Interval Estimation for the Parameters of the Modified Weibull Distribution Model with Censored Data: a Simulation Study

G.C. PERDONÁ[1], DMS-FMRP and FAEPA-RP, Universidade de São Paulo, 14049-900 Ribeirão Preto, SP, Brazil.

F. LOUZADA-NETO[2], DEs-UFSCar, Cx.P. 676, 13565-905 São Carlos, SP, Brazil.

Abstract. Expressing the lifetime behavior through its hazard enables us to derive special classes of failure distributions according to the hazard pattern. The usual lifetime distributions, as both exponential and Weibull models, accommodate constant (exponential) and increasing/decreasing (Weibull) hazard functions. Nevertheless, in practice, it is common to find lifetime data with hazard function of different types, for example, a U-shaped hazard function. In the present paper we investigate the properties of the modified Weibull model [8], a three-parameter model which allows U-shaped hazards to be accommodated. Inferences for this model's parameters based on both complete and censored samples are presented. We discuss different parametrizations as well as the interval estimation for the parameters of this model.

Keywords: Interval estimation, maximum likelihood estimation, Monte Carlo simulation, U-shaped hazard function, modified Weibull.

1. Introdution

The most commonly used models for accommodating U-shaped hazard functions have initial structure based on the Weibull model, which can traditionally handle constant, increasing, and decreasing hazard functions. These models can fit data with U-shaped form hazard functions as well as specify a simple Weibull model. Models presenting such property can be found in studies by [12], [11], [2], [10] for, respectively, generalized Weibull model, exponentiated Weibull model, poly-Weibull model, and polyhazard model. Lai, Xie and Murthy [8], considering a modified Weibull distribution, showed, among other properties, its characteristic of modelling U-shaped hazard functions. Both modified Weibull [8] and bi-Weibull models [2] accommodate situations where the failure rate has a U-shaped form. However, such models present distinctive structures. Regarding the bi-Weibull model, it is

---

[1] pgleici@fmrp.usp.br
[2] dfln@power.ufscar.br

assumed that each hazard function requires two Weibull components, thus allowing a flexibility for fitting decreasing hazard shape at first (by the first Weibull component) and the increasing hazard shape later (by the second Weibull component). The modified Weibull model contains a smaller number of parameters (three) than the bi-Weibull (four), where the U-shaped failure rates can be naturally and parsimoniously accommodated. The objective of the present study is to investigate the properties of the modified Weibull model based on complete and random censored samples considering two different parametrizations. In Section 2 we show the model formulation, in Section 3 we investigate the Coverage Probabilities (CP) of the confidence intervals produced by considering asymptotic theory as well as the bias and Mean Squared Error (MSE) [4] of the MLEs via a Monte Carlo simulation study. Where CP is the chance of the procedure to produce an interval including the true parameter value. The original and the logarithmic parametrizations are considered. In Section 4, we describe an application to real data, whereas some conclusions have been drawn in Section 5.

2.   Model Formulation

Consider a sample of independent random variables, $T_1, ..., T_n$, representing lifetimes with a distribution given by the modified Weibull model, where the hazard function at time t, is given by [8]

$$h(t) = \alpha(\beta + \lambda t)t^{\beta-1}\exp(\lambda t), \qquad (2.1)$$

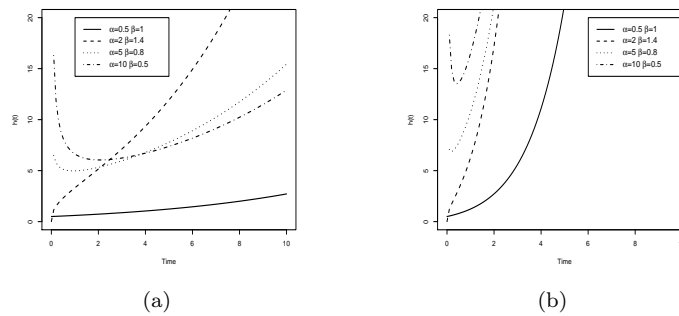where $\alpha > 0, \beta \geq 0$ and $\lambda \geq 0$.

