Data transformations and goodness-of-fit tests for type-II right censored samples

Christian Goldmann · Bernhard Klar · Simos G. Meintanis

Abstract We suggest several goodness-of-fit (GOF) methods which are appropriate with Type-II right censored data. Our strategy is to transform the original observations from a censored sample into an approximately i.i.d. sample of normal variates and then perform a standard GOF test for normality on the transformed observations. A simulation study with several well known parametric distributions under testing reveals the sampling properties of the methods. We also provide theoretical analysis of the proposed method.

Keywords Empirical characteristic function · Empirical distribution function · Goodness-of-fit test · Censored data

---

Simos G. Meintanis: On sabbatical leave from the University of Athens.

C. Goldmann
Mathematical Methods in Dynamics and Durability, Fraunhofer Institute for Industrial Mathematics ITWM, Frauhofer Platz 1, 67663 Kaiserslautern, Germany

B. Klar (✉)
Department of Mathematics, Karlsruhe Institute of Technology (KIT),
Kaiserstraße 89, 76133 Karlsruhe, Germany
e-mail: bernhard.klar@kit.edu

S. G. Meintanis
Department of Economics, National and Kapodistrian University of Athens,
8 Pesmazoglou Street, 105 59 Athens, Greece

S. G. Meintanis
Unit for Business Mathematics and Informatics, North-West University,
Potchefstroom, South Africa

1 Introduction

In a sample of size n from the distribution, assume that only the first $r \leq n$ order statistics $X_{1:n} < X_{2:n} < \ldots < X_{r:n}$, are observed. This censoring scheme is referred to as Type-II censoring. Let X be the underlying random variable and denote by $F(x)$ the distribution function (DF) of X. We are interested in the goodness-of-fit (GOF) null hypothesis

$$H_0 : F \equiv \mathcal{F}_\vartheta, \quad \text{for some } \vartheta \in \Theta, \tag{1.1}$$

with $\Theta \subseteq \mathbb{R}^p$, $p \geq 1$, where $\mathcal{F}_\vartheta$ denotes a specific family of distributions indexed by a parameter ϑ. Typically the null hypothesis in (1.1) is tested by modifications of the standard GOF tests. Early works include the Cramér–von Mises statistic by Pettitt (1976,1977), and the Kolmogorov–Smirnov test by Barr and Davidson (1973) and Dufour and Maag (1978), all with the assumption that the parameter ϑ is known, and the chi–squared tests with estimated parameter in Mihalko and Moore (1980). Other less conventional approaches are the regression tests for exponentiality of Brain and Shapiro (1983), the smooth tests of fit of Bargal and Thomas (1983) and tests based on normalized spacings suggested by D'Agostino and Massaro (1992). A standard reference for GOF tests, including tests with censored data, is D'Agostino and Stephens (1986), while Thode (2002, Chapter 8) contains a nice overview of various methods of testing normality with type–I and type–II censored data. For recent approaches to GOF tests with censored data the reader is referred to Grané (2012), Castro-Kuriss (2011), Castro-Kuriss et al. (2010), Glen and Foote (2009), and Peña (1995), among others.

As already noted the standard approach for the testing problem in (1.1) has been to consider test statistics for the case of no censoring $r = n$, and modify them accordingly in order to make them applicable for the case $r < n$. At the same time however, there exist methods which, given the r order statistics in a random sample of size n from the uniform (0,1) distribution, and based on specific transformations of the data, yield order statistics in a random sample of size r from the same distribution. Then of course any GOF statistic for uniformity may be applied as if we had a full sample of size r to begin with. Such 'transformations-to-uniformity' appear in Michael and Schucany (1979), O'Reilly and Stephens (1988), and more recently in Lin et al. (2008), and Fischer and Kamps (2011). Clearly testing uniformity is not a restriction since these statistics naturally extend to the current setting of arbitrary null hypothesis H_0 by use of the probability integral transform.

There is also a line of research which can be combined effectively with the aforementioned transformations-to-uniformity. In particular, and since under H_0 the parameter ϑ needs to be estimated from the data, we essentially have a quasi-probability integral transform, with extra variability introduced during the estimation step. Consequently, the corresponding GOF statistics will depend on the unknown value of the parameter and the method of estimation used in estimating this parameter. In this connection, and in order to make the GOF statistics independent of these choices, Chen and Balakrishnan (1995) proposed a novel transformation-to-normality for this problem. As a result, and by combining the Chen–Balakrishnan transformation-to-normality with any of the aforementioned transformations-to-uniformity, we can conveniently reduce any given testing problem with Type-II censoring to a GOF test for normality with complete samples, which of course is a well studied problem with many solutions. We note that the empirical process underlying the Chen–Balakrishnan transformation was first analysed in the PhD thesis by Chen (1991). Here we provide further theoretical as well as empirical results justifying the general validity of this transformation.

In this paper we apply these transformations-to-uniformity in conjunction with the Chen–Balakrishnan transformation to several GOF tests. The rest of the paper unfolds as follows. In Sects. 2 and 3 we present the transformations and indicate how to implement them in the corresponding GOF statistics. Section 4 deals with the issue of estimating the parameter ϑ under Type-II censoring. In Sect. 5 a Monte Carlo study is drawn in which several combinations of test statistics and transformations are studied in their sampling properties. Finally Sect. 6 contains the conclusions of this study. A theoretical analysis of the basic process involved in the Chen–Balakrishnan transformation, assisted by simulations, is provided in an appendix. A longer version of the paper is contained in a technical report which is available at `arXiv:1312.3078 [stat.ME]`. In what follows we will refer to this report as TR.

2 Transformations

Denote by $U(0, 1)$ the uniform distribution on (0,1) and suppose that $U_{1:n} < U_{2:n} < \cdots < U_{r:n}$, are the first r order statistics in a random sample of size n from $U(0, 1)$ distribution. Further, put $U_{0:n} \equiv 0$. Let $\mathbf{U} := (U_{1:n}, U_{2:n} \ldots U_{r:n})^{\mathrm{T}}$ and denote by $\mathbf{u} := (u_{1:r}, u_{2:r} \ldots u_{r:r})^{\mathrm{T}}$ the set of order statistics in a random sample of size r from $U(0, 1)$. We seek transformations of the type $\mathcal{T} : \mathbf{U} \mapsto \mathbf{u}$, that is transformations which from the censored set of order statistics $\mathbf{U}$ in a sample of size n from $U(0, 1)$, lead to a complete set $\mathbf{u}$ of order statistics in a sample of size r from $U(0, 1)$. The following transformations have appeared in the literature:

- (1). Michael and Schucany (1979)

$$u_{i:r} = \frac{U_{i:n}}{U_{r:n}} \left[B_{r,n-r+1}(U_{r:n})\right]^{1/r}, \quad i = 1, 2 \ldots r,$$

  where $B_{r,n-r+1}(u) = \sum_{k=r}^{n} \frac{n!}{k!(n-k)!} u^k (1-u)^{n-k}$, denotes the DF of the beta distribution with parameters r and $n - r + 1$.
- (2). O'Reilly and Stephens (1988)

$$u_{i:r} = 1 - \prod_{j=1}^{i} \left[\frac{1 - U_{j:n}}{1 - U_{j-1:n}}\right]^{\frac{n-j+1}{r-j+1}}, \quad i = 1, 2 \ldots r.$$

- (3). Lin et al. (2008), see also Fischer and Kamps (2011, Theorem 2(3.)). Let

$$u_i = \left[\frac{1 - U_{i:n}}{1 - U_{i-1:n}}\right]^{n-i+1}, \quad i = 1, 2 \ldots r.$$

Then set $u_{i:r} = u_{(i)}, \quad i = 1, 2 \ldots r$, where $u_{(1)} < u_{(2)} < \cdots < u_{(r)}$ denotes the ordered set of u_i.

– (4). Fischer and Kamps (2011, Theorem 2 (cases 4. and 5.))

$$u_{i:r} = \prod_{j=i}^{r} \left[1 - \left(\frac{1 - U_{j:n}}{1 - U_{j-1:n}} \right)^{n-j+1} \right]^{1/j}, \quad i = 1, 2 \ldots r.$$

– (5). Fischer and Kamps (2011, Theorem 2 (cases 2. and 6.))

$$u_{i:r} = 1 - \left[1 - B_{r,n-r+1}(U_{r:n})\right]^{1/r} \prod_{j=2}^{i} \left[1 - \left(\frac{U_{r-j+1:n}}{U_{r-j+2:n}} \right)^{r-j+1} \right]^{\frac{1}{r-j+1}},$$
$$i = 1, 2 \ldots r.$$

As it has already been mentioned, the transformations above are to be combined with a transformation-to-normality. The aim with this combination is to produce transformed values, say z_j, which are stochastically equivalent under the null hypothesis H_0 to standardized values which would have been produced in a complete random sample of size r from the standard normal distribution. The latter transformation, which is presented below for the case of a complete sample, was shown to be effective for a wide variety of distributions under testing with uncensored samples; see Meintanis (2009). It has also been applied successfully to the case of testing for the error distribution in generalized linear models by Klar and Meintanis (2012).

