Peer-reviewed version available at Stats 2018, 1, 13; doi:10.3390/stats1010013

Article

A Non-mixture Cure Model for Right Censored Data with Fréchet Distribution

Durga Kutal [1*] and Lianfen Qian [2+]

[1] Department of Mathematics and Statistics, Wake Forest University; kutaldh@wfu.edu
[2] Department of Mathematical Sciences, Florida Atlantic University; lqian@fau.edu
* Correspondence: kutaldh@wfu.edu; Tel.: +1-561-419-4604

Abstract: This paper considers a non-mixture cure model for right censored data. It utilizes the maximum likelihood method to estimate model parameters in the non-mixture cure model. The simulation study is based on Fréchet susceptible distribution to evaluate the performance of the method. Comparing with Weibull and exponentiated exponential distributions, the non-mixture Fréchet distribution is shown to be the best in modeling a real data on allogeneic marrow HLA-matched donors and ECOG phase III clinical trial e1684 data.

Keywords: Non-mixture model, Fréchet Distribution, Right Censored Survival Data, Maximum Likelihood Method.

1. Introduction

The cure fraction models are broadly used for analyzing survival data. In the literature, there are two major models to fit survival data with cure fraction. The first one is the mixture cure rate model, also known as standard cure rate model. This model was initially introduced by Boag [1] in 1949 and further developed by Berkson and Gage [2] in 1952 and later extensively studied by other authors. In this model, it is assumed that certain proportion of population is cured and the remaining is not cured. In order to estimate cure fraction, parametric, semi-parametric and non-parametric methods have been studied by several authors. Farewell [3] in 1982 used a Weibull distribution for uncured subjects and logistic regression for cure probability. Goldman [4] in 1984 discussed parametric survivorship analyses using maximum likelihood estimation and the likelihood ratio test. Taylor [5] in 1995 proposed semi-parametric method to mixture model using logistic regression for incidence part and kaplan-meier for latency part. Peng and Dear [6] in 2000 investigated non-parametric approach to mixture model to estimate parameter of interest in the model using EM algorithm, marginal likelihood approach, and multiple imputations. They also extended to Cox's proportional hazards cure model. Kuk and Chen [7] in 1992 also proposed semi-parametric cure model using the logistic regression for cure probability and the proportional hazard regression model for failure time. Zhang and Peng [8] in 2009 proposed a mixture cure model where the covariate effects on the proportion of cure and the distribution of the failure time of uncured patients are separately modeled. Moreover, Kim and Jhun [9] in 2008 investigated interval censored data based on mixture cure model. They used to derive the likelihood in interval censored data based on an approximate likelihood approach suggested by Goetghebeur and Ryan [10] in 2000.

The second one is non-mixture cure rate model, also known as bounded cumulative hazard model and promotion time cure model. In cancer study, this model was developed based on assumption

+ Dr.Lianfen Qian: Distinguished Guest Professor of Wenzhou University, Wenzhou, China.

Peer-reviewed version available at Stats 2018, 1, 13; doi:10.3390/stats1010013

that the number of cancer cells that remain active after cancer treatment and that may grow slowly and produce a detectable cancer, which assumed to follows a Poisson distribution. This model was first proposed by Yakovlev *et al*. [11] in 1993 and was further discussed by Chen *et al*. [12] in 1999. The semi-parametric approaches of estimation for survival data with a cure fraction have been discussed by Chen *et al*. [13] in 2001. Tsodikov *et al*. [14] in 2003 provided a review of existing methodology of statistical inference based on the non-mixture model. They have highlighted that there are the distinct advantages of the non-mixture cure model: the non-mixture rate model has proportional hazard model structure, the non-mixture cure model present a much more biologically meaningful interpretation of the results of the data analysis and the non-mixture cure model is easy in computations due to its simple structure for the survival function which can provide a certain technical advantage when developing maximum likelihood estimation procedures. Herring and Ibrahim [15] in 2002 studied the parametric estimation of random effects for non-ignorable missing covariates in non-mixture cure model. Uddin *et al*. [16,17] in 2006 approached both non-parametric and parametric methods in non-mixture model for uncensored data. Liu *et al*. [18] in 2009 investigated the semi-parametric non-mixture cure model for interval censored data using the EM method. Lopes *et al*. [19] in 2012 studied both bayesian and clasical approaches to long-term survivors and random effects based on non-mixture cure model.

The survival models with long term survivors have been studied for decades. The survival data with long term survivors categorize into two groups: susceptible individuals and long term survivors. In this paper, we considered generalized extreme value type II distribution for long term survivors. The generalized extreme value distribution was introduced by Jenkinson [20] in 1955 and its sub-models are widely used for modeling extreme events. The generalized extreme value distribution sub-models are Gumbel, Fréchet and Weibull distributions. An extreme value distributions play an important role in statistics. Ramos *et al*. [21] in 2017, discussed some results for long term Fréchet distribution. We study a long term survival non-mixture cure model with Fréchet distribution for uncured population. Simulation study is conducted to show the performance of the maximum likelihood estimators. Moreover, we compare the our proposed model with other established models using the real data sets of Lukemia for 46 allogeneic marrow transplantation patients presented by Kersey *et al*. [22] in 1987 and ECOG phase III clinical trial e1684 available in R package *smcure*.

2. Likelihood Function of a Mixture Cure Model

A mixture cure model, also known as standard cure rate model, refers to a class of model for survival data with long term survivors when some of them will not develop the event of interest. In a mixture cure model, the population consists of two types of subjects: uncured (susceptible) subjects who experience the event of interest and cured (non-susceptible) subjects who will never experience it. Those who are not going to develop the event of interest are referred to as "cured subjects,"or "long-term survivors ".

