MIXTURE AND NON-MIXTURE CURE FRACTION MODELS BASED ON GENERALIZED GOMPERTZ DISTRIBUTION UNDER BAYESIAN APPROACH

Prafulla Kumar Swain — Gurprit Grover — Komal Goel

ABSTRACT. The cure fraction models are generally used to model lifetime data with long term survivors. In a cohort of cancer patients, it has been observed that due to the development of new drugs some patients are cured permanently, and some are not cured. The patients who are cured permanently are called cured or long term survivors while patients who experience the recurrence of the disease are termed as susceptibles or uncured. Thus, the population is divided into two groups: a group of cured individuals and a group of susceptible individuals. The proportion of cured individuals after the treatment is typically known as the cure fraction. In this paper, we have introduced a three parameter Gompertz (viz. scale, shape and acceleration) or generalized Gompertz distribution in the presence of cure fraction, censored data and covariates for estimating the proportion of cure fraction through Bayesian Approach. Inferences are obtained using the standard Markov Chain Monte Carlo technique in openBUGS software.

1. Introduction

In lifetime data analysis, the standard survival analysis techniques inherently assume that all the subjects have the same susceptibility to the disease and will eventually experience the event over a sufficiently long period of follow-up. But the situation in which the studied population is a mixture of uncured (susceptible individuals- who may experience the event of interest), and cured (non--susceptible individuals- who will never experience the event), the standard survival models are usually not appropriate because they do not account for the possibility of cure. Many patients with disease like cancer can be long-term survivors, and thus cure models can be a useful tool to analyze and describe their

survival data. Progress in the treatment of cancer has led to a spate of statistical research to develop cure models. These models are generally used to model lifetime data with long term survivors with an objective to estimate the cure rate, survival distribution and the effect of covariates. The cure fraction is a useful measure to monitor trends and differences in survival of curable diseases. The first created cure model, which is still widely used in survival analysis, had been constructed by [3] and later developed by [2]. The model proposed by him is formulated in terms of a mixture model, which introduces a component representing the proportion of immunes in the population and a distribution representing the survival experience of the susceptible, called the latency distribution. There are many choices for latency distribution like Weibull, gamma, lognormal, Gompertz, exponential etc. Many analyses of cancer survival data are based on overall survival or progression free survival (PFS). No patient can be "cured" of death, so in these situations cure models can be used to model long--term survivors rather than cured patients. These models can be used to investigate heterogeneity between cured and uncured patients. Also they are suitable for modeling censored and uncensored lifetime data. This model extends several distributions widely used in the lifetime data analysis allowing flexibility in modeling monotone and non-monotone shape hazard rates and it serves as a good alternative for the analysis of real datasets. Generally, in a population cure is said to occur when the mortality rate in the diseased group of individuals return to same level as that of general population. A straight forward way to identify the presence of long term survivors in a dataset is to look at the pattern of survival curve of patients in dataset. If the survival curve has a plateau at the end, then a cure model may be an appropriate model for analyzing that dataset. These models can be a useful alternative to Cox-proportional hazard models as these can be used in situations where the assumption of proportionality fails. Also, they are helpful in determining the covariates associated with long-term as well as short term effects. Cure models provide simultaneous estimates of the proportion of the patients cured from the disease and the distribution of the survival times for uncured patients (latency distribution). A c h c a r et al. (2012) [1] estimated the cure fraction by using a two parameter Weibull distribution in the presence of cure fraction under both mixture and non mixture cure models. A Bayesian analysis of the four-parameter generalized modified Weibull (GMW) distribution in presence of cure fraction, censored data and covariates had been presented by [10]. K a n n a n et al. [7] used the generalized exponential cure rate model with covariates to estimate cure fraction under mixture and non mixture setup. Several authors [11] used Poisson distribution as latent distribution in analyzing long term survivors. Y a m a g u c h i (1992) [13] used the generalized gamma to model the latency distribution and the logistic function to model the cure fraction in terms of covariates. Y u et al. [14] established that among the various

distributions namely lognormal, loglogistic, Weibull and generalized gamma, the estimate of cure fraction was robust with generalized gamma distribution. As most of the authors used Weibull and exponential distributions mostly for estimating cure rate, we here used the generalized Gompertz distribution for the analysis purpose.