– (6). Chen and Balakrishnan (1995)
  (i) Efficiently estimate ϑ by $\widehat{\vartheta}_n$ based on $X_{j:n}, \ j = 1, 2 \ldots n$.
  (ii) Calculate $Y_j = \Phi^{-1}\left(\mathcal{F}_{\widehat{\vartheta}_n}(X_{j:n})\right)$, $\Phi(\cdot)$ being the standard normal DF.
  (iii) Compute $Z_j = (Y_j - \bar{Y})/s_Y$, where $\bar{Y} = n^{-1}\sum_{j=1}^{n} Y_j$, and $s_Y^2 = (n-1)^{-1}\sum_{j=1}^{n}(Y_j - \bar{Y})^2$.

In the Appendix we provide an analysis of the process produced by the Chen–Balakrishnan transformation in an effort to justify the documented validity of this approach under so diversified sampling situations. There, the process $\hat{\beta}_{n,2}$ (see in Sect. 7.1 in the Appendix) is the dominating part and corresponds to testing for normality with estimated parameters, for which efficient (ML) estimators exist, and the test statistics do not depend on the estimates of the parameters nor do they depend on the values of these parameters (mean and standard deviation).

3 Test statistics

We now illustrate the combined transformation which is suitable for testing the null hypothesis H_0 with arbitrary $\mathcal{F}_\vartheta$, based on a Type–II censoring scheme.

`TRANSFORMATION (7):`

– Efficiently estimate ϑ by $\widehat{\vartheta}_r$, based on $X_{j:n}, \ j = 1, 2 \ldots r$.
– Calculate $\widehat{U}_{j:r} = \mathcal{F}_{\widehat{\vartheta}_r}(X_{j:r})$, and set $\widehat{\mathbf{U}} = (\widehat{U}_{1:n}, \widehat{U}_{2:n} \ldots \widehat{U}_{r:n})^{\mathrm{T}}$

– Transform to $u_{j:r} = \mathcal{T}(\widehat{\mathbf{U}})$, where $\mathcal{T}$ denotes anyone of the transformations (1)–(5).
– Replace n by r in transformation (6), and perform step (ii) of this transformation with $\mathcal{F}_{\widehat{\vartheta}_n}(X_{j:n})$ replaced by $u_{j:r}$.
– Perform step (iii) of transformation (6), and then apply any test statistic for normality to the values $z_j,\ \ j = 1, 2 \ldots r$, so produced.

The appropriate normality tests are with estimated parameters and amongst them we consider the classical GOF statistics based on the empirical DF. Specifically, the Cramér–von Mises and the Anderson–Darling are given by

$$W^2 = \sum_{j=1}^{r} \left(\Phi(z_j) - \frac{2j-1}{2r} \right)^2 + \frac{1}{12r}, \tag{3.1}$$

and

$$A^2 = -r - \frac{1}{r} \sum_{j=1}^{r} \left[(2j-1) \log \Phi(z_j) + (2r+1-2j) \log(1 - \Phi(z_j)) \right], \tag{3.2}$$

respectively. Asymptotic percentage points and modifications of the statistics for finite sample size can be found in Table 4.7 in D'Agostino and Stephens (1986).

We also consider a test for normality which utilizes the characteristic function (CF) and takes the form

$$C^2 = r \int_{-\infty}^{\infty} |\widehat{\varphi}_r(t) - e^{-(1/2)t^2}|^2 \, w(t) dt, \tag{3.3}$$

where $\widehat{\varphi}_r(t) = r^{-1} \sum_{j=1}^{r} e^{itz_j}$ is the empirical CF of $z_j,\ \ j = 1, \ldots, r$, and $w(t)$ denotes a weight function introduced in order to smooth out the periodic behavior of $\widehat{\varphi}_r(t)$. Note that the test statistic C^2 compares the empirical CF of z_j to the CF of the standard normal distribution. For $w(t) = e^{-at^2},\ a > 0$, we have from (3.3) after some straightforward algebra,

$$C^2 := C_a^2 = \frac{1}{r} \sqrt{\frac{\pi}{a}} \sum_{j,k=1}^{r} e^{-(z_j - z_k)^2/4a} \tag{3.4}$$

$$- 2\sqrt{\frac{2\pi}{1+2a}} \sum_{j=1}^{r} e^{-z_j^2/(2+4a)} + r\sqrt{\frac{\pi}{1+a}}. \tag{3.5}$$

Epps and Pulley (1983) proposed this test statistic and showed that C_a^2 is very competitive to the classical tests W^2 and A^2. Despite the fact that the asymptotic null distribution of this statistic is complicated, there exist some approximations thereof; see for instance Henze (1990) for an approximation based on Johnson distributions.

In fact, by using a simple transformation of C_a^2, the test can be easily carried out for finite samples provided that the sample size is larger than or equal to 10 (Henze 1990, p. 17).

In connection with the weight function we point out that the choice $w(t) = e^{-at^2}$ has become something of a standard for the CF statistic in (3.3); see for instance Epps and Pulley (1983), Epps (2005) and Henze and Wagner (1997).

It is well-known that the choice of the weight parameter a has a considerable impact on the power properties of the CF statistic as it has been related to the choice of the bandwidth in density estimation. The only analytic treatment available on objective optimal values of a is provided in Tenreiro (2009) by relating this value to the local Bahadur slopes of the test statistic. Even with these analytical results, specific quantitative suggestions require consideration of specific deviations from the null hypothesis of normality. Nevertheless Tenreiro (2009) recovers what was already a common practice in simulations, namely that smaller values (resp. larger values) of a are appropriate for detecting short-tailed (resp. long-tailed) alternatives. Based on these theoretical considerations as well as on extensive simulations, he suggests a bandwith of 0.71, corresponding to $a = 0.5$, as an overall compromise choice. This conclusion agrees with the results described in Epps and Pulley (1983); the same weight was also chosen in other studies like those of Baringhaus et al. (1989) and Arcones and Wang (2006). As a consequence, we also employed the weight $a = 0.5$ in our simulations.

We close this section by noting that performing a test for uniformity after the second step of transformation (7), which would in fact seem as a reasonable and simpler approach, does not lead to a valid testing procedure since the test statistics would then depend on the parameter estimate and the parameter value. We will take up and further clarify this point again in the Appendix. Also note that Chen and Balakrishnan (1995) suggested transformation (6) (and provided partial justification for), in the case of the tests W^2 and A^2. Nevertheless, the Kolmogorov–Smirnov statistic would have also been a reasonable competitor in this study, however we opt not to included it as it is generally known to be less powerful compared to the Cramér–von Mises and the Anderson–Darling tests; see for instance Castro-Kuriss (2011) or Glen and Foote (2009).

4 Estimation of parameters

In order to implement transformation (7) we require an efficient estimator $\widehat{\vartheta}_r$ of the parameter ϑ, such as the maximum likelihood estimator (MLE). This estimator employs the censored data $X_{j:r},\ j = 1, 2 \ldots r$, and will depend on the specific parametric form of $\mathcal{F}_\vartheta$ under the null hypothesis H_0. The following parametric distributions are of special interest:

– The exponential distribution $Exp(\sigma)$ with DF, $F(x) = 1 - e^{-x/\sigma}$. Then the MLE is given by

$$\widehat{\sigma} = \frac{\sum_{j=1}^{r} X_{j:n} + (n-r)X_{r:n}}{r}.$$

– The gamma distribution $\gamma(\theta, \sigma)$ with density, $(\sigma^\theta \Gamma(\theta))^{-1} x^{\theta-1} e^{-x/\sigma}$. A simplified form of the MLE equations is given by (see Wilk et al. 1962 or Johnson et al. 1994),

$$r \log \mathcal{P}_r = n \left[\frac{\Gamma'(\widehat{\theta})}{\Gamma(\widehat{\theta})} - \log \frac{X_{r:n}}{\widehat{\sigma}} \right] - (n-r) \frac{\partial \log J(\widehat{\theta})}{\partial \widehat{\theta}},$$

$$\frac{X_{r:n} \mathcal{S}_r}{\widehat{\sigma}} = \widehat{\theta} - \frac{(n-r) e^{-X_{r:n}/\widehat{\sigma}}}{r J(\widehat{\theta})},$$

where $J(\widehat{\theta}) := J(\widehat{\theta}, X_{r:n}/\widehat{\sigma})$, with $J(x, y) = \int_1^\infty t^{x-1} e^{-yt} dt$, and

$$\mathcal{P}_r = \frac{\left(\prod_{j=1}^r X_{j:n} \right)^{1/r}}{X_{r:n}}, \quad \mathcal{S}_r = \frac{\sum_{j=1}^r X_{j:n}}{r X_{r:n}}.$$

– The normal distribution $\mathcal{N}(\mu, \sigma^2)$ with mean μ and variance σ^2. We employ the estimates

$$\widehat{\mu} = \sum_{j=1}^r b_j X_{j:r}, \quad \widehat{\sigma} = \sum_{j=1}^r c_j X_{j:r},$$

suggested by Gupta (1952). This author provided the values of the coefficients (b_j, c_j) for $n \leq 10$. For larger sample sizes the values suggested are

$$b_j = \frac{1}{r} - \frac{\bar{m}(m_j - \bar{m})}{\sum_{j=1}^r (m_j - \bar{m})^2}, \quad c_j = \frac{m_j - \bar{m}}{\sum_{j=1}^r (m_j - \bar{m})^2},$$

where m_j denotes the expected value of the jth order statistic in a sample of size n from the standard normal distribution, and $\bar{m} = r^{-1} \sum_{j=1}^r m_j$. There is also an approximate method whereby m_j is replaced by $\Phi^{-1}((j - 0.375)/(n + 0.125))$; see D'Agostino and Stephens (1986).
We have also implemented as an alternative estimation method the modified maximum likelihood estimation proposed by Tiku and co-authors. This method uses a suitable linearization of the likelihood function; see Tiku (1967) or Tiku et al. (1986).

