Selecting Models in both Location-Scale and Non-Location-Scale Families with Goodness-of-Fit Techniques

Claudia Castro-Kuriss*
Dto. Matematicas, Instituto Tecnologico de Buenos Aires, Argentina - ccastro@itba.edu.ar

V ctor Leiva
Facultad de Ingenier a y Ciencias, Universidad Adolfo Ibaez, Chile - victorleivasanchez@gmail.com

Abstract

Goodness-of- t (GOF) techniques are frequently used for assessing whether a distribution is appropriate to describe a data set or not. New distributions have been proposed to enhance the properties of the well-known models, but such distributions in general do not belong to the location-scale (LS) family. This is the case of the Birnbaum-Saunders and the generalized inverse Gaussian distributions. GOF tests for distributions in the non-location-scale (NLS) family are not easy to nd. For testing distributions in the LS family, the well-known Kolmogorov-Smirnov and Anderson-Darling statistics are frequently used. Therefore, it is important to have new tests or adapt the existing ones to the new variety of distributions now available considering NLS distributions. In the present work, we present GOF tests for distributions in both LS and NLS families based on di erent statistics and discuss the possibility of selecting models based on hypothesis testing considering such distributions. We apply these tests and model selection to real-world data sets.

Keywords: GOF tests; Non-Location-Scale Family.

1. Introduction

Goodness-of- t (GOF) tests have been developed for establishing the tting of a distribution to a data set coming from di erent areas. In particular, in reliability analysis, parametric life distributions are commonly used. Among these life distributions, we can mention the classical Birbaum-Saunderns (BS), gamma, inverse Gaussian (IG), log-normal (LN), generalized BS (GBS), truncated BS (TBS), truncated normal (TN), truncated Weibull (TW) and Weibull, among others; see Lawless (2003) and Marshall & Olkin (2007). GOF tests establish if the null hypothesis H_0 cannot be rejected based on empirical evidence.There are mainly two possible options:(1)the distribution under H_0 is completely speci ed and (2)some or all the parameters of the distribution are unknown. The most common case is the second one, where it is necessary to nd a distribution in the family proposed under H_0 by means of proper parameter estimates. We propose to use the maximum likelihood (ML) method to estimate the corresponding parameters. However, problems of existence and convergence may arose; see Castillo & Puig (1997). GOF tests can be of di erent types. We focus on those based on statistics that measure the distance between the empirical cumulative distribution function (ECDF) and the theoretical cumulative distribution function (CDF) established under H_0. Among these statistics, the most used and known are Anderson-Darling (AD), Cramer-von Mises (CM), Kolmogorov-Smirnov (KS) and Michael (MI); see D'Agostino & Stephens (1986). The KS and MI statistics can be related to graphical plots, which allow us to see how well the speci ed theoretical distribution ts a data set. Such graphs are the quantile-quantile (QQ) and probability-probability (PP) plots; see Marden(2004). As the points of the PP plot associated with the KS test may have a high variability, Michael(1983) proposed the stabilized probability (SP) plot, whose MI statistic is based on a modi cation of the KS statistic using an arcsin transformation. Basically, Michael (1983) studied a completely speci ed distribution under the null hypothesis and compared the power of the MI test with the KS test, for di erent distributions in the alternative hypothesis. Michael (1983) proved that the MI test is more powerful than the KS test in certain cases. Some authors also compared the power of AD, CM, KS, MI and other tests for testing normality; see e.g. Castro-Kuriss et al.(2010) and Stelh k et al.(2014).

Other interesting and common situation in real-world applied statistics is found when not all the individuals under test experienced the event of interest before the end of the study. Samples with that kind of data are

called censored. In these cases and when a parametric model is utilized, it is useful to have inferential and graphical GOF procedures to validate the model; see Lawless (2003) for definition of censored schemes and types of censoring. The mentioned tests need to be adapted to the special case of censored samples. For example, one option consists of adapting the statistic of the GOF test to censored data. Castro-Kuriss et al. (2010) proposed GOF tests with type II censored samples for the normal and log-normal distributions. GOF tests for the LS family and different censored schemes can be found in Balakrishnan et al. (2004), Lin et al. (2008) and Pakyari & Balakrishnan (2012). Castro-Kuriss (2011) proposed GOF tests for LS distributions with type-II right censored data and unknown parameters. Some new available GOF tests for the NLS family can be considered as derivations from the proposal of Chen & Balakrishnan (1995), which were also extended to type II censored samples; see Barros et al.(2014) and Castro-Kuriss et al. (2014).
A well-known approach of doing model selection is based on the Kullback-Leibler (KL) "distance" between the probability distribution of two models. Another famous criterion is the Akaike's information criterion (AIC), that also provided a tool for model selection. The result obtained by this or similar methods may differ from the selection based on hypothesis testing. See Burnham & Anderson (2002) and Rad et al. (2011).

