CURE FRACTION USING MIXTURE MODELS ON THE MODIFIED WEIBULL DISTRIBUTION WITH AN APPLICATION TO GASTRIC CANCER DATA

Marcos Vinicius de Oliveira Peres
University of Maringá (UEM), Maringá, Paraná, Brazil – marcosperes1991@hotmail.com

Edson Z. Martinez
University of São Paulo (USP), Ribeirão Preto, São Paulo, Brazil– edson@fmrp.usp.br

Isolde T. S. Previdelli
University of Maringá (UEM), Maringá, Paraná, Brazil – isoldeprevidelli@gmail.com

Abstract

In the survival analysis it is common that the event of interest is related to the death or the occurrence of a certain disease in individuals. All in all, at the end of the study, it is possible that one part of the sample does not come to suffer from the event of interest. These individuals may have been cured or it may be that they are immune to the event of interest. The traditional models aren't adequate in order to estimate this proportion of curing, being necessary that the statistical model embraces the curing proportion in the model. Nowadays there are many techniques to estimate the curing proportion, and one of these techniques is the mixture with cure fraction's model. We present an analysis based in the modified Weibull distribution, a three parameter distribution, in the presence of a curing fraction. Inferences for the proposed models are obtained under the Bayesian approach, using standard MCMC (Markov Chain Monte Carlo) methods. To verify this model's adequacy in the real data analysis, we considered an application to gastric adenocarcinoma patients'' data, in which the estimated model was well adjusted to the data and accordingly estimated the curing proportion.

Keywords: Analysis of survival, cure fraction, mixture model, modified Weibull distribution.

1. Introduction

In a survival analysis, the event of interest can many times be related to the death of na individual, to the occurrence or the recurrence of a disease, among others. However, it may happen that a major number of censored observations occur in the population in study, while or in the end of the study, thus indicating that there is a fraction of individuals in this population which is not subjected to the event of interest. These individuals may be immune or cured from the event of interest, considering that they do not suffer from the event of interest during the defined observation time. This fraction of immune individuals or that heal (cure) their selves is called the "cure fraction".

In traditional survival models it is not possible to estimate which is the population's curing proportion. This way, the use of statistical models which are capable of incorporating the curing fraction in its statistical modeling is necessary. This way many models emerged as a way of overcoming the traditional models' limitations, since in this new class of models it is admitted that one share of the population is immune to the event of interest. Currently, various methods to adjust such models have arisen, due to the diversity of applications found in several areas such as biomedical studies, financial, criminological, demographical, industrial reliability, market, and others. For example, in biomedical studies an event of interest may be the death of a pacient, which may have occurred due to a tumor occurrence. Working with financial data, an event of interest may be the closure of a bank client's contract due to several reasons. In criminological data, the event of interest may be the repeated incidence in a crime. In the industrial reliability area, these models can be used to verify the life proportion of the components which are put into test. And in market area, the immunes can be considered the individuals which are not going to buy a certain product (CALSAVARA, 2011, pg. 15).

One of the techniques to model the curing proportion is to adjust a parametric model, considering a mixture of two distributions: one that represents the time of survival to the susceptible individual and the other a distribution that allows to calculate the time of survival to the immune individuals. This model assumes that certain proportion π of the population are cured (or not susceptible to the event of interest) and its complement $(1 -$

2.2. The Modified Weibull Distribution

Suppose a modified Weibull distribution (WM) for susceptible individuals, with probability density function defined as

$$f_0(t) = \alpha t^{\beta-1}(\beta + \lambda t)\exp(\lambda t - \alpha t^{\beta} e^{\lambda t})$$

wherer $\alpha, \beta, \lambda > 0$ e $t > 0$. Consequently, the survival function for exposed individuals is given by

$$S_0(t) = \exp\left(-\alpha t^{\beta} e^{\lambda t}\right).$$

Note that when $\lambda = 0$, we have the WM distribution comes down to a standard Weibull distribution with two parameters (LAI et al, 2003). In Figure 1, we have graphics for survival distribution function WM according to some different choices for the values of the parameters.