5 Simulations

Tables 1, 2, 3, 4, 5 and 6 show parts of the results of extensive simulation studies with the testing procedures presented in Sect. 3. Specifically we employ the Cramér–von Mises test W^2, the Anderson–Darling test A^2, and the characteristic function test with weight function $e^{-t^2/2}$, denoted by C^2. We used transformations (1) to (5) (see Sect. 2), abbreviated as MS, OS, LHB, FK1 and FK2.

Table 1 Percentage of rejection of tests for exponentiality based on 10,000 replications (part 1)

Distribution	$\frac{r}{n} * 100\,\%$	n	MS			OS			LHB			FK1			FK2			DS	
			A^2	W^2	C^2	A^2	W^2	C^2	A^2	W^2	C^2	A^2	W^2	C^2	A^2	W^2	C^2	$A^2_{r,n}$	$W^2_{r,n}$
$Exp(1)$	50	40	4	4	3	6	6	5	5	5	4	5	5	4	4	4	3	5	5
	50	100	4	4	3	5	5	5	5	5	5	5	5	4	4	4	3	4	4
	75	40	4	4	4	5	5	5	5	5	5	5	5	5	4	4	4	5	5
	75	100	4	5	4	6	5	6	5	5	5	5	5	5	4	4	4	5	5
$\gamma(2, 1)$	50	40	4	4	3	6	6	6	7	6	7	7	6	6	0	0	0	53	53
	50	100	4	4	4	6	6	6	12	11	15	12	9	14	0	0	1	95	92
	75	40	5	4	4	6	5	6	8	8	9	8	7	8	0	0	0	70	68
	75	100	6	5	5	6	6	6	17	14	20	16	13	19	0	1	1	99	98
$\gamma(4, 1)$	50	40	5	5	3	7	6	6	28	24	33	7	6	7	0	0	0	99	99
	50	100	6	5	4	6	6	6	74	63	80	15	10	18	0	0	0	100	100
	75	40	5	5	4	6	6	6	42	34	49	9	7	10	0	0	0	100	100
	75	100	7	6	6	5	5	5	89	80	93	23	16	29	0	0	0	100	100
$IG(4, 1)$	50	40	15	14	15	13	12	11	5	5	5	4	3	4	2	2	2	19	16
	50	100	36	32	41	38	34	42	10	9	12	5	4	4	4	5	6	62	37
	75	40	39	35	43	36	33	39	6	6	6	5	4	5	15	14	17	23	26
	75	100	79	72	83	86	80	89	13	12	15	7	5	6	36	32	44	64	58
$IG(1, 4)$	50	40	5	5	3	6	6	5	59	48	67	3	2	3	0	0	0	100	100
	50	100	11	10	9	10	9	10	99	94	99	7	5	10	0	0	0	100	100
	75	40	8	8	6	8	7	7	74	61	82	5	3	6	0	0	0	100	100
	75	100	18	16	18	16	15	18	100	99	100	12	7	17	0	0	1	100	100

Table 2 Percentage of rejection of tests for exponentiality based on 10,000 replications (part 2)

Distribution	$\frac{r}{n} * 100\,\%$	n	MS			OS			LHB			FK1			FK2			DS	
			A^2	W^2	C^2	A^2	W^2	C^2	A^2	W^2	C^2	A^2	W^2	C^2	A^2	W^2	C^2	$A^2_{r,n}$	$W^2_{r,n}$
$Wei(2,1)$	50	40	5	5	4	10	9	10	12	11	14	8	7	8	0	0	0	84	85
	50	100	5	5	5	12	11	13	30	25	36	19	14	21	0	0	0	100	100
	75	40	8	7	7	12	11	13	20	17	23	12	10	13	0	0	0	98	98
	75	100	10	8	11	17	15	21	50	41	56	30	22	35	0	0	0	100	100
$Wei(4,1)$	50	40	11	10	10	14	13	15	65	57	72	7	5	7	0	0	0	100	100
	50	100	15	12	14	18	16	21	99	97	99	21	14	27	0	0	0	100	100
	75	40	20	17	20	19	17	21	88	81	92	11	8	14	0	0	0	100	100
	75	100	41	31	43	31	28	36	100	100	100	40	28	51	0	0	0	100	100
$\mathcal{L}\gamma(2,1)$	50	40	30	27	31	29	26	28	19	16	23	2	2	2	1	1	2	73	61
	50	100	75	67	77	80	73	82	73	52	80	2	1	2	5	6	12	100	99
	75	40	71	65	73	70	64	71	16	14	19	3	2	3	20	19	26	55	46
	75	100	99	97	99	100	99	99	50	40	57	5	4	6	57	51	69	100	93
$\mathcal{L}\gamma(4,1)$	50	40	22	20	24	19	17	18	6	5	5	4	3	3	8	8	8	13	15
	50	100	51	45	57	58	51	63	8	8	9	6	5	5	16	16	20	32	33
	75	40	58	51	63	57	52	60	19	17	20	5	4	5	48	44	47	76	78
	75	100	93	87	94	97	95	98	37	34	42	11	9	11	82	76	84	98	99
$\mathcal{LN}(0,1)$	50	40	6	6	5	6	6	4	5	5	5	6	4	5	0	1	1	34	32
	50	100	12	11	12	10	10	10	10	8	12	7	5	8	1	1	2	84	70
	75	40	11	10	12	9	9	9	5	5	5	6	5	6	2	2	2	26	22
	75	100	26	23	30	25	23	28	8	8	10	7	5	7	3	4	5	73	53

Table 3 Percentage of rejection of tests for gamma distribution based on 10000 replications (part 1)

Distribution	$\frac{r}{n} * 100\,\%$	n	MS			OS			LHB			FK1			FK2		
			A^2	W^2	C^2	A^2	W^2	C^2	A^2	W^2	C^2	A^2	W^2	C^2	A^2	W^2	C^2
$\gamma(2, 1)$	50	40	4	4	3	6	6	5	5	5	4	3	4	3	3	4	3
	50	100	4	4	4	6	6	6	5	5	5	4	4	4	4	4	4
	75	40	4	4	4	6	6	5	5	5	5	4	4	3	4	4	4
	75	100	5	5	4	6	6	6	5	5	5	4	5	4	4	5	4
$\mathcal{N}(3, 1)$	50	40	16	15	17	25	23	26	10	10	9	9	9	9	9	9	8
	50	100	31	28	34	46	42	49	17	17	17	19	18	19	18	17	19
	75	40	22	20	24	31	28	34	10	10	10	11	10	11	10	10	11
	75	100	41	38	46	56	52	61	18	17	18	22	19	21	21	19	22
$\mathcal{LN}(0, 1)$	50	40	6	6	6	5	5	4	5	5	4	4	4	3	4	5	4
	50	100	9	9	10	9	8	9	5	5	5	4	4	4	5	6	6
	75	40	11	11	12	9	9	9	5	5	5	4	4	4	6	6	7
	75	100	22	20	26	24	22	27	5	5	5	5	5	5	10	10	12
$IG(4, 1)$	50	40	15	14	16	12	11	11	4	5	4	4	4	3	6	7	6
	50	100	32	29	37	37	32	41	6	6	6	5	5	5	12	12	14
	75	40	39	35	44	37	33	39	6	6	6	5	4	4	15	14	17
	75	100	77	70	81	85	80	88	11	10	13	7	6	7	38	35	47
$IG(1, 4)$	50	40	5	5	4	4	4	3	5	5	5	3	4	3	4	4	3
	50	100	6	6	7	6	6	6	6	5	5	4	4	4	5	5	5
	75	40	7	7	7	6	6	5	5	5	5	4	4	4	5	6	5
	75	100	11	11	13	11	10	12	5	5	5	5	4	4	7	7	8

Table 4 Percentage of rejection of tests for gamma distribution based on 10,000 replications (part 2)

Distribution	$\frac{r}{n} * 100\,\%$	n	MS			OS			LHB			FK1			FK2		
			A^2	W^2	C^2	A^2	W^2	C^2	A^2	W^2	C^2	A^2	W^2	C^2	A^2	W^2	C^2
$Wei(2, 1)$	50	40	4	5	4	9	8	9	5	5	4	3	4	3	3	4	3
	50	100	5	6	6	13	11	14	5	5	5	5	5	4	4	5	4
	75	40	6	6	7	12	10	13	5	5	4	4	5	4	4	4	4
	75	100	9	8	10	18	16	21	5	5	5	4	5	4	5	6	6
$Wei(4, 1)$	50	40	7	7	7	14	13	14	5	5	4	4	4	3	4	4	4
	50	100	9	8	10	22	19	24	5	5	5	4	5	4	5	5	6
	75	40	11	10	12	20	18	22	5	5	5	4	5	4	5	5	6
	75	100	20	18	23	37	32	43	5	5	5	6	6	5	8	8	10
$\mathcal{L}\gamma(2, 1)$	50	40	30	27	33	26	23	25	5	6	5	4	4	3	11	11	12
	50	100	70	62	73	77	69	79	12	11	13	7	5	6	29	26	36
	75	40	71	64	73	70	63	71	12	11	13	4	3	4	32	29	37
	75	100	98	95	98	99	99	99	36	31	42	8	6	8	77	69	84
$\mathcal{L}\gamma(4, 1)$	50	40	22	19	24	18	16	18	5	5	4	4	3	3	7	7	8
	50	100	51	45	57	58	52	63	8	7	8	6	5	5	17	17	22
	75	40	58	52	62	57	51	60	9	9	10	4	3	4	20	19	24
	75	100	94	89	95	98	96	98	25	22	29	6	5	6	54	47	63