Figure 1: Hazard function for the modified Weibull model (1) $\lambda = 0.1$ (on the left) $\lambda = 0.5$ (on the right).

This model has interesting properties regarding its formulation. The factor "$\exp(\lambda t)$" can be viewed as an accelerator factor in function of time, meaning that the parameter λ works as a lifetime fragility factor for a given individual as time elapses. In addition, for particular values of the parameters λ and β, we retrieve well-known models (in terms of their hazard functions). For $\lambda = 0$ in (2.1), we

obtain the hazard function of the Weibull distribution in, whereas for $\beta = 0$ we obtain the hazard function of extreme value distribution. Such characteristics enable greater flexibility for this model. Also, by the hazard function (2.1), the model accommodates increasing, decreasing, and non monotonous hazard curves, such as the U-shaped one (Figure 1). The shape of $h(t)$ depends only on the parameter β in $t^{\beta-1}$. For $\beta \geq 1$ the function (2.1) is increasing in t whereas for $0 < \beta < 1$, the function (2.1) initially decreases and then increases according to t, thus implying a U-shaped hazard function. If $\beta > 1$, $h(0) = 0$, and $h(0) = \alpha\beta$ if $\beta = 1$. For $0 < \beta < 1$, the hazard function (2.1) tends to infinite when time t tends to either infinite or zero and has a minimum value at $t^* = \frac{\sqrt{\beta}-\beta}{\lambda}$.Given a dataset, all information regarding the parameters is contained in the likelihood function, which is derived by considering a sample of independent random variables, $X_1, ..., X_n$, associated with survival times, and, $C_1, ..., C_n$ associated with censored times. By defining $T_i = min(X_i, C_i)$, $\delta_i = I(X_i \leq C_i)$ denoting the censoring indicator variable and considering the parameter vector $\theta' = (\alpha, \beta, \lambda)$, the likelihood function is given by [9]

$$L(\theta) = \prod_{i=1}^{n} h(t_i; \theta)^{\delta_i} S(t_i; \theta), \tag{2.2}$$

where $S(t; \theta)$ is the survival function. From (2.1) and since $S(t) = \exp(-\alpha t^\beta \exp(\lambda t))$, the logarithm of the likelihood function (2.2) is given by

$$l(\theta) = \sum_{i=1}^{n} \delta_i \left[\log(\alpha) + \log(\beta + \lambda t_i) + (\beta - 1)\log(t_i) + \lambda t_i\right] - \alpha \sum_{i=1}^{n} t_i^\beta e^{(\lambda t_i)}. \tag{2.3}$$

The first partial derivatives regarding α, β and λ are given by

$$\frac{\partial l(\theta)}{\partial \alpha} = \sum_{i=1}^{n} \frac{\delta_i}{\alpha} - \sum_{i=1}^{n} t e^{(\lambda t_i)}, \tag{2.4}$$

$$\frac{\partial l(\theta)}{\partial \beta} = \left[\sum_{i=1}^{n} \frac{\delta_i}{\beta + \lambda t_i} + \sum_{i=1}^{n} \delta_i \log(t_i)\right] - \alpha \sum_{i=1}^{n} t_i^\beta \log(t_i) e^{(\lambda t_i)} \tag{2.5}$$

and

$$\frac{\partial l(\theta)}{\partial \lambda} = \left[\sum_{i=1}^{n} \frac{\delta_i t_i}{\beta + \lambda t_i} + \sum_{i=1}^{n} \delta_i t_i\right] - \alpha \sum_{i=1}^{n} t_i^{\beta+1} e^{(\lambda t_i)}. \tag{2.6}$$

If the derivatives are equalled to zero, we can directly obtain the estimator for α, $\hat{\alpha}$, as well as for β and λ, thus solving the equations numerically.

