ON A GENERAL CURE RATE FRAILTY MODEL AND ASSOCIATED INFERENCE

Zijie (Gavin) Zhao

April 16, 2021

Acknowledgements

I would like to express my most sincere appreciation for my supervisor, Dr. Narayanaswamy Balakrishnan, for his patient guidance and support throughout the year. I am grateful for all the advice and explanations as well as clear ideas he has given to me. Thank you for this opportunity.

Also, I would like to say thank you to all faculty members and staff in Math and Statistics department at McMaster for their continued help and support during my undergraduate study.

Table of Contents

1 Introduction

Survival Functions and Proportional Hazards

The **survival function** is a function that gives the probability that the time of death is later than some specific time t.

Let T denote survival time, and let f be its probability density function.

The survival function S is defined as

$$S(t) = \mathrm{P}(T > t) = \int_t^\infty f(u)du = 1 - F(t)$$

Properties:

- Every survival function S(t) is non-increasing, i.e. $S(u) \leq S(t)\ for\ all\ u \geq t.$
- $S(0) = 1$, can be less than 1 if there is the possibility of immediate death or failure.

The **hazard function** is defined as

$$\lambda(t) = \lim_{dt \to 0} \frac{P(t \leq T < t + dt | T \geq t)}{dt}$$

The numerator of this expression is the conditional probability that the event will occur in the time interval given that it has not occurred before, and the denominator is the width of the interval. We obtain a rate of event occurrence per unit of time. Taking the limit to zero, we get an instantaneous rate of occurrence.

$$\begin{aligned} \lambda(t) = {} & \lim_{dt \to 0} \frac{P(t \leq T < t + dt | T \geq t)}{dt} \\ = {} & \lim_{dt \to 0} \frac{P(t \leq T < t + dt)}{dt * P(T \geq t)} \\ = {} & \frac{P(T = t)}{P(T \geq t)} \\ = {} & \frac{f(t)}{S(t)} \end{aligned}$$

The hazard function is not a density or a probability, but we can think of it as the probability of failure in an infinitesimally small time period given that it is survived till time t. It is then a measure of risk. The greater the hazard between the period, the greater the risk of failure in it.

The connection between the hazard and survivor functions

$$\begin{aligned}\lambda(t) &= \frac{f(t)}{S(t)} \\ &= \frac{f(t)}{1-F(t)} \\ &= -\frac{d}{dt}\log\bigl(1-F(t)\bigr) \\ &= -\frac{d}{dt}\log(S(t))\end{aligned}$$

Introduce cumulative hazard, denoted as

$$\Lambda(t) = \int_0^t \lambda(x)dx.$$

Then we have

$$S(t) = \exp\,(-\Lambda(t)).$$

The **proportional hazards** model has the form

$$\lambda(t|x_i) = \lambda_0(t)\exp\bigl(\beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}\bigr) = \lambda_0(t)\exp(x_i\beta)$$

where

- $x_i = \{x_{1p}, x_{2p}, \dots, x_{ip}\}$ is the realized values of the covariates for subject i.
- t represents the survival time
- the coefficients β measure the impact of covariates
- $\lambda_0(t)$ is called the baseline hazard. It corresponds to the value of the hazard if all the x_i are equal to zero. The hazard may vary over time.

The proportional hazards condition states that covariates are multiplicatively related to the hazard. In the case of a continuous covariate x, it is typically assumed that the hazard responds exponentially.

Frailty Models

The hazard function plays a central role in survival analysis. In a homogeneous population, the distribution of the time to event, described by the hazard, is the same for each individual. Heterogeneity in the distributions can be accounted for by including covariates in a model for the hazard, for instance a proportional hazards model. In this model, individuals with the same value of the covariates will have the same distribution. It is natural to think that not all covariates that are thought to influence the distribution of the survival outcome are included in the model. This implies that there is unobserved heterogeneity; individuals with the same value of the covariates may have different distributions. One way of accounting for this unobserved heterogeneity is to include random effects in the model. In the context of hazard models for time to event outcomes, such random effects are called **frailties**, and the resulting models are called **frailty models**.

Let $\left(t_{ij}, \delta_{ij}, x_{ij}\right), i = 1, \dots, n, j = 1, \dots, m_i$, be the failure time, censoring indicator, and the covariate of the jth individual in the ith cluster, where $\delta_{ij} = 1 \; if \; t_{ij}$ is not censored and 0 otherwise. Let y_i denote the unobserved frailty shared by the individuals in the ith cluster. Given y_i, the frailty model specifies that t_{ij} are independent with a proportional hazards function

$$h\left(t_{ij} \middle| y_i\right) = y_i h_0\left(t_{ij}\right) \exp\left(\beta' x_{ij}\right)$$

The frailties y_i are usually assumed to be independent and identically distributed with a distribution referred to as the frailty distribution.

Cure Rate Models

Cure rate models are a special case of survival models where a portion of subjects in the population never experience event of interest. Such subjects are called immune or cured.

