ON THE INTERVAL ESTIMATION OF THE PARAMETERS OF A GENERALIZED TIME-DEPENDENT LOGISTIC MODEL

Francisco LOUZADA NETO[1]
Gilbert MACKENZIE[2]
Caroline Pires CREMASCO[3]
Paulo Henrique FERREIRA-SILVA[3]

- *ABSTRACT: In order to accommodating crossing hazard curves, which are non-proportional hazards, we consider in this paper a generalized time-dependent logistic hazard survival model, which has a time-dependent term. The model is a wholly parametric competitor for the Cox proportional hazard model. We compare different procedures to compute confidence intervals for the model parameters in presence of random censoring. Our simulation study focus on the study of the coverage probabilities of these different confidence intervals and on the significance levels of some hypothesis tests. We discovered that parametric and non-parametric resampling methods can be successfully used for hypothesis testing and generating precise confidence intervals for the parameters even on small and moderate sized samples.*

- *KEYWORDS: Bootstrap method; coverage probability; Monte Carlo simulation; non-PH model; generalized time-dependent logistic.*

1 Introduction

As a wholly parametric competitor to the PH model, MacKenzie (1996, 2002) proposed the generalized time-dependent logistic (GTDL) regression model, which is a non-proportional hazards (non-PH) model and has a logistic baseline hazard

[1]Universidade de São Paulo – USP, Instituto de Ciências Matemáticas e de Computação – ICNC, CEP: 13566-590, São Carlos, SP, Brazil. E-mai: *louzada@icmc.usp.br*
[2]University of Limerick, Department of Mathematics and Statistics, Ireland. E-mail: *gilbert.mackenzie@ul.ie*
[3]Universidade Federal de São Carlos – UFSCar, Departamento de Estatística, CEP: 13565-905, São Carlos, SP, Brazil

function. The model generalizes the relative risk in the proportional hazard (PH) Cox's model (1972) to time-dependent form. In this paper we compare different procedures to compute confidence intervals for the model parameters in presence of random censoring. We study, via Monte Carlo simulation, the coverage probabilities of these different confidence intervals as well as the significance levels of some hypothesis tests.

The paper is organized as follows. The GTDL model is presented in Section 2. Interval estimation is described in Section 3. Hypothesis tests are discussed in Section 4. The methodology is illustrated in Section 5. Final remarks in Section 6 conclude the paper.

2 The GTDL model

Letting $h(t|x)$ denote the hazard function at time t for an individual with covariate vector x, the GTDL model assumes that

$$h(t|x) = \frac{\lambda \exp(t\alpha + x'\beta)}{1 + \exp(t\alpha + x'\beta)}, \tag{1}$$

where $\lambda > 0$ is a scalar, α is a scalar measuring the effect of time and $\beta' = (\beta_1, ..., \beta_p)'$ is a vector of p parameters measuring the influence of the p covariates $x' = (x_1, ..., x_p)'$.

Consequently, the density function is given by $f(t|x) = \lambda p(\alpha, \beta) \{q(\alpha, \beta) g(\beta)\}^{\frac{\lambda}{\alpha}}$, where the individual components are simple function of the time-dependent multiple-logistic function, with $p(\alpha, \beta) = \exp(\alpha t + x\acute{}\beta) / \{1 + \exp(\alpha t + x\acute{}\beta)\}$, $q(\alpha, \beta) = 1 / \{1 + \exp(\alpha t + x\acute{}\beta)\}$ and $g(\beta) = 1 + \exp(x\acute{}\beta)$.

Intrinsically, equation (1) is neither a proportional hazards model nor an accelerated failure-time model. For instance, consider two observations 1 and 2 that differ in their x-values. The hazard ratio for these two observations,

$$\frac{h(t|x_1)}{h(t|x_2)} = \exp((x_1' - x_2')\beta) \frac{1 + \exp(t\alpha + x_2'\beta)}{1 + \exp(t\alpha + x_1'\beta)} \tag{2}$$

is dependent of time t. Consequently, the model (1) is a non-PH model. When $\alpha = 0$, model (1) reduces to the exponential PH model with a logistic baseline hazard function.

