Brazilian Journal of Probability and Statistics (2001), 15, pp. 147–154.

DETECTING NON-IDENTIFIABILITY ON THE POLY-WEIBULL MODEL

Francisco Louzada–Neto and **Christiano Santos Andrade**

Departamento de Estatística, Universidade Federal de São Carlos, C. P. 676, 13565-905 – São Carlos - SP, Brasil. Emails: dfln@power.ufscar.br and christi.ufscar@bol.com.br

Summary
The poly-Weibull model is a general family which can accommodate not only constant, increasing or decreasing hazard curves with zero or non zero asymptotes, but also nonmonotones ones, including bathtub-shaped. In this paper we consider the non-identifiability problem, which arises when the shape parameters of a poly-Weibull model are close. A graphical method based on the total-time-on-test plot and its simulated envelop is considered for detecting when a poly-Weibull model is likely to be identifiable. We also provide a general framework for constructing hypothesis tests for non-identifiability by using parametric bootstrap-based methods. We set up a simulation study and show that the bootstrap tests have desirable properties with respect to size and power.

Key words: Bi-Weibull model; hazard models; hypothesis tests; non-identifiability, TTT plot.

1 Introduction

Bathtub hazard curves are not uncommon in practice. They correspond to an initial high failure rate followed by lower odds of failure which increase with age. Real examples with such characteristic can be found in Ebrahim (1996) and Lagakos and Louis (1988) in the context of cancer survival, among others. In a recent paper Davison and Louzada-Neto (2000) considered some approaches to inference for the poly-Weibull model which was first considered by Canfield and Borgman (1975) and Berger and Sun (1993). The advantage of such model is its ability of accommodating increasing, constant and decreansing hazard functions and also nonmonotone ones, such as a bathtub-shaped (Rajarshi and Rajarshi, 1988).

The lifetime T is said to have a poly-Weibull hazard model if the overall hazard function is given by (Davison and Louzada-Neto, 2000),

$$h(t) = \sum_{j=1}^{m} h_j(t) = \sum_{j=1}^{m} \frac{\beta_j t^{\beta_j - 1}}{\mu_j^{\beta_j}}, \tag{1.1}$$

where $\mu_j > 0$ and $\beta_j > 0$ are unknown parameters. It is easy to show that, when $m = 2$ in (1.1) (we call this a bi-Weibull hazard model), the hazard is decreasing if $\max(\beta_1, \beta_2) < 1$, it is increasing if $\min(\beta_1, \beta_2) > 1$ and it is bathtub-shaped if $\beta_1 < 1$ and $\beta_2 > 1$.

The MLEs of the parameters in (1.1) can be obtained by direct maximization of the log-likelihood function, $\log L = \sum_{i=1}^{n}\{\delta_i \log h(t_i) - H(t_i)\}$, where $H(t_i) = \int_0^{t_i} h(x_i)dx_i$ is the cumulative baseline hazard function (Lawless, 1982) and δ_i an indicator variable defined by $\delta_i = 1$ if $T_i = t_i$ is an observed failure time and $\delta_i = 0$ if it is a right-censored observation. Those estimates can also be obtained by solving the system of nonlinear equations given by the partial derivatives of $\log L$ with respect to the parameters. The advantage of direct maximization approach is however that it runs immediately using existing statistical packages.

The purpose of this paper is to investigate the non-identifiability problem which arises when the shape parameters of the model (1.1) are close for the simplest poly-Weibull model (the bi-Weibull). The point is that, though the number of components m may be larger than two availability of data capable of detecting more than two causes of failure is unusual and the bi-hazard model seems likely to be the most useful poly-hazard model (Davison and Louzada-Neto, 2000).

The non-identifiability problem is decribed in Section 2, where we show that a graphical method, based on the total-time-on-test plot (Aarset, 1987) and its simulated envelope, can be used for detecting when a poly-Weibull model is likely to be identifiable in practice. A parametric bootstrap-based hypothesis tests for non-identifiability is presented in Section 3 where we also present the results of a simulation study on the size and power of the test statistics. We finish the paper with some concluding remarks in Section 4. The methodology is illustrated by some generated and real datasets.