Let T denote the survival time of a subject and let Δ be the cure indicator with $\Delta = 0$ when the subject is cured and $\Delta = 1$ when the subject is uncured. Let p be the proportion of subjects those are cured, and $1-p$ be the proportion of subjects those are not cured. That is, $P(\Delta = 1) = 1-p$ and $P(\Delta = 0) = p$. So, $P(T \leq t|\Delta = 0) = 0$ and $P(T \leq t|\Delta = 1) = F(t)$, the cumulative distribution function of uncured subjects. Then the cumulative distribution function of the overall population T is

$$\begin{aligned} F_T(t) &= P(T \leq t) = P(T \leq t|\Delta = 0)P(\Delta = 0) + P(T \leq t|\Delta = 1)P(\Delta = 1) \\ &= (1-p)F(t) \end{aligned}$$

and the survival function of T is

$$S_T(t) = 1 - F_T(t) = 1 - (1-p)F(t) = p + (1-p)S(t), \tag{1}$$

where $S(t)$ denotes a proper survival function for the uncured patients. Notice that $S_T(t)$ is not proper. The corresponding density function of T is

$$f_T(t) = (1-p)f(t), \tag{2}$$

where $f(t)$ is the probability density function of uncured patients.

Suppose y_i be the right censored survival time for subject i, then $y_i = \min(T_i, C_i)$, where T_i be the failure time of the ith subject and C_i be the right censored variable of the ith subject, $i = 1, \cdots, n$. The observed survival time of ith subject is y_i and the censoring indicator is δ_i, where $\delta_i = I(T_i \leq C_i)$. Then the likelihood function of the mixture model is

$$L_m = \prod_{i=1}^{n} f_T^{\delta_i}(y_i) S_T^{1-\delta_i}(y_i) = \prod_{i=1}^{n} [(1-p)f(y_i)]^{\delta_i} [p + (1-p)S(y_i)]^{1-\delta_i}$$

and the log-likelihood function is

$$l_m \equiv \ln(L_m) = \ln(1-p) \sum_{i=1}^{n} \delta_i + \sum_{i=1}^{n} \delta_i \ln(f(y_i)) + \sum_{i=1}^{n} (1-\delta_i) \ln(p + (1-p)S(y_i)). \tag{3}$$

3. Likelihood Function of a Non-mixture Cure Model

In this section, we introduce a non-mixture cure model developed by [11] in 1993 for survival data with long-term survivors in cancer study. Let N be the number of cancer cells for a patient after cancer treatment. Since the number of cancer cells may grow rapidly and may later produce a detectable cancer disease, the number of cancer cells, N, is assumed to have a Poisson distribution with mean λ. Let Z_j be the random time for the jth cancer cell to produce a detectable cancer mass. Then, the time to relapse of cancer can be defined by the random variable T such that $T = min\{Z_j, j = 1, 2, ..., N\}$. Furthermore, we assume that Z_j are independently and identically distributed with distribution and survival functions $F(\cdot)$ and $S(\cdot)$, respectively. Then the survival function of T is given by:

$$\begin{aligned} S_T(t) &= P[\text{No cancer by time t}] \\ &= P[N = 0] + P[Z_1 > t, Z_2 > t, ..., Z_N > t, N \geq 1] \\ &= e^{-\lambda} + \sum_{N=1}^{\infty} S^N(t) \frac{\lambda^N}{N!} e^{-\lambda} = e^{-\lambda + \lambda S(t)} = e^{-\lambda F(t)} = p^{F(t)}, \end{aligned}$$

where $p = e^{-\lambda}$ is the probability of cure or cure fraction in this model since

$$\lim_{t \to \infty} S_T(t) = p = e^{-\lambda} > 0,$$

which implies that $S_T(\cdot)$ is an improper survival function. The cumulative distribution, density and hazard functions of T are given below respectively,

$$F_T(t) = 1 - p^{F(t)}, \quad f_T(t) = -\ln(p) f(t) p^{F(t)}, \quad h_T(t) = -\ln(p) f(t).$$

Since $S_T(t)$ is not proper survival function, then we can derive the following theorem:

Theorem 1. *Let $0 \leq p \leq 1$. For $T \sim F_T(\cdot) = 1 - p^{F(\cdot)}$ and $F(\cdot)$ is a strictly increasing proper distribution function, then $Y = F_T(T)$ has an improper uniform distribution over $[0, 1-p]$.*

Proof of Theorem 1. Notice that as $t \to \infty$, $p^{F(t)} \to p$. So,

$$\lim_{t\to\infty} F_T(t) = \lim_{t\to\infty} \left(1 - p^{F(t)}\right) = 1 - p.$$

Since $F(\cdot)$ is strictly increasing which implies that $F_T(\cdot)$ is also strictly increasing, it has an inverse, denoted by $F_T^{-1}(\cdot)$. Then for $0 \le y \le 1-p$,

$$F_Y(y) = P(Y \le y) = P(F_T(T) \le y) = P(T \le F_T^{-1}(y)) = F_T(F_T^{-1}(y)) = y.$$

Therefore, $Y = F_T(T)$ has an improper uniform distribution over $[0, 1-p]$.
Case-1: If $p = 0$, this theorem is equivalent to the Inverse Transform Method.
Case-2: If $p = 1$, the result is not realistic for real life. □