The Gompertz distribution is one of the commonly used distributions in survival analysis, as it is a flexible distribution which can be skewed to the right and to the left and is also a continuous probability distribution on $(0, \infty)$ that has an increasing failure rate. It has been used as a growth model to fit the tumor growth in cancer studies. C h i e n- L i n S u et al. had used a 2-parameter Gompertz distribution (scale and shape parameters) for the survival analysis of smoking-cessation data in the presence of cure fraction. In addition to shape and scale parameter of Gompertz distribution, we have considered the acceleration parameter (which works as a factor of fragility in the survival of the individual when the time increases) for determining the changes in the hazard rate with respect to time. The main advantage of this new distribution is that it has increasing or constant or decreasing or bathtub curve failure rate depending upon the shape parameter. Also, we have made an attempt to estimate the cure fraction using Bayesian approach while assuming beta and gamma priors on parameters under mixture and non mixture cure models. Bayesian methods make it easier to estimate and analyze complicated problems, while classical inference methods are quite cumbersome. Also, the Bayesian approach allows us to include any prior information that we have on the parameters in the model and hence helps in obtaining a much refined set of posterior estimates. An illustration of proposed methodology is applied to a real dataset of melanoma cancer patients in the presence of cure data, censored observations and covariates.

The rest of the paper is organized as follows. In the next section we describe (i) classes of cure models namely, mixture and non mixture cure models, (ii) the 3-parameter Gompertz distribution and (iii) its likelihood under both mixture and non-mixture models. Then, in Section 3, results have been given and finally the discussion has been presented in Section 4.

2. Materials and methods

2.1. Mixture cure fraction model

A mixture cure fraction model [9], as the name suggests, is a mixture of two types of individuals. In this model, population is divided into two segments, viz. cured or long term survivors and uncured or susceptibles.

Let $p(0 < p < 1)$ be the probability of an individual being cured and $(1 - p)$ be the probability of an individual being susceptible.

Then the survival function at time t can be defined as

$$S(t) = p + (1 - p)S_0(t), \tag{1}$$

where $S_0(t)$ is the baseline survival function for the susceptible individuals, which is assumed to follow generalized Gompertz distribution with the parameters λ, c, θ, i.e.,

$$S_0(t) \sim GGD(\lambda, c, \theta).$$

The cumulative distribution function for the lifetime T can be defined as

$$F(t) = 1 - S(t) = (1 - p)\big(1 - S_0(t)\big) = (1 - p)F_0(t)$$

implying that the probability density function is

$$f(t) = (1 - p)f_0(t),$$

where $f_0(t)$ is the baseline probability density function for the susceptible individuals.

Let (t_i, δ_i) be a random sample of size n $(i = 1, 2, \ldots, n)$ from the cancer data set, where t_i is the survival time for the ith cancer patient and δ_i is the indicator variable defined as:

$$\delta_i = \begin{cases} 1, & \text{for uncensored lifetime,} \\ 0, & \text{for censored lifetime} \end{cases}, \quad i = 1, 2, \ldots, n.$$

Therefore, the contribution of ith cancer patient to the likelihood is given by

$$L_i = \big[f(t_i)\big]^{\delta_i}\big[S(t_i)\big]^{1-\delta_i} = \big[(1-p)f_0(t_i)\big]^{\delta_i}\big[p + (1-p)S_0(t_i)\big]^{1-\delta_i}. \tag{2}$$

2.2. Non-mixture cure fraction model

A non mixture cure fraction model [1] defines an asymptote for the cumulative hazard and hence for the cure fraction. In this case, the survival function is defined as

$$S(t) = p^{F_0(t)} = \exp[(\log p)]F_0(t), \tag{3}$$

where

$$F_0(t) = 1 - S_0(t).$$

Under this model, the contribution of the ith subject to the likelihood function is given by

$$L_i = \big[h(t_i)\big]^{\delta_i} S(t_i) = \big[-(\log p)f_0(t_i)\big]^{\delta_i} \exp\big[(\log p)F_0(t_i)\big], \tag{4}$$

where $h(t) = \frac{f(t)}{S(t)}$ is the hazard function.

