50	49	26	45	45	37	30	37	42
	44	52	54	38	55	56	56	39	45	54
$\chi^2(8)$	27	32	34	22	38	40	42	21	26	37
	27	32	34	18	37	38	39	21	26	36
	24	30	31	13	30	31	27	20	23	28
$\chi^2(10)$	31	38	40	26	44	47	49	24	30	43
	23	26	27	22	29	30	31	19	22	28
	21	24	25	19	26	27	27	18	20	26
$\beta(1, 1)$	87	95	96	84	96	97	97	93	95	97
	87	94	96	81	96	96	97	92	94	97
	85	85	91	0	23	8	0	78	68	5
$\beta(0.5, 1)$	86	91	92	73	93	93	93	89	92	94
	86	89	89	49	87	86	80	86	87	86
	85	85	91	0	24	8	0	78	68	5
$Lind(0.5)$	10	11	13	9	9	9	9	9	8	8
	10	11	13	6	10	9	9	9	9	9
	10	11	13	6	9	8	8	8	8	8

Table 9 continued

F	KS_n	CM_n	LS_n	KR_n	$S_{n,0.75}^{(1)}$	$S_{n,1}^{(1)}$	$S_{n,2}^{(1)}$	$S_{n,0.75}^{(2)}$	$S_{n,1}^{(2)}$	$S_{n,2}^{(2)}$
$Lind(0.5)$	9	10	10	7	7	7	7	8	8	7
	9	9	10	6	7	7	7	8	7	7
	9	10	10	6	7	7	7	8	8	6

When compiling the numerical results, we also considered a wider range of values for the tuning parameter a than those reported in the table. Although some power variation is evident when varying a , the powers achieved by the newly proposed classes of tests are not particularly sensitive to the choice of the tuning parameter a . However, based on the observed numerical powers, we recommend using $S_{n,2}^{(1)}$ when testing the hypothesis in question in the presence of censoring.

4 Practical applications and conclusion

In this section, we apply the tests used in Sect. 3 to test the hypothesis in (1) based on two real-world data sets. The first data set, reported in Table 10, contains the survival

Table 10 Survival times after leukemia diagnosis, in days

7, 47, 58, 74, 177, 232, 273, 285, 317, 429, 440, 445, 455, 468, 495, 497, 532, 571, 579, 581, 650, 702, 715, 779, 881, 900, 930, 968, 1077, 1109, 1314, 1334, 1367, 1534, 1712, 1784, 1877, 1886, 2045, 2056, 2260, 2429, 2509

Table 11 p -values associated with the various tests used in the full sample case

$Test$	KS_n	CM_n	LS_n	KR_n	$S_{n,1}^{(1)}$	$S_{n,2}^{(1)}$	$S_{n,5}^{(1)}$	$S_{n,1}^{(2)}$	$S_{n,2}^{(2)}$	$S_{n,5}^{(2)}$
p -value	0.49	0.62	0.41	0.34	0.68	0.25	0.18	0.32	0.36	0.19

Table 12 Initial remission times of leukemia patients, in days

4, 5, 8, 8, 9, 10, 10, 10, 10, 10, 11, 12, 12, 12*, 13, 14, 20, 20*, 23, 23, 25, 25, 25, 28, 28, 28, 28, 29, 31, 31, 31, 32, 37, 40, 41, 41, 48, 48, 57, 62, 70, 74, 75, 89, 99, 100, 103, 124, 139, 143, 159*, 161*, 162, 169, 190*, 195, 196*, 197*, 199*, 205*, 217*, 219*, 220, 245*, 258*, 269*

Table 13 p -values associated with the various tests used in the censored case

$Test$	KS_n	CM_n	LS_n	KR_n	$S_{n,1}^{(1)}$	$S_{n,2}^{(1)}$	$S_{n,5}^{(1)}$	$S_{n,1}^{(2)}$	$S_{n,2}^{(2)}$	$S_{n,5}^{(2)}$
p -value	0.03	0.04	0.01	0.28	0.13	0.28	0.41	0.66	0.82	0.73

times, in days, of 43 Leukemia patients. For a discussion of the original data set see Kotze and Johnson (1983) as well as Allison et al. (2017). This data set is not subject to censoring, i.e. all lifetimes are observed. The second data set contains the initial remission times of leukemia patients, in days; for more details see Lee and Wang (2003), this data set can be found in Table 12. These data contains censored observations, indicated using an asterisk. The original data were segmented into three treatment groups. However, Lee and Wang (2003) showed that the data do not display significant differences among the various treatments. As a result we treat the data as i.i.d. realisations from a single, censored, lifetime distribution. All reported p-values are estimated using one hundred thousand bootstrap replications; these results are displayed in Tables 11 and 13, respectively.