Population survival function of the time-to-event T can be given by the mixture model

$$S_p(t) = P(T > t) = p_0 + (1 - p_0) S_s(t)$$

where

- I is an indicator random variable where I = 0 if and only if the individual is cured and 1 otherwise
- $p_0 = P(I = 0)$ is the probability of an individual to be cured
- $S_s(t) = P(T > t | I = 1)$ is the survival function of susceptibles

Inverse Gaussian Distribution

In probability theory, the **inverse Gaussian** distribution is a two-parameter family of continuous probability distributions with support on $(0, \infty)$.

Its probability density function is given by

$$f(y; \mu, \lambda) = \sqrt{\frac{\lambda}{2\pi y^3}} e^{-\frac{\lambda(y-\mu)^2}{2\mu^2 y}}$$

for $x > 0, where\ \mu > 0\ is\ the\ mean\ and\ \lambda > 0\ is\ the\ shape\ parameter.$

We also have $E[Y] = \mu\ and\ Var(Y) = \frac{\mu^3}{\lambda}$

Weibull Distribution

The probability of a **Weibull** random variable is:

$$f(x; \lambda, k) = \begin{cases} \frac{k}{\lambda} (\frac{x}{\lambda})^{k-1} e^{-(\frac{x}{\lambda})^k}, x \geq 0 \\ 0, x < 0 \end{cases}$$

where $k > 0\ is\ the\ shape\ parameter\ and\ \lambda > 0$ is the scale parameter of the distribution.

We also have $E[X] = \lambda \Gamma\left(1 + \frac{1}{k}\right)\ and\ Var(X) = \lambda^2 [\Gamma\left(1 + \frac{2}{k}\right) - \left(\Gamma\left(1 + \frac{1}{k}\right)\right)^2].$

Assume now a competing cause scenario with M being a Bernoulli random variable denoting the number of competing causes related to the occurrence of an event of interest.

Let T be the population time-to-event, therefore the population survival function is given by

$$S_p(t) = P(T > t) = p_0 + (1 - p_0) S_s(t)$$

where $S_s(t) = P(T > t | M = 1)$ is the survival function of susceptibles.

In this article, we assume frailty model for the distribution of $S_s(t)$, with a parametric assumption on the baseline function. Specifically, the hazard function of $S_s(t)$ is taken as

$$h_s(t) = yh_0(t; k, \lambda)e^{\beta X}$$

where $X = \left(x_1, x_2, \dots, x_p\right)$ is a vector of p covariates, $\beta = \left(\beta_1, \dots, \beta_p\right)$ is the proportional hazards regression coefficients, $h_0(t; k, \lambda)$ is the baseline hazard function of a two-parameter (k, λ) Weibull distribution, y is the unobserved frailty assumed to be distributed with inverse Gaussian distribution (μ, γ). For estimating the parameters, the maximum likelihood method will be used.

Organization of the report

The rest of the article is organized as follows. The cure rate frailty model is described in Section 2. The form of the available data and the likelihood function are given in Section 3. In Section 4, we use the proposed model and analyze a real dataset on cutaneous melanoma. Some concluding remarks are presented in Section 5.

2 The Bernoulli Cure Rate Frailty Model

If the number of competing causes M follows a Bernoulli distribution, its probability mass function is

$$P(M = 0) = p_0 = 1 - P(M = 1), p_0 \text{ is the cure rate}$$

with $E(M) = 1 - p_0$.

The population survival function becomes

$$S_p(t) = P(T > t) = p_0 + (1 - p_0)S_s(t).$$

Note that $S_p(0) = p_0 + 1 - p_0 = 1 \; and \; S_p(\infty) = p_0$.

Now we bring frailty into $S_s(t)$. The hazard function of $S_s(t)$ is given by

$$h_s(t) = yh_0(t; k, \lambda)e^{\beta X}$$

where y is a frailty variable and $h_0(t)$ is the baseline hazard function.

The cumulative hazard function of $S_s(t)$ is

$$H_s(t) = yH_0(t)e^{\beta X}$$

$where \; H_0(t)$ is the cumulative baseline hazard function.

Since $S_s(t) = e^{-H_s(t)}$ as we proved in the previous section, $S_s(t)$ becomes

$$S_s(t) = e^{-yH_0(t)e^{\beta X}}$$

Thus $S_p(t)$ becomes

$$S_p(t) = p_0 + (1 - p_0)e^{-ye^{\beta x_i}H_0(t)}$$

Y is the unobserved frailty. $Y_1, Y_2, \ldots Y_n$ are independent and identically distributed and following inverse Gaussian distribution (μ, γ) and $h_0(t; k, \lambda)$ is the baseline hazard function of a two-parameter (k, λ) Weibull distribution. To make the parameters in the frailty distribution and the baseline distribution identifiable, it usually requires that the mean of the frailty distribution to be 1. Hence, we choose E[Y]=1. So $\mu = 1$.

Since there is no intercept term β_0,

Hence the frailty variable $Y \sim IG(1, \gamma)$ and $h_0 = Weibull(t; k, \lambda)$.

Therefore, we have

$$f(y) = \sqrt{\frac{\gamma}{2\pi y^3}} e^{-\frac{\gamma(y-1)^2}{2y}},$$

$$h_0(t; k, \lambda) = \frac{k}{\lambda} (\frac{t}{\lambda})^{k-1} e^{-(\frac{t}{\lambda})^k},$$

and

$$H_0(t) = 1 - e^{-\left(\frac{t}{\lambda}\right)^k}.$$

3 Likelihood Function and Estimation of Parameters

In lifetime data analysis, right censoring in data are commonly used because of the limitations of duration of the study. Therefore, we are subject to right censoring in data. Note that the censored group includes the susceptibles who have their lifetimes to be larger than the censoring time as well as all the cured individuals.

