Analysis of Survival Data With a Cure Fraction Under Generalized Extreme Value Distribution

Fangfang Liu* Vivekananda Roy†

Abstract

This paper introduces generalized extreme value (GEV) distribution to analyze right-censored survival data for populations with a surviving fraction. Our proposed GEV model leads to extremely flexible hazard functions. We show that our Bayesian model has several nice properties. For example, we prove that even when improper priors are used, the resulting posterior distribution could still be proper under some weak conditions. We further provide theoretical and numerical results showing that our GEV models offer a richer class of models than the widely used Weibull models.

Key Words:

Cure rate model, Surviving fraction, Generalized extreme value distribution, Posterior distribution, Bayesian

1. Introduction

Cure rate model is popularly used for modeling time-to-event survival data where a significant proportion of observation survived. Perhaps the most popular type of cure rate model is the mixture model introduced by Berkson and Gage (1952). For this model, the survival function for the entire population, denoted by $S_1(t)$, is given by $S_1(t) = \pi + (1-\pi)S^*(t)$, where π is the cure rate, $S^*(t)$ is the survivor function for the noncured group in the population. Chen et al. (1999) shows that the standard cure rate model has several drawbacks, e.g., when including covariates through π, we might get improper posterior distributions for many types of noninformative improper priors. And a different kind of cure rate model is proposed in Chen et al. (1999). The proposed model is given by $S_p(t) = \exp(-\theta F(t))$, where $S_p(t)$ denotes the survival function for the population and $\exp(-\theta)$ is the cure rate. This model has a proportional hazards structure through the cure rate parameter, and thus has an appealing interpretation. After introducing latent variables, posterior samples for parameters could be efficiently computationally sampled using Gibbs algorithm. Also, the model yields proper posterior distributions under a wide class of non-informative improper priors for the regression coefficients, including an improper uniform prior.

In the paper of Chen et al. (1999), the proposed model uses Gamma and Weibull distributions for survival time, which are quite popular with monotone hazard rates. But, it is long known that in practice for survival analysis, often the hazard function is not monotone and it is either upside-down shaped or bathtub shaped or a combination of upside-down and bathtub shape. For example, when studying the entire life span of a biological entity, it is quite possible that a three-phase behavior of the failure rate will be observed. A model with a bathtub or "U" shaped failure rate would be appropriate to describe the population's survival capacity. To build an extremely flexible hazard function, in this article, we propose the GEV distribution for log T where T denotes the failure time. We show that by changing the shape parameter of the GEV distribution, we obtain a variety of shape for the hazard function including upside-down and bathtub shape.

In this article, we use Bayesian methodology for making the inference about parameters. An important issue in Bayesian analysis is the specification of a prior distribution.

*Iowa State University, Ames, IA 50010

†Iowa State University, Ames, IA 50010

This is especially true in survival analysis when one wants to assess the importance of certain prognostic factors such as age, gender,etc. It may be difficult to specify informative priors for all possible candidate models, especially if little prior information is available. When such prior elicitation is difficult, or when little prior information is available, one may consider analyses with conventionally chosen priors, such as the improper uniform prior, which reflect little prior information.We provide sufficient conditions under which the posterior distribution is proper when improper uniform prior is used for regression parameters. We assume that covariates are modeled through the cure rate parameter. This prior is analytically and computationally attractive, and facilitates a direct comparison with maximum likelihood.

The article is organized as follows. In Section 2 we provide a short introduction to generalized extreme value distribution for minima. In Section 3 we derive the likelihood function and propose a useful class of prior distributions and derive some of its theoretical properties. We also derive several properties of the resulting posterior distribution. In Section 4, we perform some numerical simulation for our logGEV model. We also compare the logGEV model with the Weibull distribution model. Proofs of the theorems appear in the Appendix.