From the results of the practical example in Table 11 it is clear that none of the tests reject the null hypothesis that the survival times after a leukemia diagnosis are Weibull distributed at the 5% or 10% levels of significance. As a result, we conclude that the Weibull distribution is an appropriate model for these data.

The results associated with the initial remission times, in Table 13, indicate that KS_n , CM_n and LS_n reject the hypothesis in (1) at a 5% significance level. However, none of the remaining 7 tests considered result in a rejection of the null hypothesis at the 5% or 10% levels. We conclude that the Weibull distribution is likely to be an appropriate model for the observed times. The data set under consideration was also analysed in Bothma et al. (2020), where the null hypothesis of exponentiality of the remission time was strongly rejected. The mentioned paper recommended that a more flexible distribution be used when modelling these data. The results above indicate that the additional flexibility of the Weibull (compared to the exponential) distribution indeed ensures that the Weibull distribution is a more appropriate model than the exponential for the initial remission times considered.

A number of interesting numerical phenomena are evident when considering the powers of the various tests. It is clear that the achieved powers and, therefore, the null distribution of the test statistic, is influenced by the shape of the censoring distribution. The effect of the censoring distribution on the critical values of the tests seem not to have been investigated in the literature to date. Some authors perform goodness-of-fit testing by enforcing a parametric assumption on the censoring distribution, see e.g., Kim (2017). An additional consideration that seems to have been neglected in the literature is the effect on the null distribution of the test statistic, and hence the power of the test, of a specific assumption made in the Kaplan-Meier estimate of the distribution function. Some authors, in order to ensure that G_n satisfies the requirements of a distribution, defines $G_n(t) = 1$ for all $t > X_{(n)}$ regardless of whether or not the sample maximum is censored. We are currently investigating these open questions.

Funding Not applicable.

Availability of data and material All data used are in the public domain.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Code availability Custom code was written and are available upon request.

References

- Allison J, Milošević B, Obradović M, Smuts M (2021) Distribution-free goodness-of-fit tests for the pareto distribution based on a characterization. *Comput Stat* pp 1–16
- Allison J, Santana L (2015) On a data-dependent choice of the tuning parameter appearing in certain goodness-of-fit tests. *J Stat Comput Simul* 85(16):3276–3288
- Allison JS, Betsch S, Ebner B, Visagie IJH (2019) New weighted L^2 -type tests for the inverse Gaussian distribution. *arXiv preprint arXiv:1910.14119*
- Allison JS, Huskova M, Meintanis SG (2017) Testing the adequacy of semiparametric transformation models. *Test* 27:1–25
- Allison JS, Santana L, Smit N, Visagie IJH (2017) An “apples-to-apples” comparison of various tests for exponentiality. *Comput Stat* 32(4):1241–1283
- Balakrishnan N, Chimitova E, Vedernikova M (2015) An empirical analysis of some nonparametric goodness-of-fit tests for censored data. *Commun Stat Simul Comput* 44(4):1101–1115
- Baringhaus L, Ebner B, Henze N (2017) The limit distribution of weighted L_2 -goodness-of-fit statistics under fixed alternatives, with applications. *Ann Inst Stat Math* 69(5):969–995
- Baringhaus L, Henze N (1988) A consistent test for multivariate normality based on the empirical characteristic function. *Metrika* 35(1):339–348
- Baringhaus L, Henze N (1991) A class of consistent tests for exponentiality based on the empirical laplace transform. *Ann Inst Stat Math* 43(3):551–564
- Betsch S, Ebner B (2018) Testing normality via a distributional fixed point property in the Stein characterization. *TEST* pp. 1–34. <https://doi.org/10.1007/s11749-019-00630-0>
- Betsch S, Ebner B (2019) A new characterization of the gamma distribution and associated goodness-of-fit tests. *Metrika* 82(7):779–806
- Bothma E, Allison JS, Cockeran M, Visagie IJH (2020) Kaplan-Meier based tests for exponentiality in the presence of censoring. *arXiv preprint arXiv:2011.04519*
- Breslow N, Crowley J (1974) A large sample study of the life table and product limit estimates under random censorship. *Ann Stat* pp. 437–453
- Cabaña A, Quiroz AJ (2005) Using the empirical moment generating function in testing for the Weibull and the type I extreme value distributions. *Test* 14(2):417–431
- Cox DR, Oakes D (1984) *Analysis of survival data*, vol 21. CRC Press, Boca Raton
- Cuparić M, Milošević B (2021) New characterization-based exponentiality tests for randomly censored data. *TEST*. <https://doi.org/10.1007/s11749-021-00787-7>
- D’Agostino RB, Stephens MA (1986) *Goodness-of-fit techniques*, vol 68. CRC Press, Boca Raton
- Döring M, Cramer E (2019) On the power of goodness-of-fit tests for the exponential distribution under progressive Type-II censoring. *J Stat Comput Simul* 89:2997–3034
- Efron B (1967) The two sample problem with censored data. In: *Proceedings of the fifth berkeley symposium on mathematical statistics and probability*, vol 4, pp 831–853
- Fernández T, Rivera N (2020) Kaplan-Meier V-and U-statistics. *Electron J Stat* 14(1):1872–1916
- Fernández VA, Jiménez Gamero MD, García M (2008) A test for the two-sample problem based on empirical characteristic functions. *Comput Stat Data Anal* 52:3730–3748
- Feuerverger A, Mureika RA (1977) The empirical characteristic function and its applications. *Ann Stat* 5(1):88–97
- Giacomini R, Politis DN, White H (2013) A warp-speed method for conducting Monte Carlo experiments involving bootstrap estimators. *Econom Theory* 29(3):567–589
- Gupta RD, Richards DSP (1997) Invariance properties of some classical tests for exponentiality. *J Stat Plann Inference* 63(2):203–213
- Henze N, Visagie IJH (2020) Testing for normality in any dimension based on a partial differential equation involving the moment generating function. *Ann Inst Stat Math* 72:1109–1136
- Hlavac M (2018) stargazer: well-formatted regression and summary statistics tables. <https://CRAN.R-project.org/package=stargazer>
- Jacobson A, Wilson V, Pileggi S (2018) *parmsurvfit: Parametric Models for Survival Data*. <https://CRAN.R-project.org/package=parmsurvfit>. R package version 0.1.0