Denote the censoring time as $C_i, and\ A_i$ is the actual lifetime for the ith individual. Then, $T_i = \min(C_i, A_i)\ while\ \delta_i = I(A_i \leq C_i)$ indicates whether the ith individual is censored (0) or not (1), for i=1,2,…,n. Let us also define the sets $\Delta_1 = \{i: \delta_i = 1\}\ and\ \Delta_0 = \{i: \delta_i = 0\}$.

The observed data are of the form $(t_i, \delta_i, x_i)\ with\ x_i = \left(x_{i1}, x_{i2}, \dots x_{ip}\right)'\ and\ \beta_i = \left(\beta_{i1}, \beta_{i2}, \dots \beta_{ip}\right)'$ is the vector of the regression coefficients for the ith individual.

The likelihood function conditioned on frailty $\{y_i\}$ is given by

$$L(\theta; t, \delta, x|\{y_i\}) \propto \prod_{i\in\Delta_1} y_i h_0(t_i) e^{\beta x_i} S_p(t_i, x_i; \theta) \prod_{i\in\Delta_0} S_p(t_i, x_i; \theta)$$

where $\theta = (k, \gamma, \lambda, \beta)'$.

The unconditional likelihood is given by

$$L(\theta; t, \delta, x) \propto \int \left(\prod_{i\in\Delta_1} (y_i) h_0(t_i) e^{\beta X} S_p(t_i, x_i; \theta) \prod_{i\in\Delta_0} S_p(t_i, x_i; \theta) \right) f(y_i) dy_i$$

The complete data is given by $\{(t_i, x_i, \delta_i, I_i): i = 1,2,..,n\}$ where I_is are observed for $i \in \Delta_1$ and unobserved for $i \in \Delta_0$.

The complete data likelihood function is given by

$$L(\theta; t, x, \delta, I)$$

$$\propto \int \prod_{i\in\Delta_1} y_i h_0(t_i) e^{\beta_i X_i} S_p(t_i, x_i; \theta) \prod_{i\in\Delta_0} {p_0}^{1-I_i} \left[(1-p_0) S_s(t_i, x_i; \theta)\right]^{I_i} f(y_i) dy_i$$

$$= \prod_{i\in\Delta_1} h_0(t_i) e^{\beta_i X_i} \prod_{i\in\Delta_0} {p_0}^{1-I_i} (1-p_0)^{I_i}] \int \prod_{i\in\Delta_1} S_p(t_i, x_i; \theta) y_i \prod_{i\in\Delta_0} S_s(t_i, x_i; \theta)^{I_i} f(y_i) dy_i$$

$$= \prod_{i\in\Delta_1} h_0(t_i)e^{\beta_i X_i} \prod_{i\in\Delta_0} {p_0}^{1-I_i}(1-p_0)^{I_i}$$

$$* \int {y_i}^m \exp[-y_i E_i(\beta,\lambda)]\, {p_0}^{m-mI_i}(1-p_0)^{mI_i} f(y_i)dy_i$$

$$= \prod_{i\in\Delta_1} h_0(t_i)e^{\beta_i X_i} \prod_{i\in\Delta_0} {p_0}^{1-I_i}(1-p_0)^{I_i}] * {p_0}^{m-mI_i}(1-p_0)^{mI_i}$$

$$* \int {y_i}^m \exp\,[-y_i E(k,\beta,\lambda)] f(y_i)dy_i$$

where $m = \sum \delta \ and \ E(\beta,\lambda) = \sum_{i\in\Delta_1} H_0(t_i;\lambda,k)\exp\,(\beta x_i)$

$$\int {y_i}^m \exp\,[-y_i E_i(k,\beta,\lambda)] f(y_i)dy_i$$

$$= \int {y_i}^m \exp\,[-y_i E_i] \sqrt{\frac{\gamma}{2\pi y_i^3}}\, e^{-\frac{\gamma(y_i-1)^2}{2y_i}} dy_i$$

$$= \int {y_i}^m \exp\,[-y_i E_i] \sqrt{\frac{\gamma}{2\pi y_i^3}}\, e^{-\frac{\gamma y_i}{2}-\frac{\gamma}{2y_i}+\gamma} dy_i$$

$$= \exp\,(\gamma) \int {y_i}^m \exp\,[-y_i E_i] \sqrt{\frac{\gamma}{2\pi y_i^3}}\, e^{-\frac{\gamma y_i}{2}-\frac{\gamma}{2y_i}+\gamma} dy_i$$

$$= \exp\,(\gamma) \int {y_i}^m \exp\,[-y_i E_i - \frac{\gamma}{2} y_i - \frac{\gamma}{2y_i}] \sqrt{\frac{\gamma}{2\pi y_i^3}}\, dy_i$$

$$= \exp\,(\gamma) \int {y_i}^m \exp\,[-(y_i E_i + \frac{1}{2})\gamma y_i - \frac{\gamma}{2y_i}] \sqrt{\frac{\gamma}{2\pi y_i^3}}\, dy_i$$