2.3. Generalized Gompertz distribution for the susceptible individuals

As a special case, let us assume a three parameter Gompertz distribution introduced by [6] for the susceptible individuals with probability density function given by

$$f_0(t) = \theta \lambda e^{ct} \exp\left[-\frac{\lambda}{c}\left(e^{ct}-1\right)\right]\left[1 - \exp\left[-\frac{\lambda}{c}\left(e^{ct}-1\right)\right]\right]^{\theta-1}, \qquad (5)$$

where $t, \lambda, \theta, c > 0$. Here λ is a scale parameter, c is an acceleration parameter and θ is the shape parameter. The main property of this new distribution is that it has decreasing or unimodal probability density function and has increasing, constant or decreasing shaped hazard function. Figure 1, represents the hazard plot of this distribution. This property makes this distribution very useful in survival analysis. The advantage of this distribution over exponential distribution is that it has increasing or decreasing failure rate, while exponential distribution has constant hazard rate. Also, the Weibull distribution has disadvantage over the generalized Gompertz distribution and it is that the maximum likelihood estimators of the parameters of Weibull distribution may not behave property for all parameter values even when the location parameter is set to zero.
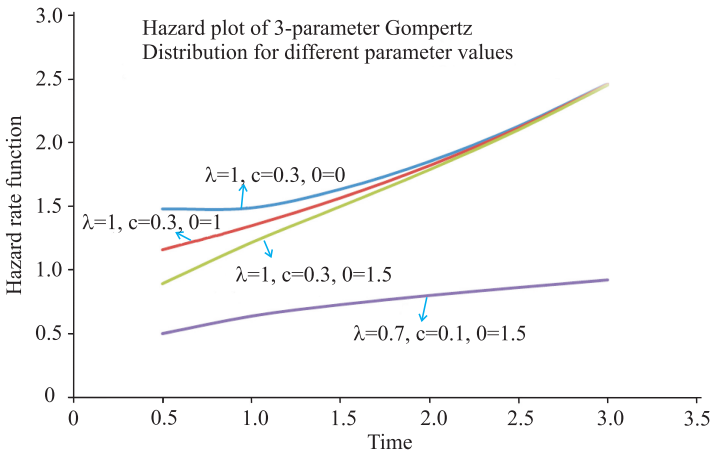

FIGURE 1. Hazard plot of 3-parameter Gompertz distribution.

Its survival function is given as

$$S_o(t) = 1 - \left[1 - \exp\left[-\frac{\lambda}{c}\left(e^{ct}-1\right)\right]\right]^{\theta}. \qquad (6)$$

Then, the cumulative density function is

$$F_0(t) = 1 - S_0(t).$$

125

The 2-parameter Gompertz distribution is a particular case of this distribution with shape parameter $\theta = 1$.

Therefore, the log likelihood functions of this distribution under mixture and non-mixture cure fraction models can be written as

$$l(\beta) = \log(1-p)\sum_i \delta_i + \sum_i \delta_i \log \theta + \sum_i \delta_i \log \lambda + c\sum_i \delta_i t_i$$

$$-\frac{\lambda}{c}\sum_i \delta_i(e^{ct_i}-1) + (\theta-1)\sum_i \delta_i \log\left(1-\exp\left[-\frac{\lambda}{c}(e^{ct_i}-1)\right]\right)$$

$$+\sum_i(1-\delta_i)\left[\log\left[p+(1-p)\left\{1-\left(1-\exp\left[-\frac{\lambda}{c}(e^{ct_i}-1)\right]\right)^\theta\right\}\right]\right] \quad (7)$$

(under mixture cure models),

and

$$l(\beta) = \sum_i \delta_i \log(-\log p) + \sum_i \delta_i \log \theta + \sum_i \delta_i \log \lambda + c\sum_i \delta_i t_i$$

$$-\frac{\lambda}{c}\sum_i \delta_i(e^{ct_i}-1) + (\theta-1)\sum_i \delta_i \log\left(1-\exp\left[-\frac{\lambda}{c}(e^{ct_i}-1)\right]\right)$$

$$+\left[\sum_i(\log p)\left(1-\exp\left[-\frac{\lambda}{c}(e^{ct_i}-1)\right]\right)^\theta\right] \quad (8)$$

(under non-mixture cure models),

where $\beta = (p, \lambda, c, \theta)$.