- Jiang R, Murthy D (2011) A study of Weibull shape parameter: properties and significance. *Reliab Eng Syst Saf* 96(12):1619–1626
- Kalbfleisch JD, Prentice RL (2011) *The statistical analysis of failure time data*, vol 360. Wiley, New York
- Kaplan EL, Meier P (1958) Nonparametric estimation from incomplete observations. *J Am Stat Assoc* 53(282):457–481
- Kim N (2017) Goodness-of-fit tests for randomly censored Weibull distributions with estimated parameters. *Commun Stat Appl Methods* 24(5):519–531
- Klar B, Meintanis SG (2005) Tests for normal mixtures based on the empirical characteristic function. *Comput Stat Data Anal* 49(1):227–242
- Kotz S, Nadarajah S (2000) *Extreme value distributions: theory and applications*. World Scientific, Singapore
- Kotze S, Johnson N (1983) *Encyclopedia of statistical sciences*, vol 3. Wiley, New York
- Kozioł JA, Green SB (1976) A Cramér-von Mises statistic for randomly censored data. *Biometrika* 63(3):465–474
- Krit M (2014) Goodness-of-fit tests for the Weibull distribution based on the Laplace transform. *Journal de la Société Française de Statistique* 155(3):135–151
- Lee ET, Wang J (2003) *Statistical methods for survival data analysis*, vol 476. Wiley, New York
- Liao M, Shimokawa T (1999) A new goodness-of-fit test for type-i extreme-value and 2-parameter Weibull distributions with estimated parameters. *Optimization* 64(1):23–48
- Mann NR, Schneuer EM, Fertig KW (1973) A new goodness-of-fit test for the two-parameter Weibull or extreme-value distribution with unknown parameters. *Commun Stat Theory Methods* 2(5):383–400
- Mazucheli J, Fernandes LB, de Oliveira RP (2016) *LindleyR: The Lindley Distribution and Its Modifications*. <https://CRAN.R-project.org/package=LindleyR>. R package version 1.1.0
- Meintanis SG, Iliopoulos G (2003) Tests of fit for the Rayleigh distribution based on the empirical Laplace transform. *Ann Inst Stat Math* 55(1):137–151
- Meintanis SG, Ngatchou-Wandji J, Allison JS (2018) Testing for serial independence in vector autoregressive models. *Stat Papers* 59(4):1379–1410
- Mijburgh PA, Visagie IJH (2020) An overview of goodness-of-fit tests for the Poisson distribution. *South African Stat J* 54(2):207–230
- R Core Team (2019) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Tiku ML, Singh M (1981) Testing the two parameter Weibull distribution. *Commun Stat Theory Methods* 10(9):907–918

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.