2. Generalized Extreme Value Distribution

2.1 Introduction of Generalized extreme value distribution

Suppose $Y_1, Y_2, ...$ is a sequence of independent and identically distributed random variables each having the distribution function $F(y)$. Let M_n be the minima of the n random variables $Y_1, Y_2, ..., Y_n$, i.e. $M_n = \min\{Y_1, Y_2, ..., Y_n\}$. If the distribution of Y_i is specified then the exact distributions of M_n are known. Besides, extreme value theory also considers the existence of the limit distribution of M_n. If there exists a non-degenerate distribution function $H_\xi(x)$, and a pair of sequence $a_n,\ b_n$,with $a_n > 0$, such that

$$\lim_{n\to\infty} P\{a_n^{-1}(M_n - b_n) \le x\} = H_\xi(x) \tag{1}$$

holds for all continuity points of $H_\xi(\cdot)$, we say that $H_\xi(x)$ is a generalized extreme value distribution for minima. The possible forms of $H_\xi(x)$ are completely specified as follows:

$$H_\xi(x) = \begin{cases} 1 - \exp[-(1 + \xi\frac{x-\mu}{\sigma})_+^{\frac{1}{\xi}}] & \text{if } \xi > 0 \text{ or } \xi < 0 \\ 1 - \exp(-\exp(\frac{x-\mu}{\sigma})) & \text{if } \xi = 0, \end{cases}$$

where $\mu \in R$, $\sigma \in R^+$, and $\xi \in R$ are the location, scale, and shape parameters, respectively, and $x^+ = \max(x, 0)$.

A more detailed discussion on extreme value distribution for minima could be found in the paper Hirose (2007). Since the GEV distribution for minima is the limiting distribution of minima, in the electrical engineering literature the GEV distribution for minima is often used to describe some systems that have many components that fail when the weakest link fails. Note that there is also generalized extreme value distribution for maxima. Interested readers could find more details about generalized extreme value distribution for maxima (see, e.g., Coles (2001) and Hirose (2007)). In this paper we just consider the generalized extreme value distribution for minima.

2.2 Comparing LogGEV for Minima and Weibull Distribution

Suppose T is a positive random variable, and if we assume a minima GEV distribution for log T, i.e. log$T \sim$ GEV(μ, σ, ξ) , then the corresponding cdf and pdf for T are

$$F(t;\mu,\sigma,\xi) = \Psi_{GEV}\Big(\frac{\log(t)-\mu}{\sigma}\Big),$$

$$f(t;\mu,\sigma,\xi) = \frac{1}{\sigma t}\psi_{GEV}\Big(\frac{\log(t)-\mu}{\sigma}\Big),$$

where Ψ_{GEV} and ψ_{GEV} are pdf and cdf for the standardized minima GEV distribution. When $\mu = 0$ and $\sigma = 1$, the density of the failure time T could be written as:

$$f(t|\xi) = \begin{cases} \frac{1}{t}(1+\xi\log t)^{\frac{1}{\xi}-1}\exp[-(1+\xi\log t)^{\frac{1}{\xi}}] & t > \exp(-\frac{1}{\xi}) \text{ if } \xi > 0 \\ \frac{1}{t}(1+\xi\log t)^{\frac{1}{\xi}-1}\exp[-(1+\xi\log t)^{\frac{1}{\xi}}] & t < \exp(-\frac{1}{\xi}) \text{ if } \xi < 0 \\ \exp(-t) & 0 < t < \infty \text{ if } \xi = 0. \end{cases}$$

Lemma 1. *If $T \sim$ Weibull(α, λ), then log$T \sim$ GEV$(\xi = 0, \mu = \log(\lambda), \sigma = \frac{1}{\alpha})$.*

Proof. The cumulative distribution function of T is

$$F(t|\alpha,\lambda) = 1 - \exp\{-(\frac{t}{\lambda})^{\alpha}\}. \tag{2}$$

Thus we have $P(\log T \le y) = P(T \le \exp(y)) = 1 - \exp\{-\frac{\exp(\alpha y)}{\lambda^{\alpha}}\}$.

Assuming $\mu = \log\lambda, \sigma = \frac{1}{\alpha}$, then $P(\log T \le y) = 1 - \exp\{-\frac{\exp(\alpha y)}{\lambda^{\alpha}}\} = 1 - \exp(-\exp(\frac{x-\mu}{\sigma}))$.

Thus, log$T \sim$ GEV$(\xi = 0, \mu = \log(\lambda), \sigma = \frac{1}{\alpha})$.

□

From Lemma 1, we see that the Weibull distribution is a special case of logGEV distribution.