Table 5 Percentage of rejection of tests for normality based on 10,000 replications (part 1)

Distribution	$\frac{r}{n} * 100\,\%$	n	MS			OS			LHB			FK1			FK2			DS	
			A^2	W^2	C^2	A^2	W^2	C^2	A^2	W^2	C^2	A^2	W^2	C^2	A^2	W^2	C^2	$A^2_{r,n}$	$W^2_{r,n}$
$\mathcal{N}(0, 1)$	50	40	6	5	5	6	5	5	5	5	5	3	3	3	5	5	5	5	5
	50	100	5	5	5	6	6	6	5	5	5	4	4	3	5	5	5	5	5
	75	40	6	5	6	6	6	6	5	5	5	4	4	3	5	5	5	5	5
	75	100	5	5	5	6	5	5	5	5	4	4	4	3	5	5	5	5	5
$\mathcal{LN}(0, 1)$	50	40	50	44	53	33	30	32	9	8	10	9	8	8	42	39	41	65	66
	50	100	87	78	87	85	78	86	19	18	21	15	12	13	81	76	83	98	98
	75	40	85	79	85	79	73	79	20	19	22	12	9	11	76	72	78	94	92
	75	100	100	99	100	100	100	100	55	52	61	21	15	17	99	99	99	100	100
$IG(4, 1)$	50	40	75	70	77	60	55	58	16	14	17	12	9	11	65	61	64	88	88
	50	100	99	97	99	99	97	99	45	43	49	21	15	17	98	96	98	100	100
	75	40	99	98	98	98	96	97	49	47	52	20	14	18	96	95	97	100	100
	75	100	100	100	100	100	100	100	93	92	95	29	20	23	100	100	100	100	100
$IG(1, 4)$	50	40	18	16	19	9	9	8	6	6	6	6	5	5	16	15	16	28	30
	50	100	33	28	37	28	25	31	7	6	7	9	8	7	32	28	33	59	62
	75	40	34	30	38	25	23	27	6	6	7	8	6	7	29	26	30	49	46
	75	100	66	56	69	69	62	74	10	10	12	12	10	10	59	53	64	89	87
t_2	50	40	50	47	52	58	55	60	10	10	11	2	2	1	5	4	7	37	27
	50	100	82	78	84	88	86	90	24	22	26	25	15	24	14	8	20	79	74
	75	40	56	53	59	61	58	64	13	12	14	17	11	16	7	5	9	53	51
	75	100	87	85	89	90	88	91	33	29	35	47	33	45	19	11	21	88	87
t_4	50	40	20	18	22	26	23	28	6	6	5	3	3	2	4	4	4	13	8
	50	100	37	32	40	46	41	50	7	6	7	6	5	6	6	5	9	35	28
	75	40	23	20	25	27	25	30	5	5	5	5	4	5	4	4	5	20	18
	75	100	41	36	44	47	42	51	9	8	9	11	7	11	8	6	9	44	41

Table 6 Percentage of rejection of tests for normality based on 10,000 replications (part 2)

Distribution	$\frac{r}{n} * 100\,\%$	n	MS			OS			LHB			FK1			FK2			DS	
			A^2	W^2	C^2	A^2	W^2	C^2	A^2	W^2	C^2	A^2	W^2	C^2	A^2	W^2	C^2	$A^2_{r,n}$	$W^2_{r,n}$
$\mathcal{L}(0, 1)$	50	40	9	9	10	13	12	14	5	5	5	3	3	2	4	4	4	5	4
	50	100	13	11	15	20	17	22	5	5	5	4	4	3	5	5	6	11	8
	75	40	11	10	11	13	11	15	5	5	5	4	4	3	4	4	5	8	7
	75	100	15	13	17	19	16	22	5	5	5	4	4	4	5	5	6	16	14
$Wei(2, 1)$	50	40	14	12	14	8	7	6	5	5	5	5	5	4	13	12	12	20	22
	50	100	24	20	27	20	17	22	6	6	6	8	7	7	24	21	26	46	48
	75	40	17	14	18	12	11	11	5	5	5	6	6	5	15	14	16	27	24
	75	100	33	26	36	34	29	38	6	6	6	8	8	8	29	26	33	59	53
$Exp(1)$	50	40	61	53	60	44	39	40	10	9	10	10	8	8	57	53	57	78	77
	50	100	95	90	94	95	90	94	26	25	30	22	19	19	95	92	96	100	100
	75	40	83	75	81	77	70	74	16	15	17	12	9	10	78	74	80	94	91
	75	100	100	99	99	100	99	100	49	45	55	27	22	25	100	99	100	100	100
$\gamma(2, 1)$	50	40	27	23	29	15	14	13	6	6	6	7	6	5	23	22	23	40	42
	50	100	55	45	58	51	43	54	9	8	10	12	11	10	52	46	55	82	83
	75	40	45	38	47	35	30	36	7	7	7	9	7	7	37	35	40	62	57
	75	100	82	71	83	86	78	88	15	14	17	14	12	12	76	70	80	97	95
$\gamma(4, 1)$	50	40	14	12	15	7	7	5	6	6	5	6	6	5	13	13	13	20	22
	50	100	23	19	26	19	16	20	6	6	6	8	7	6	22	20	23	44	47
	75	40	22	19	24	15	14	15	6	6	6	7	6	5	18	17	19	32	30
	75	100	42	35	47	44	38	50	7	6	7	10	8	9	37	33	41	68	64
$\mathcal{L}\gamma(2, 1)$	50	40	80	74	81	66	61	64	19	17	21	13	10	12	72	68	71	90	90
	50	100	99	98	99	99	98	99	51	49	56	23	17	18	99	98	99	100	100
	75	40	99	99	99	99	98	99	60	58	63	23	17	23	98	97	98	100	100
	75	100	100	100	100	100	100	100	97	96	98	32	23	28	100	100	100	100	100

As hypothetical models, the exponential, gamma and normal distribution have been used, with estimation of parameters carried out by the methods outlined in Sect. 4.

As alternatives we employed the following distributions which are often used in lifetime and failure analysis: Weibull distribution $Wei(\alpha, \beta)$ with density $\frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1} \exp\left(-(\frac{x}{\beta})^{\alpha}\right)$, inverse Gaussian distribution $IG(\mu, \lambda)$ with density $\left(\frac{\lambda}{2\pi x^3}\right)^{1/2} \exp\left(\frac{-\lambda(x-\mu)^2}{2\mu^2 x}\right)$, logarithmic gamma distribution $\mathcal{L}\gamma(\alpha, \beta)$ with density $\frac{\beta^{\alpha}}{\Gamma(\alpha)x^{\beta+1}}(\log x)^{\alpha-1}$, logistic distribution $\mathcal{L}(\alpha, \beta)$ with density $\frac{1}{\beta}(1+\exp(-\frac{x-\alpha}{\beta}))^{-2}\exp(-\frac{x-\alpha}{\beta})$, lognormal distribution $\mathcal{LN}(\mu, \sigma)$ with density $(\sqrt{2\pi}\sigma x)^{-1}\exp(\frac{-(\log x-\mu)^2}{2\sigma^2})$, Student's t-distribution with m degrees of freedom t_m, and the three distributions which were used as hypothetical models.

The censoring proportions considered are 50 % and 25 %, which corresponds to $r/n = 0.50$ and 0.75, respectively, with sample sizes $n = 40$ and $n = 100$. The entries in the tables give the percentage of rejection of the respective hypothesis based on 10,000 repetitions, at nominal level of significance $\alpha = 0.05$. A rejection rate of 100 % is indicated by $*$.

All tests have been also performed at nominal levels of significance $\alpha = 0.01$ and $\alpha = 0.1$. Since however the relative performance was unchanged, the corresponding results are omitted. Likewise, the results for the (very high) censoring proportion of 75 % are omitted, but some of the following remarks also apply to this case.

All simulations have been done using the statistical computing environment R (R Core Team 2012).

5.1 Testing for exponentiality

Simulation results for testing the hypothesis of exponentiality are given in Tables 1 and 2. The conclusions drawn from these results are as follows:

Level: Most procedures maintain the nominal level very well, with the tests based on the MS and FK2 transforms being somewhat conservative.

$\boldsymbol{\gamma(\alpha, \beta)}$ **and** $\boldsymbol{\mathcal{L}\gamma(\alpha, \beta)}$ alternatives: (Fig. 1a, b) Due to the similarity of the gamma to the exponential distribution, detecting this hypothesis is difficult. In fact, transformations MS, OS and FK2 seem completely unsuitable for this purpose. On the other hand, the tests based on the LHB transformation seem to work best against gamma alternatives, while against logarithmic gamma alternatives, the tests based on MS and OS are also competitive.