Inferences regarding $\theta' = (\alpha, \beta, \lambda)$ can be based on the properties of the MLE for large samples given by [4]

$$\hat{\theta} \sim N\left(\theta, I^{-1}(\theta)\right), \tag{2.7}$$

where $I(\theta)$ is the Fisher information matrix, which is estimated by , $I^{-1}(\hat{\theta})$, for θ evaluated on the MLE $\hat{\theta}$.

The observed Fisher information matrix for, $\hat{\alpha}, \hat{\beta}, \hat{\lambda}$, is given by

$$I^* = \left[\begin{array}{ccc} \sum_{i=1}^n \delta_i/\alpha^2 & \sum_{i=1}^n u_i(t_i;1)e^{\lambda t_i} & \sum_{i=1}^n u_i(t_i;0)e^{\lambda t_i} t_i \\ & \sum_{i=1}^n \frac{\delta_i}{A_i^2} + \alpha u_i(t_i;2)e^{\lambda t_i} & \sum_{i=1}^n \frac{\delta_i t_i}{A_i^2} + \alpha u_i(t_i;1)e^{\lambda t_i} t_i \\ & & \sum_{i=1}^n \frac{\delta_i t_i^2}{A_i^2} + \alpha u_i(t_i;0)e^{\lambda t_i} t_i^2 \end{array} \right]_{\theta=\hat{\theta}} ,$$

$$(2.8)$$

where $u_b(t_i;b) = t_i^\beta \ln^b(t_i)$, for $b = 0, 1, 2$, and $A_i = \beta + \lambda t_i$ with $i = 1, ..., n$.

When we consider survival studies involving small and moderate datasets the asymptotic theory may not be suitable and the MLE may be biased. In this context, we investigate, in the next section, the asymptotic theory by using Monte Carlo simulation.

3.   Simulation Study

In this section we investigate the CP of the confidence intervals produced by considering asymptotic theory as well as the bias and MSE [4] of the MLEs by Monte Carlo simulation for complete and randomly 10% and 20% censored samples derived from (2.1) with sizes 20, 40, 60, and 80 for two cases in particular. The first case takes as parameter values $\alpha = 2$, $\beta = 0.8$ and $\lambda = 0.1$, and corresponds to an U-shaped hazard rate. The second case takes as parameter values $\alpha = 1$, $\beta = 1.4$ and $\lambda = 0.1$ related to data with an increasing hazard rate. A particular study is characterized by the sample size, the percentage of censoring and the shape of hazard function. Overall 24 studies were considered and, for each one, 999 datasets were simulated. According to [5] however 999 is the number of replications required to get a critical level of 0.05 from the 0.95 percentile of the empirical distribution of the test statistics (page 156 and 202). The nominal CP's were fixed at 95%. For each study, we considered three different parametrizations, the original (α, β, λ), the logarithmic parametrization $(\log(\alpha), \log(\beta), \log(\lambda))$ and the logarithmic parametrization in α $(\log(\alpha), \beta, \lambda)$.

Table 1 presents the empirical CP's for the first case ($\alpha = 2$, $\beta = 0.8$ and $\lambda = 0.1$) for complete, 10% and 20% randomly censored samples. It was found that, original parametrization (α, β, λ), the only parameter for which the the nominal CP is under and over and under estimated is the parameter α, particularly, for 20% censored samples the underestimation is striking. As for logarithmic parametrization $(\log(\alpha), \log(\beta), \log(\lambda))$, which results are in parenthesis in Table 1, the nominal CP's are under estimated for the λ parameter. The results for the logarithmic parametrization in α $(\log(\alpha), \beta, \lambda)$ are not shown because they are equal to the results of the original parametrization if we look at the CP's for the parameter β and λ. But they lead to evidence that the empirical CP's are close to the nominal ones when the logarithmic parametrization in α is considered, particulary in the presence of small samples with censored observations. Table 2 presents the bias and MSE for the first case ($\alpha = 2$, $\beta = 0.8$ and $\lambda = 0.1$) for complete, 10% and 20% randomly censored samples, considering the original parametrization (α, β, λ) and logarithmic parametrization $(\log(\alpha), \log(\beta), \log(\lambda))$, in parenthesis. The results concerning β and λ for the logarithmic parametrization were omitted since

