Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
http://siba-ese.unisalento.it/index.php/ejasa/index
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v10n2p463

The defective generalized Gompertz distribution and its use in the analysis of lifetime data in presence of cure fraction, censored data and covariates

By Martinez, Achcar

Published: 14 October 2017

Electronic Journal of Applied Statistical Analysis
Vol. 10, Issue 02, October 2017, 463-484
DOI: 10.1285/i20705948v10n2p463

The defective generalized Gompertz distribution and its use in the analysis of lifetime data in presence of cure fraction, censored data and covariates

Edson Zangiacomi Martinez*and Jorge Alberto Achcar

Ribeirão Preto Medical School, Universidade de São Paulo (USP), Brazil

Published: 14 October 2017

Survival analysis methods are widely used in studies where the variable of interest is related to the time until the occurrence of an event. The usual methods assume that all individuals under study are subject to this event, but there are practical situations where this assumption is unrealistic. In some cases it is possible that a percentage of individuals are immune to the event of interest or, especially in cancer clinical trials, they were cured from their disease after a given treatment. In the literature, this percentage is usually referred as "cure fraction". In the present paper, we have proposed a model based on a modification of the generalized Gompertz distribution introduced by El-Gohary et al. (2013) to account for the presence of a cure fraction. We also considered the presence of censored data and covariates. Maximum likelihood and Bayesian methods for estimation of the model parameters are presented. A simulation study is provided to evaluate the performance of the maximum likelihood method in estimating parameters. In the Bayesian analysis, posterior distributions of the parameters are estimated using the Markov chain Monte Carlo (MCMC) method. An example involving a real data set is presented.

keywords: Survival analysis, Gompertz distribution, censored data, maximum likelihood estimation, Bayesian inference, defective distributions.

*Corresponding author:edson@fmrp.usp.br, achcar@fmrp.usp.br

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

1 Introduction

In some studies using statistical methods of survival analysis, we can observe in a sample a fraction of individuals who appear not to be subject to the occurrence of the event of interest. For example, in clinical trials, a proportion of patients who respond to treatment appear to be free of signs of the disease for a long time being possibly cured (Morbiducci et al., 2003). We can infer their presence of "immune" or "cured" individuals in a data set if many of the largest failure times are censored (Ghitany et al., 1994). This is visually suggested when a Kaplan-Meier plot of the survival function shows a long and stable plateau with heavy censoring at the right of the plot.

Let T be a random variable representing the time to an event of interest. The survival function, defined by $S(t) = P(T > t)$, is the probability that an individual survives longer than a certain time, and traditional methods of survival analysis assume that $S(t)$ converges to zero as t tends to infinity. This assumption is interpreted to mean that all individuals are subject to the occurrence of the event of interest, and this is unrealistic when there are "immune" or "cured" individuals in the population. The mixture model (Farewell, 1982) is commonly used to handle data with such characteristics. It explicitly includes a parameter accounting for the cure rate and thus assumes that the survival function is given by

$$S(t) = p + (1 - p)S_0(t),$$

where p is the proportion of "cured" individuals ($0 < p < 1$) and $S_0(t)$ is the baseline survival function for the susceptible individuals. In this case, if $S_0(t)$ is a function that converges to zero as t tends to infinity, then $S(t)$ converges to p as t tends to infinity. The choice for $S_0(t)$ should be thus based on a proper distribution (with a *pdf* that integrates to one) and common choices includes the Gompertz, exponential and Weibull distributions.

Alternatively, models based on defective distributions are useful in studies which include data describing survival times in the presence of "immune" or "cured" individuals. A defective distribution is defined as an improper distribution that is not normalized to one for some values of their parameters. In this case, the corresponding survival function $S(t)$ converges to a value η as t tends to infinity. Examples of defective distributions in the analysis of time-to-event data include the defective Gompertz distribution (Gieser et al., 1998; Rocha et al., 2014; Santos et al., 2017), the defective inverse Gaussian distribution (Balka et al., 2011), and the exponentiated-Weibull distribution (Cancho and Bolfarine, 2001). More recently, new families of defective distributions have been introduced in the literature, such as those based on the Kumaraswamy (Rocha et al., 2017b) and the Marshall-Olkin families (Rocha et al., 2017a).

The generalized Gompertz distribution (GGD), introduced by El-Gohary et al. (2013), has probability density function (*pdf*) given by

$$f(t) = \theta\alpha e^{\gamma t} \exp\left[-\frac{\alpha}{\gamma}\left(e^{\gamma t} - 1\right)\right]\left\{1 - \exp\left[-\frac{\alpha}{\gamma}\left(e^{\gamma t} - 1\right)\right]\right\}^{\theta-1},$$

where $t \geq 0$, $\alpha > 0$, $\gamma > 0$ and $\theta > 0$. In the present article we propose a modification of the GGD that allows for the presence of a proportion of "cured" patients, by setting $\gamma =$

$-\beta$. We will call this new distribution as the modified generalized Gompertz distribution (MGGD). The remainder of the article is organized as follows. Section 2 introduces the MGGD and presents maximum likelihood and Bayesian inference procedures for its parameters. An extension to a model with covariates is outlined. Section 2 also describes a simulation method for a random variable following a MGGD. Section 3 presents a simulation study and applications to simulated and real data. Section 4 concludes the article.

2 Methods

2.1 The modified generalized Gompertz distribution (MGGD)

Let T be a random variable representing the time to some event. The survival function of the MGGD with three parameters is given by

$$S(t) = 1 - \left\{1 - \exp\left[\frac{\alpha}{\beta}\left(e^{-\beta t} - 1\right)\right]\right\}^{\theta}, \tag{1}$$

where $t \geq 0$, $\alpha > 0$, $\beta > 0$ and $\theta > 0$, and the proportion η of "immune" or "cured" individuals is thus given by

$$\eta = \lim_{t\to\infty} S(t) = 1 - \left[1 - \exp\left(-\frac{\alpha}{\beta}\right)\right]^{\theta}, \tag{2}$$

where $0 < \eta < 1$. A random variable that follows a MGGD will be denoted by $T \sim MGGD(\alpha, \beta, \theta)$. The parameters α and β are scale parameters and θ is a shape parameter. The modified Gompertz distribution (MGD), first introduced by Cantor and Shuster (1992) and after extended by Gieser et al. (1998), is a special case of (1) in which $\theta = 1$. Figure 1 illustrates the survival curves (1) for the MGGD, considering different values for the parameters α, β and θ. The *pdf* of the MGGD is given by

$$f(t) = \theta\alpha e^{-\beta t} \exp\left[\frac{\alpha}{\beta}\left(e^{-\beta t} - 1\right)\right]\left\{1 - \exp\left[\frac{\alpha}{\beta}\left(e^{-\beta t} - 1\right)\right]\right\}^{\theta-1} \tag{3}$$

and the corresponding hazard function $h(t)$ takes the following form

$$h(t) = \frac{f(t)}{S(t)} = \frac{\theta\alpha e^{-\beta t} \exp\left[\frac{\alpha}{\beta}\left(e^{-\beta t} - 1\right)\right]\left\{1 - \exp\left[\frac{\alpha}{\beta}\left(e^{-\beta t} - 1\right)\right]\right\}^{\theta-1}}{1 - \left\{1 - \exp\left[\frac{\alpha}{\beta}\left(e^{-\beta t} - 1\right)\right]\right\}^{\theta}}.$$

From the cumulative function $F(t) = 1 - S(t) = P(T \leq t)$, we have that the quantiles of the MGGD can be obtained by using the expression

$$F^{-1}(u) = -\frac{1}{\beta}\ln\left[1 + \frac{\beta}{\alpha}\ln\left(1 - u^{\frac{1}{\theta}}\right)\right], \tag{4}$$

where $0 < u < 1$. As a special case, the median of the MGGD is obtained from (4) setting $u = 0.5$.