In this work, by means of graphical and theoretical methods based on GOF tests for both LS and NLS families, we propose to select models among those that can be possible candidates to describe a data set. We focus on some truncated distributions taking into account data from the literature; see Cohen (1991), Ahmed et al. (2010) and Leiva et al. (2012).
The article is structured as follows. In Section 2, we present some well-known life distributions. In Section 3, we present some GOF tests. In Section 4, we illustrate the model selection using the tests and graphical tools derived, analyzing one real-world data set. Finally, we sketch some conclusions.

2. Life Distributions

We recall here two life distributions with $\alpha > 0$ and $\beta > 0$ denoting shape and scale parameters, respectively. In the following section μ will denote the mean of the corresponding distribution.
A random variable (RV) T with BS distribution of shape $\alpha > 0$ and scale $\beta > 0$ parameters is denoted by $T \sim \mathrm{BS}(\alpha, \beta)$. In this case, the CDF of T is

$$F(t; \alpha, \beta) = \Phi\left([1/\alpha]\xi(t/\beta)\right), \quad t > 0,$$

where $\xi(y) = \sqrt{y} - 1/\sqrt{y} = 2\sinh(\log(\sqrt{y}))$ and $\Phi(\cdot)$ is the N(0, 1) CDF. The corresponding quantile function (QF) is $F^{-1}(q; \alpha, \beta) = \beta[\alpha z(q)/2 + \sqrt{(\alpha z(q)/2)^2 + 1}]^2$, for $0 < q < 1$, where $z(\cdot)$ is the N(0, 1) QF and $F^{-1}(\cdot)$ is the inverse CDF. Note that $F^{-1}(0.5; \alpha, \beta) = \beta$, that is, β is also the median or 50th percentile of the BS distribution. If $T \sim \mathrm{BS}(\alpha, \beta)$, then $X \sim \mathrm{TBS}_{\kappa}(\alpha, \beta)$ denotes the truncated version at κ of T and its CDF is

$$F(t; \alpha, \beta, \kappa) = \frac{\Phi\left([1/\alpha]\xi\left(t/\beta\right)\right) - \Phi\left([1/\alpha]\xi\left(\kappa/\beta\right)\right)}{\Phi\left(-[1/\alpha]\xi\left(\kappa/\beta\right)\right)}, \quad t \geq \kappa > 0.$$

The corresponding QF is $F^{-1}(q; \alpha, \beta, \kappa) = \beta[\alpha z_{\eta}(q)/2 + \sqrt{(\alpha z_{\eta}(q)/2)^2 + 1}]^2$, for $0 < q < 1$, where $z_{\eta}(\cdot)$ is the QF of the truncated N(0, 1) (TN) distribution at $\eta = [1/\alpha]\xi(\kappa/\beta)$.

3. GOF tests with censored and uncensored data

Consider the hypotheses: H_0: "the data come from a RV T with CDF $F(\cdot)$" versus H_1: "the data do not come from this RV". The hypothesized distribution with CDF $F(\cdot)$ can depend on a parameter vector θ that can contain location (μ), scale (β), shape (α) parameters or any other parameter not necessarily of location and scale, in this case, T belongs to the NLS family. If the CDF is completely specified in H_0, that is, θ is known, the data must be transformed for testing uniformity (using the probability integral transformation. On the contrary, the parameters must be consistently estimated and the data transformed for testing normality from the distribution under H_0.
In order to test H_0, when $F(\cdot)$ is completely specified, and then to assess GOF of a distribution to a censored or uncensored data set, we consider test statistics using the ECDF $F_n(\cdot)$. The most common statistic based on the ECDF uses vertical differences between $F_n(t)$ and $F(t)$ by means of the supremum and it is KS statistic given by the following expression.

$$\mathrm{KS} = \sup_t \left|F_n(t) - F(t)\right| = \max\left\{\sup_t\left\{F_n(t) - F(t)\right\}, \ \sup_t\left\{F(t) - F_n(t)\right\}\right\}.$$