For the $\mathcal{L}\gamma(2, 1)$ distribution, all tests based on the LHB transformation show an astonishing behaviour: The power is sharply decreasing when r is increasing, i.e. with a lower degree of censoring. The same can be seen for further alternatives. This behaviour of the LHB transform can also be observed when testing a simple hypotheses, see Tables 1, 2, 3 and 4 in Lin et al. (2008): there, the power of the Anderson-Darling test combined with the LHB transform, called $\underset{T}{*}A_r^2$, decreases with r for several alternatives (in particular for alternatives F_{41} and F_{51} which are defined on p. 634 in Lin et al. 2008). Similar patterns also occur with other transformations, see the results for the Anderson-Darling test combined with the

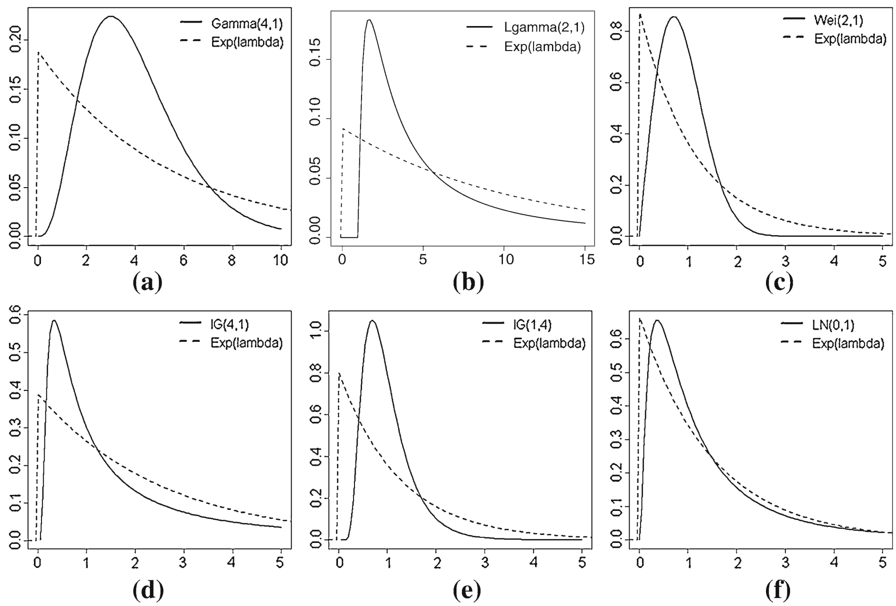

Fig. 1 Densities of some alternative distributions compared with the density of a fitted exponential distribution. Parameter estimation is based on a sample from the corresponding alternative with $n = 100$ and $r = 75$. **a** $\gamma(4, 1)$, **b** $\mathcal{L}\gamma(2, 1)$, **c** $Wei(2, 1)$, **d** $IG(4, 1)$, **e** $IG(1, 4)$, **f** $\mathcal{LN}(0, 1)$

MS transform, called ${}_T A_r^2$, in Lin et al. (2008), and Table 1 in O'Reilly and Stephens (1988) for tests based on the OS transformation.

$\boldsymbol{Wei(\alpha, \beta)}$ alternatives: (Fig. 1c) Again, tests based on the LHB transform are the best, followed by FK1, while the tests based on the FK2 transformation is clearly unable to detect the Weibull alternatives used in the simulations.

$\boldsymbol{IG(\mu, \lambda)}$ and $\boldsymbol{\mathcal{LN}(\mu, \sigma)}$ alternatives: (Fig. 1d, e, f) For these distributions, the tests based on FK1 and FK2 do not work. Interestingly, the MS and OS-based tests work much better against $IG(4, 1)$ than LHB, while the converse holds true for the $IG(1, 4)$ distribution. The power results against the $\mathcal{LN}(0, 1)$ distribution comply with the fact that this distribution is hard to distinguish from a suitable exponential distribution (see Fig. 1f).

Summary: The tests based on the transformation of Lin, Huang and Balakrishnan should be used, while the tests based on FK1 and FK2 do not work well. Also within one transformation, the difference between the four test statistics is not always particularly noticeable. Nevertheless, in many cases the characteristic function based test has slightly higher power than the tests based on the empirical distribution function.

In the simulations for the exponential distribution, we added results for the two most common direct statistics (DS), i.e., statistics which are modifications of corresponding full–sample versions and may be applied directly to the original censored data, without transformations. These tests are the Cramér-von Mises and the Anderson-Darling test; see Sect. 4.9.5 of D'Agostino and Stephens (1986). Corre-

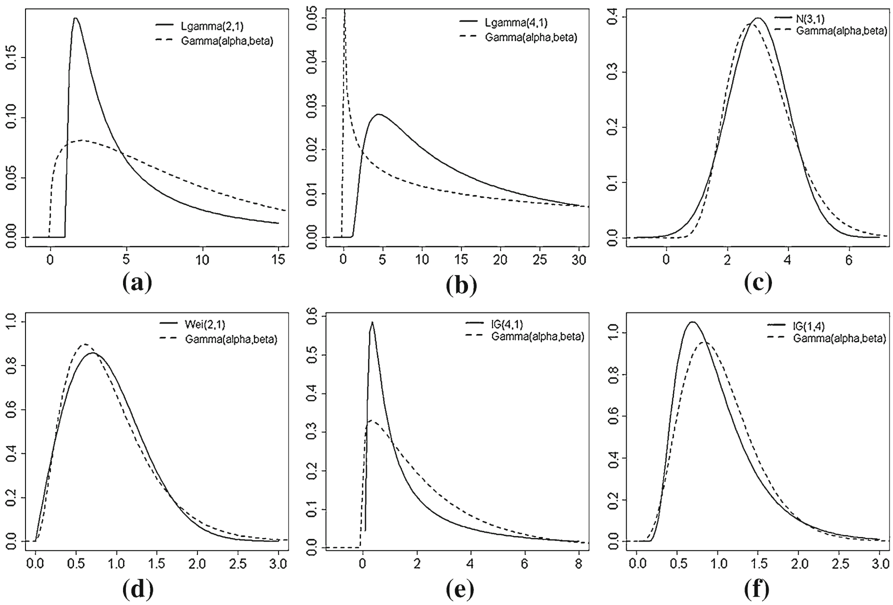

Fig. 2 Densities of some alternative distributions compared with the density of a fitted gamma distribution. Parameter estimation is based on a sample from the corresponding alternative with $n = 100$ and $r = 75$ (**a**) $\mathcal{L}\gamma(2, 1)$ (**b**) $\mathcal{L}\gamma(4, 1)$ (**c**) $\mathcal{N}(3, 1)$ (**d**) $Wei(2, 1)$ (**e**) $IG(4, 1)$ (**f**) $IG(1, 4)$

sponding results are given in obvious notation in the last two columns in Tables 1 and 2. Generally, the newly proposed tests have inferior power. Recall however their advantage of general applicability: We do not need new critical values for each distribution, sample size and censoring proportion, not to mention dependence on parameters and corresponding estimators. In this connection note that since even for this standard distribution critical values for our censoring proportions 25 % and 50 % are not available in the literature, they are provided in Table 1 of the TR.

5.2 Testing for gamma distribution

Simulation results for testing the gamma hypothesis are given in Tables 3 and 4. The conclusions drawn from these results are as follows:

Level: The tests based on the OS and LHB transformations maintain the nominal level very well, while those based on the MS, FK1 and FK2 are somewhat conservative.

$\boldsymbol{\mathcal{L}\gamma(\alpha, \beta)}$ alternatives: (Fig. 2a, b) All variants of MS and OS based tests have good power against the logarithmic gamma distribution, while the FK1-based tests are worst in this case.

$\boldsymbol{\mathcal{N}(3, 1)}$ alternative: (Fig. 2c) The tests based on the MS and the OS transformation (in this order) have higher powers against the $\mathcal{N}(3, 1)$ distribution.

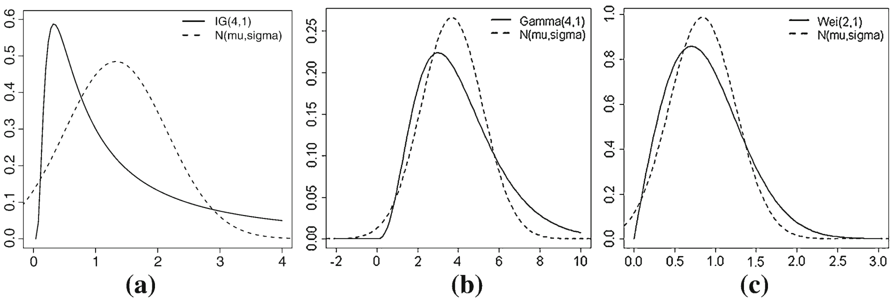

Fig. 3 Densities of some alternative distributions compared with the density of a fitted normal distribution. Parameter estimation is based on a sample from the corresponding alternative with $n = 100$ and $r = 75$ **a** $IG(4, 1)$, **b** $\gamma(4, 1)$, **c** $Wei(2, 1)$

$\boldsymbol{Wei(\alpha, \beta)}$ alternatives: (Fig. 2d) Given the fact that Weibull distributions can be approximated quite well by suitable gamma distributions it is not suprising to see that power is low, uniformly over all tests and transformations. Clearly however the OS-based tests stand out as best.
$\boldsymbol{IG(\mu, \lambda)}$ and $\boldsymbol{\mathcal{LN}(\mu, \sigma)}$ alternatives: (Fig. 2e, f) The $\mathcal{LN}(0, 1)$ and the $IG(1, 4)$ distributions can be well fitted by a gamma distribution. Hence, power against these alternatives is generally low. Against the $IG(4, 1)$ alternative the tests based on MS and OS work best.
Summary: several of the alternatives are difficult to distinguish from a fitted gamma distribution. The best results are observed with the tests based on the transformations of Michael and Schucany and O'Reilly and Stephens, and within these transformations the test based on the characteristic function have a certain edge.

