UNIVERSITI TEKNOLOGI MARA Cawangan Kedah Kampus Sungai Petani

2021 iCMS INTERNATIONAL CONFERENCE ON COMPUTING, MATHEMATICS AND STATISTICS

e-PROCEEDINGS

of The 5th International Conference on Computing, Mathematics and Statistics (iCMS2021)

4-5 August 2021

Driving Research Towards Excellence

e-Proceedings
of the 5th International Conference on Computing, Mathematics and Statistics (iCMS 2021)

Driving Research Towards Excellence

e-ISBN: 978-967-2948-12-4
DOI

Library of Congress Control Number:

Publication by
Department of Mathematical Sciences
Faculty of Computer & Mathematical Sciences
UiTM Kedah

TABLE OF CONTENT

PART 1: MATHEMATICS

PART 2: STATISTICS

PART 3: COMPUTER SCIENCE & INFORMATION TECHNOLOGY

PART 4: OTHERS

ACCELERATED FAILURE TIME (AFT) MODEL FOR SIMULATION PARTLY INTERVAL-CENSORED DATA

Ibrahim El Feky[1] and Faiz Elfaki[2,*]
[1,2]Department of Mathematics, Statistics and Physics, College of Arts and Sciences, Qatar University, P. O. Box 2713, Doha, Qatar
([1]ie1107474@student.qu.edu.qa, [2,*]felfaki@qu.edu.qa)

The performance of maximum likelihood estimators of the parameters of Accelerated Failure Time (AFT) regression model based on Weibull distribution with simple imputations methods under Partly-Interval Censored (PIC) data is studied. The proposed model is tested through the simulation data and the result were compare with semiparametric model. The result indicates that the AFT with Weibull distribution is comparable with Cox model under simulated PIC data via breast cancer data in the present of imputation techniques. In additional to that the result suggested that the parameters of our model are stable, and the treatments are significant for simulated breast cancer patients.

Keywords: Accelerated Failure Time model, partly interval-censored, imputation techniques.

1. Introduction

One of the important methods in statistics that used for analyzing time to event is survival analysis. There is common parametric model in survival analysis underline the exponential, Weibull, Gompers, Lognormal and Log logistic distributions. The AFT and Proportional Hazards Regression Model (PHRM) are derived from a Log logistic (Wei, 1992). There are important nonparametric methods in analysis of survival such as: Kaplan Meier estimator, PHRM and log-rank test. The purpose of this study is to study one of the important models in industrial fields and clinical trials which AFT model that provides an alternative to the commonly used PHRM (Wei, 1992; Saikia and Barman, 2017). In additional to that the data for survival analysis must involve censoring observation that is; right, left and interval-censoring. When the subject involved exact and interval censored data is called a PIC. Several authors used PIC in their research such as; Kim (2003), Alharpy and Ibrahim (2013), Zyoud et al. (2016), Gao et al. (2017) and Saeed and Elfaki (2020). However, in this study PIC will be use based on the simulation data sets via median and random imputations methods. These simple imputations techniques are used to handle the missing data as well to create the exact observations in PIC data.

2. Accelerated Failure Time model

It's a huge challenge when estimating the regression parameters of the AFT model in the presence of PIC since the actual failure times are not observed directly and their exact values are unknown. This literature suggested that the parameter estimating methods are mainly the basis of the estimated distribution function of censored observations through the censored intervals and assumed the examination times that to be laid between censored failure time. However, if the actual failure times are not available, parametric method for estimating the parameters is very appealing to researchers since it does not involve complicated statistical assumptions about the distribution of the variables in the model and the error terms of the model. Several researches used AFT model in their studies such as; Pike (1966) studied the AFT model based on the carcinogenesis data. Nelson and Hann (1972) determine the relationship between temperature and failure time data using AFT. Kalbfleisch and Prentice (2002) introduced of survival semi-parametric model, that is a class of log-linear models for time. The explanatory variable act multiplicatively on survival time in AFT.