Table 1: Coverage probability observed (95%) for approximate confidence intervals based on original parametrization and on the logarithmic parametrization, results in parenthesis, for censored samples and fixed $\alpha = 2$, $\beta = 0.8$ e $\lambda = 0.1$.

	Cens		95%				
n	(%)	$\hat{\alpha}$		$\hat{\beta}$		$\hat{\lambda}$	
	0	92.0	(98.7)	96.3	(99.8)	99.1	(73.8)
20	10	96.4	(97.7)	96.3	(99.5)	98.9	(72.3)
	20	78.8	(96.1)	96.8	(99.7)	98.6	(71.2)
	0	94.6	(98.3)	96.8	(98.7)	98.3	(76.8)
40	10	86.5	(97.2)	96.2	(98.9)	97.4	(74.3)
	20	77.6	(91.9)	95.4	(98.9)	97.8	(74.8)
	0	96.8	(99.9)	96.0	(98.8)	99.4	(74.3)
60	10	88.7	(96.2)	96.4	(98.6)	97.9	(74.5)
	20	76.3	(90.4)	96.1	(98.6)	97.8	(75.5)
	0	96.5	(99.8)	99.2	(99.9)	97.9	(79.6)
80	10	91.1	(97.2)	98.3	(99.8)	97.9	(79.4)
	20	77.1	(88.8)	98.6	(99.5)	97.8	(79.6)

Table 2: Bias and MSE of $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\lambda}$ for original parametrization as $\alpha = 2$, $\beta = 0,8$ and $\lambda = 0,1$. Results in parenthesis correspond to logarithmic parametrization.

n	Cens	$\hat{\alpha}$		$\hat{\beta}$		$\hat{\lambda}$	
	(%)	Bias	MSE	Bias	MSE	Bias	MSE
	0	-0.26 (-0.20)	0.43 (0.16)	-0.055	0.029	0.34	0.24
20	10	-0.45 (-0.32)	0.53 (0.24)	-0.057	0.031	0.37	0.28
	20	-0.65 (-0.47)	0.70 (0.37)	-0.057	0.036	0.40	0.34
	0	-0.15 (-0.11)	0.22 (0.08)	-0.025	0.017	0.21	0.11
40	10	-0.37 (-0.24)	0.33 (0.14)	-0.033	0.018	0.24	0.13
	20	-0.56 (-0.37)	0.48 (0.22)	-0.035	0.020	0.24	0.14
	0	-0.14 (-0.09)	0.15 (0.05)	-0.018	0.010	0.17	0.07
60	10	-0.35 (-0.21)	0.24 (0.09)	-0.022	0.011	0.18	0.08
	20	-0.54 (-0.34)	0.40 (0.17)	-0.024	0.013	0.19	0.09
	0	-0.09 (-0.05)	0.10 (0.11)	-0.008	0.001	0.12	0.05
80	10	-0.31 (-0.18)	0.17 (0.06)	-0.014	0.001	0.14	0.06
	20	-0.50 (-0.31)	0.33 (0.13)	-0.016	0.001	0.15	0.06

they were equal to the results for the original parametrization. The results for the logarithmic parametrization in α $(\log(\alpha), \beta, \lambda)$ are equal to the ones obtained by considering the logarithmic parametrization to α and the original parametrization to β and λ, and are also not shown here. In general, the MLEs present smaller bias and smaller MSE when the logarithmic parametrization in α is considered: The bias increase according to the percentage of censoring and decrease according to the sample size. We observed that the bias was greater for the λ parameter. Mid-range of asymptotic intervals for $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\lambda}$ in the original parametrization and their standard deviations (SD) are presents on Table 3. Tables 4, 5 and 6 show the results of the same analyzes presented in Tables 1, 2 and 3, that is, the CP, bias,