5.3 Testing for normality

Simulation results for testing the hypothesis of normality using the estimates suggested by Gupta (1952) are given in Tables 5 and 6. The conclusions drawn from these results are as follows:

Level: Apart from the tests based on FK1, which are somewhat conservative, all tests maintain the nominal level very well.
$\boldsymbol{IG(\mu, \lambda)}$ and $\boldsymbol{\mathcal{LN}(\mu, \sigma)}$ alternatives: (Fig. 3a) For these alternatives, the MS, OS and FK2–based tests have higher powers than LIN and FK1-based tests.
$\boldsymbol{t_m}$ alternatives: The tests based on MS and OS transforms give the best results, with powers being significantly lower for the t_4 alternative.
$\boldsymbol{Exp(\lambda)}$ alternative: The tests based on the MS, OS and FK2 transformations detect this alternative reliably for medium and low censoring ($r \geq n/2$) and are clearly preferable to tests based on LHB and FK1.
$\boldsymbol{\gamma(\alpha, \beta)}$ and $\boldsymbol{\mathcal{L}\gamma(\alpha, \beta)}$ alternatives: (Fig. 3b) The tests based on MS, OS and FK2 transformation have high power against logarithmic gamma distributions for medium and low censoring. They are also preferable against gamma alternatives.

***Wei*(α, β)** alternatives: (Fig. 3c) Due to the similarity of the $Wei(3, 1)$ and $Wei(3, 5)$ and suitable normal densities, power is only slightly above the nominal level.

$\mathcal{L}(\alpha, \beta)$ alternatives: Again, this alternative is hard to distinguish from a normal distribution. Nevertheless, MS and OS based tests show some power.

Summary: The highest power is observed with the tests based on the transformations of Michael and Schucany and O'Reilly and Stephens. However, the tests based on transformation FK2 also show a comparable behavior. As in the case of the exponential distribution, we added results for the direct statistics (DS), the Cramér-von Mises and the Anderson-Darling test; see Sect. 4.8.4 of D'Agostino and Stephens (1986) for the normal distribution with censored data. Corresponding results are given in the last two columns in Tables 5 and 6. Critical values for our censoring proportions 25 % and 50 % and sample sizes $n = 40$ and 100 are provided in Table 2 of the TR. All comments given at the end of Sect. 5.1 for the exponential case also apply here, although the transformed–based tests are more competitive in this case.

We repeated the simulations using modified maximum likelihood estimation as suggested by Tiku (1967). The results have been quite similar but power was slightly worse compared to the method of Gupta. Therefore, the results have been omitted.

5.4 Summary of simulation results

In this subsection, we try to convey a qualitative message based on the simulation results. Table 7 has the three tests as entries for the lines, and the five transformations to uniformity as entries for the columns. At each cell, i.e. at each specific combination of test and transformation, there are two entries which show order: The left digit is the order w.r.t. the lines (tests), while the right digit shows the order w.r.t. the columns (transformations). For example the entry 2\1 in the first cell means that the Anderson-Darling test is the second best for the MS transformation, while the MS transformation ranks 1st for the Anderson–Darling test.

To get such an overall assessment for, say, a fixed transformation, we assigned ranks to the three tests. The test with the highest percentage of rejection has rank one, and so on. Then, we summed up the ranks for all listed combinations of null hypotheses,

Table 7 Overall assessment for the tests and transformations. Left digit: order w.r.t. the lines (tests). Right digit: order w.r.t. the columns (transformations)

Test \ Transformation	MS	OS	LHB	FK1	FK2
A^2	2**1**	2\\2	2\\3	**1**\\5	2\\4
W^2	3\\2	3**1**	3\\3	3\\5	3\\4
C^2	**1**\\2	**1****1**	**1**\\3	2\\5	**1**\\4

alternatives and censoring proportions r. The test with the lowest sum score has the entry one in Table 7 for this transformation.

This overview gives a clear picture: For any given test, the MS and OS transformations perform best. In this connection, a direct look at the sum scores shows that there is not much to choose between these two transforms. On the other hand, the LHB transform follows at a clear distance, having the edge over FK2. By far the lowest sum score for all tests has the FK1 transformation. Also, for the MS and OS transformations, the best test is the characteristic function test C^2. C^2 also performs best for LHB and FK2 transform. Finally, the best combination of test and transformation is C^2 with OS.

6 Conclusion

We have applied a series of transformations to original type-II censored data with the aim of rendering corresponding full-sample test statistics for the parent population, applicable and approximately distribution-free. The conclusions from our Monte Carlo study show that these transformations generally work well across all goodness-of-fit tests studied in terms of recovering the nominal level of significance. On the other hand, the best transformation in terms of power depends on the distribution under test, with the transformation of Lin et al. (2008) being best for testing exponentiality, and the transformations of Michael and Schucany (1979) and O'Reilly and Stephens (1988) for testing normality and testing for the gamma distribution. Also, and within each transformation, the characteristic function based test seems to yield the best power for the majority of alternatives. However, this superiority is generally not significant compared to the large differences in power corresponding to different transformations.

Before closing however we wish to reiterate once more that, despite the fact that the transformed based tests often show good power, their advantage lies not so much in the power but in the general applicability of the method: The user does not need new critical values for each distribution under test, each combination of sample size and censoring proportion, and each possible choice of parameter value/parameter estimate, but, following the transformation suggested, essentially faces the simplified problem of testing for normality with estimated parameters.

7 Appendix

In this appendix we shall investigate the reasons underlying the eventual validity of the Chen–Balakrishnan transformation. In doing so, first in Sect. 7.1 of the appendix we report results on the process corresponding to goodness-of-fit testing for the normal distribution with estimated parameters. Then in Sects. 7.2 and 7.3 of the appendix we study in detail the process produced by the Chen–Balakrishnan transformation and compare it with the process in Sect. 7.1 both theoretically and by simulation. The figures referred to in this appendix can be retrieved from the technical report (TR) at `arXiv:1312.3078 [stat.ME]`.

7.1 The empirical process under normality

Suppose that $Z_j, \ j = 1, \ldots, n$, are iid normal with unknown mean and variance. Then most standard goodness-of-fit tests are merely functionals of the empirical process

$$\hat{\alpha}_n(t) = \frac{1}{\sqrt{n}} \sum_j \left[I \left\{ \Phi \left(\frac{Z_j - \bar{Z}}{s_Z} \right) \leq t \right\} - t \right]$$
$$= \frac{1}{\sqrt{n}} \sum_j \left[I \left\{ U_j \leq \Phi \left(\bar{Z} + s_Z \, \Phi^{-1}(t) \right) \right\} - t \right],$$

where $\bar{Z}$ and s_Z^2 are the sample mean and sample variance of $Z_1, \ldots, Z_n$, and $U_j = \Phi(Z_j)$ and $t \in [0, 1]$. This process has been studied by Durbin (1973) and showed that under regularity conditions,

$$\hat{\alpha}_n \Rightarrow \alpha$$

where α is a centered Gaussian process with covariance function

$$C(\alpha(s), \alpha(t)) = \min(s, t) - st - \varphi(\Phi^{-1}(s))\varphi(\Phi^{-1}(t))$$
$$-\frac{1}{2}\Phi^{-1}(s)\varphi(\Phi^{-1}(s))\Phi^{-1}(t)\varphi(\Phi^{-1}(t)),$$

where Φ^{-1} and φ are the quantile and density function of the standard normal distribution (We note that Loynes (1980) extended the analysis from the iid setting to the case of generalized linear models). *Clearly the process α_n is identical to the process involved in the Chen–Balakrishnan transformation only in the case of testing for normality with estimated parameters.*

7.2 The empirical process under non-normality

In this section, we consider iid random variables X_j with DF $\mathcal{F}_\vartheta(x)$ (assumed to be continuous and strictly increasing) and the standardized quantile residuals $Z_j = \frac{Y_j - \bar{Y}}{s_Y}$ with $Y_j = \Phi^{-1}\left(\mathcal{F}_{\hat{\vartheta}}(X_j)\right)$ (concerning the term *standardized quantile residual*, refer to Klar and Meintanis (2012), sec 2.1). We shall study the following empirical process based on the Z_j:

$$\begin{aligned}
\hat{\beta}_n(t) &= \frac{1}{\sqrt{n}} \sum_j \left[I\left\{ \Phi(Z_j) \le t \right\} - t \right] \\
&= \frac{1}{\sqrt{n}} \sum_j \left[I\left\{ \Phi\left(\frac{Y_j - \bar{Y}}{s_Y} \right) \le t \right\} - t \right] \\
&= \frac{1}{\sqrt{n}} \sum_j \left[I\left\{ X_j \le \mathcal{F}_{\hat{\vartheta}}^{-1}(\Phi(\bar{Y} + s_Y\, \Phi^{-1}(t))) \right\} - t \right] \\
&= \frac{1}{\sqrt{n}} \sum_j \left[I\left\{ U_j \le \mathcal{F}_{\vartheta}\left(\mathcal{F}_{\hat{\vartheta}}^{-1}\left(\Phi(\bar{Y} + s_Y\, \Phi^{-1}(t)) \right) \right) \right\} - t \right],
\end{aligned}$$

where $\mathcal{F}_{\vartheta}^{-1}(p)$ denotes the quantile function of $\mathcal{F}_{\vartheta}(\cdot)$, and $U_j = \mathcal{F}_{\vartheta}(X_j)$ are iid uniformly distributed on [0, 1]. *This is the empirical process actually produced by the Chen–Balakrishnan transformation*. Now define $c_Y(t) = \Phi(\bar{Y} + s_Y \Phi^{-1}(t))$, and, similarly, $c_N(t) = \Phi(\bar{N} + s_N \Phi^{-1}(t))$, where $N_j = \Phi^{-1}(U_j)$ are iid standard normal random variates, and $\bar{N}$ and s_N^2 are the arithmetic mean and sample variance of $N_1, \ldots, N_n$.