2.3 Shape of Hazard Function for LogGEV Distribution

If log$T \sim$ GEV(μ, σ, ξ), so the hazard function of T is

$$\lambda(t) = \begin{cases} \frac{1}{\sigma t}(1+\xi\frac{\log t-\mu}{\sigma})^{\frac{1}{\xi}-1} & \text{if } \xi \neq 0 \\ \frac{1}{\sigma t}\exp(\frac{\log t-\mu}{\sigma}) & \text{if } \xi = 0. \end{cases}$$

Next we check the behavior of the hazard function for the logGEV distribution. Figure 1 shows the plot of hazard function for logGEV$(\mu = 0, \sigma = 1, \xi)$ with three different values of $\xi = -0.03, 0, 0.03$. Note that a small change in the value of ξ may lead to a huge change in the hazard functions especially when the time-to-failure is very small. Figure 2 shows the plot of the hazard function for logGEV$(\mu = 0, \sigma = 1.5, \xi)$ with three different values of $\xi = -0.03, 0, 0.03$. When the value of the scale parameter σ is changed from 1 to 1.5, we observe that the shape of the hazard function in Figure 2 is quite different from the plot of hazard functions in Figure 1. This means that the logGEV distribution is flexible enough in the sense that we could obtain a variety of shapes for the hazard functions as we change the parameters of this distribution. So we could consider using logGEV distribution rather than Weibull distribution if a model with a upside down shaped or basetub shaped hazard rate is needed since the hazard function for Weibull distribution is monotone.

3. GEV model for survival data with a surviving fraction

Cure rate models have been used for modeling time-to-event data for various types of cancers. For these diseases, a significant proportion of patients are being "cured". One of the most popular model is called standard cure rate model. In this model, the surviving function for the entire population, denoted by $S_1(t)$, is given by $S_1(t) = \pi + (1 - \pi)S^*(t)$, where a fraction π of the population are considered "cure", and the remaining $1 - \pi$ are not "cured", $S^*(t)$ denotes the surviving function for the non-cured group in the populations. This model is attractive, but it still has several drawbacks. For example, when including covariates through π, we might get improper posterior distributions for many types of noninformative improper priors. This is a serious drawback, since if we want to obtain Bayesian inference proper prior must be required. To overcome the drawbacks of the standard cure rate model, in Chen et al. (1999) a different type of cure rate model is introduced and the specified distribution for survival time is Weibull distribution. In section 2.2, it is proved that if T has Weibull distribution then logT has generalized extreme value distribution with $\xi = 0$. In this paper, we apply the proposed model given in Chen et al. (1999) to the generalized extreme value distribution to incorporate a larger class of models.

For the cancer data, suppose that we have n subjects, and let N_i denote the number of carcinogenic cells for the ith subject. Further, assume that the N_i's are iid Poisson random variables with mean θ_i, $i = 1, \ldots, n$. We emphasize here that the N_i's are not observed and can be viewed as latent variables in the incubation times $z'_{ij}s$ $(j = 1, \ldots, N_i)$ for the N_i carcinogenic cells for the ith subject, which are unobserved, and all have cdf $F(\cdot)$, $i = 1, \ldots, n$. Let t_i denote the time to relapse of cancer for subject i, where t_i is right censored, so $t_i = \min\{z_{ij}, 0 \leq j \leq N_i\}$. Let c_i denote the censoring time, so that we observe $y_i = \min(t_i, c_i)$, where the censoring indicator $\delta_i = I(t_i < c_i)$ equals 1 if t_i is a failure time and 0 if it is right censored. Represent the observed data by the vector $(n, \mathbf{y}, \boldsymbol{\delta})$, where $\mathbf{y} = (y_1, \cdots, y_n)$ and $\boldsymbol{\delta} = (\delta_1, \cdots, \delta_n)$. Also, let $\mathbf{N} = (N_1, \cdots, N_n)$, $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_n)$. The complete data are given by $\mathbf{D} = (n, \mathbf{y}, \boldsymbol{\delta}, \mathbf{N})$, where $\mathbf{N}$ is an unobserved vector of latent variables. In this Section and Section 4, the standard GEV distribution for minima will be used. So we assume the density for y_i is $f(y_i|\xi)$ and $S(y_i|\xi)$ is the corresponding survival function, where ξ is the shape parameter, the form of $f(y_i|\xi)$ is shown in section 2.2. This model could also be applied to other failure time data with surviving fraction, including time to death, time to infection and so on.