Table 3: Mid-range of asymptotic intervals for $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\lambda}$ in the original parametrization and their standard deviations(SD) for $\alpha = 2$, $\beta = 0.8$ e $\lambda = 0.1$. Results in parenthesis correspond to logarithmic parametrization

		$\widehat{\alpha}$		$\widehat{\beta}$		$\widehat{\lambda}$	
n	$Cens(\%)$	Mean	SD	Mean	SD	Mean	SD
	0	3.20 (1.81)	1.58 (0.38)	0.78	0.20	1.77	0.71
20	10	3.00 (1.92)	1.56 (0.42)	0.83	0.22	1.88	0.77
	20	2.82 (2.06)	1.55 (0.50)	0.90	0.27	2.03	0.87
	0	2.26 (1.22)	0.70 (0.20)	0.55	0.11	1.11	0.35
40	10	2.10 (1.29)	0.69 (0.22)	0.58	0.12	1.18	0.37
	20	1.95 (1.36)	0.67 (0.24)	0.61	0.13	1.24	0.40
	0	1.80 (0.97)	0.43 (0.15)	0.45	0.07	0.87	0.25
60	10	1.70 (1.03)	0.43 (0.16)	0.47	0.08	0.92	0.26
	20	1.59 (1.09)	0.43 (0.17)	0.50	0.09	0.97	0.28
	0	1.56 (0.82)	0.30 (0.12)	0.38	0.06	0.71	0.20
80	10	1.47 (0.87)	0.31 (0.13)	0.40	0.06	0.76	0.21
	20	1.38 (0.92)	0.32 (0.14)	0.43	0.07	0.80	0.23

Table 4: Coverage probability observed (95%) for approximate confidence intervals based on original parametrization and on the logarithmic parametrization, results in parenthesis, for censored samples and fixed $\alpha = 1$, $\beta = 1.4$ e $\lambda = 0.1$.

	$Cens$	95%					
n	(%)	$\hat{\alpha}$		$\hat{\beta}$		$\hat{\lambda}$	
	0	87.5	(98.3)	96.2	(99.0)	98.0	(75.0)
20	10	82.6	(98.3)	96.5	(99.0)	97.7	(73.5)
	20	77.6	(96.8)	97.2	(99.0)	97.7	(74.0)
	0	89.0	(99.5)	97.4	(99.0)	98.7	(76.3)
40	10	81.2	(96.5)	95.6	(98.8)	97.1	(73.3)
	20	74.5	(94.0)	95.6	(98.6)	96.8	(73.1)
	0	91.5	(98.5)	95.2	(98.1)	97.8	(72.3)
60	10	83.6	(96.5)	95.2	(98.1)	96.7	(72.8)
	20	72.2	(90.5)	93.7	(97.1)	95.5	(70.4)
		90.8	(97.5)	94.0	(98.1)	96.0	(73.5)
80	10	82.4	(93.3)	93.1	(96.4)	93.9	(74.3)
	20	70.1	(87.9)	92.8	(95.7)	94.5	(72.0)

MSE and Mid-range but considering the second case ($\alpha = 1$, $\beta = 1.4$ and $\lambda = 0.1$). We observed almost the same results as to the first case, with a major exception. In the first case, the bias and MSE for the parameter α are better if we use the logarithmic parametrization in α $(\log(\alpha), \beta, \lambda)$; in the second case, they are better if we use the original parametrization (α, β, λ).