The joint posterior distribution for the parameters of the model is obtained by combining the joint prior distribution with the likelihood function of β. Although the joint posterior distribution for the parameters of the proposed model is of great complexity, samples of the joint posterior distribution can be generated using some existing MCMC (Markov Chain Monte Carlo) simulation methods. A great computational simplification to simulate these samples is obtained using the OpenBUGS software, where we only need to specify the distribution for the data and the prior distributions for the parameters. For a Bayesian analysis of the mixture and non-mixture models not including covariates, we assume a beta prior distribution Beta $(1, 1)$ for the proportion p of the long-term survivors because p is a probability and can only take values in the interval $(0, 1)$. We also assume gamma prior distribution for the parameters of generalized Gompertz distribution where Gamma (a, b) denotes a gamma distribution with mean (a/b) and variance (a/b^2) with $a = b = 1$. In the presence of covariates, we assume $N(e, f^2)$ with mean e and variance f^2 for all the regression parameters. For all cases, we assume prior independence among the parameters included in the model.

Posterior summaries of interest are obtained from simulated samples for the joint posterior distribution using standard Markov Chain Monte Carlo (MCMC) procedures. For all the cases, we assume a burn-in-sample of 10,000 samples, taking every 100th sample to have approximately uncorrelated values. We generated 10,10,000 samples for each parameter of interest. The 10000 first simulated samples were discarded as a burn-in period, which is usually used to eliminate the effect of the initial values.

2.4. Model comparison criteria

Comparison between mixture and non-mixture models assuming different distributions was assessed using the deviance information criteria (DIC) proposed by [12], where a lower DIC value indicates better model fit.

The DIC is defined as follows:

$$DIC = \hat{D} + 2(\bar{D} - \hat{D}),$$

where $\bar{D}$ is the average of deviance (-2LogLikelihood) over the posterior distribution, and $\hat{D}$ is the deviance calculated at the posterior mean parameters.

To obtain inferences regarding the predictors, we have consider the following regression model

$$\lambda_i = \beta_o \exp(\beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki}) \qquad \text{for overall survival}$$

and

$$\log\left(\frac{p_i}{1 - p_i}\right) = \alpha_o + \alpha_1 x_{1i} + \cdots + \alpha_k x_{ki} \qquad \text{for cure fraction,}$$

where $x = (x_1, \ldots, x_k)'$ is a set of predictors affecting the parameters. Assuming the mixture and non-mixture models based on the generalized Gompertz distribution; let us consider normal prior distributions $N(0, 100)$ for the parameters of regression models. Thus, we are assuming approximately non-informative priors for these parameters. Note that the parameter α_2 is related to the effect of the treatment on the cure fraction. If the credible interval for α_2 includes zero, we can conclude that there is no evidence of treatment effect on cure fraction.

3. Results

To illustrate the methodology, mixture and non-mixture cure models are fitted to the melanoma dataset from the ECOG phase III clinical trial e1684, which is also illustrated by PSPMCM SAS macro [5]. This trial was a two-arm clinical trial involving patients randomized to one of two treatment arms: high-dose interferon (IFN) or observation. The aim of this trial is to evaluate the high dose interferon alpha-2b (IFN) regimen against the placebo as the postoperative adjuvant therapy. After deleting missing data, a total number of 284 observations is used in the analysis. Three covariates viz. treatment (0 control group;

1 IFN group), gender (0 for male, 1 for female) and age are taken into account both in the incidence and latency parts. A total of 69 % patients were found to be censored in the trial. The Kaplan-Meier estimate of the survival function is given in Figure 2, where the presence of a plateau near 0.3 suggests that cure model is suitable for this dataset.
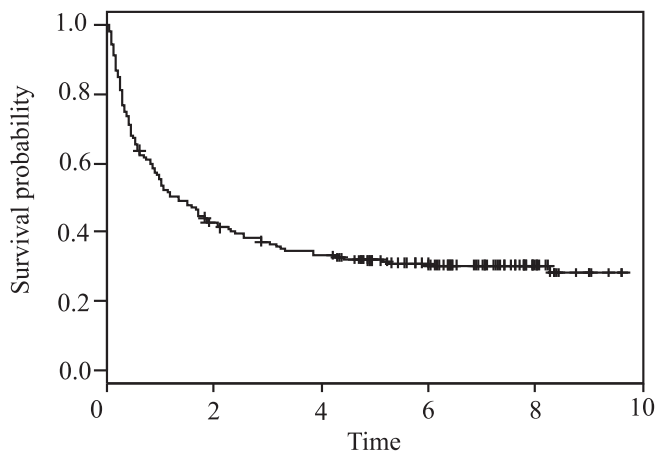

FIGURE 2. Kaplan–Meier survival curve.

We first obtained the posterior summaries under standard Weibull, standard Gompertz and generalized Gompertz distribution. The parameter estimates from all these three distributions are quite similar (Table 1). Moreover, from the DIC values, it has been observed that the generalized Gompertz distribution has the smallest DIC value, concluding this distribution to be best amongst all.