In addition, MI statistic is defined by

$$\mathrm{MI} = \max\left\{ \sup_t \left\{ \frac{2}{\pi}\arcsin(\sqrt{F_n(t)}) - \frac{2}{\pi}\arcsin(\sqrt{F(t)}) \right\}, \sup_t \left\{ \frac{2}{\pi}\arcsin(\sqrt{F(t)}) - \frac{2}{\pi}\arcsin(\sqrt{F_n(t)}) \right\} \right\}.$$

Quantiles of the distribution of the statistics KS and MI must be obtained under H_0. However, if the distribution under this hypothesis is not completely specified, its parameters must be properly estimated and the KS and MI statistics must be modified for the distribution under H_0. These modified statistics are denoted by $\mathrm{KS}^\star$ and $\mathrm{MI}^\star$ and their calculated values by $\mathrm{ks}^\star$ and $\mathrm{mi}^\star$, respectively. Thus, new quantiles of the distribution of $\mathrm{KS}^\star$ and $\mathrm{MI}^\star$ must be computed under H_0. Some authors employed the estimated parameters as the true ones applying then tests for a completely known distribution, but this option derives in conservative tests. Therefore, we do not recommend such practice as a GOF technique.
Consider now a sample with censoring proportion p, which conducts to r uncensored data and $n-r$ censored data. From now on, the ordered sample $t_1, \ldots, t_m$ is denoted as $t_{1:m}, \ldots, t_{m:m}$. Assume a type-II right censorship, so that, in this case, r is fixed and $n-r$ observations are greater than the censoring point $t_{r:n}$. Then, $U_{1:n} = F(T_{1:n}) \leq \cdots \leq U_{r:n} = F(T_{r:n})$ are the smallest r order statistics (OS) of the type-II censored sample of size n and the $n-r$ censored observations are greater than $u_{r:n} = F(t_{r:n})$. To test H_0 when $F(\cdot)$ is completely specified with a type-II right censored data and r complete and $n-r$ censored observations, we use the results presented in D'Agostino & Stephens (1986) and adapt the KS and MI statistics by

$$\mathrm{KS}_{r,n} = \max_{1\leq j\leq r}\left\{ \left| \frac{j-0.5}{n} - U_{j:n} \right| \right\}.$$

$$\mathrm{MI}_{r,n} = \max_{1\leq j\leq r}\left\{ \frac{2}{\pi}\left| \arcsin\left(\sqrt{\frac{j-0.5}{n}}\right) - \frac{2}{\pi}\arcsin(\sqrt{U_{j:n}}) \right| \right\}.$$

The quantiles of the distribution of $\mathrm{KS}_{r,n}$, and $\mathrm{MI}_{r,n}$ statistics must be obtained under H_0. If the distribution under H_0 is not completely specified, its parameters must be properly estimated, taking into account the censorship, and the statistics must be modified for each case under H_0. We denote these statistics by $\mathrm{KS}^\star_{r,n}$ and $\mathrm{MI}^\star_{r,n}$ and their calculated values by $\mathrm{ks}^\star_{r,n}$ and $\mathrm{mi}^\star_{r,n}$. Also, new quantiles of the distribution of $\mathrm{KS}^\star_{r,n}$ and, $\mathrm{MI}^\star_{r,n}$ must be computed under H_0. For more details see Castro-Kuriss et al. (2010, 2011).
If the hypotheses of interest H_0 is $F(t) = \Phi([t-\mu]/\beta)$ with unknown parameters, we can consider Algorithm 1.

Algorithm 1. GOF test for normality with uncensored data.

1. Estimate μ and β of $\Phi([t-\mu]/\beta)$ by $\widehat{\mu}$ and $\widehat{\beta}$, respectively, with $t_1, \ldots, t_n$;
2. Obtain $\widehat{u}_{j:n} = \Phi(\widehat{z}_j)$, with $\widehat{z}_j = [t_{j:n} - \widehat{\mu}]/\widehat{\beta}$, for $j = 1, \ldots, n$;
3. Evaluate $\mathrm{KS}^\star$ and $\mathrm{MI}^\star$ statistics at $\widehat{u}_{j:n}$;
4. Compute the p-values of the $\mathrm{KS}^\star$ and $\mathrm{MI}^\star$ statistics;
5. Reject H_0: $F(t) = \Phi([t-\mu]/\beta)$ for a specified significance level based on the obtained p-values.