We now consider the right censored life time data and assume there are n patients under cancer study. Let y_i refer to the right censored survival time for individual i and $T_i = min\{Z_{ij}, j = 1, 2, ..., N_i\}$ where C_i be the right censored variables. Then we observe $y_i = min(T_i, C_i)$ and the censoring indicator $\delta_i = I(T_i \le C_i)$. The likelihood function of $(y_1, ..., y_n)$ is

$$L = \prod_{i=1}^{n} h_T^{\delta_i}(y_i) S_T(y_i) = \prod_{i=1}^{n} [-ln(p) f(y_i)]^{\delta_i} p^{F(y_i)}$$

and the log-likelihood function is

$$ln(L) = \sum_{i=1}^{n} \delta_i ln(-ln(p)) + \sum_{i=1}^{n} \delta_i ln f(y_i) + \sum_{i=1}^{n} ln(p) F(y_i). \tag{4}$$

4. Fréchet Susceptible Distribution

In this paper, we consider to use a generalized extreme value distribution as the susceptible distribution, i.e., the type II Fréchet distribution. The class of extreme value distributions essentially consists of three types of extreme value distributions: type I (Gumbel distribution), type II (Fréchet distribution) and type III (Weibull distribution). The extreme value distribution is useful in modeling and measuring the events which occur with very low probability. Moreover, extreme value distributions are widely used in finance, insurance, risk management, economics, material sciences and many other subjects those dealing with extreme events.

More specifically, we consider the Fréchet distribution, which was named after the French mathematician Maurice Fréchet (1878-1973), to fit the survival time of uncured individuals. The probability density function of Fréchet distribution is

$$f(y) = \frac{\alpha}{\beta} \left(\frac{\beta}{y}\right)^{\alpha+1} e^{-\left(\frac{\beta}{y}\right)^{\alpha}}, \quad y \ge 0$$

where $\alpha > 0$ is the shape parameter and $\beta > 0$ is the scale parameter and the survival and hazard functions are

$$S(y) = 1 - e^{-\left(\frac{\beta}{y}\right)^{\alpha}} \quad and \quad h(y) = \frac{\alpha}{\beta} \left(\frac{\beta}{y}\right)^{\alpha+1} e^{-\left(\frac{\beta}{y}\right)^{\alpha}} \left[1 - e^{-\left(\frac{\beta}{y}\right)^{\alpha}}\right]^{-1}, \; y \ge 0.$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 September 2018 doi:10.20944/preprints201809.0588.v1

Peer-reviewed version available at Stats 2018, 1, 13; doi:10.3390/stats1010013

Denote $\theta = (\alpha, \beta, p)'$. Substituting the above functions into (4) of the non-mixture model with Fréchet susceptible distribution, we obtain the log-likelihood:

$$\begin{aligned} l(\theta) = lnL(\theta) = & \sum_{i=1}^{n} \delta_i ln[-ln(p)] + \sum_{i=1}^{n} ln(p) e^{-\left(\frac{\beta}{y_i}\right)^{\alpha}} \\ & + \sum_{i=1}^{n} \delta_i \left[ln(\alpha) + \alpha ln(\beta) - (\alpha+1) ln(y_i) - \left(\frac{\beta}{y_i}\right)^{\alpha} \right]. \end{aligned} \tag{5}$$

The maximum likelihood estimator (MLE) of $\theta = (\alpha, \beta, p)'$ is one that maximizes $l(\theta)$, denoted by $\hat{\theta}$. The score functions are given below by taking partial derivatives of log-likelihood function (5) with respect to θ:

$$\begin{aligned} \frac{\partial l(\theta)}{\partial \alpha} &= \sum_{i=1}^{n} \delta_i \left[\frac{1}{\alpha} + \ln(\beta) - \ln(y_i) - \left(\frac{\beta}{y_i}\right)^{\alpha} \ln\left(\frac{\beta}{y_i}\right) \right] - \ln(p) \sum_{i=1}^{n} e^{-\left(\frac{\beta}{y_i}\right)^{\alpha}} \left(\frac{\beta}{y_i}\right)^{\alpha} \ln\left(\frac{\beta}{y_i}\right) \\ \frac{\partial l(\theta)}{\partial \beta} &= \frac{\alpha}{\beta} \left\{ \sum_{i=1}^{n} \delta_i \left[1 - \left(\frac{\beta}{y_i}\right)^{\alpha} \right] - \ln(p) \sum_{i=1}^{n} e^{-\left(\frac{\beta}{y_i}\right)^{\alpha}} \left(\frac{\beta}{y_i}\right)^{\alpha} \right\} \\ \frac{\partial l(\theta)}{\partial p} &= \frac{1}{p \ln(p)} \sum_{i=1}^{n} \delta_i + \frac{1}{p} \sum_{i=1}^{n} e^{-\left(\frac{\beta}{y_i}\right)^{\alpha}} \end{aligned}$$

If the log-likelihood function has a global maximizer, then the MLE is the solution of the score equations. Since it is non-linear system, numerical solution is computed by Newton-Raphson method.

The second partial derivative of the maximum likelihood function are given as follows:

$$\begin{aligned} \frac{\partial^2 l(\theta)}{\partial \alpha^2} &= \sum_{i=1}^{n} \left(\frac{\beta}{y_i}\right)^{\alpha} \left(\ln\left(\frac{\beta}{y_i}\right) \right)^2 \left[\ln(p) e^{-\left(\frac{\beta}{y_i}\right)^{\alpha}} \left(\left(\frac{\beta}{y_i}\right)^{\alpha} - 1 \right) - \delta_i \right] - \frac{1}{\alpha^2} \sum_{i=1}^{n} \delta_i \\ \frac{\partial^2 l(\theta)}{\partial \beta^2} &= \frac{\alpha}{\beta^2} \left\{ \sum_{i=1}^{n} \ln(p) e^{-\left(\frac{\beta}{y_i}\right)^{\alpha}} \left(\frac{\beta}{y_i}\right)^{\alpha} \left[\alpha \left(\frac{\beta}{y_i}\right)^{\alpha} - (\alpha - 1) \right] \right\} \\ & \quad - \frac{\alpha}{\beta^2} \left\{ \sum_{i=1}^{n} \delta_i \left[1 + (\alpha - 1) \left(\frac{\beta}{y_i}\right)^{\alpha} \right] \right\} \\ \frac{\partial^2 l(\theta)}{\partial p^2} &= - \frac{1 + \ln(p)}{(p \ln(p))^2} \sum_{i=1}^{n} \delta_i - \frac{1}{p^2} \sum_{i=1}^{n} e^{-\left(\frac{\beta}{y_i}\right)^{\alpha}} \\ \frac{\partial^2 l(\theta)}{\partial \alpha \partial \beta} &= \frac{1}{\beta} \sum_{i=1}^{n} \left\{ \delta_i \left[1 - \left(\frac{\beta}{y_i}\right)^{\alpha} \left(1 - \ln\left(\frac{\beta}{y_i}\right) \right) \right] \right\} \\ & \quad + \frac{1}{\beta} \sum_{i=1}^{n} \left\{ \ln(p) e^{-\left(\frac{\beta}{y_i}\right)^{\alpha}} \left(\frac{\beta}{y_i}\right)^{\alpha} \left[\alpha \ln\left(\frac{\beta}{y_i}\right) \left(\left(\frac{\beta}{y_i}\right)^{\alpha} - 1 \right) - 1 \right] \right\} \\ \frac{\partial^2 l(\theta)}{\partial \alpha \partial p} &= - \frac{1}{p} \sum_{i=1}^{n} e^{-\left(\frac{\beta}{y_i}\right)^{\alpha}} \left(\frac{\beta}{y_i}\right)^{\alpha} \ln\left(\frac{\beta}{y_i}\right) \\ \frac{\partial^2 l(\theta)}{\partial \beta \partial p} &= - \frac{1}{p} \sum_{i=1}^{n} e^{-\left(\frac{\beta}{y_i}\right)^{\alpha}} \left(\frac{\beta}{y_i}\right)^{\alpha} \left(\frac{\alpha}{\beta}\right). \end{aligned}$$

The asymptotic normality of maximum likelihood estimations of parameters are given by the inverse of the Fisher information matrix. The Fisher information is defined by the following relation

$$\mathcal{I}(\theta) = -E\left[\frac{\partial^2 l(\theta)}{\partial \theta \partial \theta^T} \right].$$

In practical application, the observed Fisher information matrix is used when the expected Fisher information matrix is difficult to compute.

5. Simulation Study

We conducted simulation studies to examine the performance of the maximum likelihood estimator of $\theta = (\alpha, \beta, p)'$ in finite samples. The right censored survival times were generated by using the inverse transform method to the survival function $S_T(t) = p^{F(t)}$. The following algorithm used to simulate a sample of size n from the non-mixture Fréchet distribution with right censored data:

Step 1: Generate a simple sample of $u_1, \cdots, u_n$ from $\sim U(0,1)$.
Step 2: Suppose p is a cure fraction. The random survival time can be calculated from equation

$$t_i = \beta \left[-\ln \left(\frac{\ln(1-u_i)}{\ln(p)} \right) \right]^{-\frac{1}{\alpha}},$$

if $u_i \leq 1-p$, otherwise t_i is infinity.
Step 3: We generate the simple sample of the censoring times $c_1, \cdots, c_n$ from a Fréchet distribution. We adjust the parameters of the Fréchet distribution to obtain the desired censoring rates.
Step 4: The right censored data is obtained from minimum of censoring time and survival time. That is $y_i = \min(t_i, c_i), \quad \delta_i = I(t_i \leq c_i), \quad i = 1, 2, \cdots, n.$
Step 5: The observed data set is $D = \{(y_i, \delta_i), \quad i = 1, \cdots, n\}$.
Step 6: Maximize likelihood function with respect to θ to obtain $\hat{\theta}$. The standard optimization method, optim() in R is used.

We consider various simulation settings with different proportion of cure fractions and different censoring rates. Censoring variable follows a Fréchet distribution with parameter $\alpha > 0$ and $\beta > 0$, where the value of α and β would be adjusted to get the desired censoring rate in the right censored survival data.

In this simulation study, we are interested in the bias, standard error and root mean square error as the performance measures. The bias is the difference between the expected value of an estimator and true parameter and the standard error is a measure of the dispersion of the values in the sampling distribution, which is a statistical term that measures the accuracy. The mean square error (MSE) of an estimator is the expected squared deviation of the estimator of a parameter from the true parameter. The root mean square error (RMSE) is the squared root of MSE.

The simulation results are based on 500 replications with the sample sizes 100, 200, 300, and 500 for each parameter setting. Results are presented in Tables 1-4, which show the values of mean, bias, standard error (SE) and root mean square error (RMSE) of MLE. The simulation results suggest that the proposed method has a good performance overall. The average of the estimates are very closed to respective true parameter values in all different settings of simulations. However, the simulation results for both cure fractions of 1% and 2% with lower censoring rates perform better than the higher censoring rates. The biases of estimates are small, and the estimates of the standard error (SE) and root mean square error (RMSE) of all examined parameters decreased with increasing sample size for all settings. The estimates of all examined parameters perform better for low level of censoring than high level of censoring. Moreover. The estimates of all examined parameters performed well for low rate of cure proportion in comparing to high rate of cure proportion. Finally, we can conclude that the MLE works well for non-mixture cure model with Fréchet susceptible distribution.