Figure 3 represents the Cox-Snell residual curves for testing the goodness of fit of Weibull and Gompertz distributions. For this, we computed the Kaplan-Meier estimates of the cumulative hazard function and plotted it against the Cox-Snell

TABLE 1. Posterior summaries not including the cure fraction p.

Model	Parameter	Posterior mean (SD)	95 % Credible interval	DIC
Standard Weibull (2-pmts.)	λ	8.347(81.34)	(0,37.9)	393.2
	θ	8.6(81.86)	(0,38.4)	
Standard Gompertz (2-pmts.)	λ	8.961(89.01)	(0,44.07)	580.7
	θ	9.945(102.6)	(0,47.76)	
Generalized Gompertz (3-pmts.)	λ	8.63(84.32)	(0,36.9)	200.6
	θ	10.54(110.6)	(0,44.3)	
	c	9.19(91.16)	(0,49.9)	

residuals. The best fitted model is one in which Cox-Snell residuals have an exponential distribution, which is being showed by Gompertz distribution. In Table 1, we have the posterior summaries of the estimates considering Bayesian approach for each of these probability distributions. The Bayesian estimates were obtained using OpenBUGS software.
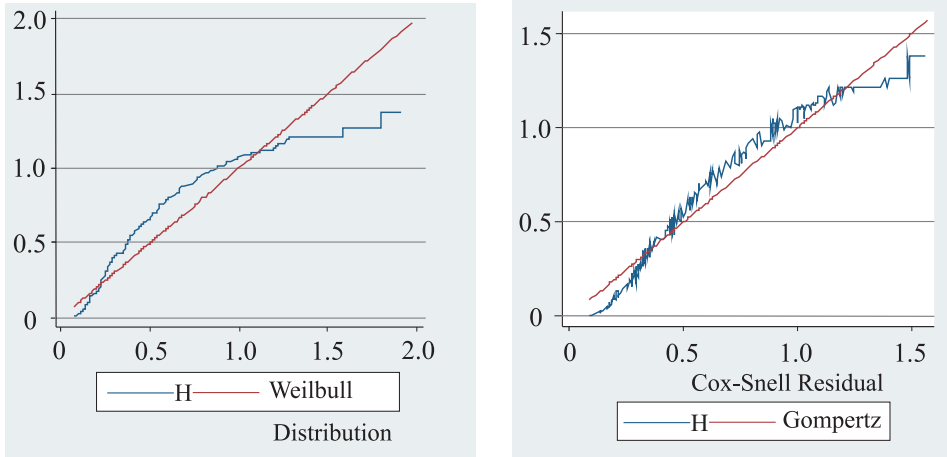

FIGURE 3. Cox-Snell plot for Weibull and Gompertz distribution.

As the result from Table 1 indicates GGD to be the best, we now fit cure models on this distribution in our dataset of melanoma cancer patients. To analyze this dataset, we consider the mixture and non mixture cure fraction models defined earlier in the presence of and not of covariates. As a first analysis, we assume the cure fraction models not in presence of covariates. Table 2 presents the posterior summaries of parameters based on GGD under mixture and non-mixture cure models in the absence of covariates.

From the fitted cure models in the absence of covariates, it has been observed that the mixture and non mixture GGD fits well to the survival times. The results indicate that the cure fraction (p) is significant under both the models. Also the DIC value of mixture cure model (282.8) is less than the non-mixture cure model (328.7). In the presence of covariates (treatment, age, sex), we assume the following two regression models,

$$\lambda_i = \beta_o \exp(\beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i}) \qquad \text{for overall survival}$$

and,

$$\log\left(\frac{p_i}{1-p_i}\right) = \alpha_o + \alpha_1 x_{1i} + \alpha_2 x_{2i} + \alpha_3 x_{3i} \qquad \text{for cure fraction,}$$

where x_{1i} is the sex of the patient (0=male, 1=female), x_{2i} denotes the treatment (1=IFN treatment group, 0=control group), x_{3i} is the patient age for $i = 1, 2, \ldots, n$. Table 3 gives the estimates of the regression models considering Bayesian approach under both the models in the presence of covariates.

TABLE 2. Posterior summaries including the cure fraction p
(in the absence of covariates).