Let $c^2 = 2\gamma^2\left(\frac{1}{2}+\frac{E}{\gamma}\right), s = \left(\frac{1}{2}+\frac{E}{\gamma}\right)\gamma y = \frac{c^2 y}{2\gamma} \ and \ s = \frac{c^2}{2\gamma} y$

integral$=\frac{\mathrm{c}}{2\sqrt{\pi}} \int \frac{y^m \exp\left(-s-\frac{c^2}{4s}\right)ds}{s^{\frac{3}{2}}}$

$$=\frac{c}{2\sqrt{\pi}} (\frac{2\gamma}{c^2})^m \int \frac{s^m \exp\left(-s-\frac{c^2}{4s}\right)ds}{s^{\frac{3}{2}}}$$

Recalling the integral representation of the modified Bessel function Kv(z)

$$K_v(z) = \frac{1}{2}\left(\frac{z}{2}\right)^v \int \frac{\exp\left(-t-\frac{z^2}{4t}\right)dt}{t^{v+1}}$$

$$K_{\frac{1}{2}}(z) = \left(\frac{\pi}{2z}\right)^{\frac{1}{2}} e^{-z}, z \geq 0 \; and \; k_n(z) = \sqrt{\frac{\frac{1}{2}\pi}{z}} k_{n+\frac{1}{2}}(z)$$

Thus, we have

$$\text{Integral=}\frac{\mathrm{c}}{2\sqrt{\pi}} (\frac{2\gamma}{c^2})^m \int \frac{s^m \exp\left(-s-\frac{c^2}{4s}\right) ds}{s^{\frac{3}{2}}}$$

$$=\frac{\mathrm{c}}{2\sqrt{\pi}} \left(\frac{2\gamma}{c^2}\right)^m 2 * \left(\frac{2}{c}\right)^{-m+\frac{1}{2}} k_{\frac{1}{2}-m}(c)$$

$$=\frac{\mathrm{c}}{2\sqrt{\pi}} \left(\frac{2\gamma}{c^2}\right)^m 2 * \left(\frac{2}{c}\right)^{-m+\frac{1}{2}} k_{\frac{1}{2}}(c) \left(\frac{c}{2}\pi\right)^m$$

$$=e^{-c} \left(\frac{2\gamma}{c^2}\right)^m \pi^m \left(\frac{c}{2}\right)^{2m}$$

Hence, the whole integral becomes

$$\exp(\gamma\left(1-\gamma\sqrt{1+\frac{2E}{\gamma}}\right)) (\frac{\pi\gamma}{2})^m$$

Let $l(\theta; t, x, \delta, I) = \log\big(L(\theta; t, x, \delta, I)\big)$ be the log-likelihood function.