Brown and Wang (2007) proposed weighted rank estimators that involves smoothing method based on semi-parametric AFT models. Johnson and Strawderman (2009) extended the method of Brown and Wang

(2007) for semiparametric AFT model based on Newton Raphson algorithm and other common numerical methods. However, in this research we will use AFT model based on Weibull distribution for cancer PIC data. Liu and Lim (2018) introduced that the Weibull AFT model for a subject ($i = 1,2, \ldots, n$) which takes the form;

$$logT_i = \acute{\theta} z_\iota + \sigma\varepsilon_i \tag{1}$$

where random variables $T_1, T_2, \ldots, T_n$ denote the failure times of n independent subjects, $z_1, z_2, \ldots, z_n$ denote their associated $p \times 1$ vector of covariates, θ is a $p \times 1$ vector of unknown regression parameters, σ is a scale parameter and ε_i's are the error terms of the model, which are independent and identically distributed according to a Gumbel.

Presently the work of survival for the ith individual (during that point) related $p \times 1$ vector of covariates which is equivalent to the work of the survival for the outrageous worth distribution ε_i.

$$S_z(z_i) = S_{\varepsilon\mathrm{i}}(z_i) \tag{2}$$

where

$$z_i = \frac{\mathrm{logT} - \mu - \beta_1 z_1 - \cdots - \beta_p z_p}{\sigma} \tag{3}$$

Also, S_Z is a survival for the ith individual, $S_{\varepsilon\mathrm{i}}$ is a survival for the outrageous worth distribution and $\beta_1, \ldots\ldots\beta_p$ are the regression coefficients of interest, z_i is called the ith individual at the time t_i (which is associated vector of covariates).

3. Simulation Study

Simulation studies are computational tests including results creation through pseudo-random sampling from built up likelihood appropriations. This studies are an irreplaceable statistical analysis asset, particularly in the evaluation of current tactics and the assessment of other methodologies.

Under certain circumstances, simulation tests are utilized to deliver logical information on the statistical strategies productivity, instead of algebraic tests which is a more common analytical that can incorporate different circumstances. It is not easy to achieve empirical tests for a lot of people. This problem however is solved by our simulation experiments.

They are independent where messy systems settle on inaccurate choices or results so they could decide the sturdiness of strategies under these conditions. It is not true for observational discoveries, as the outcomes may be duplicated where the aspects are removed from a specific model.

A simulation study was done dependent on the true information of breast malignant featured in this hypothesis to look at the impact of the AFT via Weibull model and to assess the covariates in the informational indexes.

Adding up to that sample, we used 20000 times for each treatment (Chemotherapy, Hormone, Radiotherapy (RT) and Surgery). Using mean and standard deviation of 0.0759168 & 0.0227825 for chemotherapy treatment, in light of (0%, 25%, half, and 75%) as levels of accurate perception for the PIC data based on the real data set which is not addressed here in this paper (reader may refer to El Feky, 2021). We acquired the measurements of endurance in every simulation data for the two gatherings of every treatment that depends on the precise observation contrasted with the one evaluated by imputation strategies that is; median and random point through our model. In each simulation data we obtained the function of survival for the two groups (with treatment and without treatment) of each treatment that based on the exact observation compared to the one estimated by mentioned imputations methods.

Table 1: Results from chemotherpay and hormone obtanied by AFT model with median imputaion based on simulation data

Type of Parameter	Percentage of Exact observation	Estimate	SE of Estimation	P-value
		Chemotherapy		
Coefficient	0	0.094042	0.004943	2e-10
Shape		2.903226	0.081959	
Scale		972.7186	3.479775	
Coefficient	25	0.094244	0.004944	2e-16
Shape		2.904272	0.016458	
Scale		972.2322	3.478308	
Coefficient	75	0.094223	0.004945	2e-16
Shape		2.902097	0.016454	
Scale		972.3008	3.479340	
		Hormone		
Coefficient	0	0.167267	0.005038	2e-10
Shape		2.880782	0.016654	
Scale		920.6089	3.389513	
Coefficient	25	0.167590	0.005046	2e-16
Shape		2.876051	0.016633	
Scale		920.1505	3.392943	
Coefficient	75	0.167824	0.005063	2e-16
Shape		2.866203	0.016587	
Scale		919.7937	3.401836	

Table 1 display the outcomes from AFT model based on median point for chemotherapy and hormone treatment with different percentages of exact and interval censored data. It showed significant results with respect Standard Deviation (SD) of the estimators and their p-value. These results indicate that for more exact observation in the data the result is better (as value of AIC=282573.8 when 75% exact compared to AIC= 282597.3 for 0% exact). Moreover, the chance of survival increases significantly when patient used both treatments compared with a patient who have not gone through either treatment while fighting with breast cancer (Figures 1 & 2).