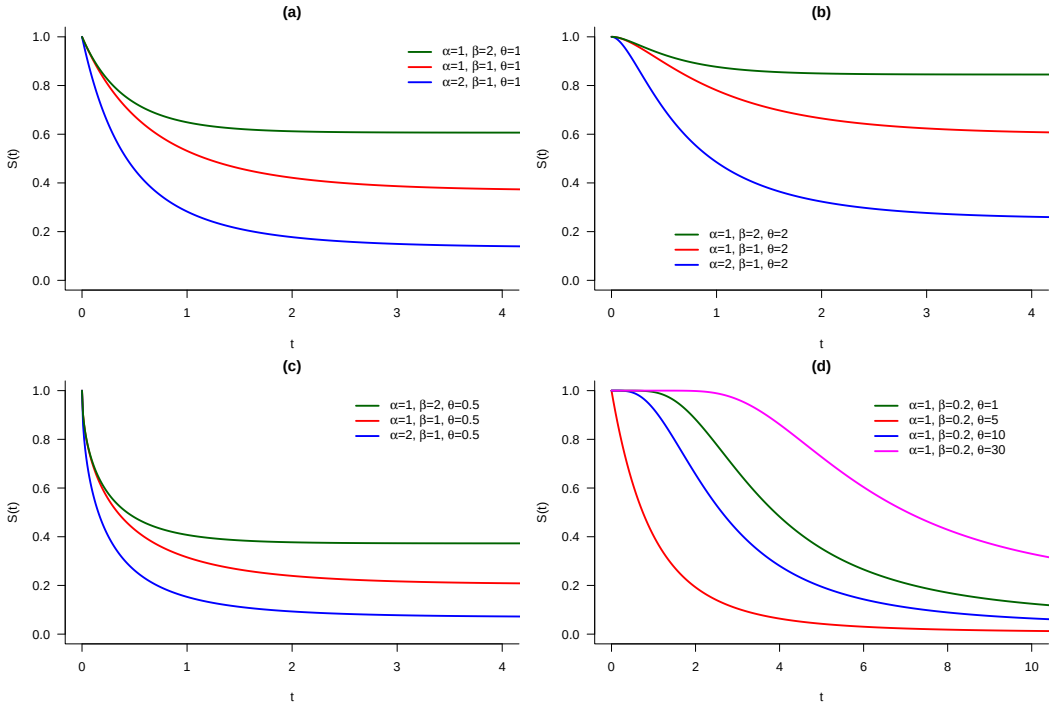

Figure 1: The survival function of the modified generalized Gompertz distribution $MGGD(\alpha, \beta, \theta)$ for some values of α, β and θ

2.2 Maximum likelihood estimation

Let $t_1, t_2, ..., t_n$ be a random sample of size n from a given continuous probability distribution with vector of parameters λ, where each value t_i is greater to 0. The corresponding likelihood function for λ is given by

$$L(\lambda) = \prod_{i=1}^{n} [f(t_i)]^{\delta_i} [S(t_i)]^{1-\delta_i},$$

where δ_i is a censoring indicator variable, that is, $\delta_i = 1$ for an observed survival time and $\delta_i = 0$ for a right-censored survival time. From the expressions (3) and (1), the likelihood function for the model based on the MGGD is thus given by

$$\begin{aligned} L(\lambda) &= (\theta\alpha)^{\sum_{i=1}^{n} \delta_i} \exp\left(-\beta \sum_{i=1}^{n} \delta_i t_i\right) \\ &\quad \times \prod_{i=1}^{n} g(\alpha, \beta, t_i) [1 - g(\alpha, \beta, t_i)]^{\delta_i(\theta-1)} \left\{1 - [1 - g(\alpha, \beta, t_i)]^{\theta}\right\}^{1-\delta_i} \end{aligned} \quad (5)$$

and the corresponding log-likelihood function is

$$\begin{aligned} l(\lambda) &= \ln(\theta\alpha)\sum_{i=1}^{n}\delta_i - \beta\sum_{i=1}^{n}\delta_i t_i + \frac{\alpha}{\beta}\sum_{i=1}^{n}\delta_i\left(e^{-\beta t_i}-1\right) \\ &\quad + (\theta-1)\sum_{i=1}^{n}\delta_i \ln\left[1-g(\alpha,\beta,t_i)\right] + \sum_{i=1}^{n}(1-\delta_i)\ln\left\{1-\left[1-g(\alpha,\beta,t_i)\right]^{\theta}\right\}, \end{aligned}$$

where

$$g(\alpha,\beta,t_i) = \exp\left[\frac{\alpha}{\beta}\left(e^{-\beta t_i}-1\right)\right]$$

and $\lambda = (\alpha,\beta,\theta)$. By deriving $l(\lambda)$ with respect to α, β and θ, we have the following equations:

$$\begin{aligned} \frac{\partial l(\lambda)}{\partial\alpha} &= \frac{1}{\alpha}\sum_{i=1}^{n}\delta_i + \frac{1}{\beta}\sum_{i=1}^{n}\delta_i\left(e^{-\beta t_i}-1\right) \\ &\quad -\frac{(\theta-1)}{\beta}\sum_{i=1}^{n}\frac{\delta_i\left(e^{-\beta t_i}-1\right)g(\alpha,\beta,t_i)}{1-g(\alpha,\beta,t_i)} \\ &\quad +\frac{\theta}{\beta}\sum_{i=1}^{n}(1-\delta_i)\frac{\left(e^{-\beta t_i}-1\right)g(\alpha,\beta,t_i)\left[1-g(\alpha,\beta,t_i)\right]^{\theta}}{\left\{\left[1-g(\alpha,\beta,t_i)\right]^{\theta}-1\right\}\left[g(\alpha,\beta,t_i)-1\right]}, \end{aligned}$$

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

and

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

Setting these expressions equal to zero and solving them simultaneously we get the maximum likelihood estimators (MLE) of the parameters α, β and θ. Although we cannot obtain explicit expressions for the MLEs for these parameters, they can be estimated numerically using iterative algorithms such as the Newton-Raphson method and its variants. Asymptotic variances of MLEs are provided by the diagonal elements of the inverse Fisher information matrix.

The expected Fisher information matrix is given by

$$I(\lambda) = \begin{bmatrix} -E\left[\frac{\partial^2 l(\lambda)}{\partial \alpha^2}\right] & -E\left[\frac{\partial^2 l(\lambda)}{\partial \alpha \beta}\right] & -E\left[\frac{\partial^2 l(\lambda)}{\partial \alpha \theta}\right] \\ -E\left[\frac{\partial^2 l(\lambda)}{\partial \alpha \beta}\right] & -E\left[\frac{\partial^2 l(\lambda)}{\partial \beta^2}\right] & -E\left[\frac{\partial^2 l(\lambda)}{\partial \beta \theta}\right] \\ -E\left[\frac{\partial^2 l(\lambda)}{\partial \alpha \theta}\right] & -E\left[\frac{\partial^2 l(\lambda)}{\partial \beta \theta}\right] & -E\left[\frac{\partial^2 l(\lambda)}{\partial \theta^2}\right] \end{bmatrix},$$

but in practical applications we could use the observed information matrix given the difficulties to get the expected values in $I(\lambda)$. Wald-type confidence intervals and hypotheses tests for the parameters of the model based on the MGGD can be thus obtained from the respective asymptotic estimates of the standard errors.

A Wald-type confidence interval for the proportion η of "immune" or "cured" individuals can be obtained by applying the delta method using the observed information matrix for λ (Oehlert, 1992). From (2), let us consider η as a function of the parameters α, β and θ, or say,

$$w(\lambda) = 1 - \left[1 - \exp\left(-\frac{\alpha}{\beta}\right)\right]^{\theta}.$$

Using a first-order Taylor series approximation, an approximation for the variance of $w(\lambda)$ is given by

$$Var\left[w(\lambda)\right] \approx \begin{bmatrix} \frac{\partial w(\lambda)}{\partial \alpha} & \frac{\partial w(\lambda)}{\partial \beta} & \frac{\partial w(\lambda)}{\partial \theta} \end{bmatrix} \mathbf{\Sigma}(\lambda) \begin{bmatrix} \frac{\partial w(\lambda)}{\partial \alpha} \\ \frac{\partial w(\lambda)}{\partial \beta} \\ \frac{\partial w(\lambda)}{\partial \theta} \end{bmatrix},$$

where

$$\frac{\partial w(\lambda)}{\partial \alpha} = -\frac{\theta}{\beta}\exp\left(-\frac{\alpha}{\beta}\right)\left[1 - \exp\left(-\frac{\alpha}{\beta}\right)\right]^{\theta-1},$$

$$\frac{\partial w(\lambda)}{\partial \beta} = \frac{\alpha\theta}{\beta^2}\exp\left(-\frac{\alpha}{\beta}\right)\left[1 - \exp\left(-\frac{\alpha}{\beta}\right)\right]^{\theta-1},$$

$$\frac{\partial w(\lambda)}{\partial \theta} = -\left[1 - \exp\left(-\frac{\alpha}{\beta}\right)\right]^{\theta}\ln\left[1 - \exp\left(-\frac{\alpha}{\beta}\right)\right]$$

and $\mathbf{\Sigma}(\lambda)$ is the respective maximum-likelihood estimated variance-covariance matrix. Therefore, the Wald-type asymptotic $100(1-\varphi)\%$ confidence limits for η are given by

$$w(\widehat{\lambda}_{ML}) - z_{\varphi/2}\sqrt{Var\left[w(\widehat{\lambda}_{ML})\right]} \quad \text{and} \quad w(\widehat{\lambda}_{ML}) + z_{\varphi/2}\sqrt{Var\left[w(\widehat{\lambda}_{ML})\right]},$$

where $\widehat{\lambda}_{ML} = (\widehat{\alpha}_{ML}, \widehat{\beta}_{ML}, \widehat{\theta}_{ML})$ is the vector of maximum likelihood estimates for α, β and θ,

$$w(\widehat{\lambda}_{ML}) = 1 - \left[1 - \exp\left(-\frac{\widehat{\alpha}_{ML}}{\widehat{\beta}_{ML}}\right)\right]^{\widehat{\theta}_{ML}} \tag{6}$$

is the MLE for η and $z_{\varphi/2}$ is the upper $(\varphi/2)$th percentile of a standard normal distribution.