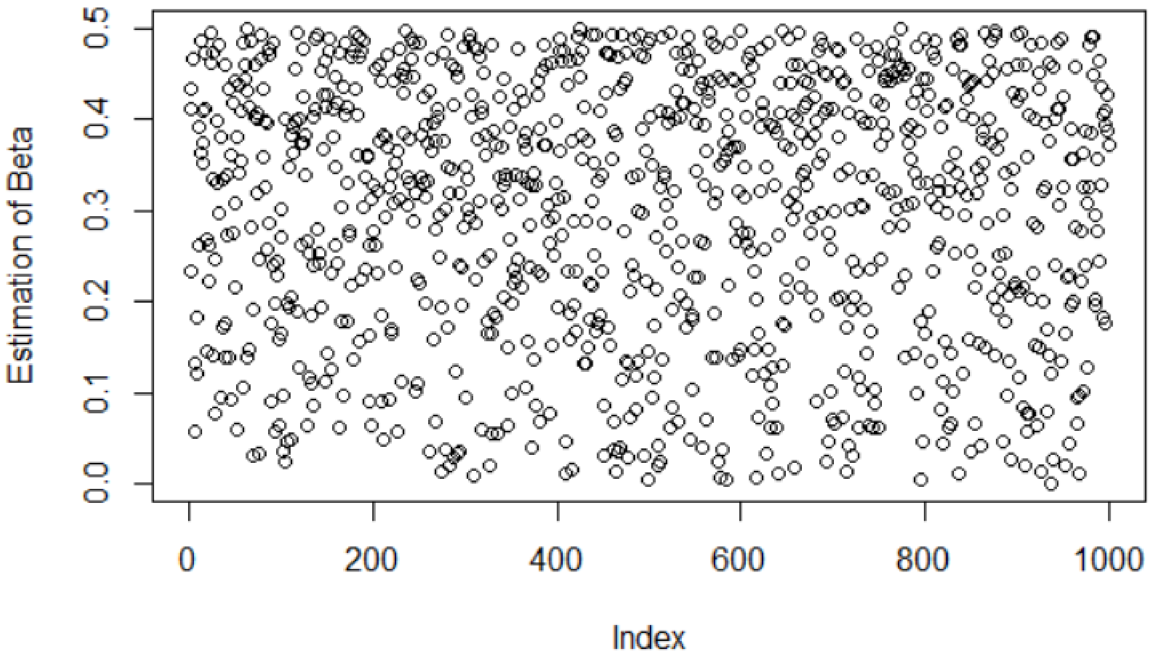
Hence, we have

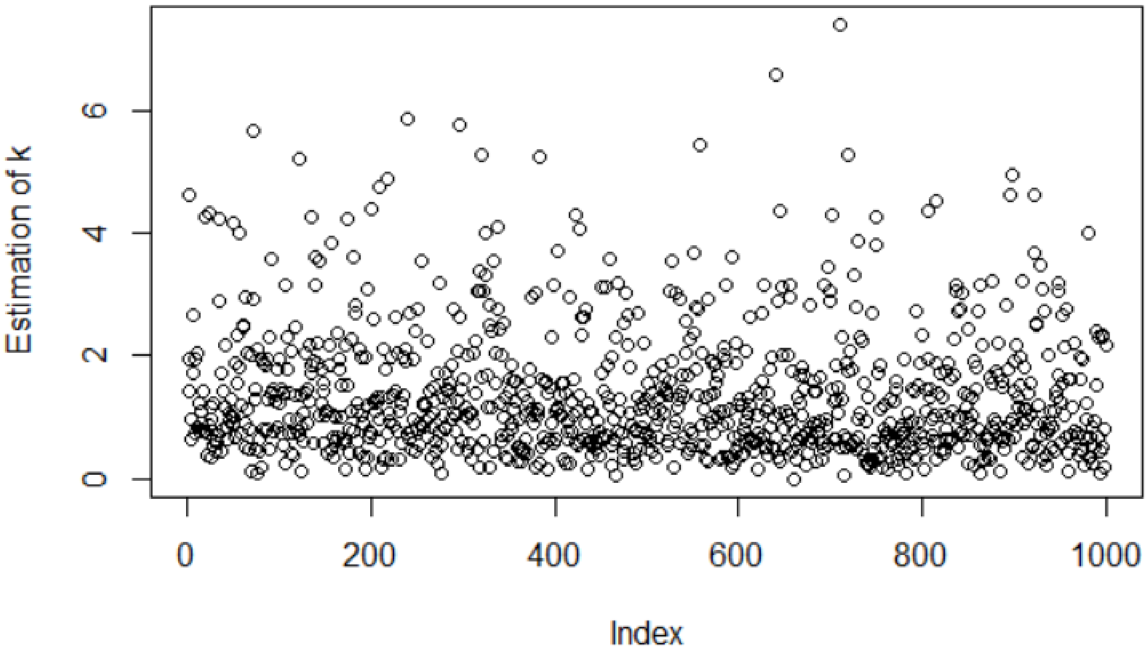
$$\begin{aligned} l(\theta; t, x, \delta, I) = \sum_{i\in\Delta_1} [\beta x_i + \log\big(h_0(t_i)\big)] + \sum_{i\in\Delta_0} [(1 - I_i) logp_0 + I_i log(1 - p_0)] \\ + m(1 - I_i) logp_0 + mI_i \log(1 - p_0) + \left[\gamma\left(1 - \sqrt{1 + \frac{2E_i}{\gamma}}\right)\right] + \mathrm{mlog}(\frac{\pi\gamma}{2}) \end{aligned}$$