Table 2 display the results for the Radiotherapy (RT) and surgery treatments based on random imputation with and without both treatments through different exact observations of PIC with 0%, 25% and 75%. The patient treated with both treatments have a long survival compared to those without treatment while fighting with breast cancer.

These results indicate that when we have more exact observation the result is more reasonable for PIC data. Moreover, the chances of survival increase significantly when patient use the above-mentioned treatments compared with a patient who haven’t gone through any of treatment while fighting with breast cancer as showed in Table 1 & 2 and Figure 1, 2, 3 and 4. In additional to that, Figure 5 showed that the AFT with Weibull distribution via PIC breast cancer data under median imputation is comparable with Cox model based on Chemo treatment.

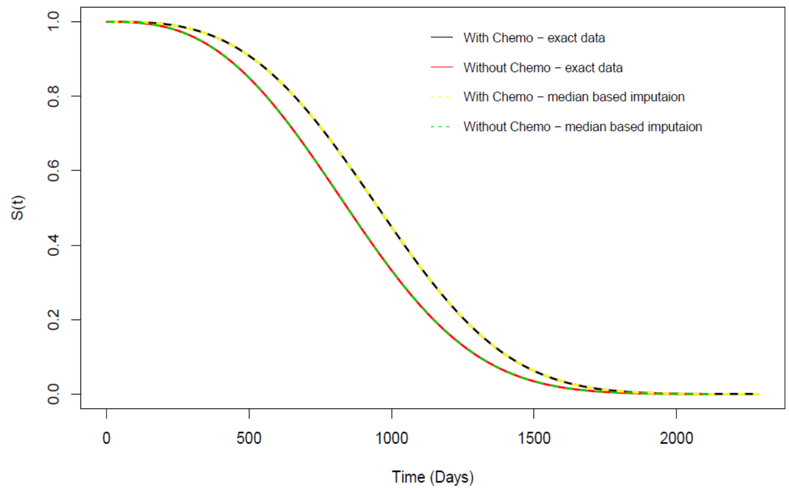

Figure 1: The function of survival based on chemotherpy with 0% exact observation from median imputaion via AFT model.

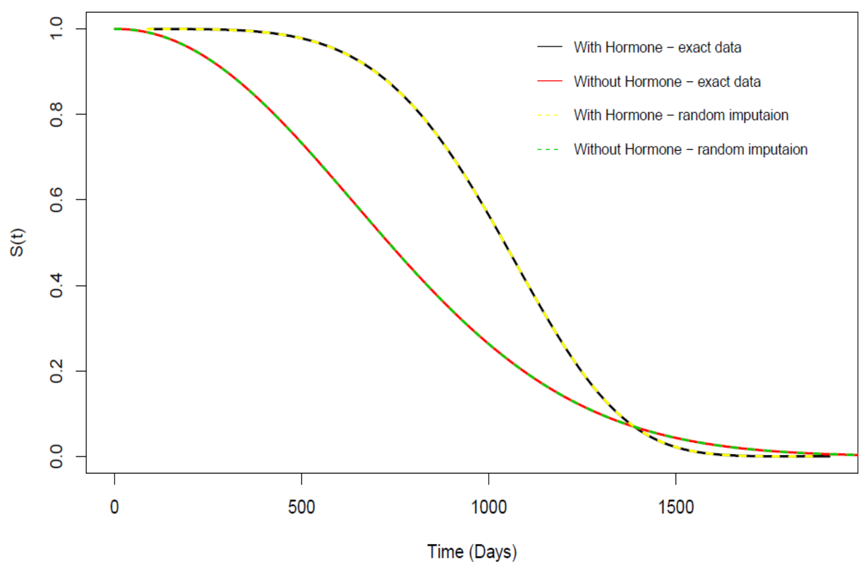

Figure 2: The function of survival based on hormone with 25% exact observation from random point imputation via AFT model.