Model	Parameter	Posterior mean (SD)	95 % Credible interval	DIC
Mixture cure model	λ	9.407(103)	(0,30.97)	282.8
	θ	8.757(86.15)	(0,41.12)	
	c	10.63(98.85)	(0,67.23)	
	p	0.499(0.291)	(0.024,0.976)	
Non-mixture cure model	λ	6.785(63.32)	(0,23.52)	
	θ	7.137(81.58)	(0,32.1)	328.7
	c	9.037(86.89)	(0,48.85)	
	p	0.498(0.291)	(0.023,0.975)	

TABLE 3. Posterior summaries including cure fraction p
(in the presence of covariates).

Model	Parameter	Posterior mean	SD	95 % Credible interval
Mixture cure model	$\alpha_{0(\text{intercept})}$	0.0002	0.102	(-0.192,0.195)
	$\alpha_{1(\text{sex})}$	0.0013	0.099	(-0.196,0.191)
	$\alpha_{2(\text{treatment})}$	-0.0005	0.100	(0.189,0.199)
	$\alpha_{3(\text{age})}$	0.0029	0.098	(-0.189,0.198)
	$\beta_{0(\text{intercept})}$	0.0009	0.101	(-0.199,0.199)
	$\beta_{1(\text{sex})}$	-0.0032	0.098	(-0.198,0.186)
	$\beta_{2(\text{treatment})}$	-0.0011	0.101	(0.194,0.201)
	$\beta_{3(\text{age})}$	-0.0007	0.098	(-0.188,0.192)
	c	11.55	113.10	(0,55.97)
	θ	7.037	60.71	(0,29.75)
Non-mixture cure model	$\alpha_{0(\text{intercept})}$	-0.0010	0.100	(-0.194,0.193)
	$\alpha_{1(\text{sex})}$	0.0007	0.100	(-0.196,0.198)
	$\alpha_{2(\text{treatment})}$	-0.0008	0.099	(0.196,0.194)
	$\alpha_{3(\text{age})}$	0.0007	0.100	(-0.195,0.201)
	$\beta_{0(\text{intercept})}$	-0.0015	0.101	(-0.199,0.199)
	$\beta_{1(\text{sex})}$	0.0002	0.099	(-0.195,0.196)
	$\beta_{2(\text{treatment})}$	-0.0002	0.100	(0.197,0.200)
	$\beta_{3(\text{age})}$	0.0008	0.100	(-0.194,0.195)
	c	10.11	96.03	(0,43.94)
	θ	12.95	130.8	(0,64.01)

Dynamic trace and posterior Density plot of cure fraction under mixture model
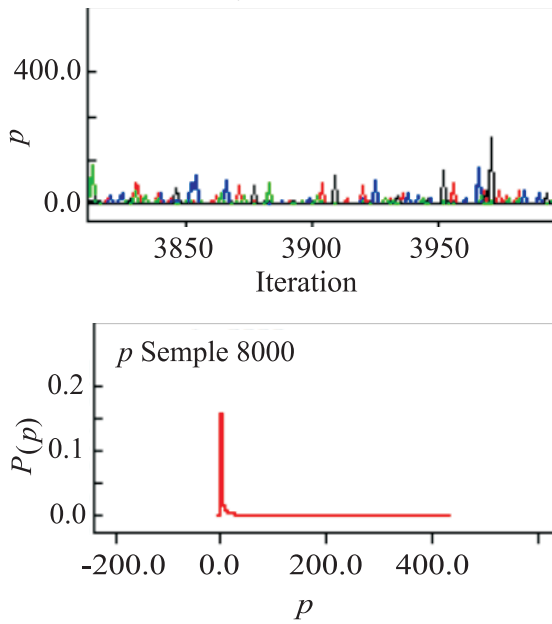
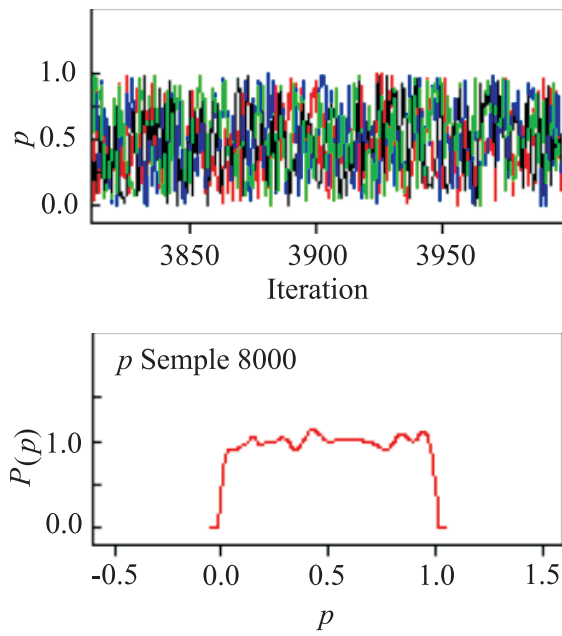Dynamic trace and posterior Density plot of cure fraction under non-mixture model