Consider a sample of independent random variables $T_1, \ldots, T_n$ denoting the lifetimes of n units. Assume that T_i has associated 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. The likelihood function for the parameters λ, α and β indexing (1) is given by $L(\lambda, \alpha, \beta) = \prod_{i=1}^{n} h(t_i)^{\delta_i} \; S(t_i)$ (Lawless, 2002), where $h(t_i)$ is given in (1) and $S(t_i)$ is the survival function given (from (1)) by

$$S(t_i|x) = \left\{ \frac{1 + \exp(x_i'\beta)}{1 + \exp(t_i\alpha + x_i'\beta)} \right\}^{\lambda/\alpha}. \tag{3}$$

Then, the likelihood function for λ, α and β is given by

$$L(\lambda, \alpha, \beta) = \prod_{i=1}^{n} \left\{ \lambda \frac{\exp(t_i\alpha + x_i'\beta)}{1 + \exp(t_i\alpha + x_i'\beta)} \right\}^{\delta_i} \left\{ \frac{1 + \exp(x_i'\beta)}{1 + \exp(t_i\alpha + x_i'\beta)} \right\}^{\lambda/\alpha}. \quad (4)$$

The maximum likelihood estimates (MLEs) of the parameter vector $\theta = (\lambda, \alpha, \beta)$ are obtained by direct maximization of (4). The advantage of this procedure is that it runs immediately using existing statistical packages. The maximization procedure can be performed by solving the system of nonlinear equations given by the partial derivatives of $\log L(\lambda, \alpha, \beta)$ with respect to the parameters, but in our experience pure Newton-Raphson schemes are extremely susceptible to failure to converge. An important aspect of implementing the estimation procedure concerns parametrization. In our numerical examples and simulation studies we have not faced numerical problems, such as evidence of failure of convergence or end on multiple maxima, from parameters with unbounded ranges. That is, since $\lambda > 0$ and $\alpha > 0$, we have considered the following parametrization $\varphi = \log \lambda$, $\phi = \log \alpha$ and β.

3 Interval Estimation

Inference for the parameter vector $\theta = (\lambda, \alpha, \beta)$ can be based on large sample properties of the MLEs, which lead to $\left(\widehat{\theta} - \theta\right) / I^{-1/2}(\widehat{\theta}) \xrightarrow{D} N(0, I_3)$, where $I^{-1/2}(\widehat{\theta})$ denotes the observed information matrix of θ evaluated at the MLEs and I_3 is the identity matrix of dimension three (Sen and Singer, 1993). However, in reliability and survival studies, it is common to find small or moderate datasets. In order to check the behavior of the asymptotic theory for small and moderate sized samples, we performed a small-scale simulation study for examining the coverage probabilities of the asymptotical confidence intervals for the parameters. The study was based on $1,000$ samples which were generated according to the scheme that follows. Each lifetime t_i was given by $t_i = \min(y_i, c_i)$, for $i = 1, ..., n$, where y and c were two independent random variables representing the lifetimes and the censoring times, respectively. Both of them were generated according to (1) with $\lambda = \alpha = \beta = 0.5$ and x_i was generated according to a Bernoulli distribution with success probability equals to 0.5. In order to have approximately 50% of censoring, we controled the parameters of the distribution of c. The censoring variable was given by $\delta_i = 1$, if $y_i < c_i$ and $\delta_i = 0$, otherwise, characterizing a type I censoring scheme. We have considered $n = 15, 30, 50, 100, 300$ and 1000.