Table 2: Results from Radiotherapy (RT) and Surgery obtanied by AFT model with median imputaion based on simulation data

Type of Parameter	% of Exact observation	Estimate	SE of Estimation	P-value
		Radiotherapy		
Coefficient	0	0.134629	0.004918	2e-10
Shape		2.941726	0.016868	
Scale		947.5181	3.400072	
Coefficient	25	0.134742	0.004925	2e-16
Shape		2.937418	0.016850	
Scale		947.2983	3.403819	
Coefficient	75	0.135145	0.004928	2e-16
Shape		2.935213	0.016841	
Scale		946.8905	3.404387	
		Surgery		
Coefficient	0	0.164277	0.00531	2e-10
Shape		2.704938	0.01545	
Scale		887.0654	3.42914	
Coefficient	25	0.164408	0.00531	2e-16
Shape		2.703855	0.01545	
Scale		886.9470	3.42988	
Coefficient	75	0.164177	0.00532	2e-16
Shape		2.699769	0.01543	
Scale		886.9672	3.43458	

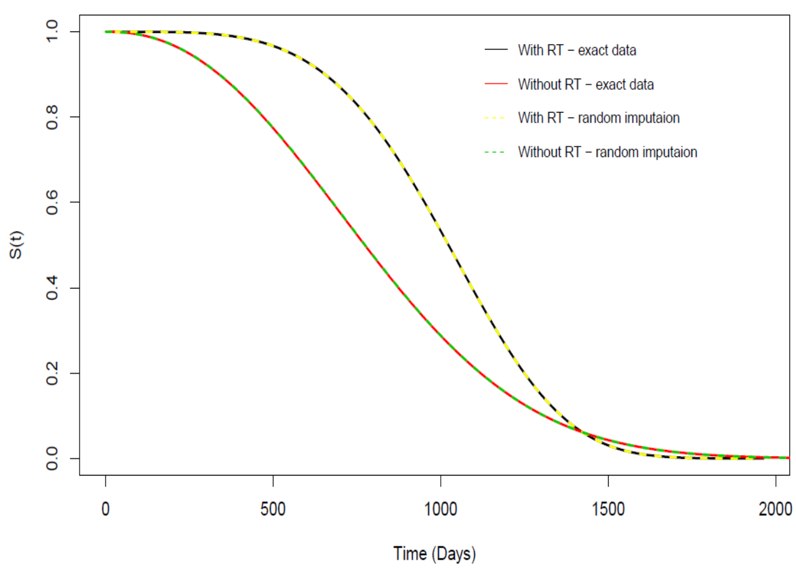

Figure 3: The function of survival based on radiotherapy with 25% exact observation from random point imputation via AFT model.

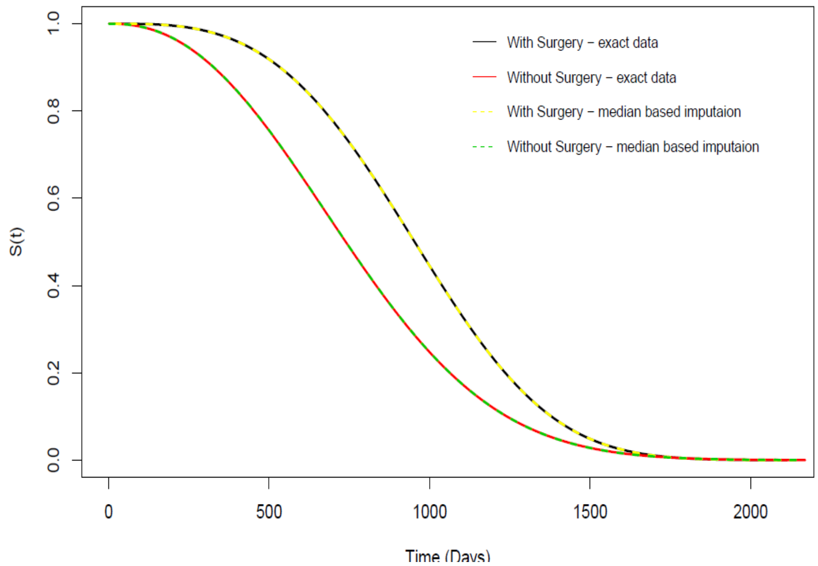

Figure 4: The function of survival based on surgery with 75% exact observation from median point imputation via AFT model.