FIGURE 4.   Trace plots for convergence diagnostics and marginal posterior kernel density plots.

From Table 3, we observe that the similar results are obtained for both the models. A significant comparison reveals that 95 % Credible Interval for α_2 and β_2 does not include zero suggesting that the IFN treatment has a significant effect on the survival and cure probability, and the covariates namely age and sex have no significant effect on the survival and cure probability. Figure 4 shows the dynamic trace and posterior density plots of cure fraction under mixture and non mixture cure models. The trace plot indicates that the Markov chain has stabilized with good mixing and hence MCMC algorithm converged, and the kernel density plot estimates the posterior marginal distribution.

4. Discussion

The purpose of this study is to show the utility and flexibility of generalized Gompertz distribution under cure models. These models have been well developed in the statistical literature, but are not as common in the clinical literature. For diseases like cancer in which patients are long-term survivors, cure models can provide an interesting way to characterize and study their survival. They are generally used to model lifetime data with long term survivors. Even though the cure models are first proposed by [3] but they are still widely used in survival analysis, as they provide a measure to monitor trends and differences in survival of curable diseases. These are the only models which provide an estimate for proportion of cure patients and distribution of survival times of uncured patients simultaneously. These models are very appropriate when the studied population is a mixture of cure (who does not experience the event) and susceptible individuals (who experience the event). There are two classes of cure models, mixture and non-mixture cure models, both of which can describe short-term and long-term effects. One advantage of these models besides estimating cure rate is to reduce the cure model to a common survival model in absence of cure patients, i.e., if the study or follow-up time is long enough, then also these models are reliable. Also, they incorporate several distributions which are widely used in the lifetime data analysis, allowing flexibility in modeling their monotone and non-monotone shape hazard rates.

Our aim in this study is to show the utility of generalized Gompertz distribution under cure models. Based on this distribution we developed a cure fraction regression model suitable for modeling censored and uncensored lifetime data. Aiming to show its flexibility, practical relevance and applicability of the regression model, we provided an application to melanoma cancer patient's dataset. As the Kaplan-Meier survival curve shows a long and stable plateau with heavy censoring at the tail, it may be taken as evidence of a cured fraction. Mixture and non-mixture cure models are fitted over a dataset of 284 melanoma cancer

patients with baseline survivor function as generalized Gompertz distribution. A descriptive comparison reveals that mixture cure models are best fitted than non-mixture cure models which match with the findings of [1]. As treatment effect is significant across both the models, implying that IFN treatment improve the cure rate and survival of patient, which has also been proved by [8].

As discussed above, some alternative parametric distributions could also be considered that provide more flexibility in the shape of excess mortality/relative survival functions, while still giving reliable estimates of cure fraction.

REFERENCES

[1] ACHCAR, J. A.—COELHO-BARROS, E.—MAZUCHELI, J.: *Cure fraction models using mixture and non-mixture models*, Tatra Mt. Math. Publ. **51** (2012), 1–9.

[2] BERKSON, J.—GAGE, R.: *Survival curve for cancer patients following treatment*, J. Amer. Statist. Assoc. **47** (1952), 501–515.

[3] BOAG, J. W.: *Maximum likelihood estimates of the proportion of patients cured by cancer therapy*, J. R. Stat. Soc. Ser. B Stat. Methodol. **11** (1949), 15–44.

[4] CHIEN-LIN, S.—ADAM DING, A.—WANG, W.: *Analysis of clustered survival data in the presence of cure with the Gompertz distribution model*, 2013, http://www.wseas.us/e-library/conferences/2013/CambridgeUSA/MATHECO/MATHECO-26.pdf

[5] CORBIÈRE, F.—JOLY, P.: *A sas macro for parametric and semiparametric mixture cure models*, Computer Methods and Programs in Biomedicine **8** (2007), 173–180.