The maxLik library in R software can be used to find the maximum likelihood estimates (Henningsen and Toomet, 2011), where the likelihood equations were solved by a method based on the Newton-Raphson algorithm. The R code used in the present article is provided in an Appendix at the end of the paper.

2.3 Bayesian analysis

The Bayesian approach considers the unknown model parameters as random variables, where for each parameter it is assigned a suitable prior *pdf* (Gelman et al., 2014). According to the Bayes theorem, we can write the joint posterior density by combining the joint prior distribution with the likelihood function for the parameters α, β and θ. Bayesian Markov chain Monte Carlo (MCMC) is a useful alternative to obtain the posterior summaries of interest of the joint posterior distribution of the parameters, using simulation techniques. Considering that α, β and θ are positive parameters, we can assume inverse gamma (IG) prior distributions for these parameters. In this case, we have $\alpha \sim IG(a_1, b_1)$, $\beta \sim IG(a_2, b_2)$ and $\theta \sim IG(a_3, b_3)$, where a_1, a_2, a_3, b_1, b_2, and b_3 are known positive hyperparameters and $IG(a, b)$ denotes a inverse gamma distribution with mean $b(a-1)^{-1}$ and variance $b^2\left[(a-1)^2(a-2)\right]^{-1}$. Note that an improper prior distribution is assumed if $a < 2$. Supposing prior independence between α, β and θ, the joint prior distribution for these parameters is thus given by

$$\pi(\lambda) \propto \alpha^{-a_1-1}\beta^{-a_2-1}\theta^{-a_3-1}\exp\left(-\frac{b_1}{\alpha}-\frac{b_2}{\beta}-\frac{b_3}{\theta}\right),$$

where $\propto$ denotes "proportional to" and $\lambda = (\alpha, \beta, \theta)$. The joint posterior density is given by the expression $\pi(\lambda|\mathbf{y}) \propto \pi(\lambda) L(\mathbf{y}|\lambda)$, where the likelihood function $L(\mathbf{y}|\lambda)$ is given by (5) and $\mathbf{y}$ denotes the set of observed pairs (t_1, δ_1), (t_2, δ_2), ..., (t_n, δ_n).

In order to get information about the posterior distributions of the parameters of interest, we used the MCMC method with a Gibbs sampling algorithm available in the OpenBUGS software. OpenBUGS is a program for Bayesian analysis of complex statistical models using MCMC simulation methods, only requiring the specification of the likelihood function and the prior distributions for the parameters in the model (Lunn et al., 2000). The model was ran for 1,005,000 iterations with a burn-in phase of 5,000 simulated samples and a thinning interval of size 100. Bayesian estimates of the parameters were obtained as the mean of samples drawn from the respective posterior distributions, and 95% credible intervals ($95\%CrI$) were given by the $0.025th$ and $0.975th$ percentiles of their posterior distributions. Convergence of the MCMC samples was assessed by visual examination of traceplots of the simulated samples. The OpenBugs code used to specify the model is given in an Appendix.

2.4 A simulation method for a random variable with a MGGD

A sample of size n from a MGGD with right-censored data can be randomly generated according to following steps:

Step 1: Fix values of $\alpha > 0$, $\beta > 0$ and $\theta > 0$.

Step 2: Generate n random samples from $M_i \sim Bernoulli(0, 1-\eta)$, where η is given by replacing the values for the parameters α, β and θ chosen in the previous step in the expression (2).

Step 3: For $i = 1, ..., n$, consider $t_i^{'} = \infty$ if $M_i = 0$ and $t_i^{'} = F_Y^{-1}(U_i)$ if $M_i = 1$, where the inverse of the cumulative function (4) of the MGGD is given by

$$F_Y^{-1}(U_i) = -\frac{1}{\beta} \ln \left[1 + \frac{\beta}{\alpha} \ln \left(1 - U_i^{\frac{1}{\theta}} \right) \right]$$

and $U_i \sim Uniform(0, 1-\eta)$.

Step 4: Generate n random samples from $u_i^{'} \sim Uniform(0, \max(t_i^{'}))$, considering only the finite $t_i^{'}$.

Step 5: Let $t_i = \min(t_i^{'}, u_i^{'})$.

Step 6: Pairs of simulated values (t_i, δ_i), $i = 1, ..., n$, are thus obtained, where $\delta_i = 1$ if $t_i < u_i^{'}$ and $\delta_i = 0$ otherwise.

These steps are similar to those described by Rocha et al. (2017) and Santos et al. (2017). An R function based on these steps is available in the Appendix.

2.5 Model with covariates

In order to include covariates in the analysis, the parameter α in the likelihood function (5) can be replaced by a function $\alpha(\mathbf{x}_i)$ such as

$$\ln \alpha(\mathbf{x}_i) = \mathbf{x}_i \alpha^*,$$

where $\mathbf{x}_i = (1, x_{1i}, x_{2i}, ..., x_{pi})$ is a vector containing observations on p independent variables and $\alpha^* = (\alpha_0, \alpha_1, ..., \alpha_p)$ is a vector of unknown parameters. Analogously, the parameter β in $L(\lambda)$ can be replaced by an function $\beta(\mathbf{w}_i)$ such as

$$\ln \beta(\mathbf{w}_i) = \mathbf{w}_i \beta^*,$$

where $\mathbf{w}_i = (1, w_{1i}, w_{2i}, ..., w_{qi})$ is a vector which may or may not be equal to $\mathbf{x}_i$ and $\beta^* = (\beta_0, \beta_1, ..., \beta_q)$ is a vector of unknown parameters.

In the Bayesian analysis, consider the following prior distributions for the parameters:

$$\alpha_j \sim N(c_j, d_j), \qquad j = 0, 1, ..., p,$$

and

$$\beta_k \sim N(e_k, f_k), \qquad k = 0, 1, ..., q,$$

where $N(c, d)$ denotes a normal distribution with mean c and variance d, and c_j, d_j, e_k and f_k ($j = 0, 1, ..., p$ and $k = 0, 1, ..., q$) are known hyperparameters.

For simplicity, we can consider only the effect of the covariates on the parameters α and β. However, in the analysis of more complex data sets, we also can consider that the shape parameter θ in the likelihood function (5) is replaced by an function $\gamma(\mathbf{z}_i)$ such as

$$\ln \gamma(\mathbf{z}_l) = \mathbf{z}_l \gamma^*,$$

where $\mathbf{z}_l = (1, z_{1l}, z_{2l}, ..., z_{rl})$ is a vector containing observations on r independent variables and $\gamma^* = (\gamma_0, \gamma_1, ..., \gamma_r)$ is a vector of unknown parameters.

When using the maximum likelihood method for parameter estimation, comparisons between different model formulations can be based on the Akaike's information criterion (AIC) (Akaike, 1998). The model with the lowest AIC value suggest a better fit to the data. Under the Bayesian framework, the deviance information criterion (DIC) can be used for this same purpose. DIC is analogous to AIC and can be used for model comparison when posterior distributions are approximated via MCMC (Spiegelhalter et al., 2014). Models with lower DIC values are considered to fit the data better.

3 Results

3.1 Simulation study

In order to examine the performance of the maximum likelihood method of estimation, we conducted a simulation study to estimate the coverage probability of the Wald-type confidence intervals for the parameters α, β, θ and η, and the corresponding bias and mean squared errors (MSE). The coverage probability is the observed percent of times the confidence interval includes the respective parameter. The bias in the estimation of a parameter λ is estimated by

$$\widehat{Bias}(\widehat{\lambda}_{ML}) = \frac{1}{B} \sum_{b=1}^{B} \widehat{\lambda}_{ML}^{(b)} - \lambda_N$$

and the corresponding MSE is estimated by

$$\widehat{MSE}(\widehat{\lambda}_{ML}) = \frac{1}{B} \left(\sum_{b=1}^{B} \widehat{\lambda}_{ML}^{(b)} - \lambda_N \right)^2,$$

where $\widehat{\lambda}_{ML} \in (\widehat{\alpha}_{ML}, \widehat{\beta}_{ML}, \widehat{\theta}_{ML}, \widehat{\eta}_{ML})$ is the MLE for a given parameter, $\widehat{\lambda}_{ML}^{(b)}$ is the maximum likelihood estimate obtained for λ considering the b-th simulated sample, λ_N is the corresponding nominal value for λ, $\lambda \in (\alpha, \beta, \theta, \eta)$, and B is the number of simulated samples.