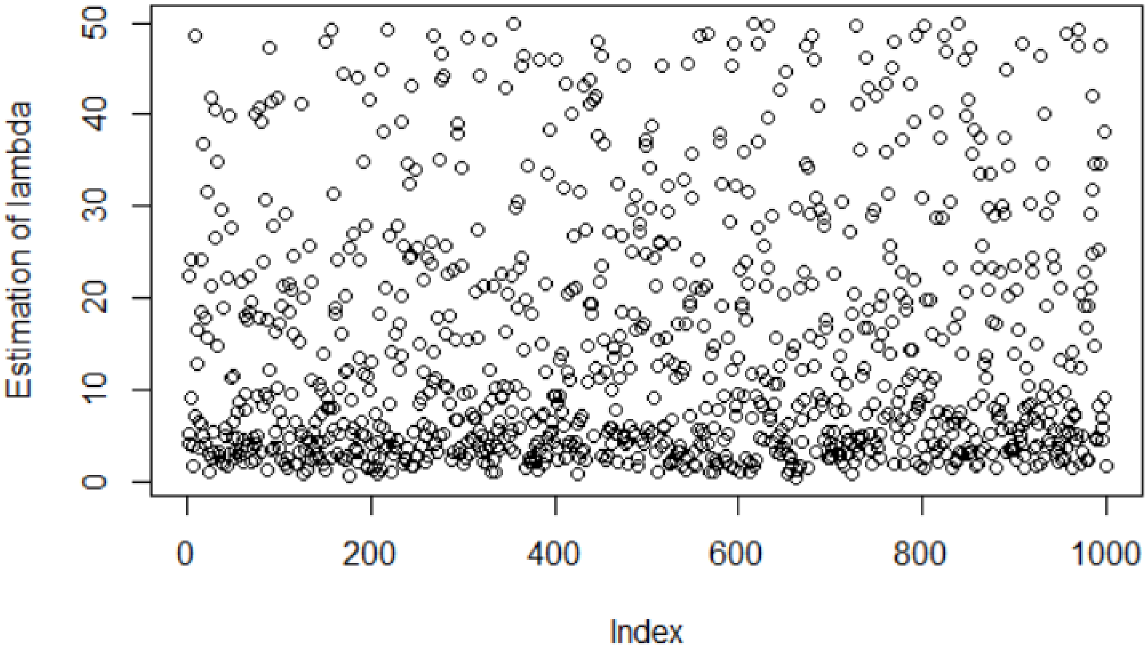
4 Analysis of Cutaneous Melanoma Data

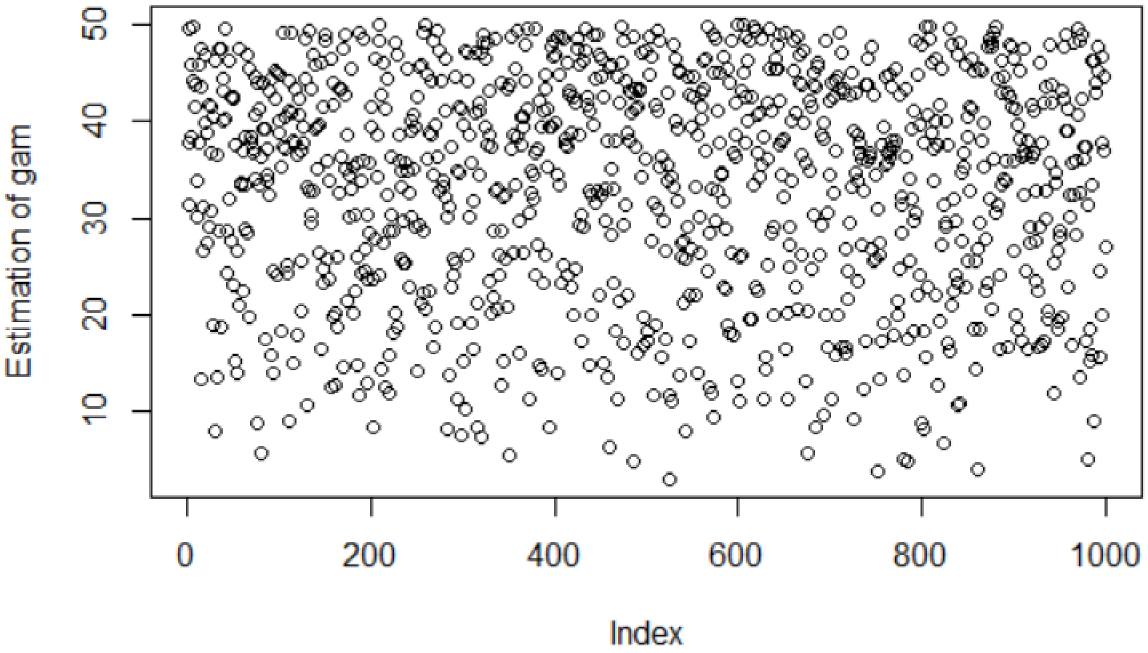
The proposed model is illustrated with a dataset on cancer recurrence taken from Ibrahim et al; the data are part of a study on cutaneous melanoma (a cancer that starts in the pigment-producing cells of the skin) for the evaluation of post-treatment with a high dose of interferon alpha-2b as medicine to prevent recurrence. There were 417 patients in the study divided into four nodule categories based on tumor thickness and this will be the only covariate (x=1,2,3,4) in the analysis. The sample sizes for the four nodule categories are 111, 137, 87 and 82, respectively. The patients have been observed for the period 1991-1995 and followed until 1998. The overall percentage of censored observation is 56%. We considered the observed censoring proportion of each group to be its cure rate. What was observed was either the exact lifetimes (time till patients' death) or the censoring times, in years; the observed lifetimes had mean 3.18 and standard deviation 1.69.

To approximate the parameters, we maximized the log-likelihood function. We uniformly generated k, γ, λ in range from 0 to 50. For β, since 1/(1+ $e^{\beta x}$) is the cure rate, we generated β in range from 0 to 0.5 instead. Under 100 iterations, we recorded $(k, \gamma, \lambda, \beta)$ with the maximum likelihood estimation. Then we iterated the whole process 1000 times and generated samples for $(k, \gamma, \lambda, \beta)$.




Estimation of k
0
2
4
6
0
200
400
600
800
1000
Index


Estimation of lambda
0
10
20
30
40
50
0
200
400
600
800
1000
Index



We then noticed that most of k in the sample were captured within (0,3), most of lambda in the sample were captured within (0,15), whereas beta and gamma distributed relatively uniform in the range.

Mean, median and standard error of beta are 0.299, 0.326 and 0.142, respectively. Mean, median and standard error of k are 1.364, 1.070 and 1.035, respectively. Mean, median and standard error of lambda are 14.577, 9.123 and 13.408, respectively. Mean, median and standard error of gamma are 33.520, 35.864 and 11.478, respectively.