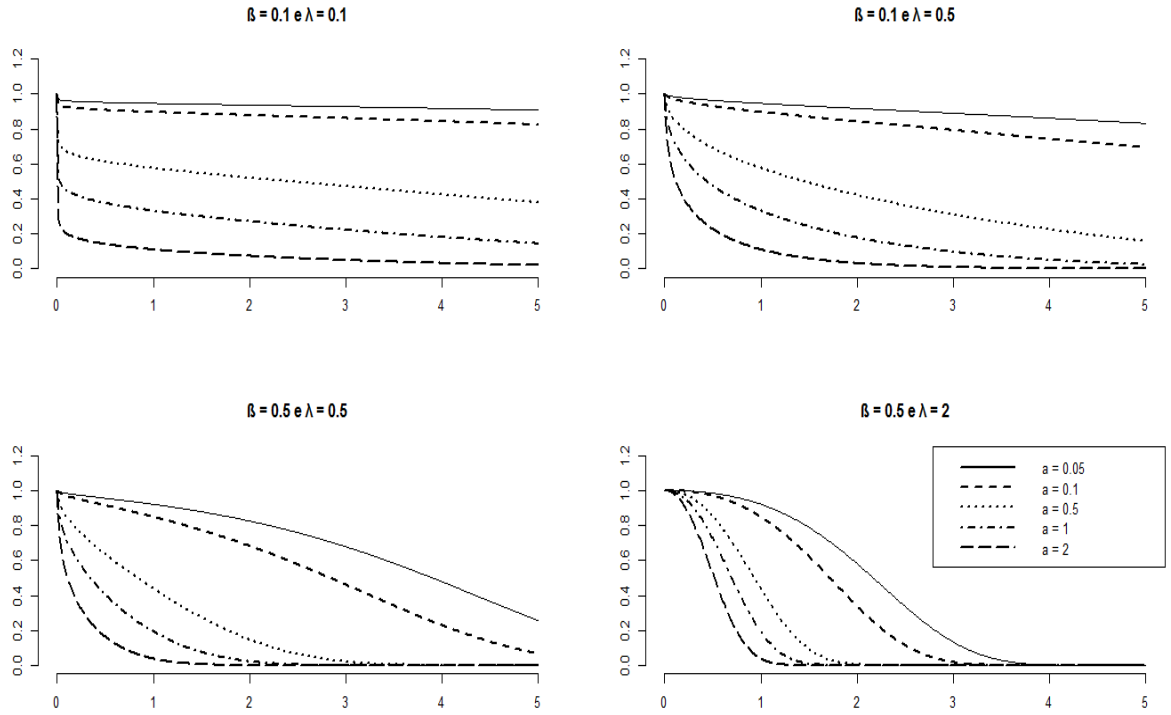

Figure 1 Graphics the survival function of a modified Weibull distribution, for some parameter values.

Through the graphs of Figure 1, we can verify how the shape of the survival distribution function WM is flexible, thereby making it an attractive distribution modeling in the survival analysis.

2.3. Bayesian inference

For a Bayesian analysis of the mixture model, we assume a uniform U(0, 1) prior distribution for the probability of cure π and Gamma(0.001, 0.001) prior distribution for the parameter α, β and γ, where Gamma (a, b) is the gamma distribution with mean a/b and variance a/b^2. We further assume prior independence among π, α, β and γ. Observe that we are using approximately non-informative priors for the parameters models.

In this paper, we assume a "burn-in-sample" of size 10,000 to eliminate the effect of the initial values used in the simulation approach; after this "burn-in-sample" period, we simulated another

200,000 Gibbs Samples, taking every 100th sample, which gives a final sample of size 2, 000. Monte Carlo estimates for the random quantities of interest are based on this final Gibbs sample of size 2,000. Convergence of the algorithm was monitored using standard methods, as the test of half-width test and stationarity test.