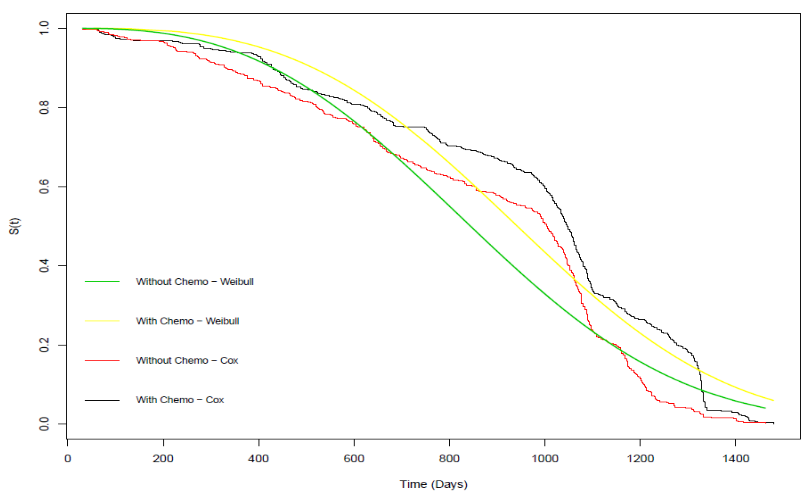

Figure 5: The function of survival based on Chemo from median point imputation via AFT model compare with Cox model.

4. Concluding

In this study, AFT model is used based on Weibull distribution via simple imputation technique to simplify the procedure for PIC data which are the median and random imputations. The estimated survival function were computed based on the maximum likelihood estimation. The simulation data was used based on the real breast cancer data (which not addressed here in this paper). The data generated for 2000 times from the every medicines that is hormone, surgery, chemotherapy and RT.

It can be concluded that AFT model with various attributions techniques fits the data well particularly when the data is PIC. In additional, when there is more exact observations in data the result are more accurate which similarly to the finding by other researchers such as Kim (2003), AL harpy and Ibrahim (2013), Zyoud et al. (2016).

References

Alharphy and Ibrahim (2013). Parametric Test for Partly Interval-censored Failure Time Data under Weibull Distribution via Multiple Imputation. *Journal of Applied Science*, 13(4):621-626.

Brown B. M and Wang Y. G. (2007). Induced smoothing for rank regression with censored survival times. *Statist Med.* 26:828–836.

El Feky, I. (2021). *Accelerated Failure time for Weibull distribution based partly interval censored data*. Master Thesis, Qatar University, Qatar.

Gao, F., Zeng, D., and Lin, D. Y. (2017). Semiparametric estimation of the accelerated failure time model with partly interval-censored data. *Biometrics*, 73(4):1161-1168.

Johnson L. M, and Strawderman R. L. (2009). Induced smoothing for the semiparametric accelerated failure time model: asymptotic and extensions to clustered data. *Biometrika*, 96(3):577–590.

Kalbfleisch J. D, and Prentice R. L.(2002) *The Statistical Analysis of Failure Time Data.* (2nd ed.) Hoboken, New Jersey: Wiley-Interscience.

Kim. J. S. (2003). Maximum Likelihood Estimation for the Proportional Hazards Model with Party Interval-Censored Data. *J. R. Statist. Soc.* Series B 65:489-502.

Liu, E. and Lim, K (2018). Using the Weibull accelerated failure time regression model to predict time to health events. Cold Spring Harbor Laboratory. https://doi.org/10.1101/362186

Nelson, W.B. and Hahn, G.J.(1972). Linear Estimation of a Regression Relationship from Censored data: 1,Simple Methods and Their Applications (With Discussion). *Technometrics*,14:247-276.

Pike, M. C. (1966). A method of analysis of a certain class of experiments in carcinogenesis, *Biometrics*, 22:142-61.

Saeed N. M, and Elfaki F. A (2020). Parametric Weibull Model Based on Imputations Techniques for Partly Interval Censored Data. *Austrian Journal of Statistics*, 73(49):30-37.

Saikia, R., and Barman, M. P. (2017). A Review on Accelerated Failure Time Models. *International Journal of Statistics and Systems*, 12(2):311-322.

Wei, L. J. (1992) The Accelerated Failure Time Mode: A Useful Alternative to the Cox Regression Model in Survival Analysis. *Statistics in Medicine*, 11:1871-1879.

Weibull, W. (1939). *A statistical theory of the strength of materials*. Generalstabens Litografiska Anstalts Förlag, Stockholm.

Zyoud, Elfaki and Hrairi (2016). Non-parametric estimate based in imputations techniques for interval and partly interval censored data. *Science International (Lahore)*, 28:879-884.

2021 ICMS

INTERNATIONAL CONFERENCE ON COMPUTING, MATHEMATICS AND STATISTICS

e ISBN 978-967-2948-12-4

9 7 8 9 6 7 2 9 4 8 1 2 4