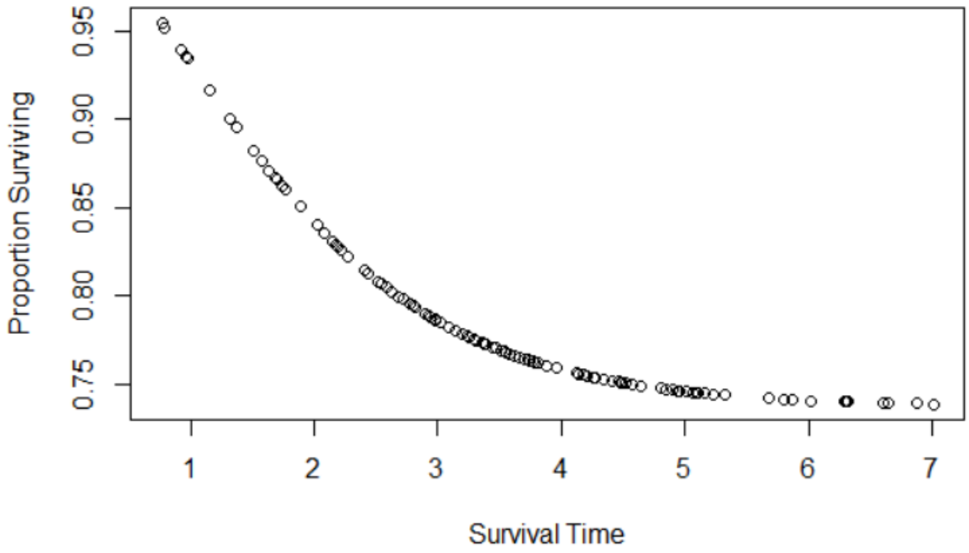
Then we reduced the estimation range for k and lambda and redo the process above.

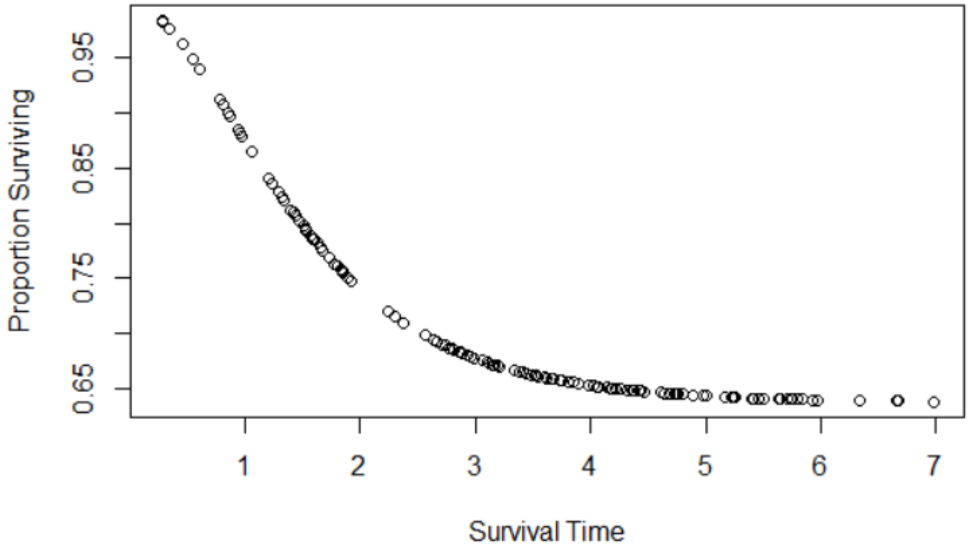
We then obtain that mean and standard error of beta are 0.40, 0.42 and 0.07, respectively. Mean, median and standard error of k are 1.76, 1.71 and 0.520, respectively. Mean, median and standard error of lambda are 2.901, 2.669 and 1.028, respectively. Mean, median and standard error of gamma are 43.255, 44.814 and 5.494, respectively. We can see that standard errors for each parameter get much smaller.