The complete-data likelihood function of the parameters $(\xi, \boldsymbol{\theta})$ can be written as

$$
\begin{aligned}
L(\boldsymbol{\theta}, \xi|\mathbf{D}) &= \Big(\prod_{i=1}^{n} S(y_i|\xi)^{N_i - \delta_i}(N_i f(y_i|\xi))^{\delta_i}\Big) \\
&\times \; exp\Big\{\sum_{i=1}^{n}(N_i \log(\theta_i) - \log(N_i!) - \theta_i)\Big\}. \qquad (3)
\end{aligned}
$$

In addition, for each subject i $(i = 1, 2, \ldots, n)$, the survival function is given by $S_i(t) = exp(-\theta_i F(t))$, so the cure fraction is given by $S_i(\infty) = exp(-\theta_i)$. Now we incorporate covariates into the model through $\boldsymbol{\theta}$. For each $i = 1, \cdots, n$, let $\mathbf{x}'_i = (x_{i1}, \cdots, x_{ik})$ denote the $k \times 1$ vector of covariates for the ith subject, and let $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_k)$ denote the corresponding vector of regression coefficients, $\mathbf{D} = (n, \mathbf{y}, \mathbf{X}, \boldsymbol{\delta})$. We relate $\boldsymbol{\theta}$ to the covariates by $\theta_i = \exp(\mathbf{x}'_i\beta)$, so the cure rate for subject i is $\exp(-\exp(\mathbf{x}'_i\beta)),\ i = 1, \cdots, n$.

Thus we could write the complete-data likelihood of $(\boldsymbol{\beta}, \xi)$ as

$$L(\boldsymbol{\beta},\xi|\mathbf{D}) = \Big(\prod_{i=1}^{n} S(y_i|\xi)^{N_i-\delta_i}(N_i f(y_i|\xi))^{\delta_i}\Big) \times \exp\Big\{\sum_{i=1}^{n} N_i\mathbf{x}_i'\boldsymbol{\beta} - \log(N_i!) - \exp(\mathbf{x}_i'\boldsymbol{\beta})\Big\}. \quad (4)$$

Since $\mathbf{N}$ is an unobserved latent vector, by summing out $\mathbf{N}$, the complete data likelihood given in 4 is reduced to

$$\sum_{\mathbf{N}} L(\beta,\xi|\mathbf{D}) = \prod_{i=1}^{n}(\theta_i f(y_i|\xi))^{\delta_i}\exp\left\{-\theta_i(1-S(y_i|\xi))\right\}.$$

Let $\mathbf{D_{obs}} = (n,\mathbf{y},\mathbf{X},\boldsymbol{\delta})$ and assume that the prior distribution for $(\boldsymbol{\beta},\xi)$ is $\pi(\boldsymbol{\beta},\xi)$, then the posterior distribution $\pi(\boldsymbol{\beta},\xi|\mathbf{D}_{obs})$ is given by:

$$\pi(\beta,\xi|\mathbf{D}_{obs}) \propto \sum_{\mathbb{N}} L(\boldsymbol{\beta},\xi|\mathbf{D})\pi(\boldsymbol{\beta},\xi). \quad (5)$$

Now we assume that $\log y_i \sim H_\xi(\mu=0,\sigma=1,\xi)$ and use the following prior on $\boldsymbol{\beta}$ and ξ, $\pi(\boldsymbol{\beta}) \propto 1$ and $\pi(\xi) = \frac{1}{2}I_{[-1,1]}(\xi)$ and assume that $\pi(\boldsymbol{\beta},\xi) = \pi(\boldsymbol{\beta})\cdot\pi(\xi)$. Since we are using improper noninformative prior on $\boldsymbol{\beta}$, (5) is not necessarily a proper density, however, theorem 4 below provides conditions under which (5) is proper.