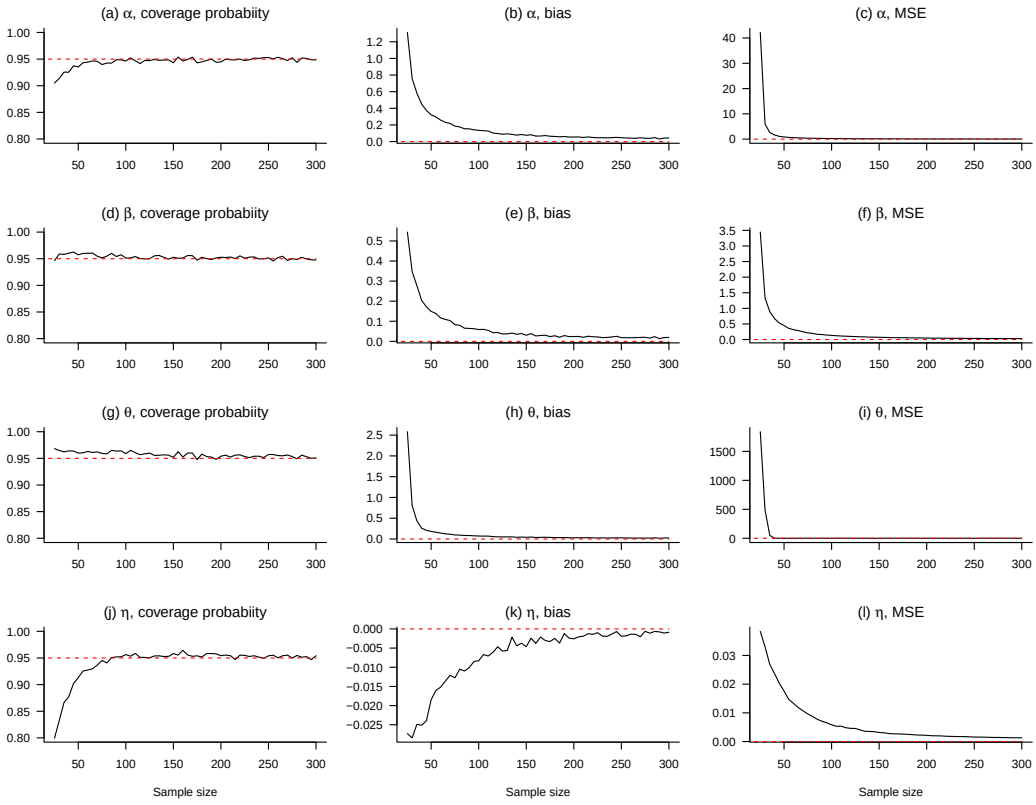

Figure 2: Plots of the coverage probability, biases and MSE of $\widehat{\alpha}_{ML}, \widehat{\beta}_{ML}, \widehat{\theta}_{ML}$ and $\widehat{\eta}_{ML}$ versus n for simulated data from $T \sim MGGD(1, 1, 1)$.

We generated $B = 5,000$ random samples each of size $n = 25, 30, 35, 40, ..., 300$. The random samples were taken to come from (a) $T \sim MGGD(1, 1, 1)$, (b) $T \sim MGGD(2, 1, 2)$ and (c) $T \sim MGGD(0.5, 1, 0.7)$. The nominal confidence level is taken as 95%. We computed the maximum likelihood estimates $\widehat{\alpha}_{ML}, \widehat{\beta}_{ML}, \widehat{\theta}_{ML}$ and $\widehat{\eta}_{ML}$ and the corresponding standard errors for each simulated sample. The proportion η of "immune" individuals was estimated based on the expression (6) and the corresponding standard error was obtained from an application of delta method. These quantities were used to compute the bias, the MSE and the coverage probability for each sample size n. Figures 2 to 4 show the plots of the coverage probability, the biases and the MSE of $\widehat{\alpha}_{ML}, \widehat{\beta}_{ML}, \widehat{\theta}_{ML}$ and $\widehat{\eta}_{ML}$ versus n for simulated data from the MGGD. In these plots, we can observe that the MSE for all parameters generally decrease to zero with increasing n and the respective biases generally approach zero with increasing n. The coverage probabilities for all parameters generally approach the nominal level (95%) with increasing n. The Panels (j) of Figures 2 to 4 show that the coverage probabilities for the confidence intervals for η generally approach the nominal level when $n \geq 100$. The Panels (d) and (g) of these figures suggest that the coverage probabilities for the confidence intervals for β and θ generally approach the nominal level even when the sample size is relatively

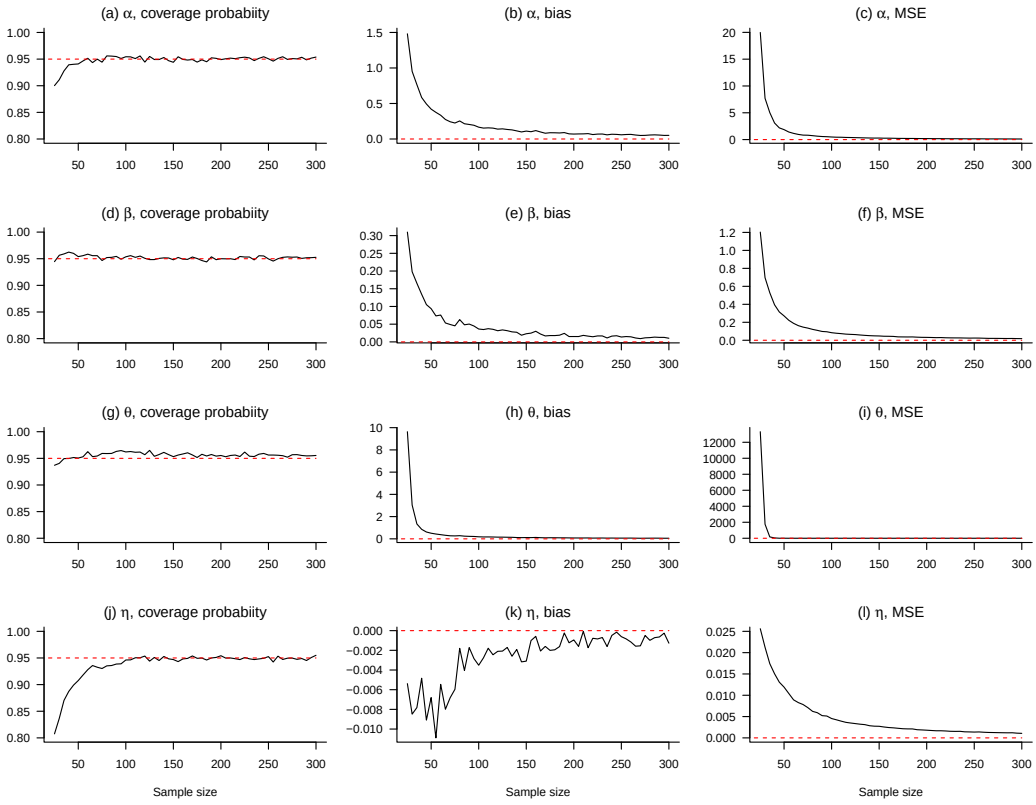

Figure 3: Plots of the coverage probability, biases and MSE of $\widehat{\alpha}_{ML}, \widehat{\beta}_{ML}, \widehat{\theta}_{ML}$ and $\widehat{\eta}_{ML}$ versus n for simulated data from $T \sim MGGD(2, 1, 2)$.

small. For all parameters, the Panels (c), (f), (i) and (l) show that the MSE can assume very high values when $n < 50$. In all simulations, the bias for any parameter was lower than 0.1 even when $n > 170$ and the MSE was lower than 0.1 even when $n > 180$.

3.2 Applications to simulated data

In order to exemplify the application of the proposed model and to compare the results from the frequentist and Bayesian approaches, samples from the MGGD were simulated for different sizes and values of the parameters.