2.4

The OpenBUGS code used in analysis of the gastric cancerdata is given below, considering the mixture model is

```
model;
{
for (i in 1:m) {
f0t[i]  <-alpha*(pow(t[i],beta-1))*(beta+lambda*(t[i]))*exp(lambda*t[i]-alpha*(pow(t[i],beta))
                    *exp(lambda*t[i]))
s0t[i]  <- exp(-alpha*(pow(t[i],beta))*exp(lambda*t[i]))
L[i]    <- pow((1-pi)*f0t[i],d[i])*pow(pi+(1-pi)*s0t[i],1-d[i])
logL[i]  <- log(L[i])
zeros[i] <- 0
zeros[i] ~ dloglik(logL[i])
}
    pi ~ dunif(0,1)
 alpha ~ dgamma(0.001,0.001)
 beta ~ dgamma(0.001,0.001)
 lambda ~ dgamma(0.001,0.001)
}
```

In this code, n is the sample size, f0[i] is the probability density function the probability distribution modified Weibull, S0[i] is the respective survival function the probability distribution modified Weibull, L[i] is the likelihood function, t[i] is the time-to-event variable and d[i] is the censoring indicator variable (denote by δ_i). Considering that the MW distribution is not available directlyas a choice in OpenBUGS we used the dloglik() distribution, which requires us to specify the logarithm of the likelihood function.

2.5. The data

Between January of 2002 and December of 2007, it was held a retrospective study in gastric adenocarcinoma patients operated in the Cancer Hospital of Barretos (Barretos, Brazil). 201 patients were accompanied in this study, from these, 125 patients did only the resection surgery and the other 76 patients besides the surgery had a complementary treatment of chemoradiotherapy (CRT). From the patients 53,2% had censored data, being 57,9% treated with surgery and CRT and 50,4% treated only with surgery (MARTINEZ, ACHCAR, JACOME, SANTOS, 2013).

The Kaplan-Meier graphic of the survival function to these data is in Figure 2. The graphic suggests that the models that ignore the proportion π of long term survivors is not suitable for these data. The horizontal asymptote of the graphics suggest that the adequate approach is based on models that include the curing fraction.

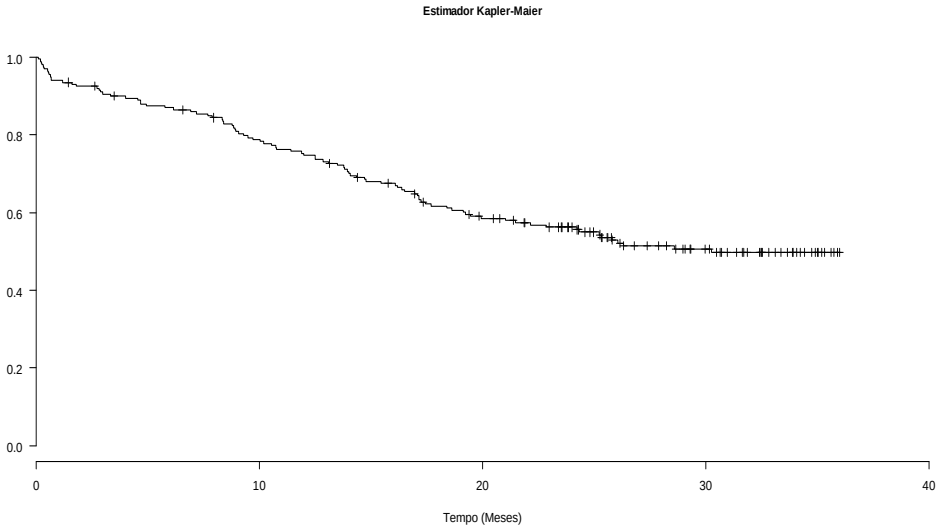

Figure 2. Graphic of Kapler-Meier's estimator for the gastric cancer data

3. Results

The following table displays the maximum likelihood estimates for the models with curing fraction based in the WM and traditional Weibull distributions. These estimates were obtained with assistance of software R, using the package "BRugs".

Parameter	Model WM		Model Weibull	
	Posterior Median	95% Credible Interval	Posterior Median	95% Credible Interval
α	0.0729	(0.0365,0.1276)	0.0481	(0.0272, 0.0834)
β	0.6475	(0.3502,0.9963)	-	-
λ	0.0564	(0.00007,0.0919)	0,9007	(0.7246, 1.1370)
π	0.4799	(0.1351,0.5621)	0,2516	(0.0151, 0.4751)
	DIC = 892,9		DIC = 898,6	

In Figure 3 we have the graphics of the survival function adjusted to the estimated parameters. In the left graphic, in blue, is the curve of WM model adjusted to the data and the horizontal line indicates the estimated π value, such value refers to the proportion of cured patients. In the right graphic, we observe that the model based on tradition Weibull distribution does not fit to the data as properly as the model based on WM distribution. The Weibull model estimates a value to the curing fraction (0, 2516) far from the one suggested by the Kaplan-Meier's graphic, it is also possible to note that the cure rate credibility interval has a value near 0, indicating that this model the cure rate is not significant, going unlike the clinical evidence. The value of the Deviance Information Criterion (DIC) is a criterion for confirming that such data gastric cancer WM fits better than the standard Weibull.