Table 1 shows the variation in coverage of nominal 90% confidence intervals according to the sample size. The 90% confidence interval for the nominal coverage probability of 0.90 based on a sample of size equals to a thousand observations is given by $(0.884, 0.9156)$. If a confidence interval has exact coverage of 0.90, roughly 90% of the observed coverages should be inside these bounds. There is clear under-coverage of the confidence intervals for small and moderate sized samples.

Such findings are evidence for the need of a more adequate procedure for small or moderate sized samples.

Tabela 1 - Coverage probabilities of the 90% asymptotical confidence intervals

n	φ	ϕ	β
15	0.781	0.762	0.833
30	0.887	0.874	0.898
50	0.891	0.889	0.872
100	0.913	0.921	0.893
300	0.899	0.881	0.911
1000	0.933	0.905	0.900

Table 2 shows the slopes obtained by regressing $\log\{\text{var}(\cdot)\}$ on $\log n$. That is, the first entry of Table 2 means that, for $15 \leq n \leq 30$, $\text{var}(\hat{\varphi}) \propto n^{-1.401}$, which corresponds to a difference in slope of 40.1% in comparison with the asymptotic slopes which are equal to 1. Overall, the asymptotic slopes are well approached only for $n \geq 100$. This is another fact that corroborates with the use of a more adequate procedure for small or moderate sized samples.

Tabela 2 - Slope of log-log relation between the variances and n

n	φ	ϕ	β
[15, 30]	-1.401	-1.965	-1.468
[30, 50]	-1.131	-1.096	-1.305
[50, 100]	-0.974	-1.000	-1.160
[100, 300]	-0.946	-0.939	-1.22
[300, 1000]	-0.992	-0.998	-0.938

An alternative direct approach is the bootstrap procedure, which aims to obtaining empirical interval estimations by resampling the original data set. There are two basic bootstrap types: the parametric bootstrap, where the simulating datasets are drown by generating observations, in our case, from model (1) with the parameters replaced by their MLEs, and the non-parametric bootstrap, where the simulating datasets are drown with replacement directly from the original sample. More details about bootstrap techniques may be found in Davison & Hinkley (1997).

Consider β the parameter of interest and suppose that we are interested in constructing a confidence interval for it. For each resample, obtained by a parametric or a non-parametric way, we calculate the MLE for β, having at the end of R resamples $\hat{\beta}_1 < ... < \hat{\beta}_R$ ordered MLE values. Then, we use $\hat{\beta}_{(R+1)(a/2)}$ and $\hat{\beta}_{(R+1)(1-a/2)}$ as the lower and upper bounds of the bootstrap percentile confidence interval $100(1-a)\%$ for β, respectively, where a is the significance level. The bootstrap percentile intervals for the other model parameters can be analogously obtained.

In order to check the adequacy of the bootstrap procedure for small and moderate sized samples when censoring is observed, we run the simulation study described above, with a thousand samples generated for each case. We further fixed R equals to 999, number which, according to Hall (1986), is bigger than the number of replications required to get a critical level of 0.10 from the 0.90 percentile of the empirical distribution of the parameters.

Table 3 shows the coverage probabilities of the 90% bootstrap confidence intervals according to the sample size. Based on the same criteria for indicating whether a confidence interval has exact coverage of 90%, that is, based on the 90% asymptotic confidence interval for the nominal coverage probability of 0.90, which is given by $(0.884, 0.9156)$, there is evidence for the adequacy of both bootstrap procedures even for samples with $n = 30$. Also, even for samples with $n = 15$ the coverage probabilities are less than 4% different from the nominal 90% asymptotic lower bound (equals to 0.884) of the confidence interval for the nominal coverage probability of 0.90.