2 Detecting non-identifiability

A problem with model (1.1) is that a singularity arises if the shape parameters β_j's are equal (Berger and Sun, 1993). In this case, the poly-Weibull model is a single-Weibull model and the scale parameters μ_j's are nonidentifiable or the parameters are only identifiable through functions $q(\mu, \beta)$'s, where $\mu' = (\mu_1, ..., \mu_m)$ and $\beta' = (\beta_1, ..., \beta_m)$. Also, in our experience, even if the β_j's are differents but sufficiently close to each other the non-identifiability problem will arise. The source of this problem lies in the early stage of the model building when a redundant vector of parameters is kept in the parametrization (Bernardo and Smith, 1994). Therefore fit of the model is sensible only if there is clear evidence from the data (or a priori) that separated β_j's are needed. In this context, as commented by a referee, it could be feasible to consider a Bayesian approach with proper priors to avoid nonidentifiability, but this is out of the scope of the paper.

The simplest approach for the non-identifiability problem, which is developed in the paper, is to detecting it graphically before the fitting. In many applications there are qualitative information about the hazard shape, which can help in selecting a particular lifetime model and detecting possible parameter noindentifiablity. A device called the total time on test (TTT) plot (Aarset, 1987) is useful in this context. The TTT plot is obtained by plotting $\mathsf{G}(r/n) = [(\sum_{i=1}^{r} T_{i:n}) + (n-r)T_{r:n}]/(\sum_{i=1}^{n} T_{i:n})$, where $r = 1, \ldots, n$ and $T_{i:n}$, $i = 1, \ldots, n$ are the order statistics of the sample, against r/n (Mudholkar, et al, 1996). It is a straight diagonal for constant hazards leading to an exponential model. It is convex for decreasing hazards and concave for increasing hazards leading to a single-Weibull model. And it is first convex and then concave if the hazard is bathtub-shaped leading to a bi-Weibull model. Figure 1 show the TTT plots for five generated dataset, with 100 observations each, from a bi-Weibull model with parameters fixed at $\mu_1 = 1$, $\mu_2 = 1, 20$ (left and right panels, respectively) and $(\beta_1, \beta_2) = (0.5, 8)$, $(0.5, 4)$, $(0.5, 1)$, $(2, 3)$, $(1, 1)$, representing two situations where the bi-Weibull model seems to be adequate, two situations where a single-Weibull model seems suffice for fitting the data and one situation where an exponential distribution is enough for fitting, respectively. We also include in the Figure 1 the limits of the 90% TTT diagonal line simulated envelop, which correspond to the simpler lifetime model among those considere here (the exponential). The envelop was obtained by generating 1000 exponential samples with mean equals to 1 and 20 (left and right panels, respectively). For each sample we caculated the $\mathsf{G}(r/n)$ vector of the TTT plot and obtained $\mathsf{G}_1(\cdot), ..., \mathsf{G}_{1000}(\cdot)$ vectors. After to order the elements of these vectors by r/n, $r = 1, ..., n$, we used the ordered values to directly determine the limits of the 90% simulated envelop (Neter, et al, 1996). If a more sofisticate hazard model is likely to be feasible, the sampling TTT plot curve should be, at least in part, out of the envelop domain. This is particularly true for all four situations considered above, where the exponential seems to be inadequate. We also note that the difference between the envelops obtained from an exponential mean equal to 1 and 20 is negligible, indicating that the simulated envelops do not depend on the exponential mean.