Chen & Balakrishnan (1995) provided an approximate GOF test that can be applied to NLS distributions. This method first transforms the data to normality and then applies Algorithm 1. Testing normality in H_0 allows us to compute the critical values of the corresponding test statistics, independently of the parameter estimators, if consistent estimators are available and the sample size is large enough. To test the hypotheses of interest, for $\alpha > 0$ and $\beta > 0$ unknown, we consider a generalization Algorithm 1, which is detailed in Algorithm 2.

Algorithm 2. GOF test for NLS distributions with uncensored data.

1. Estimate α and β of $F(t;\alpha,\beta)$ by $\widehat{\alpha}$ and $\widehat{\beta}$, respectively, with $t_1, \ldots, t_n$;
2. Compute $\widehat{v}_{j:n} = F(t_{j:n};\widehat{\alpha},\widehat{\beta})$, for $j = 1, \ldots, n$;

Table 1: Descriptive statistics for placebo data.

Mean	Median	SD	CV	CS	CK	Range	Min	Max	n
33.83	25	32.2	0.539	1.53	6.8	127	21	148	25

Table 2: ML estimates of the parameters and p-values for placebo data.

Model	Estimated parameter	Estimated value	Estimated $KS^\star$ and $MI^\star$ statistics	p-value
Gamma	shape	4.317	$ks^\star$=0.1546	$0.10 < p < 0.20$
	rate	0.063	$mi^\star$=0.0859	$025 < p < 0.40$
Weibull	shape	2.334	$ks^\star$=0.1189	$0.40 < p < 0.50$
	scale	77.771	$mi^\star$=0.066	$0.60 < p < 0.70$
LN	mean	4.11	$ks^\star$=0.1839	$0.01 < p < 0.05$
	standard deviation	0.515	$mi^\star$=0.1048	$0.05 < p < 0.10$
TEXp	scale	3.489	$ks^\star$=0.1535	$0.10 < p < 0.20$
			$mi^\star$=0.0853	$0.25 < p < 0.40$
TN	mean	58.857	$ks^\star$=0.1247	$0.25 < p < 0.40$
	standard deviation	39.64	$mi^\star$=0.0669	$0.60 < p < 0.70$
TBS	shape	0.535	$ks^\star$=0.1847	$0.01 < p < 0.05$
	scale	59.976	$mi^\star$=0.1053	$0.05 < p < 0.10$
BS-t	shape	0.525	$ks^\star$=0.1838	$0.01 < p < 0.05$
	scale	60.429	$mi^\star$=0.1047	$0.05 < p < 0.10$
BS-L	shape	0.422	$ks^\star$=0.1936	$0.01 < p < 0.05$
	scale	70.999	$mi^\star$=0.1111	$0.01 < p < 0.05$
BS-log	shape	0.308	$ks^\star$=0.1828	$0.01 < p < 0.05$
	scale	62.899	$mi^\star$=0.1041	$0.05 < p < 0.10$

3. Calculate $\widehat{y}_j = \Phi^{-1}(\widehat{v}_{j:n})$, where $\Phi^{-1}(\cdot)$ is the N(0, 1) inverse CDF;
4. Obtain $\widehat{u}_{j:n} = \Phi(\widehat{z}_j)$, with $\widehat{z}_j = [\widehat{y}_j - \bar{y}]/s_y$, $\bar{y} = \sum_{j=1}^{n} \widehat{y}_j/n$ and $s_y = [\sum_{j=1}^{n}(\widehat{y}_j - \bar{y})^2/(n-1)]^{1/2}$;
5. Repeat Steps 3-5 of Algorithm 1. with $F(t) = F(t; \alpha, \beta)$.

GOF tests for NLS distributions with censored data can be obtained adapting the GOF statistics. Graphical plots can also be derived from the above mentioned tests, basically the PP and SP plots, specially considering type II right censored samples. Extensions of these graphs and their acceptance regions for testing NLS distributions under H_0 were proposed by Castro-Kuriss et al. (2014).

4. Example: Application to a Clinical Trial

Clinical trials are aimed at shortening the time-to-discharge. Thus, hospitals and health care management, in general, need to understand the effect of drugs on patients in order to see their efficacy reducing the length of stay. In a double-blind placebo controlled drug study, Shuster et al. (2008) reported times (in hours) of 23 patients on drug and of 25 patients on placebo. None censoring occurred on the trial. The authors tested the hypotheses that, compared with an overnight continuous femoral nerve block (cFNB), a 4-day ambulatory cFNB decreases the time until three specific discharge criteria are met after total knee arthroplasty. We consider this data set as a complete sample and also as a left singly truncated one at 15 hours, because no patient could be discharged of the trial before undergoing surgery and previous inter-hospital exams. We include here only the model selection for the data of the placebo group. Adjusting a parametric model, we are able to easily estimate, for example the length of stay in this group.