[6] GOHARY, E.—ALSHAMRANI, A.—AL-OTAIBI, A.: *The generalized Gompertz distribution*, Appl. Math. Model. **37** (2013), 13–24.

[7] KANNAN, N.—KUNDU, D.—NAIR, P.—TRIPATHI, R. C.: *The generalized exponential cure rate model with covariates*, J. Appl. Stat. **37** (2010), 1625–1636.

[8] KIRKWOOD, J. M.—STRAWDERMAN, M. H.—ERNSTOFF, M. S.—SMITH, T.––BORDEN, E. C.—BLUM, R. H.: *Interferon alfa-2b adjuvant therapy of high-risk resected cutaneous melanoma: the eastern cooperative oncology group trial est 1684*, J. Clinical Oncology **14** (1996), 7–17.

[9] MALLER, R. A.—ZHOU, X.: *Survival Analysis with Long-Term Survivors.* John Wiley & Sons Ltd., Chichester, 1996.

[10] MARTINEZ, E. Z.—ACHCAR, J. A.—JÁCOMEB, A. A.—SANTOS, S. J.: *Mixture and non-mixture cure fraction models based on the generalized modified weibull distribution with an application to gastric cancer data*, Computer Methods and Programs in Biomedicine **112** (2012), 343–355.

[11] RAHIMZADEH, M.—KAVEHIE, B.—ZALI, M. R.: *Cure models in analyzing long-term survivors*, Transl. Gastrointest Cancer **3** (2014), 149–154.

[12] SPIEGELHALTER, D. J.—BEST, N. G.—CARLIN, B. P.—LINDE, V.: *A Bayesian measures of model complexity and fit*, J. R. Stat. Soc. Ser. B Stat. Methodol. **64** (2002), 583–639.

[13] YAMAGUCHI, K.: *Accelerated failure-time regression model with a regression model for the surviving fraction: an application to the analysis of 'permanent employ-ment' in Japan*, J. Amer. Statist. Assoc. **87** (1992), 284–292.

[14] YU, B.—TIWARI, R. C.—CRONIN, K. A.—FEUER, E. J.: *Cure fraction estimation from the mixture cure models for grouped survival data*, Stat. Med. **23** (2004), 1733–1747.

Appendix. OpenBUGS codes

The OpenBUGS code used for the analysis of the melanoma cancer data is given below, considering the mixture model.

```
model
{
for(i in 1: N)
    {
    f0A[i]<-theta*lambda*exp(c*t[i])*exp(-(lambda/c)*(exp(c*t[i])-1))
    f0B[i]<-pow(1-exp(-(lambda/c)*(exp(c*t[i])-1)),1-theta)
    f0[i]<-f0A[i]/f0B[i]
    S0[i]<-1-pow(1-exp(-(lambda/c)*(exp(c*t[i])-1)),theta)
    F0[i]<-1-S0[i]
    L[i]<-pow((1-p)*f0[i],d[i])*pow(p+(1-p)*S0[i],1-d[i])
                                        \# for mixture models
    log L[i]<-log(L[i])
    zeros[i]<-0
    zeros[i]~dloglik(log L[i])
    }
 p~dbeta(1,1)
 theta~dgamma(1,1)
 lambda~dgamma(1,1)
 c~dgamma(1,1)
 }
```

In this code N is the sample size, f0[i] is the baseline probability density function as given in eqn. (5), S0[i] is the baseline survivor function given in eqn. (6), L[i] is the likelihood function, t[i] is the time-to-event variable and d[i] is the censoring indicator variable (denoted by δ_{ij}). Considering that the generalized Gompertz distribution is not available directly as a choice in OpenBUGS we used the dloglik() distribution, which requires us to specify the logarithm of the likelihood function.

For the non-mixture model, we replace in the code for `L[i]` by,

```
L[i]<-pow(h[i],d[i])*exp(F0[i]*log(p))
h[i]<-(-(log(p)))*f0[i]
```

We did not encounter any problem of convergence while obtaining posterior samples in the models.

Prafulla Kumar Swain
Department of Statistics
Utkal University
Vanivihar
Bhubaneswar 751004
INDIA
E-mail: prafulla86@gmail.com

Gurprit Grover
Komal Goel
Department of Statistics
Faculty of Mathematical Sciences
University of Delhi
Delhi 110007
INDIA
E-mail: gurpritgrover@yahoo.com
　　　 komalstats@gmail.com