Table 5: Bias and MSE of $\hat{\alpha}$, $\hat{\beta}$, $\hat{\lambda}$ in original parametrization for $\alpha = 1$, $\beta = 1.4$ and $\lambda = 0.1$. Results in parenthesis correspond to logarithmic parametrization.

n	$Cens$	$\hat{\alpha}$		$\hat{\beta}$		$\hat{\lambda}$	
	(%)	Bias	MSE	Bias	MSE	Bias	MSE
	0	-0.25 (-0.39)	0.16 (0.37)	-0.17	0.14	0.40	0.30
20	10	-0.34 (-0.52)	0.20 (0.50)	-0.18	0.14	0.41	0.32
	20	-0.43 (-0.68)	0.25 (0.73)	-0.17	0.16	0.43	0.36
	0	-0.17 (-0.26)	0.10 (0.19)	-0.11	0.07	0.25	0.15
40	10	-0.28 (-0.39)	0.14 (0.29)	-0.13	0.09	0.28	0.17
	20	-0.36 (-0.52)	0.18 (0.42)	-0.13	0.10	0.29	0.18
	0	-0.17 (-0.22)	0.08 (0.23)	-0.10	0.06	0.22	0.11
60	10	-0.26 (-0.34)	0.11 (0.21)	-0.10	0.06	0.23	0.11
	20	-0.35 (-0.48)	0.16 (0.34)	-0.11	0.08	0.25	0.13
	0	-0.15 (-0.19)	0.06 (0.34)	-0.07	0.05	0.19	0.08
80	10	-0.24 (-0.31)	0.09 (0.16)	-0.08	0.06	0.20	0.08
	20	-0.33 (-0.44)	0.14 (0.28)	-0.10	0.07	0.21	0.10

Table 6: Mid-range of asymptotic intervals for $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\lambda}$ in the original parametrization and their standard deviations(SD) for $\alpha = 1$, $\beta = 1.4$ e $\lambda = 0.1$. Results in parenthesis correspond to logarithmic parametrization.

n	$Cens(\%)$	$\hat{\alpha}$		$\hat{\beta}$		$\hat{\lambda}$	
		Mean	SD	Mean	SD	Mean	SD
	0	1.93 (2.61)	0.94 (0.63)	1.89	0.58	2.17	0.69
20	10	1.79 (2.75)	0.89 (0.65)	1.99	0.62	2.28	0.71
	20	1.66 (2.94)	0.90 (0.78)	2.15	0.74	2.45	0.83
	0	1.43 (1.75)	0.49 (0.32)	1.31	0.29	1.43	0.34
40	10	1.31 (1.84)	0.47 (0.34)	1.38	0.35	1.50	0.36
	20	1.22 (1.94)	0.47 (0.38)	1.46	0.37	1.58	0.39
	0	1.15 (1.40)	0.32 (0.23)	1.07	0.20	1.14	0.24
60	10	1.09 (1.48)	0.32 (0.24)	1.13	0.22	1.20	0.26
	20	0.99 (1.56)	0.31 (0.27)	1.19	0.26	1.26	0.29
	0	1.01 (1.19)	0.24 (0.18)	0.93	0.16	0.97	0.19
80	10	0.94 (1.25)	0.25 (0.20)	0.97	0.18	1.01	0.21
	20	0.87 (1.32)	0.25 (0.22)	1.02	0.203	1.07	0.23

4. Application

4.1. Example 1

In this section, the proposed methodology is applied to a real dataset extracted from [6] referring to the ages of 18 patients classified as "other causes of death" in a cancer study. A device called the total time on test (TTT) plot [1] is a technique used for detecting which model is more suitable for data analysis according to the hazard function behavior. This graph is obtained by plotting $G(\frac{r}{n}) = [(\sum_{i=1}^{n} T_{i:n}) + (n-r)T_{r:n}] / (\sum_{i=1}^{n} T_{i:n})$ versus r/n [12], where $r = 1, .., n$ e $T_{i:n}$, $i = 1, ..n$ are