Tabela 3 - Coverage probabilities of the 90% bootstrap confidence intervals

n	Parametric			Non-Parametric		
	φ	ϕ	β	φ	ϕ	β
15	0.865	0.852	0.861	0.871	0.868	0.873
30	0.877	0.870	0.883	0.911	0.881	0.896
50	0.897	0.894	0.896	0.893	0.896	0.898
100	0.910	0.903	0.909	0.914	0.914	0.910
300	0.903	0.892	0.902	0.906	0.897	0.904

4 Hypothesis Tests

There are two major problems that should be addressed from the hypothesis tests point of view related to the GTDL model (1). The first problem is related to the test for the covariates effect, that is, $H_0 : \beta = 0$, while the second problem is related to the test for the time t effect, that is, $H_0 : \exp(\phi) = 1$.

For testing $H_0 : \beta = 0$ and $H_0 : \exp(\phi) = 1$, we can consider the likelihood ratio statistics (LRS), $w = 2(l_{alt} - l_{null})$, where l_{null} and l_{alt} are the log-likelihood functions for models under the null and alternative hypothesis, respectively. Large positive values of w give favourable evidence to model under the alternative hypothesis.

For testing $H_0 : \beta = 0$, the empirical significance levels (in percentage) according to the sample size are given by 0.122, 0.105, 0.093, 0.076 and 0.055, for $n = 15, 30, 50, 100$ and 300, respectively, indicating that the LRS, w, is not distributed as a chi-squared distribution with one degree of freedom for small and moderate sized samples.

Then, similarly to the Section 3, an alternative direct approach is to bootstrapping (parametrically or nonparametrically) the LRS w in order to obtain its empirical distribution. The parametric bootstrap technique consists of generating R datasets from the model under the null hypothesis with the parameters substituted by their MLEs obtained by using the procedure discussed in Section 2, record $w_1^* < \cdots < w_R^*$, and use $w_{(R+1)(1-a)}^*$ as the critical point to test the null hypothesis with size a. We consider here for R equals to 999.

The non-parametric bootstrap technique operates in much the same way, but instead of generating datasets from the model under the null hypothesis (model 1), with the parameters substituted by their MLEs, we draw R samples with replacement of n observations each from the original dataset $t_1, \ldots, t_n$.

Table 4 shows the empirical significance levels (in percentage) according to the sample size considering both bootstrap procedures discussed above. The empirical significance levels are near the nominal one (0.05) even for small datasets.

Tabela 4 - Empirical significance levels (in percentage) for testing $H_0 : \beta = 0$ according to the sample size considering both bootstrap procedures

n	15	30	50	100	300
$Non-Parametric$	0.056	0.051	0.062	0.056	0.054
$Parametric$	0.039	0.040	0.049	0.054	0.057

Following the same procedure described above for testing $H_0 : \exp(\phi) = 1$, the empirical significance levels (in percentage) according to the sample size considering both bootstrap procedures discussed above are given in Table 5. Also, here the empirical significance levels are near the nominal one (0.05) even for small datasets.

Tabela 5 - Empirical significance levels (in percentage) for testing $H_0 : \exp(\phi) = 1$ according to the sample size considering both bootstrap procedures

n	15	30	50	100	300
$Non-Parametric$	0.065	0.059	0.058	0.054	0.053
$Parametric$	0.053	0.049	0.047	0.052	0.056

5 The Advanced Inoperable Lung Cancer Data

Survival from lung cancer tends not to follow the proportional hazards assumption, especially in age (Blagojevic, MacKenzie and Ha, 2003). Accordingly, for instance, we consider data on the survival of males with advanced inoperable lung cancer (Louzada-Neto, Cremasco and MacKenzie, 2010). Survival time in months and several covariates are available for the 137 patients with inoperable lung cancer; 9 patients were right-censored. After preliminary investigations, it was discovered that initial performance status exerted a strong prognostic effect. The Karnofsky score is measured on a scale 0-100, with high values implying improved performance, typically among patients who are less ill. Although the original objective of this trial was to assess chemotherapy, we focus on how the Karnofsky score (henceforth performance) influences survival. Patients with performance up to 50 were assigned to Group 1, while patients with performance greater than 50 were assigned as Group 2. Table 6 shows the MLEs and respective 90% asymptotical, parametric and non-parametric bootstrap confidence intervals of the GTDL model parameters. The parameter ϕ is significantly different from one, evidenciating the time effect, which is not supported by a PH modeling.