Table 1. Summary statistics of a non-mixture model with 1% cure fraction for right censored data

	θ_0	Mean	Bias	SE	RMSE
n=100 and CR 7.71%					
α	0.50	0.5100	0.0100	0.0880	0.0886
β	1.00	1.4130	0.4130	1.4900	1.5462
p	0.01	0.0146	0.0046	0.0151	0.0158
n=200 and CR 7.68%					
α	0.50	0.5023	0.0023	0.0620	0.0621
β	1.00	1.1822	0.1822	0.6277	0.6536
p	0.01	0.0119	0.0019	0.0104	0.0106
n=300 and CR 7.81%					
α	0.50	0.5019	0.0019	0.0499	0.0499
β	1.00	1.1022	0.1022	0.4343	0.4462
p	0.01	0.0117	0.0017	0.0083	0.0085
n=500 and CR 7.82%					
α	0.50	0.5003	0.0003	0.0371	0.0371
β	1.00	1.0588	0.0588	0.2879	0.2938
p	0.01	0.0108	0.0008	0.0061	0.0062

Table 2. Summary statistics of non-mixture model with 2% cure fraction for right censored data

	θ_0	Mean	Bias	SE	RMSE
n=100 and CR 10.27%					
α	0.50	0.5110	0.0110	0.0895	0.0902
β	1.00	1.3538	0.3538	1.1763	1.2284
p	0.02	0.0253	0.0053	0.0225	0.0231
n=200 and CR 10.24%					
α	0.50	0.5019	0.0019	0.0637	0.0637
β	1.00	1.1954	0.1954	0.6277	0.6574
p	0.02	0.0218	0.0018	0.0160	0.0161
n=300 and CR 10.39%					
α	0.50	0.5019	0.0019	0.0511	0.0511
β	1.00	1.1035	0.1035	0.4269	0.4393
p	0.02	0.0218	0.0018	0.0133	0.0134
n=500 and CR 10.37%					
α	0.50	0.4992	-0.0008	0.0379	0.0379
β	1.00	1.0694	0.0694	0.2898	0.2980
p	0.02	0.0204	0.0004	0.0097	0.0097

The overall censoring rates of Table 1 is about 7.78% , while Table 2 is about 10.34%. Moreover, θ_0 is true parameter vector, CR the censoring rate, SE the standard error, and RMSE a root mean square error.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 September 2018 doi:10.20944/preprints201809.0588.v1

Peer-reviewed version available at Stats 2018, 1, 13; doi:10.3390/stats1010013

Table 3. Summary statistics of non-mixture model with 1% cure fraction for right censored data

	θ_0	Mean	Bias	SE	RMSE
n=100 and CR 14.26%					
α	0.50	0.5119	0.0119	0.0955	0.0962
β	1.00	1.5112	0.5112	1.7149	1.7895
p	0.01	0.0163	0.0063	0.0181	0.0192
n=200 and CR 14.34%					
α	0.50	0.5034	0.0034	0.0679	0.0680
β	1.00	1.2270	0.2270	0.7577	0.7909
p	0.01	0.0130	0.0030	0.0123	0.0127
n=300 and CR 14.41%					
α	0.50	0.5034	0.0034	0.0555	0.0556
β	1.00	1.1274	0.1274	0.5458	0.5605
p	0.01	0.0127	0.0027	0.0103	0.0106
n=500 and CR 14.42%					
α	0.50	0.5001	0.0001	0.0399	0.0399
β	1.00	1.0726	0.0726	0.3189	0.3271
p	0.01	0.0111	0.0011	0.0072	0.0073

Table 4. Summary statistics of non-mixture model with 2% cure fraction for right censored data

	θ_0	Mean	Bias	SE	RMSE
n=100 and CR 18.49%					
α	0.50	0.5162	0.0162	0.1001	0.1014
β	1.00	1.4419	0.4419	1.5639	1.6251
p	0.02	0.0285	0.0085	0.0276	0.0289
n=200 and CR 18.55%					
α	0.50	0.5040	0.0040	0.0702	0.0703
β	1.00	1.2361	0.2361	0.8332	0.8660
p	0.02	0.0237	0.0037	0.0193	0.0197
n=300 and CR 18.63%					
α	0.50	0.5035	0.0035	0.0588	0.0589
β	1.00	1.1442	0.1442	0.5999	0.6170
p	0.02	0.0236	0.0036	0.0171	0.0175
n=500 and CR 18.53%					
α	0.50	0.4981	-0.0019	0.0409	0.0409
β	1.00	1.0961	0.0961	0.3576	0.3703
p	0.02	0.0206	0.0006	0.0115	0.0115

The overall censoring rates of Table 3 is about 14.36%, and Table 4 is about 18.55%. Moreover, θ_0 is true parameter vector, CR the censoring rate, SE the standard error, and RMSE a root mean square error.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 September 2018 doi:10.20944/preprints201809.0588.v1

Peer-reviewed version available at Stats 2018, 1, 13; doi:10.3390/stats1010013

Figure 1 shows the relationship between the estimated bias against the censoring rate.

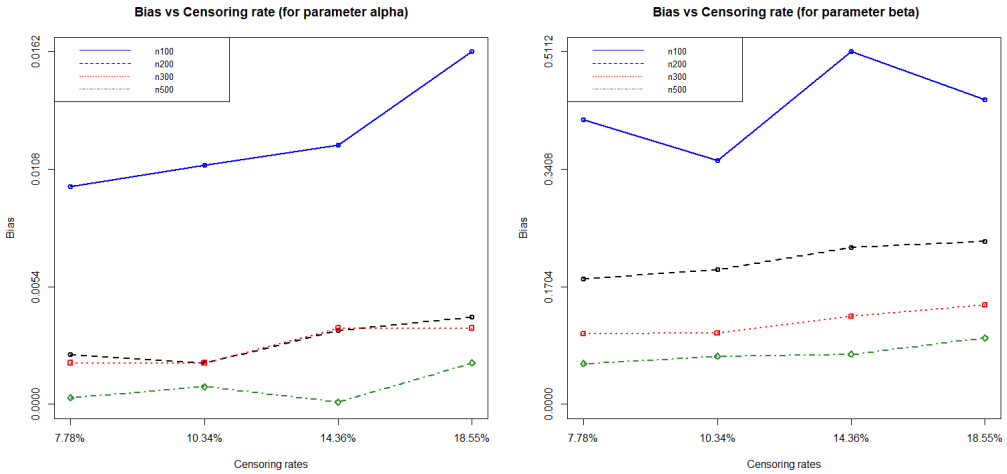

Figure 1. The relationship between the bias and the censoring rate.