the sample order statistics. A diagonal line in the resulting TTTplot indicates a constant hazard function, whereas if it is a convex or concave curve it indicates, respectively, a decreasing or increasing hazard function. In those cases, the simple Weibull model might be applied for fitting the data. However, if it is convex and then concave the hazard function is U-shaped, and a modified Weibull model could be more suitable. Figure 2, left panel, shows the convexity and concavity of the curve for the dataset, thus indicating a U-shaped hazard. Considering the logarithmic parametrization in α $(\log(\alpha), \beta, \lambda)$, we fit the modified Weibull model to the data. The MLEs and their asymptotic 95% confidence intervals, in parenthesis, are given by $\widehat{\alpha} = 0.041$ (0.0270; 0.0641), $\widehat{\beta} = 0.456$ (0.3505 ; 0.5905) and $\widehat{\lambda} = 0.0138$ (0.0012 ; 0.0165).
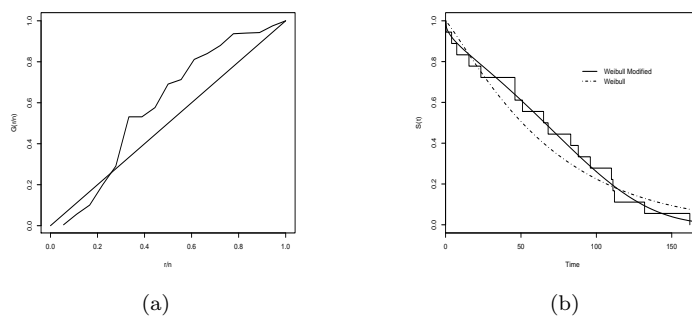

(a)   (b)

Figure 2: Empirical Scaled TTT-Tranforms (left) and empirical K-M and fitted survival for the dataset [6] (right).

Figure 2, right panel, shows the Kaplan-Meier estimator and the fits of the simple Weibull and modified Weibull models indicating the suitability of the modified Weibull model. This results are corroborated by the likelihood ratio statistic, which is equal to 6.57.

4.2.   Example 2

The data set refers to the serum-reversal time (days) of 368 children contaminated with HIV from vertical transmission at the university hospital of the Ribeirão Preto School of Medicine (Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto) from 1986 to 2001 [13] born from untreated mothers follow-up by two years. Serum-reversal is a process of disappearance of anti-HIV antibodies (antitoxin) in blood (neutralization of anti-HIV serology) in an individual who previously showed positive anti-HIV serology. Serum-reversal can occur in children born from mothers infected with HIV. Their children are born with positive anti-HIV serology (vertical HIV transmission), which can occur due to the intrauterine or intra-parturition transplacental transmission of the mother's antibodies to her baby during labor or in the period following childbirth while the infant is breastfed. Thus the presence of antibody by itself in an infant younger than 18 months is not diagnostic of infection [3] [7], because after a few months, the mother's antibodies are eliminated and the

anti-HIV serology changes from positive to negative. In this dataset the infection was confirmed with the polymerase chain reaction (PCR) assays for HIV-1 RNA in plasma.

Considering again the logarithmic parametrization in α $(\log(\alpha), \beta, \lambda)$, we fit the modified Weibull model to the data. The MLEs and their asymptotic 95% confidence intervals, in parenthesis, are given by $\widehat{\alpha} = 0.0002$ $(0.00018; 0.0025)$, $\widehat{\beta} = 0.8392$ $(0.8047; 0.8751)$ and $\widehat{\lambda} = 0.00583$ $(0.0058 ; 0.0059)$.