Tabela 6 - MLEs and respective 90% asymptotical, parametric and non-parametric bootstrap confidence intervals

Parameter	φ	ϕ	β
MLE	0.965	0.612	-0.971
Asymptotical	(0.95, 0.98)	(0.55, 0.67)	(-1.15, -0.79)
Non-Parametric	(0.91, 0.99)	(0.32, 0.70)	(-1.27, -0.83)
Parametric	(0.93, 0.99)	(0.39, 0.74)	(-1.29, -0.82)

Final Remarks

The GTDL model considered in this paper is an alternative to the PH model and allows a broad class of survival models. The model provides a reasonable physical interpretation of the phenomenon underlying survival data. Maximum likelihood inference can be implemented straightforwardly and parametric and non-parametric simulation can be successfully used for hypothesis testing and generating precise confidence intervals for the parameters even on small and moderate sized samples.

Acknowledgments

This work has been partially supported by the Brazilian Organizations FAPESP and CNPq.

LOUZADA NETO, F.; MACKENZIE, G.; CREMASCO, C. P. Estimação intervalar dos parâmetros de um modelo sobrevivência logístico generalizado dependente do tempo. *Rev. Bras. Biom.,* São Paulo, v.29, n.3, p.512-519, 2011.

▪ RESUMO: No artigo consideramos um modelo de sobrevivência logístico generalizado dependente do tempo, que acomoda curvas de risco não proporcionais. O modelo é um concorrente paramétrico do modelo de riscos proporcionais de Cox. No artigo são comparados diferentes procedimentos para calcular intervalos de confiança para os parâmetros do modelo na presença de censura aleatória. Consideramos um estudo de simulação sobre as probabilidades de cobertura desses intervalos de confiança e diferentes níveis de confiança de alguns testes de hipóteses. Descobrimos que procedimentos de amostragem paramétrica e não paramétrica podem ser utilizado com sucesso mesmo na presença de amostras pequenas e moderadas.

▪ PALAVRAS-CHAVE: Método Bootstrap; probabilidade de cobertura; simulação de Monte Carlos; modelo de riscos não proporcionais; modelo logístico generalizado dependente do tempo.

Referências

BLAGOJEVIC, M.; MACKENZIE, G.; HA, I. D.. *A comparison of PH and non-PH survival models with Gamma Frailty.* Proceedings of the 18th International Workshop on Statistical Modelling, 39-44. 2003.

COX, D. R.. Regression models and life tables (with Discussion). *Journal of the Royal Statistical Society - B*, v.34, p.187-220. 1972.

DAVISON, A. C.; HINKLEY, D. V.. *Bootstrap Methods and their Applications.* Cambridge: Cambridge University Press. 1997.

HALL, P.. On the bootstrap and confidence intervals. *Ann. Statist.*, v.14, p.1431-1452, 1986.

LAWLESS, J. F. *Statistical models and methods for lifetime data.* 2nd ed. New York: Wiley. 2002.

LOUZADA-NETO, F. ; CREMASCO, C.; MACKENZIE, G. Sampling-based inference for the generalized time-dependent logistic hazard model. *Journal of Statistical Theory and Applications*, v.9, p.169-184, 2010.

MACKENZIE, G. Regression models for survival data. *Journal of the Royal Statistical Society - B*, v.45, p.21-34. 1996.

MACKENZIE, G. *A Logistic Model for Survival Data.* In: Proceedings of the 17th International Workshop in Statistical Modelling, p.431-438. 2002.

SEN, P. K.; SINGER, J. M. *Large Sample Methods in Statistics*: an Introduction with Applications. New York: Chapman & Hall. 1993.

Recebido em 13.07.2011.

Aprovado após revisão em 31.01.2012.