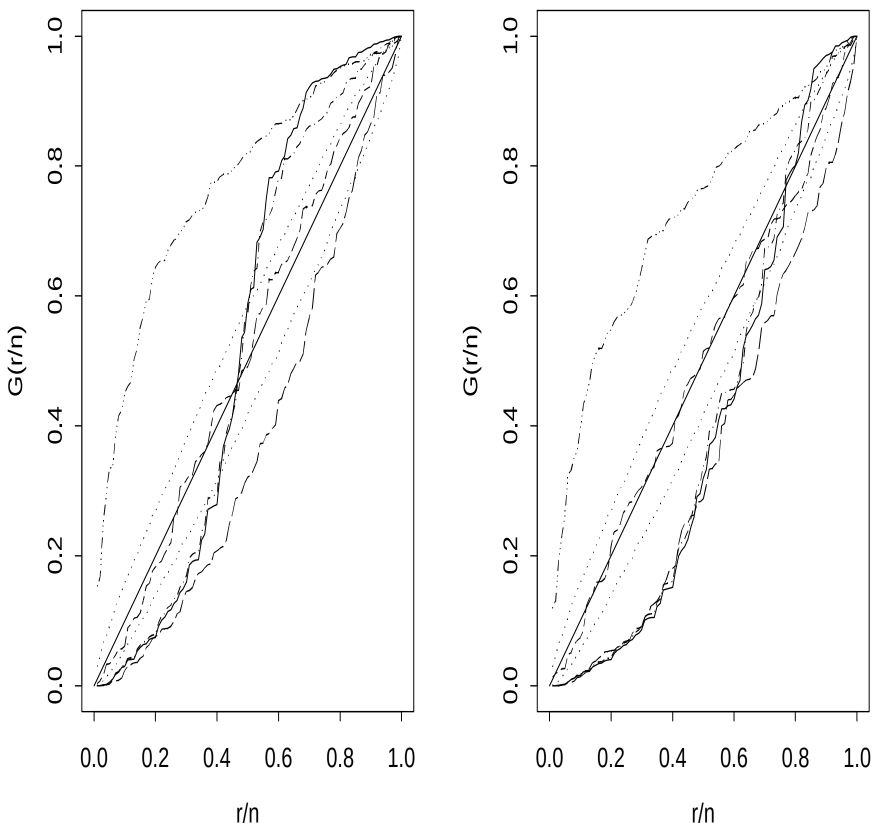

Figure 1
TTT plots for five generated dataset from a bi-Weibull model with parameters fixed at $\mu_1 = 1$ *and* $\mu_2 = 1,20$ *(left and right panels, respectively) and (—):* $(\beta_1, \beta_2) = (0.5, 8)$*,* *($- \cdot - \cdot -$):* $(\beta_1, \beta_2) = (0.5, 4)$*, (– – –):* $(\beta_1, \beta_2) = (0.5, 1)$*, ($- \cdot \cdot \cdot - \cdot \cdot \cdot -$):* $(\beta_1, \beta_2) = (2, 3)$*, (— – — – —):* $(\beta_1, \beta_2) = (1, 1)$*. The two lines ($\cdot \cdot \cdot$) represent the lower and upper limits of the 90% envelop confidence interval for the TTT plot of an exponential random variable with* $\mu_2 = 1,20$ *(left and right panels, respectively).*

3 Tests for non-identifiability

A more general procedure for dealing with the non-identifiability problem is to assume that the number of components, m, is unknown. This procedure however raises a difficult issue. For instance, the comparation of $m = 2$ with $m = 1$ in (1.1) introduces two parameters, one of which is meaningless when $m = 1$ and the regularity conditions, on which the standard asymptotic theory is based, will not hold (Cheng and Traylor 1995).