The sample shows an skewed to the right distribution. Hence, we consider the following possible distributions: gamma, generalized BS with Student-t, Laplace and logistic kernels (BS-t, BS-L and BS-log, respectively), LN, TBS, truncated exponential (TExp), TN, and Weibull. Table 1 shows the descriptive statistics (including standar deviation, SD, and coefficients of variation, CV, of skewness, CS, and of curtosis, CK) of the sample, whereas Table 2 shows the maximum likelihood estimates of the parameters from placebo group for each model under study and the corresponding observed statistics with boundaries for the p-values. According to Table 2, two models are excellent candidates to describe the data: we select the Weibull distribution. Figure 1 shows PP and SP plots of the sample, considering a Weibull distribution under the null hypothesis with the 95% acceptance bands, which show that all the points fall inside them. By means of the selected distribution, we estimate that 43.4% of the patients that received the conventional treatment stay in the hospital more than 3 days and 80.5% of this group stay at most 4 days.

5. Conclusions

We have reviewed available tests for the non-location-scale family and discuss the possibility of model selection in both location-scale and non-location-scale families based on a hypothesis test approach. We are now

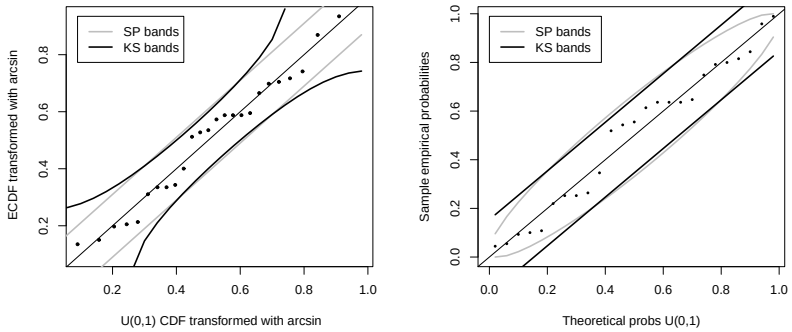

Figure 1: PP (left) and SP (right) plots with 95% acceptance bands from a Weibull distribution for placebo data.

developing a new R package that will allow the practitioners to easily reach results, thus encouraging them to use the methods here described with a wide range of possible distributions to be considered under the null hypothesis with complete or censored data. We think a discussion on goodness-of-fit (GOF) techniques is open about new possible distributions describing data, due to till today vast unexplored un9.123Tje43n

Lawless, J.F. (2003) *Statistical Models and Methods for Lifetime Data*. Wiley, NY.

Leiva, V., Ponce, M.G., Marchant, C.& Bustos, O. (2012) Fatigue statistical distributions useful for modeling diameter and mortality of trees. *Col. J. Stat.*, 35, 349-367.

Lin, C-T., Huang, Y-L.& Balakrishnan, N. (2008) A new method for goodness-of-fit testing based on type-II right censored samples. *ITR*, 57, 633-642.

Marden, J.I. (2004) Positions and QQ-Plots. *Stat. Sciences*, 19, 606-614.

Marshall, A.W.& Olkin, I. (2007). Life Distributions. Springer, NY.

Michael, J.R. (1983) The stabilized probability plot. *Biometrika*, 70, 11-17.

Pakyari, R., Balakrishnan, N. (2012) A general purpose approximate goodness-of-fit test for progressively type-II censored data. *ITR*, 61, 238-244.

R Development Core Team (2009) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rad, A., Yousefzadeh, F. & Balakrishnan, N. (2011), Goodness-of-fit test based on Kullback-Leibler information for progressively type-II censored data, *IEEE, Transactions on Reliability*, 60, 570579.

Shuster, J., Theriaque, D. & Ilfeld B. (2008) Applying Hodges-Lehmann scale parameter estimates to hospital discharge times.*Clinical Trials*, 5, 631634.

Stehlík, M., Strelec, L.& Thulin, M. (2014) On robust testing for normality in chemometrics. *Chem. Int. Lab. Syst.*, 130, 98-108.