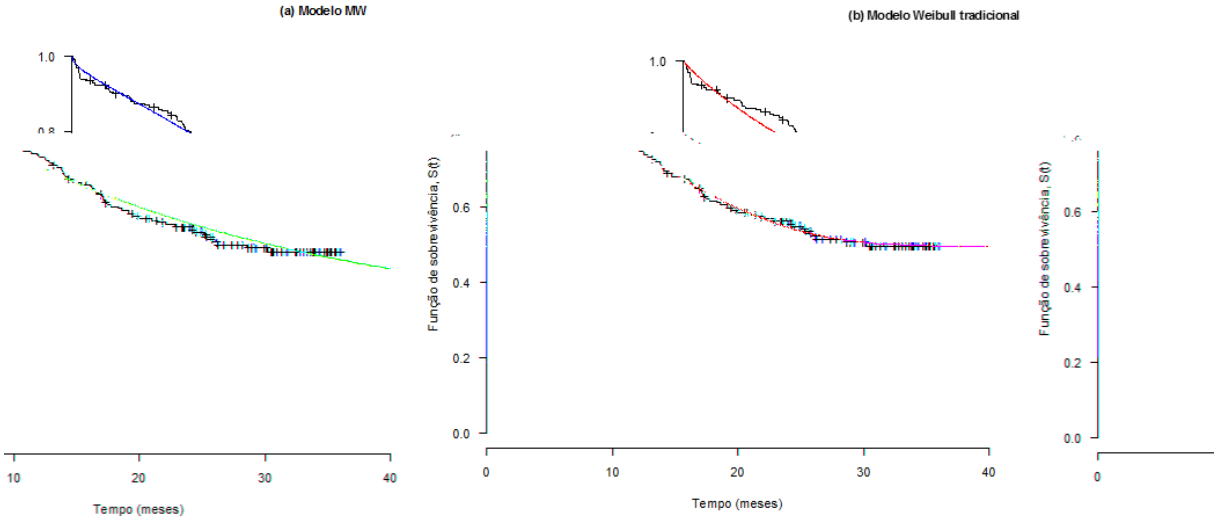

Figura 3. Kapler-Meier's estimator graphic with the estimated curve of the modified Weibull's survival function (a) and the traditional Weibull distribution (b). In the left graphic an asymptote is shown in the p proportion estimated to the gastric cancer data.

5. Conclusions

For the gastric cancer data the mixture model based on WM distribution has adjusted properly to the data and has managed to calculate considerably well the proportion of cured patients, it was the proposed objective. We observe that the model based on WM distribution is more suitable to de data than the usual Weibull, usual in cancer studies.

References

CALSAVARA, V. (2011). **Modelos de Sobrevivência com Fração de Cura usando um Termo de Fragilidade e Tempo de Vida Weibull Modificada Generalizada.** 72f. Dissertação (Mestrado) – Departamento de Estatística, Universidade Federal de São Carlos. São Carlos- SP, Brasil**.**

MALLER, R.A. & ZHOU, X. (2012) apud ACHAR, J. A., BARROS, E. D. C., & MAZUCHELI, J. Cure fraction models using mixture and non-mixture models. **Mathematical Publications**. n.51, pag. 1-9.

MARTINEZ, E.Z., ACHCAR, J.A., JACOME, A.A.A., & SANTOS, J.S. (2013)**.** Mixture and non-mixture cure fraction models based on the generalized modified Weibull distribution with an application to gastric cancer data**. Computer Methods and Programs in Biomedicine.** v.112, ed.3, pag. 343–355.

LAI, C.D., XIE, M., MURTHY, & D.N.P. (2003). **A modified Weibull distribution**. IEEE Transactions on Reliability; 52(1), 33–37.