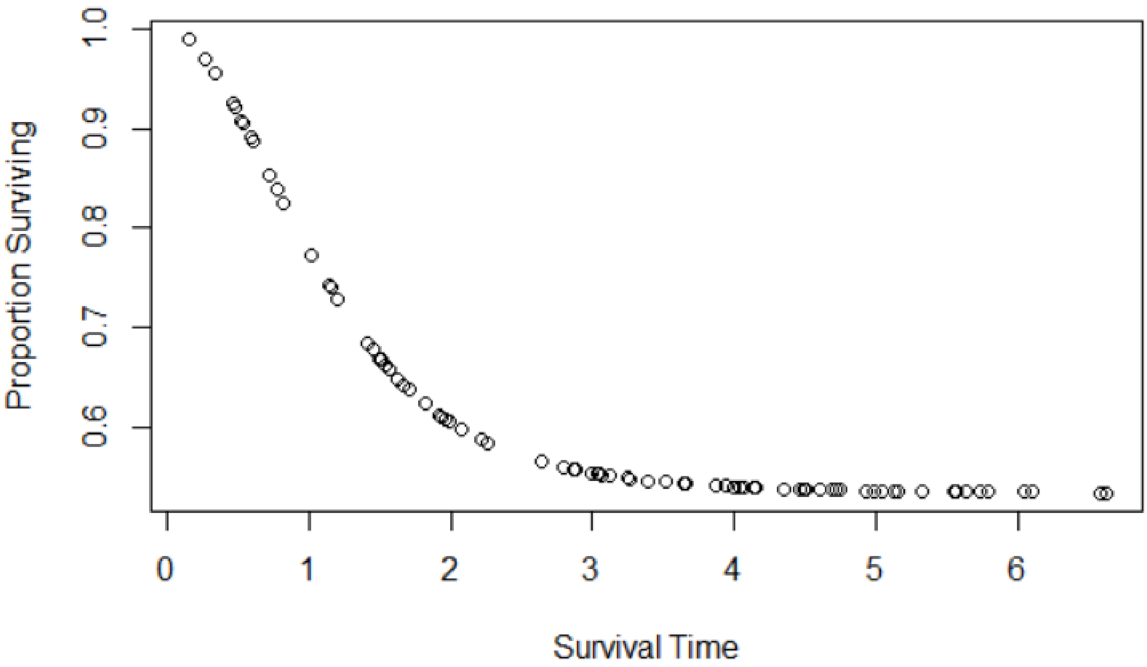
Then we used the profile likelihood approach for estimating beta over [0,0.5] with increment of 0.01 and gamma over [0,50] with increment of 0.1, and then evaluating the log-likelihood value for each beta and gamma. Since we could obtain a relatively precise estimation for k and gamma, we took k=1.76, lambda=2.90. It was observed that the maximum log-likelihood was achieved at beta=0.5 and gamma=50.

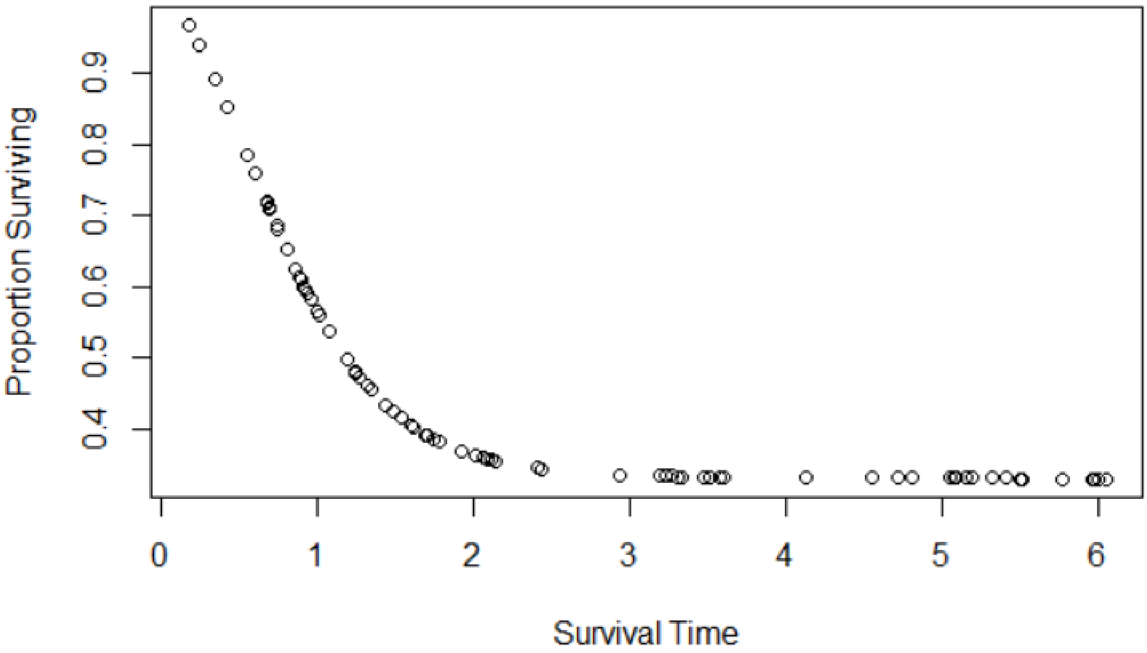
With all parameters being settled, we can plot the probability an individual to be cured, given that he/she has survived up to a specific time.

Four plots of this probability for the four nodule categories can clearly be seen. The cure probability for nodule category 1 is the highest, whereas that of nodule category 4 is the lowest.






Proportion Surviving
1.0
0.9
0.8
0.7
0.6
0
1
2
3
4
5
6
Survival Time


Proportion Surviving
0.9
0.8
0.7
0.6
0.5
0.4
0
1
2
3
4
5
6
Survival Time

5 Concluding Remarks

In this article, a cure rate model has been studied with a frailty model for the lifetime distribution of susceptibles with the baseline hazard function being Weibull distribution and frailty being Inverse Gaussian distribution. The estimation of the model parameters has been carried out by maximum likelihood estimation and profile likelihood approach.

References

1. Ibrahim JG, Chen MH and Sinha D. Bayesian survival analysis. Hoboken, NJ: John Wiley & Sons, 2005.
2. N Balakrishnan, S Barui and FS Milienos. Proportional hazards under Conway-Maxwell-Poisson cure rate model and associated inference. SMMR, 2017
3. N Balakrishnan and Yingwei Peng. Generalized gamma frailty model. Statistics in Medicine. Wiley InterScience, 11 Oct 2005.