We simulated samples of size $n = 25$, 50 and 100 from the MGGD for (a) $(\alpha, \beta, \theta) = (1, 1, 1)$, (b) $(\alpha, \beta, \theta) = (2, 1, 2)$ and (c) $(\alpha, \beta, \theta) = (0.5, 1, 0.7)$. The nominal values for η were obtained from (2). Table 1 shows that the maximum likelihood and Bayesian estimates for the parameters α, β, θ and η are satisfatory close to each other. In this Bayesian analysis, we assigned an $IG(0.01, 0.01)$ prior distribution for each parameter α, β and θ. From these results, we note that the Wald-type 95% confidence intervals can include negative values, especially when involving relatively small samples. Figure 5 shows the survival functions obtained by the Kaplan-Meier method for each of these simulated data sets and the corresponding parametric curves obtained from the fit of

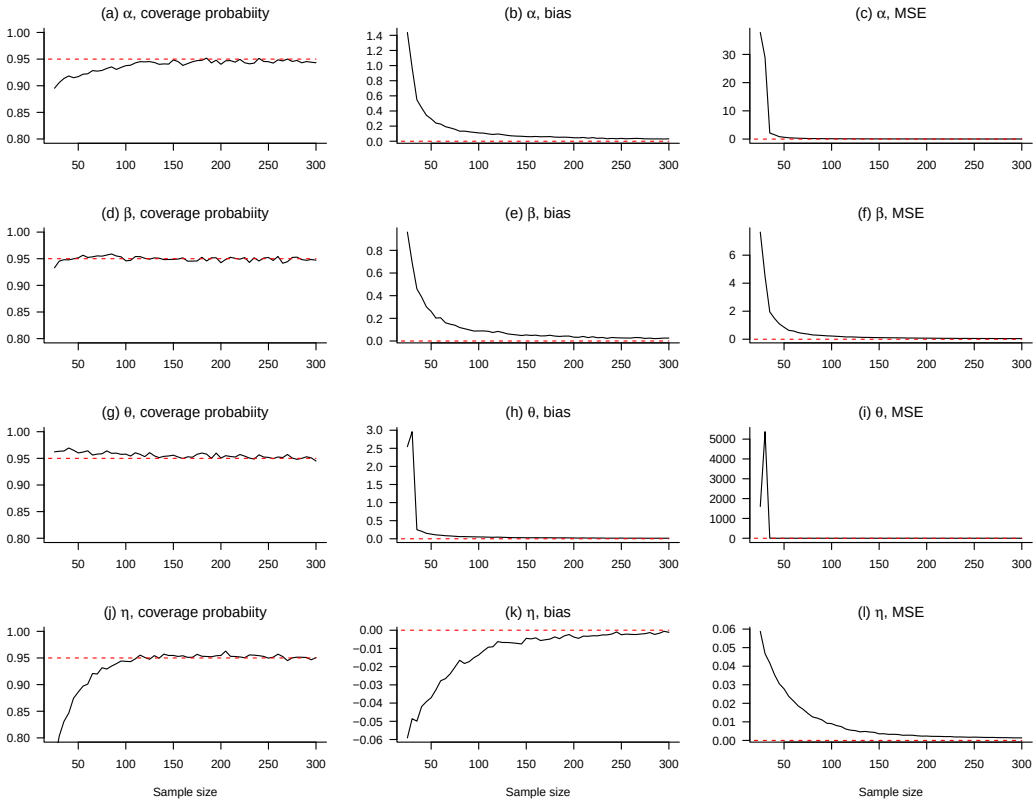

Figure 4: Plots of the coverage probability, biases and MSE of $\widehat{\alpha}_{ML}, \widehat{\beta}_{ML}, \widehat{\theta}_{ML}$ and $\widehat{\eta}_{ML}$ versus n for simulated data from $T \sim MGGD(0.5, 1, 0.7)$.

the model based on the MGGD.

In a brief sensitivity analysis, we also assumed in the Bayesian analysis a non-informative prior distribution for the parameters α, β and θ proportional to $1/(\alpha\beta\theta)$, where $\alpha > 0$, $\beta > 0$ and $\theta > 0$, and a locally uniform prior distribution for the parameters where $v_1 = log(\alpha)$, $v_2 = log(\beta)$ and $v_3 = log(\theta)$, for $-\infty < v_1 < \infty$, $-\infty < v_2 < \infty$ and $-\infty < v_3 < \infty$. However, no significant changes were observed when compared with the results presented in Table 1.

3.3 Application to a real data set

Let us consider the data introduced by Kalbfleisch and Prentice (2011) for a part of a large clinical trial carried out by the Radiation Therapy Oncology Group in the United States. The data set includes 195 patients with carcinoma of the oropharynx who were randomly assigned to one of two treatment groups, radiation therapy alone or radiation therapy together with a chemotherapeutic agent. Let us consider the survival time in years from day of diagnosis. Nearly 27% of the data are censored observations. The dataset is available at https://www.umass.edu/statdata/statdata/data/pharynx.txt.

Let us consider the following covariates:

Table 1: Maximum likelihood and Bayesian estimates for the model based on the MGGD, simulated data.

		Nominal	Maximum likelihood estimates				Bayesian estimates		
n	Parameter	values	Estimate	Std. error	95%CI	AIC	Estimate	95%CrI	DIC
25	α	1	1.6997	1.0711	(-0.3996 , 3.7989)	54.3	1.3260	(0.2441 , 3.4280)	54.3
	β	1	1.0852	0.5109	(0.0838 , 2.0865)		0.8689	(0.1154 , 1.8180)	
	θ	1	1.6558	0.7829	(0.1214 , 3.1902)		1.4090	(0.5344 , 3.0510)	
	η	0.3678	0.3215	0.1119	(0.1022 , 0.5408)		0.2895	(0.0220 , 0.5277)	
50	α	1	1.9031	0.8457	(0.2455 , 3.5607)	50.5	1.6650	(0.3318 , 3.5930)	51.2
	β	1	1.7202	0.5764	(0.5903 , 2.8499)		1.5190	(0.0948 , 2.8080)	
	θ	1	0.8396	0.2085	(0.4309 , 1.2483)		0.7771	(0.4177 , 1.2350)	
	η	0.3678	0.2862	0.0750	(0.1391 , 0.4333)		0.2706	(0.0139 , 0.4362)	
100	α	1	0.7535	0.2533	(0.2569 , 1.2501)	206.4	0.6812	(0.2561 , 1.2690)	207.1
	β	1	0.6852	0.1996	(0.2940 , 1.0764)		0.6129	(0.1680 , 1.0490)	
	θ	1	0.8907	0.1641	(0.5690 , 1.2123)		0.8449	(0.5536 , 1.2080)	
	η	0.3678	0.3028	0.0609	(0.1834 , 0.4221)		0.2821	(0.1141 , 0.4123)	
25	α	2	2.3455	1.7509	(-1.0864 , 5.7774)	44.4	2.1590	(0.5848 , 5.5830)	43.2
	β	1	0.9932	0.7167	(-0.4114 , 2.3979)		0.9013	(0.2535 , 2.0430)	
	θ	2	2.8443	1.9284	(-0.9353 , 6.6239)		2.7640	(0.8931 , 7.4910)	
	η	0.2524	0.2455	0.1694	(-0.0864 , 0.5774)		0.2285	(0.0192 , 0.4989)	
50	α	2	1.9996	0.8875	(0.2601 , 3.7390)	90.3	1.6080	(0.5068 , 3.5100)	88.9
	β	1	1.0919	0.4344	(0.2405 , 1.9434)		0.8506	(0.1088 , 1.7640)	
	θ	2	2.2112	0.8028	(0.6378 , 3.7846)		1.9020	(0.8909 , 3.6650)	
	η	0.2524	0.3203	0.1032	(0.1180 , 0.5225)		0.2571	(0.0031 , 0.4837)	
100	α	2	2.5897	0.7343	(1.1504 , 4.0289)	160.0	2.389	(1.1600 , 3.9940)	160.1
	β	1	1.1789	0.2899	(0.6108 , 1.7471)		1.086	(0.4959 , 1.6930)	
	θ	2	2.6977	0.7175	(1.2913 , 4.1039)		2.538	(1.4240 , 4.1510)	
	η	0.2524	0.2724	0.0659	(0.1431 , 0.4015)		0.2556	(0.1068 , 0.3887)	
25	α	0.5	0.5716	0.4505	(-0.3113 , 1.4545)	59.2	0.5322	(0.1222 , 1.4800)	57.8
	β	1	0.4957	0.4345	(-0.3559 , 1.3473)		0.4394	(0.0033 , 1.1900)	
	θ	0.7	0.9258	0.3743	(0.1921 , 1.6594)		0.8773	(0.4301 , 1.6490)	
	η	0.4794	0.2961	0.1735	(-0.0440 , 0.6362)		0.2491	(0.0028 , 0.5345)	
50	α	0.5	0.3708	0.3046	(-0.2263 , 0.9678)	72.8	0.3516	(0.0720 , 1.0290)	71.1
	β	1	0.8744	0.5812	(-0.2648 , 2.0135)		0.7389	(0.0741 , 1.8740)	
	θ	0.7	0.6477	0.1932	(0.2691 , 1.0264)		0.6207	(0.3641 , 1.0080)	
	η	0.4794	0.4975	0.1057	(0.2902 ,0.7048)		0.4266	(0.0682 , 0.6478)	
100	α	0.5	0.5915	0.2953	(0.0128 , 1.1703)	134.6	0.5206	(0.1530 , 1.1750)	133.5
	β	1	1.0634	0.4239	(0.2324 , 1.8943)		0.8947	(0.1454 , 1.7740)	
	θ	0.7	0.6658	0.1323	(0.4065 , 0.9250)		0.6299	(0.4227 , 0.9048)	
	η	0.4794	0.4328	0.0677	(0.3001 , 0.5656)		0.3919	(0.1208 , 0.5455)	

Note: Std. error, standard error; 95%CrI, 95% credible interval; 95%CI, 95% confidence interval.