Then we can decompose the above process as (compare Chen 1991, pp 126-128)

$$\hat{\beta}_n(t) = \hat{\beta}_{n,1}(t) + \hat{\beta}_{n,2}(t) + \hat{\beta}_{n,3}(t), \tag{7.1}$$

where

$$\begin{aligned}
\hat{\beta}_{n,1}(t) &= \frac{1}{\sqrt{n}} \sum_j \Big[I\left\{ U_j \le \mathcal{F}_{\vartheta}\left(\mathcal{F}_{\hat{\vartheta}}^{-1}(c_Y(t)) \right) \right\} - \mathcal{F}_{\vartheta}\left(\mathcal{F}_{\hat{\vartheta}}^{-1}(c_Y(t)) \right) \\
&\qquad -I\left\{ U_j \le c_N(t) \right\} + c_N(t) \Big], \\
\hat{\beta}_{n,2}(t) &= \frac{1}{\sqrt{n}} \sum_j \left[I\left\{ U_j \le \Phi(\bar{N} + s_N\, \Phi^{-1}(t)) \right\} - t \right],
\end{aligned}$$

and

$$\hat{\beta}_{n,3}(t) = \frac{1}{\sqrt{n}} \sum_j \left[\mathcal{F}_{\vartheta}\left(\mathcal{F}_{\hat{\vartheta}}^{-1}(c_Y(t)) \right) - c_N(t) \right].$$

The first part $\hat{\beta}_{n,1}(t)$ in decomposition (Sect. 7.1) is the difference of an empirical process and a random perturbation thereof, and should be $o_P(1)$ under appropriate regularity conditions (see Chen 1991, p.128; Loynes (1980), Lemma 1; and Rao and Sethuraman 1975). To check this claim, we simulated a random sample of size n from a unit mean exponential distribution and computed $\hat{\beta}_{n,1}(t),\ t \in [0, 1]$, on the basis of an equidistant grid with spacing equal to 0.005. This was repeated $B = 10000$ times. We approximated the mean function $E[\hat{\beta}_{n,1}(t)]$ and the standard deviation $\sqrt{Var[\hat{\beta}_{n,1}(t)]}$ by the arithmetic mean and empirical standard deviation based on the B replications;

Figure 4 in the TR shows the result for sample sizes $n = 10, 40, 160$ and 640. Clearly the mean function is nearly zero and decreases for increasing n, while the standard deviation is small compared to the standard deviation of $\hat{\beta}_n$ or $\hat{\beta}_{n,2}$ (see below). The corresponding variance seems to converge to zero, but rather slowly, with a speed of convergence approximately equal to $1/\sqrt{n}$.

The second part $\hat{\beta}_{n,2}$ corresponds to the normal empirical process $\hat{\alpha}_n$ in Sect. 7.1 of the appendix. Figure 5 in the TR shows the empirical mean function and standard deviation of $\hat{\beta}_{n,2}$ for an underlying exponential distribution computed in the same way as for $\hat{\beta}_{n,1}$ above. The mean function, which takes on much larger values than that of $\hat{\beta}_{n,1}$, again converges to zero, whereas the variance function is nearly constant for $n \geq 40$.

For the third part we have

$$\hat{\beta}_{n,3}(t) = \sqrt{n}\left[\mathcal{F}_{\vartheta}\left(\mathcal{F}_{\hat{\vartheta}}^{-1}(\Phi(\bar{Y} + s_Y\,\Phi^{-1}(t)))\right) - \Phi(\bar{N} + s_N\,\Phi^{-1}(t))\right].$$

In general, this process does not converge to zero in probability. However, the contribution of $\hat{\beta}_{n,3}$ seems to be negligibly small *in comparison to* $\hat{\beta}_{n,2}$ in many situations. Figure 6 in the TR shows the empirical mean function and standard deviation of $\hat{\beta}_{n,3}$ for the exponential distribution, computed as above. The mean function is very small and goes to zero. The standard deviation is small compared to the standard deviation of $\hat{\beta}_{n,2}$, and it converges, but not to zero. We stress that the crucial point for the behavior of $\hat{\beta}_{n,3}$ is the coupling between the Y_j's and the normal variates N_j which are both based on the original X_j, the first computed by using $\hat{\vartheta}$ while the second by using ϑ. In fact if we generate iid standard normal random variables $\tilde{N}_j$ independent of the X_j's and use them instead of the N_j's, the mean function is small, but does not seem to converge to zero, and the variance is much larger, even larger than that of $\hat{\beta}_{n,2}$.

From the above it follows that the values of the process $\hat{\beta}_n$ will be eventually dominated by $\hat{\beta}_{n,2}$, at least for large n. This is documented in Figure 7 of the TR where the mean and standard deviation of all four processes are plotted for $n = 40$. Note that the standard deviations of $\hat{\beta}_n$ (in red) and $\hat{\beta}_{n,2}$ (in green) are nearly identical, and therefore, visually indistinguishable.

7.3 Further analysis of the process $\hat{\boldsymbol{\beta}}_{n,3}$

To keep things simple, we assume in the following that $\vartheta \in \Theta \subset \mathbb{R}$. Let ϑ_0 denote the true parameter value, and define

$$N_j(\vartheta) = \Phi^{-1}\left(\mathcal{F}_{\vartheta}\left(X_j\right)\right),$$
$$\bar{N}(\vartheta) = \frac{1}{n}\sum_{j=1}^{n} N_j(\vartheta), \quad s_N^2(\vartheta) = \frac{1}{n-1}\sum_{j=1}^{n}\left(N_j(\vartheta) - \bar{N}(\vartheta)\right)^2.$$

Then, $N_j(\vartheta_0) = N_j, \bar{N}(\vartheta_0) = \bar{N}, s_N^2(\vartheta_0) = s_N^2$, and $N_j(\hat{\vartheta}) = Y_j, \bar{N}(\hat{\vartheta}) = \bar{Y}, s_N^2(\hat{\vartheta}) = s_Y^2$. Putting

$$h_t(\vartheta) = \Phi\left(\bar{N}(\vartheta) + s_N(\vartheta)\cdot\Phi^{-1}(t)\right),$$

we obtain $h_t(\vartheta_0) = c_N(t)$ and $h_t(\hat{\vartheta}) = c_Y(t)$. Thus, we can write

$$\hat{\beta}_{n,3}(t) = \sqrt{n}\left(g_t(\hat{\vartheta}) - g_t(\vartheta_0)\right), \tag{7.2}$$

where

$$g_t(\vartheta) = \mathcal{F}_{\vartheta_0}\left(\mathcal{F}_{\vartheta}^{-1}\left(h_t(\vartheta)\right)\right).$$

Assume now that $\sqrt{n}(\hat{\vartheta} - \vartheta_0) = O_p(1)$. Then, by using the expansion

$$g_t(\hat{\vartheta}) = g_t(\vartheta_0) + (\hat{\vartheta} - \vartheta_0)\, g_t'(\vartheta_0) + (\hat{\vartheta} - \vartheta_0)^2\, g_t''(\vartheta^*)/2,$$

with ϑ^* between $\hat{\vartheta}$ and ϑ_0, and by omitting the quadratic term, we see that $\hat{\beta}_{n,3}(t)$ can be approximated by

$$\mathring{\beta}_{n,3}(t) = \sqrt{n}\left(\hat{\vartheta} - \vartheta_0\right)\, g_t'(\vartheta_0). \tag{7.3}$$

Of course, the validity of such a Taylor expansion is not enough to justify the uniform convergence $\sup_t |\hat{\beta}_{n,3}(t) - \mathring{\beta}_{n,3}(t)| = o_P(1)$. A sufficient condition would be Fréchet differentiability of $g_t(\cdot)$ (see, e.g. van der Vaart and Wellner (2002), p. 373). However, since we do not intend to give rigorous theory here, this issue is not discussed in any detail. Further analysis of $g_t'(\vartheta_0)$ leads to the following result, the proof of which is omitted.