6. A Real Data Analysis

In this section, we would like to compare non-mixture and mixture models with different distributions, we use different susceptible assessments: the negative value of the log-likelihood function, the Akaike's information Criterion (AIC) and the corrected Akaike's information Criterion (AIC_c). The (AIC) is defined by $AIC = -2lnL(\theta) + 2k$, where $\theta = (\alpha, \beta, p)'$ is the vector of unknown parameters included in the model, $L(\theta)$ is the likelihood function and where k is the number of free parameters in the model. The (AIC_c) is defined by

$$AIC_c = -2lnL(\theta) + 2k\frac{n}{n-k-1},$$

where n is the sample size. The lower the value of selection criterion indicates the better fit.

We consider the data set of 46 patients with an HLA-matched donor received allogeneic marrow presented by Kersey [22] in 1987. Out of the 46 patients, 12 with allogeneic transplanted were died during their observed periods. Those are the censored observations (26%). Using this data, we compare the non-mixture and mixture models with Fréchet, two parameters Weibull and two parameters Exponentiated Exponential (EE) susceptible distributions using maximum likelihood method.

Tables 5, and 6 show the maximum likelihood estimates of the parameters and their standard errors of the above three susceptible distributions with both non-mixture and mixture models for allogenic marrow transplanted (leukemia) data, respectively. Table 7 shows the values of $-logL$, AIC and AIC_c of three susceptible distributions using both mixture and non-mixture models for leukemia data. From Table 7, indicates that Fréchet Distribution is the best within non-mixture and mixture cure models, respectively.

Table 5. Results of non-mixture cure models with Leukemia data

Parameter	Fréchet		Weibull		EE	
	Estimate	SE	Estimate	SE	Estimate	SE
α	0.430	0.134	1.005	0.149	1.045	0.220
β	1.526	1.781	1.221	0.356	0.863	0.318
p	0.068	0.092	0.233	0.069	0.234	0.069

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 September 2018 doi:10.20944/preprints201809.0588.v1

Peer-reviewed version available at Stats 2018, 1, 13; doi:10.3390/stats1010013

Table 6. Results of mixture cure models with Leukemia data

Parameter	Fréchet		Weibull		EE	
	Estimate	SE	Estimate	SE	Estimate	SE
α	0.608	0.138	0.902	0.132	0.910	0.211
β	0.363	0.162	0.769	0.171	1.184	0.329
p	0.054	0.150	0.239	0.068	0.242	0.067

Table 7. Model comparison by information criteria for Leukemia data

Model	Non-mixture Model			Mixture Model		
	-logL	AIC	AIC_c	-logL	AIC	AIC_c
Fréchet	47.366	100.73	101.30	47.371	100.75	101.32
Weibull	48.340	102.68	103.25	48.752	103.50	104.70
EE	48.319	102.64	103.21	48.937	103.87	104.45

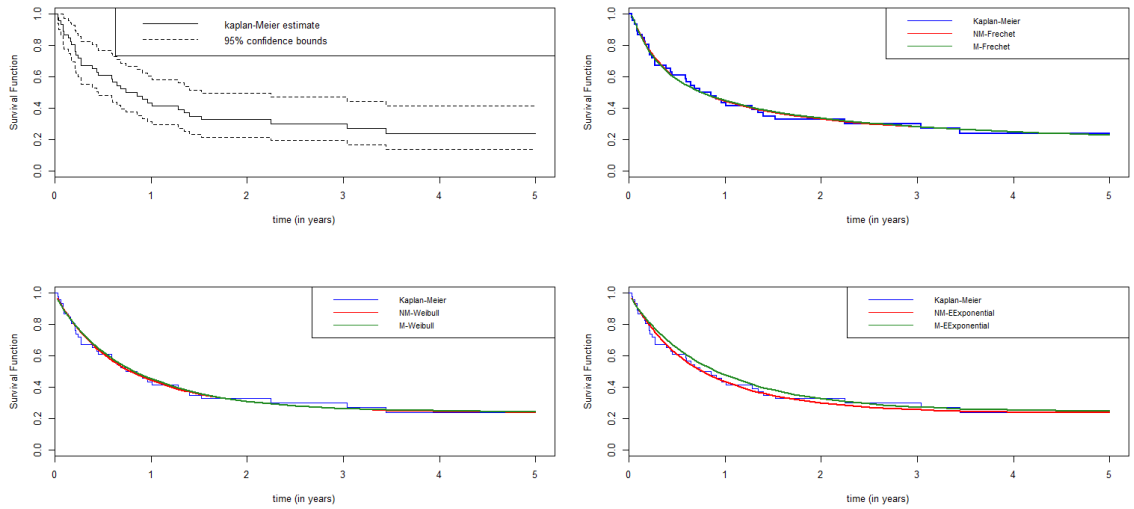

Figure 2. Fitted survival curves of the Leukemia (Allogenic) data

We also compared the Kaplan-Meier survival curves with the parametric curves. The model is better fit when they are closer to each other. Top left corner figure of Figure 2 is the Kaplan-Meier survival curve of the allogenic marrow transplanted data. Top right corner of Figure 2 is the Kaplan-Meier estimated curve overlaid with the estimated survival curves using Fréchet non-mixture and mixture model. Bottom left corner of Figure 2 is the Kaplan-Meier estimated curve overlaid with the estimated survival curves using Weibull non-mixture and mixture models. Bottom right corner of Figure 2 is the Kaplan-Meier estimated curve with survival curve of non-mixture and mixture model of Exponentiated Exponential Distribution. The curve of non-mixture Fréchet distribution is closer to the Kaplan-Meier survival curve in compare to all other distributions, and mixture model is little bit over estimating the survival rate.