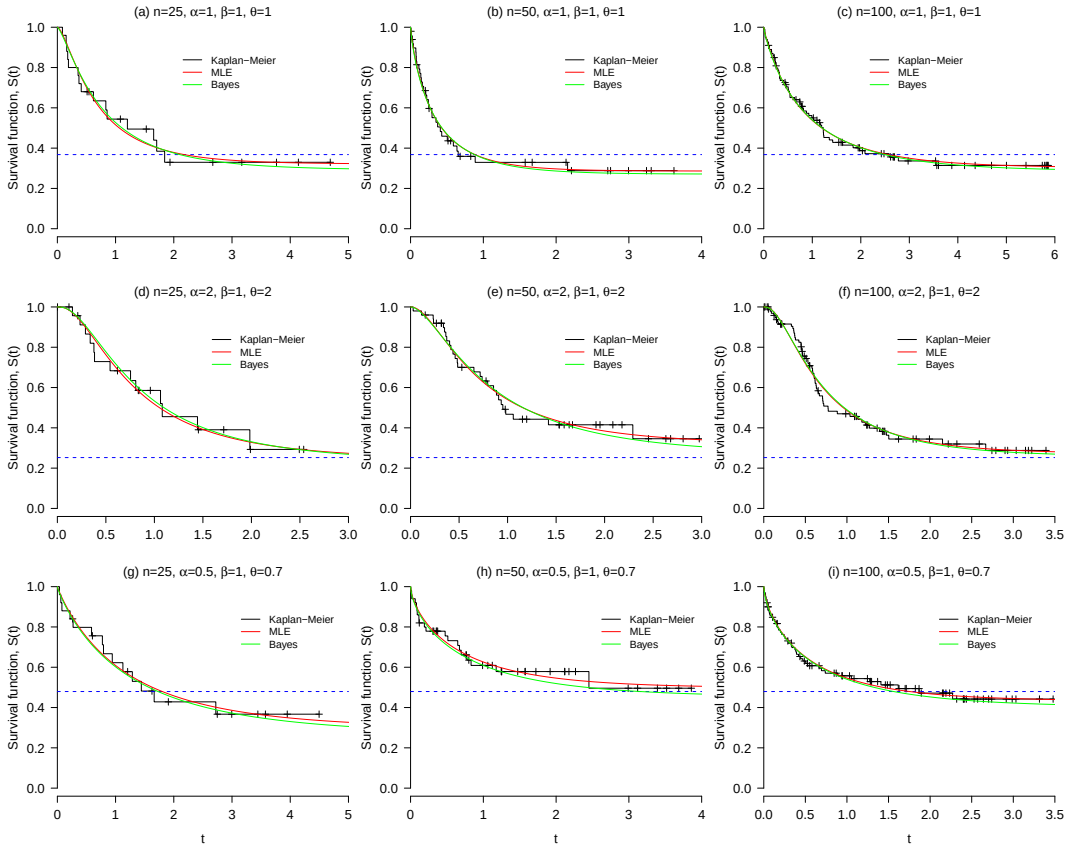

Figure 5: Survival function estimated by the Kaplan-Meier method (black lines) and assuming the model based on the MGGD under the maximum likelihood (red lines) and Bayesian (green lines) approaches. The horizontal dashed lines in each plot correspond to the nominal value for η and vertical ticks are censored observations.

- Treatment (x_1), were $x_1 = 0$ if radiation therapy alone (standard) and $x_1 = 1$ if radiation therapy together with a chemotherapeutic agent (test).
- T stage, represented by two dummy variables (x_2 and x_3), were $x_2 = 0$ and $x_3 = 0$ if primary tumour measuring 2 cm or less in largest diameter (T1), or primary tumour measuring 2 cm to 4 cm in largest diameter with minimal infiltration in depth (T2), $x_2 = 1$ and $x_3 = 0$ if primary tumour measuring more than 4 cm (T3), and $x_2 = 1$ and $x_3 = 1$ if massive invasive tumour (T4).
- Sex (x_4), classified as male ($x_4 = 0$) versus female ($x_4 = 1$).
- Age of the patients at start of follow-up (x_5), classified as less than 60 years ($x_5 = 0$) versus greater or equal to 60 years ($x_5 = 1$).

We firstly fitted models for the carcinoma of the oropharynx data based on the MGGD

Table 2: Maximum likelihood and Bayesian estimates for the models based on the MGGD and MGD, treatment of carcinoma of the oropharynx data.

	Maximum likelihood estimates				Bayesian estimates		
Parameter	Estimate	Std. error	95%CI	AIC	Estimate	95%CrI	DIC
MGGD model				473.9			473.8
α	1.5130	0.2818	(0.9607 , 2.0652)		1.4530	(0.9568 , 2.0250)	
β	0.5648	0.1202	(0.3292 , 0.8003)		0.5363	(0.3050 , 0.7745)	
θ	2.3680	0.4039	(1.5767 , 3.1601)		2.2980	(1.6210 , 3.1590)	
η	0.1550	0.0448	(0.0671 , 0.2429)		0.1463	(0.0566 , 0.2364)	
MGD model				496.6			495.7
α	0.5546	0.0707	(0.4159 , 0.6932)		0.5646	(0.4518 , 0.7045)	
β	0.1360	0.0901	(-0.0405 , 0.3126)		0.1516	(0.0431 , 0.3134)	
η	0.0169	0.0396	(-0.0606 , 0.0946)		0.0326	(8.6×10^{-6} , 0.1312)	

Note: Std. error, standard error; 95%CrI, 95% credible interval; 95%CI, 95% confidence interval.

and MGD without considering the presence of covariates. The corresponding results are showed in Table 2. In the Bayesian fit, it were considered inverse gamma prior distributions for all parameters, or say, $\alpha \sim IG(0.01, 0.01)$, $\beta \sim IG(0.01, 0.01)$ and $\theta \sim IG(0.01, 0.01)$. We found that the Bayesian estimates are quite close to the maximum likelihood estimates. The models based on the MGGD showed lower AIC and DIC values than the models based on the MGD. In addition, the confidence interval for θ do not include the value 1, also suggesting that the model based on the MGGD is found to better fit the data. Considering maximum likelihood approach, the graph in the panel (a) of Figure 6 compares the survival curves $S(t)$ estimated from the Kaplan-Meier method and from the fit of models based on the MGGD and MGD. Panels (b) and (c) of Figure 6 show plots of the Kaplan-Meier estimates versus the corresponding predict values obtained from the models based on the MGGD (Panel (b)) and MGD (panel (c)). Clearly, we observe from these plots that the predict values obtained from the model based on the MGGD are those closest to the empirical values.

Table 3 shows the results from the fit of univariate regression models based on MGGD (Models 1 to 4) taking one covariate at a time, and the results from the fit of a multiple regression model, including all the covariates (Model 5, including treatment, T stage, sex and age). All these models were based on the MGGD. For all regression coefficients, the Bayesian analysis assumed normal prior distributions with mean 0 and relatively large variance. The results in Table 3 show that some coefficient estimates are associated with relatively large standard errors. Since the covariance between the parameter estimates are not very large (results no showed), we consider that these standard errors are affected by the sample size, but no potential identifiability problems are present. Model 2 shows some evidence that patients in T4 stage have a shorter survival time than patients in

Table 3: Maximum likelihood and Bayesian estimates for the coefficients of regression models based on MGGD.