Lemma 7.1 *Let $\bar{W}$ and s_W^2 denote the arithmetic mean and sample variance of the random variables $W_{j0} := W_j(\vartheta_0)$, with $W_j(\vartheta) := dN_j(\vartheta)/d\vartheta$, while r denotes the sample correlation coefficient of $W_{10}, \ldots, W_{n0}$ and $N_1, \ldots, N_n$. Then,*

$$g_t'(\vartheta_0) = \frac{\partial\mathcal{F}_{\vartheta_0}\left(F_{\vartheta_0}^{-1}(c_N(t))\right)}{\partial x}\cdot\left(\frac{\partial\mathcal{F}_{\vartheta_0}^{-1}(c_N(t))}{\partial p}\cdot h_t'(\vartheta_0) + \frac{\partial\mathcal{F}_{\vartheta_0}^{-1}(c_N(t))}{\partial\vartheta}\right),$$

$$h_t'(\vartheta_0) = \varphi\left(\bar{N} + s_N\,\Phi^{-1}(t)\right)\cdot\left(\bar{W} + \Phi^{-1}(t)\, r\, s_W\right),$$

where $\varphi(\cdot)$ denotes the density of the standard normal distribution.

Since $N_1, \ldots, N_n$ are iid standard normal variates, $\bar{N} \to 0$ and $s_N \to 1$ almost surely. Furthermore, $\bar{W} \to \mu_W$, $s_W \to \sigma_W$, and $r \to \rho$ a.s., where (μ_W, σ_W^2) are the mean and variance of W_{10}, while ρ denotes the correlation coefficient of W_{10} and N_1. Hence, the following approximation holds for the process in (7.3).

Lemma 7.2 *The process $\mathring{\beta}_{n,3}(t)$ can be approximated by the process*

$$\tilde{\beta}_{n,3}(t) = \sqrt{n}\left(\hat{\vartheta} - \vartheta_0\right)\, \tilde{g}_t'(\vartheta_0), \tag{7.4}$$

where

$$\tilde{g}_t'(\vartheta_0) = \frac{\partial \mathcal{F}_{\vartheta_0}\left(\mathcal{F}_{\vartheta_0}^{-1}(t)\right)}{\partial x} \cdot \left(\frac{\partial \mathcal{F}_{\vartheta_0}^{-1}(t)}{\partial t} \cdot \tilde{h}_t'(\vartheta_0) + \frac{\partial \mathcal{F}_{\vartheta_0}^{-1}(t)}{\partial \vartheta} \right),$$

$$\tilde{h}_t'(\vartheta_0) = \varphi\left(\Phi^{-1}(t)\right) \cdot \left(\mu_W + \Phi^{-1}(t)\, \rho\, \sigma_W\right).$$

Figure 8 in the TR shows the simulated mean and standard deviation of $\hat{\beta}_{n,3}$ in (7.2), $\mathring{\beta}_{n,3}$ in (7.3), and $\tilde{\beta}_{n,3}$ in (7.4) for sample size $n = 40$ and $n = 640$, again for the exponential distribution. The mean functions take on very small values; the standard deviations are very similar in all cases.

Figure 9 of the TR shows the function $\tilde{h}_t'(\vartheta_0)$, the part inside the brackets in $\tilde{g}_t'(\vartheta_0)$, and $\tilde{g}_t'(\vartheta_0)$ itself. The values of $\tilde{g}_t'(\vartheta_0)$ are close to zero on the whole interval. For this reason, $\hat{\beta}_{n,3}$ is negligible in comparison to $\hat{\beta}_{n,2}$ for the exponential case at hand.

We also performed Monte Carlo experiments for other gamma distributions with shape parameter not equal to one. These experiments lead to qualitatively similar results and although not reported here they are available from the authors upon request. A reasonable overall conclusion seems to be that under different sampling scenarios the processes $\hat{\beta}_{n,1}$ and $\hat{\beta}_{n,3}$ in decomposition (Sect. 7.1) are asymptotically negligible, and hence the behavior of the process $\hat{\beta}_n$ of the Chen–Balakrishnan transformation is dominated by the values of the process $\hat{\beta}_{n,2}$. The latter process however coincides with the process $\hat{\alpha}_n(t)$ in Sect. 7.1 of the appendix which is involved in goodness-of-fit testing for normality with estimated parameters, and this fact justifies the validity of the Chen–Balakrishnan transformation.

References

Arcones MA, Wang Y (2006) Some new tests for normality based on U-processes. Stat Probab Lett 76:69–82

Bargal AI, Thomas DR (1983) Smooth goodness of fit tests for the Weibull distribution with singly censored data. Commun Stat Theory Methods 12:1431–1447

Baringhaus L, Danschke R, Henze N (1989) Recent and classical tests for normality. A comparative study. Commun Stat Simul Comput 18:363–379

Barr DR, Davidson Teddy (1973) A Kolmogorov–Smirnov test for censored samples. Technometrics 15:739–757

Brain CW, Shapiro SS (1983) A regression test for exponentiality. Technometrics 25:69–76

Castro-Kuriss C (2011) On a goodness-of-fit test for censored data from a location-scale distribution with applications. Chil J Stat 2:115–136

Castro-Kuriss C, Kelmansky DM, Leiva V, Martínez EJ (2010) On a goodness-of-fit test for normality with unknown parameters and type-II censored data. J Appl Stat 37:1193–1211

Chen, G (1991) Empirical processes based on regression residuals: Theory and Application. Ph.D. Thesis-Simon Fraser University

Chen G, Balakrishnan N (1995) A general purpose approximate goodness-of-fit test. J Qual Technol 27:154–161

D'Agostino R, Massaro JM (1992) Goodness-of-fit tests. In: Balakrishnan N (ed) Handbook of the logistic distribution. Marcel Dekker Inc, New York, pp 291–371

D'Agostino R, Stephens M (1986) Goodness-of-fit techniques. Marcel Dekker Inc, New York

Dufour R, Maag UR (1978) Distribution results for modified Kolmogorov–Smirnov statistics for truncated or censored samples. Technometrics 20:29–32

Durbin J (1973) Weak convergence of the sample distribution function when parameters are estimated. Ann Stat 1:279–290
Epps TW (2005) Tests for location-scale families based on the empirical characteristic function. Metrika 62:99–114
Epps TW, Pulley LB (1983) A test for normality based on the empirical characteristic function procedures. Biometrika 70:723–726
Fischer T, Kamps U (2011) On the existence of transformations preserving the structure of order statistics in lower dimensions. J Stat Plan Inference 141:536–548
Glen AG, Foote BL (2009) An inference methodology for life tests with complete samples or type-II right censoring. IEEE Trans Reliab 58:597–603
Grané A (2012) Exact goodness-of-fit tests for censored data. Ann Inst Stat Math 64:1187–1203
Gupta AK (1952) Estimation of the mean and standard deviation of a normal population from a censored sample. Biometrika 39:266–273
Henze N (1990) An approximation to the limit distribution of the Epps–Pulley test statistic for normality. Metrika 37:7–18
Henze N, Wagner T (1997) A new approach to the BHEP tests for multivariate normality. J Multivar Anal 62:1–23
Klar B, Meintanis SG (2012) Specification tests for the response distribution in generalized linear models. Comput Stat 27:251–267
Johnson NL, Kotz S, Balakrishnan N (1994) Continuous univariate distributions, vol 1. Wiley, New York
Lin C-T, Huang Y-L, Balakrishnan N (2008) A new method for goodness-of-fit testing based on type-II censored samples. IEEE Trans Reliab 57:633–642
Loynes RM (1980) The empirical distribution function of residuals from generalised regression. Ann Stat 8:285–298
Meintanis SG (2009) Goodness-of-fit testing by transforming to normality: comparison between classical and characteristic function-based methods. J Stat Comput Simul 79:205–212
Michael JR, Schucany WR (1979) A new approach to testing goodness of fit for censored samples. Technometrics 21:435–441
Mihalko DP, Moore DS (1980) Chi-square tests of fit for type-II censored data. Ann Stat 8:625–644
O'Reilly FJ, Stephens MA (1988) Transforming censored samples for testing fit. Technometrics 30:79–86
Pettitt AN (1976) Cramér-von Mises statistics for testing normality with censored samples. Biometrika 63:475–481
Pettitt AN (1977) Tests for the exponential distribution with censored data using the Cramér-von Mises statistics. Biometrika 64:629–632
Peña EA (1995) Residuals from type II censored samples. In: Balakrishnan N (ed) Recent advances in life-testing and reliability. CRC Press, London, pp 523–543
R Core Team (2012) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria
Rao JS, Sethuraman J (1975) Weak convergence of empirical distribution functions of random variables subject to perturbations and scale factors. Ann Stat 3:299–313
Tenreiro C (2009) On the choice of the smoothing parameter for the BHEP goodness-of-fit test. Comput Stat Data Anal 53:1038–1053
Thode HC (2002) Testing for normality. Marcel Dekker Inc, New York
Tiku ML (1967) Estimating the mean and standard deviation from a censored normal sample. Biometrika 54:155–165
Tiku ML, Tan WY, Balakrishnan N (1986) Robust inference. Marcel Dekker Inc, New York
van der Vaart AW, Wellner JA (2002) Weak convergence and empirical processes. Springer, New York
Wilk MB, Gnanadesikan R, Huyett MJ (1962) Estimation of parameters of the gamma distribution using order statistics. Biometrika 49:525–545