Theorem 1. *Let $\mathbf{X}^*$ be an $n \times k$ matrix with rows $\delta_i\mathbf{x}_i'$. If the following two conditions hold:*
(a) $\mathbf{X}^*$ is of full rank,
(b) for every i with $\delta_i = 1$, $y_i > \exp(1)$,
then the posterior distribution given in (5) is proper.

The proof of the Theorem 4 is given in Appendix. Note hat the conditions given in Theorem 1 is not sufficient but not necessary for the propriety of posterior distributions. However the conditons stated in the theorem above is quite general for the real data set.

4. Numerical Simulation

To check the propriety of using logGEV distribution for minima for the survival time in the model described in Section 3, we perform a numerical simulation of the logGEV model. The simulation process is shown below:

1. We take the sample size $n = 100$ and consider a single covariate x. We choose x by randomly obtaining a sample from $1, 2, \ldots, 100$ with replacement, and then standardize x that we have simulated to stabilize the posterior computations. Note that intercept is also included in the analysis. Thus we get the $n \times 2$ covariate matrix X with the first column as $1's$ and the second column is the standardized x. We take $\boldsymbol{\beta} = (0, 0.5)'$. Denote the ith row of X is $(1, x_i)$, so $\theta_i = \exp(\beta_0 + \beta_1 x_i^T),\ i = 1, ..., n.$

2. For every i, $i = 1, ..., n$, get a sample from Poisson(θ_i), denote it as N_i, then get a sample of size N_i from Weibull($\alpha = 1.03, \lambda = 1$) , we denote these observations as $Z_{i1}, \ldots, Z_{i,N_i}$ and set $t_i = \min(Z_{i1}, \ldots, Z_{i,N_i})$. If $N_i = 0$, then we set $t_i = \infty$.

3. For every i, $i = 1, ..., n$, let c_i denote the censoring time, we take $y_i = \min(t_i, c_i)$ and indicator $\delta_i = I(t_i < c_i)$ equals 1 if y_i is a failure time and 0 if it is right censored. Here, we take $c_i = 5$ to let the censoring percentage close to 30% .

4. Next we estimate the parameters using $\log T \sim \text{GEV}(\mu = 0, \sigma = 1, \xi)$ distribution.

From Figure 3, we observe that when fitting the simulated data set of Weibull distribution with the proposed model of $\log GEV$ distribution, the survival function estimates are quite similar to the Kaplan-Meier estimates of the survival function. This indicates that the logGEV model might be able to provide a quite good fit to the simulated data set. From table 1, note that both the maximum likelihood estimates and Bayesian estimates for β_0 and β_1 are very close to the true values of β_0 and β_1 (2 and 0.6 respectively).

Table 1: Maximum Likelihood Estimates and Bayesian estimates of the Model Parameters: $\beta_0,\ \beta_1,\ \xi$.

Parameter	MLE	Bayesian
β_0	0.0848 (0.1270)	0.0614 (0.1327)
β_1	0.5077 (0.1253)	0.5077 (0.1280)
ξ	-0.004 (0.0576)	-0.0336 (0.0631)

5. Discussion

In this paper we establish sufficient conditions for the propriety of the posterior distribution when improper uniform prior is used for the regression coefficients through cure rates. So far we only prove the case when the standard extreme value distribution is used. This result could be extended to the proposed model for the general form of generalized extreme value distribution.

Appendix:Proofs of Theorem 1

Proof. By summing out the unobserved latent vector **N**, the complete-data likelihood given in (5) reduces to

$$\sum_{\mathbf{N}} L(\beta, \xi|\mathbf{D}) = \prod_{i=1}^{n} (\theta_i f(y_i|\xi))^{\delta_i} \exp\left\{-\theta_i(1 - S(y_i|\xi))\right\}.$$

When $\delta_i = 0$, $(\theta_i f(y_i|\xi))^{\delta_i} \exp\{-\theta_i(1 - S(y_i|\xi))\} = \exp\{-\theta_i(1 - S(y_i|\xi))\} \le 1$.

When $\delta_i = 1$, we will show that there exists a constant M such that

$$(\theta_i f(y_i|\xi))^{\delta_i} \exp\{-\theta_i(1 - S(y_i|\xi))\} \le g_i(\xi)M, \tag{6}$$

where $g_i(\xi) = \frac{1}{1+\xi \log y_i}$.