The second data set we considered is the melanoma data without covariates from the Eastern Cooperative Oncology Group (ECOG) phase III clinical trial e1684 available in R package *smcure*. In this trial, we found that a total of 69% patients are censored. This trial, e1684, was a two arm clinical trial comparing high-dose interferon alpha-2b (IFN) regimen to observation. There were a total of 286 patients enrolled in the study, collected from 1984 to 1990, and the study was published in 1996. After deleting the missing data, we used a total of 284 observations without covariates for

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 September 2018 doi:10.20944/preprints201809.0588.v1

Peer-reviewed version available at Stats 2018, 1, 13; doi:10.3390/stats1010013

analyzing our models.

The statistics summary under Fréchet susceptible distribution, two-parameter Weibull and Exponentiated Exponential susceptible distributions are presented in Tables 8 and 9. Table 10 shows the values of $-logL$, AIC and AIC_c of different distributions both mixture and non-mixture models for melanoma data. From the result of Table 10, we conclude that the Fréchet distribution non-mixture and mixture cure models are better fit, since $-logL$, AIC and AIC_c are smaller when compared to the other models.

Table 8. Results of a non-mixture cure model for melanoma data

	Frechet		Weibull		EE	
Parameter	Estimate	SE	Estimate	SE	Estimate	SE
α	0.574	0.059	1.014	0.058	1.089	0.096
β	0.963	0.240	1.566	0.153	0.695	0.089
p	0.191	0.040	0.293	0.028	0.293	0.028

Table 9. Results of a mixture cure model for melanoma data

	Frechet		Weibull		EE	
Parameter	Estimate	SE	Estimate	SE	Estimate	SE
α	0.732	0.061	0.913	0.053	0.947	0.091
β	0.440	0.058	1.076	0.094	0.866	0.092
p	0.201	0.041	0.299	0.028	0.301	0.028

Table 10. Models comparison by information criteria for melanoma data

	Non-mixture Model			Mixture Model		
Model	-logL	AIC	AIC_c	-logL	AIC	AIC_c
Frechet	367.546	741.09	741.18	367.356	740.72	740.80
Weibull	378.405	762.81	762.90	381.396	768.79	768.88
EE	377.970	761.94	762.03	382.549	771.10	771.18

The Kaplan-Meier estimate of survival curve and fitted survival curves of the Fréchet, Weibull and Exponentiated Exponential distributions for the melanoma data are given in Figure 3. Top left corner figure of Figure 3 is the Kaplan-Meier survival curve of the melanoma data. Top right corner of Figure 3 is the Kaplan-Meier estimated curve overlaid with the estimated survival curves using Fréchet non-mixture and mixture model. Bottom left corner of Figure 3 is the Kaplan-Meier estimated curve overlaid with the estimated survival curves using Weibull non-mixture and mixture models. Bottom right corner of Figure 3 is the Kaplan-Meier estimated curve with survival curve of non-mixture and mixture model of Exponentiated Exponential distribution. The curve of non-mixture Fréchet distribution is closer to the Kaplan-Meier survival curve in compare to all other distributions, and mixture model is little bit over estimating the survival rate.

Peer-reviewed version available at Stats 2018, 1, 13; doi:10.3390/stats1010013

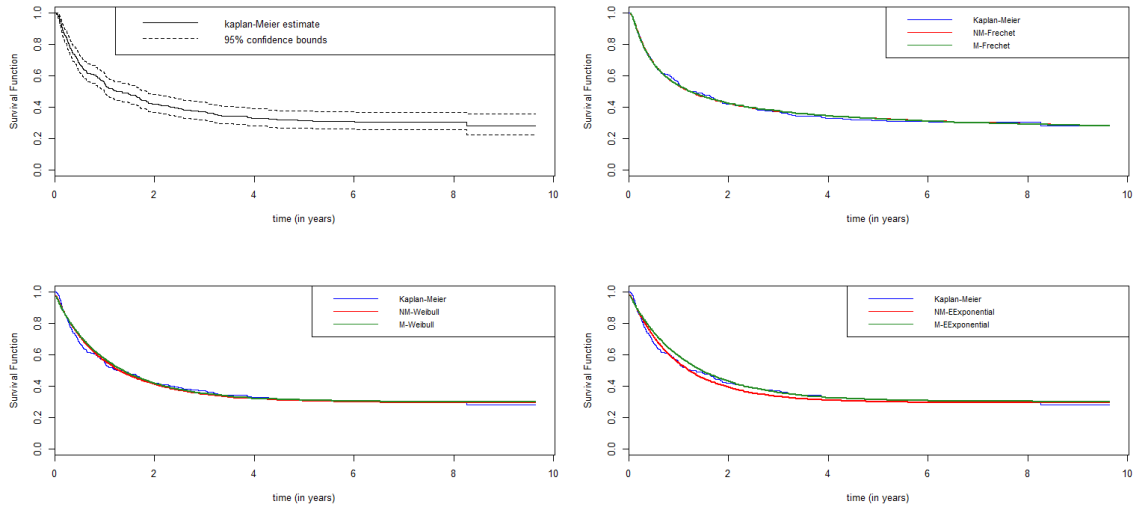

Figure 3. Fitted survival curves of the Melanoma data

7. Conclusion

In this article, we proposed a non-mixture cure model with Fréchet susceptible distribution for right censored data. Finite sample properties of maximum likelihood estimator of the model parameters are empirically illustrated by simulation study under various parameter settings. The results of the simulation studies show that the proposed estimator has a good performance and is not sensitive to model parameters, little bit sensitive to the censoring rate. Moreover, we compared the different non-mixture and mixture models using real data sets. We found that the proposed non-mixture Fréchet distribution model is the best one in comparing to other models for modeling the real data about allogeneic marrow HLA-matched donors and ECOG phase III clinical trial e1684 data.