An alternative direct approach is to bootstraping the likelihood ratio statistics (LRS) for testing $m = 1$ against $m = 2$ in order to obtain its empirical distribution (Davison and Hinkley, 1997). Let $w = 2(l_2 - l_1)$ be the LRS for testing two alternative models, denoted by 1 and 2, where l_1 and l_2 are the log-likelihoods for each model. Large positive values of w give favorable evidence to model 2. The parametric bootstrap technique consists of generating R (here assumed equals to 999) datasets from the model under the null hypothesis (model 1) with the parameters substituted by their MLEs obtained by considering the procedure discussed in Section 1, record $w_1^* < \cdots < w_R^*$, and use $w^*_{(R+1)(1-\alpha)}$ as the critical point to test the null hypothesis with size α. For instance, considering the generated datasets in Section 2 (Figure 1) with parameters $\mu_1 = \mu_2 = 1$ and $(\beta_1, \beta_2) = (0.5, 8)$, $(0.5, 4)$, $(2, 3)$, the LRSs for testing $m = 1$ against $m = 2$ are 17.3, 8.4 and 2.6, with empirical p-values for the null hypothesis of a single-Weibull model equal to 0.037, 0.062 and 0.373, respectively. These results suggest that, for the two first datasets the bi-Weibull model is requered, while the third dataset can be successfully fitted by a single-Weibull model. These results are in broad agreement with those obtained by using the graphical approach described in Section 2.

For power calculation purposes we estimate the p-value for the alternative hypothesis, in which case we adopt the reverse procedure above generating datasets from the model under the alternative (model 2) with the parameter values substituted by their fitted values.

3.1 A simulation study

For a limited assessment of the performance of the bootstrap test procedure we study its size and power for sample sizes fixed at $n = 30, 60$ e 100 with the number of bootstrap replications fixed at 999. For the empirical sizes we generate samples from single-Weibull models with parameters $\mu = 1$ e $\beta = 0.5, 1$ e 3. For the power calculations the samples were generated from bi-Weibull models with parameters $\mu_1 = \mu_2 = 1$, $\beta_2 = 0.5$, 1, 3, and five β_1 values assigned appropriately inside the interval $(0, 10]$. The overall study described above was repeated with censored samples with 10 and 30% of righ-censoring.

The empirical sizes and powers of the tests for $\alpha = 0.05$ were summarized in Tables 1 and 2 for complete and 30% of censoring cases. The results for the 10% right-censoring cases were always between the complete and 30% of censoring cases results and were omitted. The probabilities of correctly accepting the null hypothesis are near to the nominal size, except for small samples where we observe some departure from $\alpha = 0.05$, particularly if presence of censoring is observed. The empirical powers increase with the ratio β_1/β_2 and are bigger for large n. For ratios moving towards one however the powers are rather low. A phenomenon related to the non-identifiability problem that arises in such situations. The censoring effect is to reduce the empirical powers.

Table 1

Empirical sizes. The entry a/b indicates the empirical sizes a and b for complete and 30% of censoring, respectively.

	n		
β_2	30	60	100
0.5	0.122/0.158	0.065/0.087	0.048/0.065
1	0.118/0.152	0.087/0.099	0.046/0.061
3	0.136/0.161	0.073/0.089	0.053/0.064

Table 2

Empirical powers. The entry a/b indicates the empirical powers a and b for complete and 30% of censoring, respectively.

		n		
β_2	β_1/β_2	30	60	100
0.5	0.20	0.748/0.313	0.963/0.497	0.998/0.635
	1.50	0.259/0.208	0.273/0.251	0.396/0.352
	3.00	0.405/0.350	0.605/0.454	0.842/0.678
	6.00	0.791/0.523	0.988/0.621	0.989/0.732
	10.0	0.987/0.833	0.999/0.854	0.999/0.899
1	0.25	0.776/0.425	0.953/0.597	0.992/0.714
	0.50	0.489/0.197	0.587/0.207	0.649/0.336
	1.50	0.244/0.192	0.245/0.223	0.292/0.314
	3.00	0.637/0.207	0.707/0.259	0.752/0.353
	5.00	0.910/0.411	0.996/0.587	0.999/0.603
3	0.25	0.883/0.430	0.997/0.552	0.999/0.673
	0.50	0.490/0.371	0.759/0.621	0.877/0.864
	0.83	0.203/0.190	0.367/0.252	0.492/0.370
	1.67	0.381/0.221	0.526/0.384	0.768/0.437
	5.00	0.989/0.399	0.999/0.424	0.999/0.607