	Maximum likelihood estimates				Bayesian estimates		
Parameter	Estimate	Std. error	95%CI	AIC	Estimate	95%CrI	DIC
Model 1 (treatment)				474.1			474.1
α_0 (intercept)	0.2658	0.2136	(-0.1528 , 0.6845)		0.1147	(-0.4096 , 0.5544)	
α_1 (standard vs. test)	0.3294	0.1690	(-0.0018 , 0.6606)		0.3550	(0.0066 , 0.7194)	
β_0 (intercept)	-0.8229	0.3190	(-1.4481 , -0.1977)		-1.1950	(-3.1160 , -0.4279)	
β_1 (standard vs. test)	0.5071	0.3282	(-0.1361 , 1.1504)		0.7410	(-0.09741 , 2.5300)	
θ	2.4051	0.4163	(1.5891 , 3.2211)		2.2010	(1.5070 , 3.0720)	
Model 2 (T stage)				469.4			468.6
α_0 (intercept)	0.0405	0.2886	(-0.5251 , 0.6062)		0.0183	(-0.4818 , 0.4974)	
α_1 (T3 vs. T1-2)	0.3224	0.2494	(-0.1663 , 0.8112)		0.2141	(-0.2210 , 0.6384)	
α_2 (T4 vs. T1-2)	0.6732	0.2566	(0.1701 , 1.1762)		0.5507	(0.1235 , 0.9800)	
β_0 (intercept)	-0.9528	0.5124	(-1.9571 , 0.0515)		-0.9003	(-1.7500 , -0.2637)	
β_1 (T3 vs. T1-2)	0.3875	0.5184	(-0.6286 , 1.4035)		0.1541	(-0.6653, 0.9909)	
β_2 (T4 vs. T1-2)	0.4620	0.5461	(-0.6082 , 1.5323)		0.1079	(-0.9531 , 1.0060)	
θ	2.4742	0.4328	(1.6258 , 3.3225)		2.2800	(1.6020 , 3.1460)	
Model 3 (sex)				477.4			477.6
α_0 (intercept)	0.4313	0.1905	(0.0579 , 0.8047)		0.3189	(-0.1209 , 0.6966)	
α_1 (male vs. female)	-0.0955	0.2018	(-0.4909 , 0.3001)		-0.1270	(-0.5245 , 0.2752)	
β_0 (intercept)	-0.5831	0.2351	(-1.0438 , -0.1224)		-0.7448	(-1.4840 , -0.2812)	
β_1 (male vs. female)	0.0123	0.3541	(-0.6817 , 0.7064)		-0.0718	(-0.9954 , 0.7005)	
θ	2.3615	0.4032	(1.5712 , 3.1517)		2.1870	(1.5070 , 3.0200)	
Model 4 (age)				477.7			478.2
α_0 (intercept)	0.4506	0.2018	(0.0551 , 0.8461)		0.3218	(-0.1694 , 0.7345)	
α_1 (< 60 vs. ≥ 60)	-0.0728	0.1691	(-0.4040 , 0.2586)		-0.08281	(-0.4441 , 0.2722)	
β_0 (intercept)	-0.5547	0.2628	(-1.0697 , -0.0396)		-0.7518	(-1.6950 , -0.1953)	
β_1 (< 60 vs. ≥ 60)	-0.0386	0.3114	(-0.6488 , 0.5717)		-0.0758	(-1.0120 , 0.8063)	
θ	2.3666	0.4064	(1.5701 , 3.1630)		2.1750	(1.4820 , 3.0260)	
Model 5 (all covariates)				475.6			467.2
α_0 (intercept)	-0.1823	0.3327	(-0.8344 , 0.4699)		-0.1677	(-0.6956 , 0.3397)	
α_1 (std. treatment vs. test)	0.3836	0.1732	(0.0441 , 0.7229)		0.2781	(-0.0524 , 0.6017)	
α_2 (T stage 3 vs. 1-2)	0.4001	0.2742	(-0.1373 , 0.9374)		0.2383	(-0.2017 , 0.6723)	
α_3 (T stage 4 vs. 1-2)	0.7425	0.2691	(0.2149 , 1.2699)		0.5732	(0.1365 , 0.9969)	
α_4 (male vs. female)	-0.0987	0.2178	(-0.5255 , 0.3281)		-0.1528	(-0.5380 , 0.2396)	
α_5 (< 60 vs. ≥ 60)	-0.0290	0.1818	(-0.3854 , 0.3274)		-0.0645	(-0.4056 , 0.2566)	
β_0 (intercept)	-1.4581	0.7129	(-2.8554 , -0.0608)		-1.2670	(-2.2600 , -0.4119)	
β_1 (std. treatment vs. test)	0.6630	0.3877	(-0.0968 , 1.4229)		0.4781	(-0.3441 , 1.2650)	
β_2 (T stage 3 vs. 1-2)	0.5805	0.6780	(-0.7484 , 1.9094)		0.2203	(-0.7056 , 1.1770)	
β_3 (T stage 4 vs. 1-2)	0.6249	0.6541	(-0.6570 , 1.9068)		0.1386	(-0.9882 , 1.0950)	
β_4 (male vs. female)	-0.0353	0.4449	(-0.9074 , 0.8368)		-0.2720	(-1.5820 , 0.6216)	
β_5 (< 60 vs. ≥ 60)	-0.0393	0.3727	(-0.7698 , 0.6912)		-0.1367	(-1.0540 , 0.6741)	
θ	2.4755	0.4434	(1.6065 , 3.3445)		2.1190	(1.4940 , 2.9340)	

Note: Std. error, standard error; 95%CrI, 95% credible interval; 95%CI, 95% confidence interval.

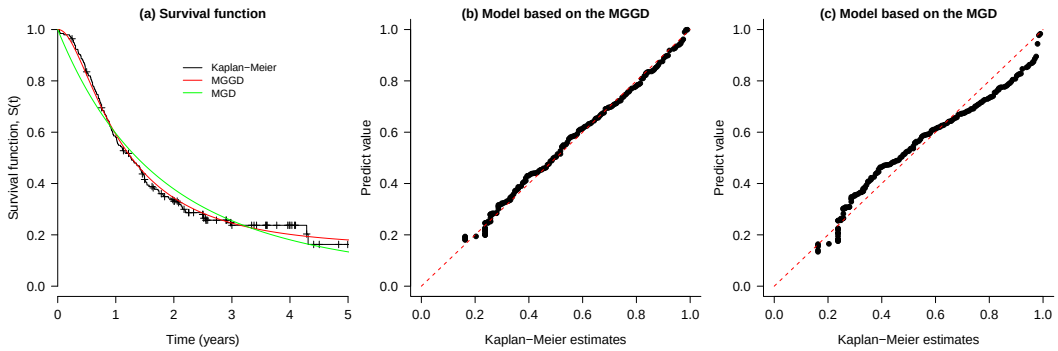

Figure 6: Plots of the survival function estimated from the Kaplan-Meier method and models based on the MGGD and MGD (panel (a)), and plots of the Kaplan-Meier estimates versus the corresponding predict values obtained from the models based on the MGGD (Panel (b)) and MGD (panel (c)). Treatment of carcinoma of the oropharynx data.

T1-T2 stages, since the confidence interval for the parameter α_2 do not include the zero value (see Table 3 and Figure 7). Plots in Figure 7 show survival curves survival curves estimated by means of the Kaplan-Meier method and from models based on MGGD, stratified by each one of the covariates. We can observe that maximum likelihood and Bayesian methods produce similar survival curves, with some differences at the right of the plots.

Models including the effect of the covariates on the shape parameter θ are also fitted to the treatment of carcinoma of the oropharynx data. However, AIC and DIC values indicated no improvement on the fit to the data (results no showed), and this leads us to opt for a more parsimonious model only including the effect of the covariates on the parameters α and β.

4 Conclusion

Based on a modification of the generalized Gompertz distribution introduced by El-Gohary et al. (2013), we proposed in this article a cure fraction model in the presence of right-censored data and covariates. This modified distribution also extends the modified Gompertz distribution introduced by Cantor and Shuster (1992) and Gieser et al. (1998). After the submission of this work for publication, we became aware of a very recent paper from Borges (2017) in which he introduces a model quite similar to ours. However, the present article brings original contributions, such as an extensive simulation study about the performance of the likelihood method and the introduction of Bayesian inference for the parameters. The simulation study showed that both the maximum likelihood and Bayesian approaches are computationally feasible to estimate the parameters of MGGD. Applications with simulated and real data have shown that the MGGD can fits the data very well, under both the maximum likelihood and Bayesian frameworks. The proposed model can be easily implemented in computational programs as R and OpenBUGS, as

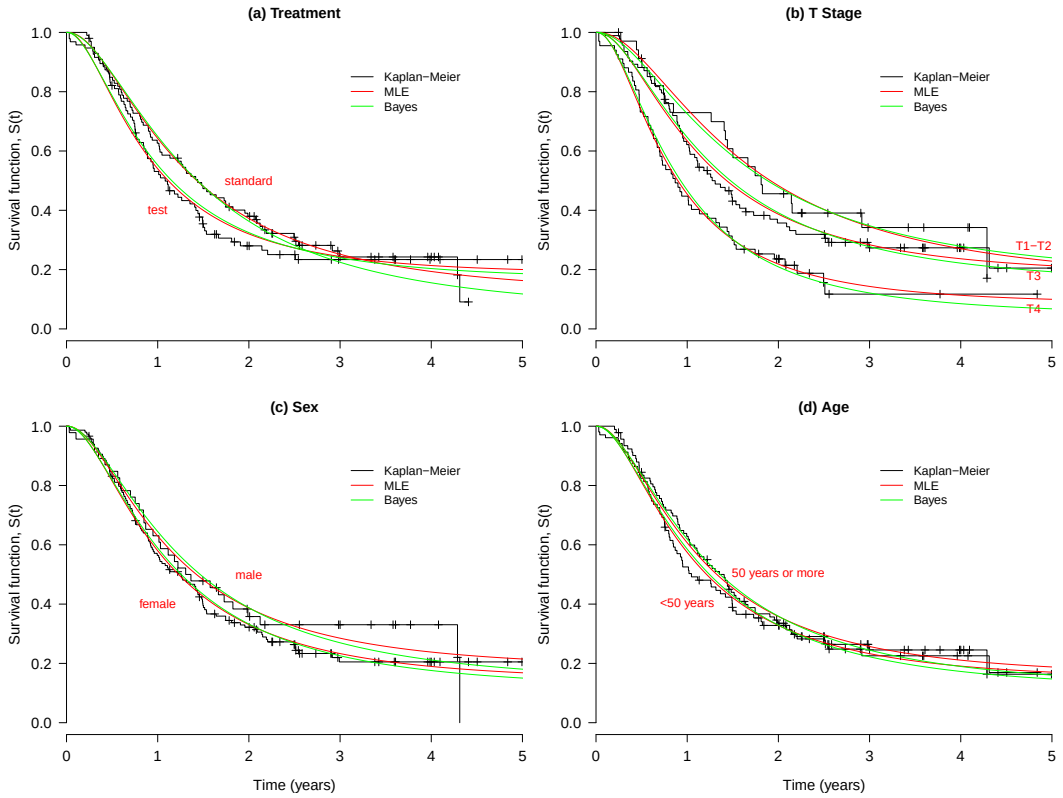

Figure 7: Survival function for each covariate (treatment of carcinoma of the oropharynx data) estimated by the Kaplan-Meier method (black lines) and assuming the model based on the MGGD under the maximum likelihood (red lines) and Bayesian (green lines) approaches. Vertical ticks are censored observations.

showed in the Appendix.