(a) (b)
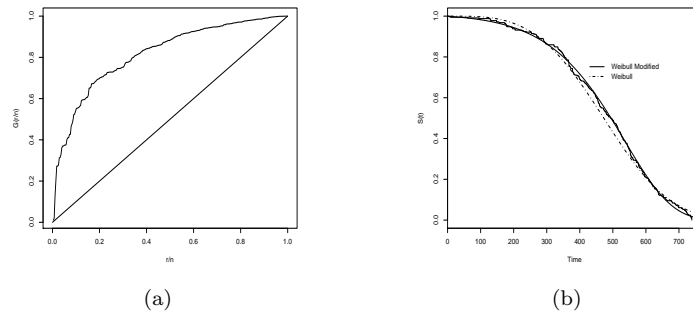
Figure 3: Empirical Scaled TTT-Tranforms (left) and empirical K-M and fitted survival for the dataset [13] (right).

Figure 3 the empirical survival function and the estimated survival function of the simple Weibull and modified Weibull models provides a good fit for the data of the modified Weibull model. This results are corroborated by the likelihood ratio statistic, which is equal to 35.45.

5.   Final Comments

The modified Weibull model was found to be an alternative for fitting U-shaped hazard data. Our Monte Carlo simulation study has shown that the asymptotic confidence interval procedure can be straightforwardly used if a logarithmic parametrization in α is considered, even in the presence of a small dataset with censored observations. The code related to the simulation study and to the application can be obtained by emailing the authors.

Acknowledgements

Resumo. Representar o tempo de vida pela função de risco nos permite trabalhar com especiais classes de distribuições de risco adequando-as a necessidade do fenômeno. As distribuições de probabilidade de sobrevivência como exponencial, Weibull, acomodam formas de risco constante (exponencial) e formas crescente e decrescente (Weibull). No entanto, na prática, é comum encontrarmos dados com

diferentes formas na representação do risco, por exemplo, uma função em forma de U. Neste trabalho nós investigamos as propriedades do modelo modificado Weibull [8], um modelo com três parâmetros que permite modelar funções de risco em forma de U. Inferências dos parâmetros do modelo baseadas em dados completos e censurados são apresentados. Discutimos diferentes parametrizações, bem como estimativas intervalares para os parâmetros desse modelo.

References

[1] M.V. Aarset, How to identify bathtub hazard rate, *IEEE Transactions on Reliability*, **36**, No. 1 (1987), 106-108.

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

[3] Biggar et al., Perinatal intervention trial in Africa: effect of a birth canal cleansing intervention to prevent HIV transmission. *Lancet 15*, **347** (1996), 1647-50.

[4] D.R. Cox, D.V. Hinkley, "Theoretical Statistics", Chapman and Hall, London, 1974.

[5] A.C. Davison, D.V. Hinkley, "Bootstrap Methods and their Application", Cambridge University Press, Cambridge, 1997.

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

[7] W.W. Hauck, L.J. McKee, B.J. Turner, Two-part survival models applied to administrative data for determiningrate of and predictors for maternal–childtransmission of HIV, *Statistics in Medicine*, **16** (1997), 1683– 1694.

[8] C.D. Lai, M. Xie, N.P. Murthy, A modified Weibull distribution, *IEEE Transactions on Reliability*, **52** (2003), 33-37.

[9] J.F. Lawless, "Statistical Models and Methods for Lifetime Data", John Wiley, New York, 1982.

[10] F. Louzada-Neto, Polyhazard regression models for lifetime data, *Biometrics*, **55** (1999), 1281-1285.

[11] G.S. Mudholkar, D.K. Srivastava, Exponentiated Weibull family for analyzing bathtub failure-rate data, *IEEE Transactions on Reliability*, **42** (1993), 99-302.

[12] G.S. Mudholkar, D.K. Srivastava, G.D. Kollia, A generalization of the Weibull distribution with application to the analysis of survival data, *Journal of the American Statistical Association*, **91**, No. 436 (1996), 1575-1583.

[13] A.N.F. Silva, "Estudo Evolutivo das Crianças Expostas ao HIV e Notificadas pelo Núcleo de Vigilância Epidemiológica do HCFMRP-USP", M.Sc. Thesis, University of São Paulo, Brazil, 2004.