Acknowledgments: Lianfen Qian's research is partially supported by the Natural Science Foundations of Zhejiang Province (LY17A010012) and the OURI CUrriculum Grant from Florida Atlantic University, USA.

Author Contributions: Lianfen Qian provided the theoretical derivation and Durga Kutal did the simulation under Lianfen Qian mentorship. Both authors contributed equally to the final version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations is used in this manuscript:
MDPI: Multidisciplinary Digital Publishing Institute

References

1. Boag, John W. Maximum likelihood estimates of the proportion of patients cured by cancer therapy *Journal of the Royal Statistical Society, Series B (Methodological)* **1949**,*11*,15-53.
2. Berkson, Joseph ; Gage, Robert P. Survival curve for cancer patients following treatment *Journal of the American Statistical Association* **1952**,*47*,501-512.
3. Farewell, V.T. The Use of Mixture Models for the Analysis of Survival Data with Long-Term survivors *Journal of the International Biometric Society* **1982**,*38*,1041-1046.
4. Goldman, Anne I. Survivorship analysis when cure is a possibility: A monte carlo study *Journal of Statistics in Medicine* **1984**,*3*,153-163.

5. Taylor, Jeremy M.G. Semi-Parametric Estimation in Failure Time Mixture Models *Journal of the International Biometric Society* **1995**,*51*,899-907.
6. Peng, Yingwei and Dear, Keith B.G. A Non-parametric Mixture model for Cure Rate Estimation *Journal of the International Biometric Society* **2000**,*56*,237-243.
7. Kuk, A. Y. C. and Chen, C. H. A Mixture Model Combining Logistic Regression with Proportional Hazards Regression *Biometrika* **1992**,*79*,531-541.
8. Peng, Y. and Zhag, J. Accelerated Hazards Mixture Cure Model *Lifetime Data Analysis* **2009**,*15*,455-467.
9. Kim, Y. and Jhun, M. Cure rate model with interval censored data *Lifetime Data Analysis* **2008***27*,3-14.
10. Goetghebeur, E. and Ryan, L. Semiparametric Regression Analysis of Interval-Censored Data *Biometrics* **2000**,*56*,1139-1144.
11. Yakovlev, AYu ; Klebanov, LB and Rachev, ST A Stochastic model of radiation carcinogenesis: latent time distributions and their properties *Journal of Mathematical Biosciences* **1993**,*113*,51-75.
12. Chen, Ming-Hui; Ibrahim, Joseph G. and Sinha, Debajyoti A new bayesian model for survival data with a survival function *Journal of American Statistical Association* **1999**,*94*,909-919.
13. Chen, Ming-Hui; Ibrahim, Joseph G. and Sinha, Debajyoti Bayessian Semiparametric Models for survival data with a Cure Fraction *Journal of the International Biometric Society* **2001**,*57*,383-388.
14. Tsodikov, A.D.; Ibrahim, J.G. and Yakovlev, A. Y. Estimating Cure Rate from Survival Data:An Alternative to Two-component Mixture Models*Journal of the American Statistical Association* **2003**,*98*,1063-1078.
15. Herring, A. H. and Ibrahim, J. G. Maximum likelihood estimation in random effects cure rate models with nonignorable missing covariates *Biostatistics* **2002**,*3*,387-405.
16. Uddin, M. T.; ,Sen, A.; Noor, M. S.; Islam, M. N. and Chowdhury, Z. I. An Analytical Approaches on Non-parametric Estimation of Cure Rate Based on Uncensored Data *Journal of Applied Sciences* **2006a**,*6* ,1258-1264.
17. Uddin, M. T.; Islam, M. N. and Ibrahim, Q. I. U. An Analytical Approaches on Non-parametric Estimation of Cure Rate Based on Uncensored Data *Journal of Applied Sciences* **2006b**,*6*,548-552.
18. Liu, Hao and Shen, Yu A Semiparametric Regression Cure Model for Interval-Censored Data *Journal of the American Statistical Association* **2009**,*104*,1168-1178.
19. Lopes, C. M. C. and Bolfarine, H. Random effects in promotion time cure models *Computational Statistics and Data Analysis* **2012**,*56*,75-87.
20. Jenkinson, A. F. The frequency distribution of the annual maximum (or minimum) values of meteorological elements *Quarterly Journal of the Royal Meteorological Society* **1955**,*81*,158-171.
21. Ramos, P.L.; Nascimento, D. and Louzda, F. The Long Term Fréchet Distribution: Estimation, Properties and Its Application *BioMetrics and Biostatistics International Journal* **2017**,*6*(3):00170, DOI: 10.15406/bbij.2017.06.00170.
22. Kersy, J.H.; Weisdorf, D.; Nesbit, M.E; LeBien; T.W. and W.G.Wood. Comparison of autologous and allogenic bone marrow transpalntation for treatment of high-risk refractory acute lymphoblastic leukemia *The New England Journal of Medicine* **1987**,*317*,461-467.