References

Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In *Selected Papers of Hirotugu Akaike*, pages 199–213. Springer.

Balka, J., Desmond, A. F., and McNicholas, P. D. (2011). Bayesian and likelihood inference for cure rates based on defective inverse gaussian regression models. *Journal of Applied Statistics*, 38(1):127–144.

Borges, P. (2017). Em algorithm-based likelihood estimation for a generalized gompertz regression model in presence of survival data with long-term survivors: an application to uterine cervical cancer data. *Journal of Statistical Computation and Simulation*, 87(9):1712–1722.

Cancho, V. G. and Bolfarine, H. (2001). Modeling the presence of immunes by using the exponentiated-weibull model. *Journal of Applied Statistics*, 28(6):659–671.

Cantor, A. B. and Shuster, J. J. (1992). Parametric versus non-parametric methods for estimating cure rates based on censored survival data. *Statistics in Medicine*, 11(7):931–937.

El-Gohary, A., Alshamrani, A., and Al-Otaibi, A. N. (2013). The generalized gompertz distribution. *Applied Mathematical Modelling*, 37(1):13–24.

Farewell, V. T. (1982). The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics*, pages 1041–1046.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2014). *Bayesian data analysis*, volume 2. CRC press Boca Raton, FL.

Ghitany, M., Maller, R. A., and Zhou, S. (1994). Exponential mixture models with long-term survivors and covariates. *Journal of Multivariate Analysis*, 49(2):218–241.

Gieser, P. W., Chang, M. N., Rao, P., Shuster, J. J., and Pullen, J. (1998). Modelling cure rates using the gompertz model with covariate information. *Statistics in Medicine*, 17(8):831–839.

Henningsen, A. and Toomet, O. (2011). maxlik: A package for maximum likelihood estimation in r. *Computational Statistics*, 26(3):443–458.

Kalbfleisch, J. D. and Prentice, R. L. (2011). *The statistical analysis of failure time data*, volume 360. John Wiley & Sons.

Lunn, D. J., Thomas, A., Best, N., and Spiegelhalter, D. (2000). Winbugs-a bayesian modelling framework: concepts, structure, and extensibility. *Statistics and computing*, 10(4):325–337.

Morbiducci, M., Nardi, A., and Rossi, C. (2003). Classification of cured individuals in survival analysis: the mixture approach to the diagnostic–prognostic problem. *Computational Statistics & Data Analysis*, 41(3):515–529.

Oehlert, G. W. (1992). A note on the delta method. *The American Statistician*, 46(1):27–29.

Rocha, R., Nadarajah, S., Tomazella, V., and Louzada, F. (2017a). A new class of defective models based on the marshall–olkin family of distributions for cure rate

modeling. *Computational Statistics & Data Analysis*, 107:48–63.

Rocha, R., Nadarajah, S., Tomazella, V., Louzada, F., and Eudes, A. (2017b). New defective models based on the kumaraswamy family of distributions with application to cancer data sets. *Statistical Methods in Medical Research*, 26(4):1737–1755.

Rocha, R., Tomazella, V. L. D., and Louzada, F. (2014). Bayesian and classic inference for the defective gompertz cure rate model. *Revista Brasileira de Biometria*, 32(1):104–114.

Santos, M. R., Achcar, J. A., and Martinez, E. Z. (2017). Bayesian and maximum likelihood inference for the defective gompertz cure rate model with covariates: an application to the cervical carcinoma study. *Ciência e Natura*, 39(2):244.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Linde, A. (2014). The deviance information criterion: 12 years on. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(3):485–493.

Appendix

The following R function `rMGGompertz` can be used to generate random samples of size n from a MGGD with parameters α, β and θ.

```
rMGGompertz <- function(n,alpha,beta,theta) {
eta <- 1-(1-exp(-alpha/beta))^theta
m <- rbinom(n,prob=1-eta,size=1)
u <- runif(n,0,1-eta)
y0 <- -(1/beta)*log(1+beta/alpha*log(1-u^(1/theta)))
t0<-ifelse(m,y0,Inf)
maxti<-max(y0*m)
w <- runif(n,0,maxti)
t <- pmin(t0,w)
d <- as.numeric(t0<w)
data <- data.frame(t,d)
return (data) }
```

The OpenBUGS code used for the model based on MGGD is given below.

```
model {
 for (i in 1:N) {
   S[i] <- 1- pow(1- exp(alpha/beta*(exp(-beta*t[i])-1)),theta)
   f[i] <- theta*alpha*exp(-beta*t[i])
           *exp(alpha/beta*(exp(-beta*t[i])-1))
           *pow(1- exp(alpha/beta*(exp(-beta*t[i])-1)),theta-1)
   L[i] <- pow(f[i],d[i])*pow(S[i],1-d[i])
   logL[i] <- log(L[i])
   zeros[i] <- 0
   zeros[i] ~ dloglik(logL[i])
 }
 prec <- 0.01
 a  ~ dgamma(prec,prec)
 b  ~ dgamma(prec,prec)
 th ~ dgamma(prec,prec)
 theta <- 1/th
 alpha <- 1/a
 beta  <- 1/b
 eta <- 1- pow(1-exp(-(alpha/beta)),theta)  }
```

In this code `N` is the sample size, `S[i]` is the survival function given in equation (1), `f[i]` is the *pdf* as given in equation (3), `L[i]` is the likelihood function given in equation (5), `t[i]` is the time-to-event variable and `d[i]` is the censoring indicator variable.

Under the frequentist approach, the following R code can be used for the model using the function `maxLik` of the maxLik package (Henningsen and Toomet, 2011) for the maximization of the likelihood function.

```
log.f <- function(parms){
 alpha <- parms[1]
 beta  <- parms[2]
 theta <- parms[3]
 if (parms[1]<0) return(-Inf)
 if (parms[2]<0) return(-Inf)
 if (parms[3]<0) return(-Inf)
 St <- 1-(1-exp(alpha/beta*(exp(-beta*t)-1)))^theta
 ft <- theta*alpha*exp(-beta*t)*exp(alpha/beta*(exp(-beta*t)-1))*
       ((1-exp(alpha/beta*(exp(-beta*t)-1)))^(theta-1))
 like <- ft^d * St^(1-d)
 L <- sum(log(like))
 if (is.na(L)==TRUE) {return(-Inf)} else {return(L)}
 }

 library(maxLik)
 mle  <- maxLik(logLik=log.f,start=c(alpha.0,beta.0,theta.0))
 summary(mle)
 alpha<-mle$estimate[1]
 beta <-mle$estimate[2]
 theta<-mle$estimate[3]
 eta <- 1-(1-exp(-alpha/beta))^theta
 Estimate<-c(alpha,beta,theta,eta)
 s<-vcov(mle)
 llim<-ulim<-rep(NA,4)
 for (j in 1:3) {
  llim[j] <- Estimate[j] - qnorm(0.975) * sqrt(s[j,j])
  ulim[j] <- Estimate[j] + qnorm(0.975) * sqrt(s[j,j]) }
 da<- -theta/beta*exp(-alpha/beta)*(1-exp(-alpha/beta))^(theta-1)
 db<- alpha*theta/(beta*beta)*exp(-alpha/beta)*(1-exp(-alpha/beta))^(theta-1)
 dw<- -(1-exp(-alpha/beta))^theta*log(1-exp(-alpha/beta))
 s4<-t(c(da,db,dw))%*%s%*%c(da,db,dw)
  llim[4] <- eta - qnorm(0.975) * sqrt(s4)
  ulim[4] <- eta + qnorm(0.975) * sqrt(s4)
 StdError<-c(sqrt(s[1,1]),sqrt(s[2,2]),sqrt(s[3,3]),sqrt(s4))
 cbind(Estimate,StdError,llim,ulim)
 print(paste("AIC = ", AIC(mle)))
```

In this code `alpha.0`, `beta.0` and `theta.0` are respectively the initial values for α, β and θ. The standard error for the parameter η is obtained based on the delta method. This code also shows the confidence intervals for each parameter and the AIC value.