3.2 A numerical example

The bootstrap test described above was applied to a survival dataset on the ages of 18 patients classified as other causes of death in a cancer study, but for whom no specific cause is know (Ebrahimi, 1996). Louzada-Neto (1999) shows the convex-concave path of the TTT plot for this data, which indicates that a bi-Weibull model is requered. The LRS, $w = 2(l_{\mathrm{bi}} - l_{\mathrm{sin}})$, for testing a single-Weibull model against a bi-Weibull hazard model, where $l.$ are the log likelihoods for each model, is 7.0. According to the bootstrap approach described above the estimated p-value is equal to 0.07, showing some evidence that the single-Weibull is inadequate. The estimate the p-value for the alternative hypothesis, obtained by considering the reverse resampling procedure, is equal to 0.37, which shows no evidence against the bi-Weibull model.

4 Concluding remarks

The model (1.1) is very attractive from the practical point of view since it accomodates not only increasing, constant and decreasing hazard curves, but also bathtub-shaped ones. Care is however needed for handling the model once the general poly-hazard model can be non-identifiable if the shape parameters are close and the use of the model is sensible only if there is clear evidence from the data (or a priori) that separated β's are needed.

A important lesson learned from this work is that these conditions leads to the need of considering a procedure for detecting non-identifiability. The simpler approach would be always to use a TTT plot for checking whether determined poly-hazard model is likely to be feasible. Tests for non-identifiability however can also be considered by bootstraping the LRS in order to obtain its empirical distribution. Our simulation study results reveal that the boostrap-based tests seem to work well, even for moderate size samples and presence of censoring. The bootstrap test is more conclusive than the indicative graphical test and can be implemented straightforwardly on existing statistical packages, with little knowledge of programming. However, from the practical point of view, in our experience, both approaches are important for handle non-identifiability and should be used concomitantly when fitting the general poly-hazard model.

Acknowledgments

The authors are grateful to David Cox and Anthony Davison for comments and suggestions on this work. The research was supported by the Brazilian Organizations CAPES and CNPq.

References

Aarset, M. V. (1987) How to identify a bathtub hazard rate. *IEEE Transactions on Reliability*, **36**, 106–108.

Berger, J.O. and Sun, D. (1993). Bayesian analysis for the poly-Weibull distribution. *Journal of the American Statistical Association*, **88**, 1412–1417.

Candfield, R.V. and Borgman, L.E. (1975). Some distributions of time to failure for reliability applications. *Technometrics*, **17**, 263–268.

Cheng, C.H. and Traylor, L. (1995). Non-regular maximum likelihood problems (with discussion). *Journal of the Royal Statistical Society B*, **57**, 3–44.

Davison, A.C. and Hinkley, D.V. (1997). *Bootstrap Methods and their Application*. Cambridge: Cambridge University Press.

Davison, A.C. and Louzada-Neto, F. (2000). Inference for the poly-Weibull model. *Journal od the Royal Statistical Society D*, **49**, 189–196.

Ebrahimi, N. (1996). The effect of misclassification of the actual cause of death in competing risks analysis. *Statistics in Medicine*, **15**, 1557–1566.

Lagakos, S.W. and Louis, T. A. (1988). Use of tumour lethality to interpret tumorigenicity experiments lacking cause-of-death data. *Applied Statistics*, **37**, 169–179.

Lawless, J.F. (1982). *Statistical Models and Methods for Lifetime Data*. New York: Wiley.

Neter, J., Kutner, M.H., Nachtsheim, C.J. and Wasserman, W. (1996). *Applied Linear Statistical Models*. Chicago: Irwin.

Mudholkar, G.S., Srivastava, D.k. and Kollia, G.D. (1996). A generalization of the Weibull distribution with application to the analysis of survival data. *Journal of the American Statistical Association*, **91**, 1575–1583.

Rajarshi, S., and Rajarshi, M.B. (1988). Bathtub distributions: a review. *Communication in Statistics – Theory Methods* **17**, 2597–2621.