The left side of (6) can be rewritten as

$$\begin{aligned} &\tfrac{f(y_i|\xi)}{1-S(y_i|\xi)} \cdot (\theta_i(1 - S(y_i|\xi)) \exp\{-\theta_i(1 - S(y_i|\xi))\} \\ =\ &\frac{\frac{1}{y_i}(1+\xi \log y_i)^{\frac{1}{\xi}-1} \exp[-(1+\xi \log y_i)^{\frac{1}{\xi}}]}{1-\exp[-(1+\xi \log y_i)^{\frac{1}{\xi}}]} (\theta_i(1 - S(y_i|\xi)) \exp\{-\theta_i(1 - S(y_i|\xi))\}, \end{aligned} \tag{7}$$

Let $\phi_1(z) = ze^{-z}$ $\phi_2(z) = \frac{ze^{-z}}{1-e^{-z}}$ $for\ z > 0$, then it can be shown that there exists a constant $g_0 > 0$ such that

$$\phi_i(z) \le g_0, \forall z > 0\ i = 1, 2. \tag{8}$$

Using (8), (7) is less than or equal to $y_i^{-1} g_0^2 g_i(\xi)$. Thus taking $M = g_0^2 \max_{i:\delta_i=1}\{y_i^{-1}\}$, we obtain (6).

Since $\mathbf{X}^*$ is of full rank, there must exist k linearly independent row vectors $\mathbf{x}'_{i_1}, \mathbf{x}'_{i_2}, \ldots, \mathbf{x}'_{i_k}$ such that $\delta_{i_1} = \delta_{i_2} = \ldots = \delta_{i_k} = 1$.

To prove the posterior given in (5) is proper, we only need to show that

$$\int_{-1}^{1} \int_{\mathbb{R}^k} \sum_{\mathbf{N}} L(\beta, \xi | D) \pi(\xi) d\beta d\xi < \infty \tag{9}$$

$$\begin{aligned}
\text{left side} &= \frac{1}{2} \int_{-1}^{1} \int_{\mathbb{R}^k} \sum_{N} L(\beta, \xi | D) d\beta d\xi \\
&= \frac{1}{2} \int_{-1}^{1} \int_{\mathbb{R}^k} \prod_{i=1}^{n} (\theta_i f(y_i|\gamma))^{\delta_i} \exp\left\{-\theta_i (1 - S(y_i|\gamma))\right\} d\beta d\xi \\
&\leq \frac{1}{2} \int_{-1}^{1} \int_{\mathbb{R}^k} (\prod_{i:\delta_i=0}^{n-d} 1) \\
&\quad \{\prod_{j=1}^{k} f(y_{i_j}|\xi) \theta_{i_j} \exp\{-\theta_{i_j}(1 - S(y_{i_j}|\gamma))\}\} \\
&\quad \{\prod_{i:\delta_i=1, i\neq i_j, j=1,\ldots,k} g_i(\xi) M\} d\beta d\xi \\
&= \frac{1}{2} \int_{-1}^{1} \int_{\mathbb{R}^k} M^{d-k} \prod_{i:\delta_i=1, i\neq i_j, j=1,\ldots,k} g_i(\xi) \\
&\quad \prod_{j=1}^{k} f(y_{i_j}|\xi) \theta_{i_j} \exp\{-\theta_{i_j}(1 - S(y_{i_j}|\gamma))\} d\beta d\xi,
\end{aligned} \tag{10}$$

where $\theta_{i_j} = \exp(\mathbf{x}'_{i_j}\beta)$, $d = \sum_{i=1}^{n} \delta_i$ and $\mathbb{R}^k$ denotes k-dimensional Euclidean space.

Now we make the transformation $u_j = x'_{i_j}\beta$ for $j = 1, 2, \ldots, k$. This is a one-to-one linear transformation from β to $\mathbf{u} = (u_1, \ldots, u_k)'$.

Thus (10) is proportional to

$$\begin{aligned}
&\int_{-1}^{1} \int_{\mathbb{R}^k} \prod_{i:\delta_i=1, i\neq i_j, j=1,\ldots,k} g_i(\xi) \prod_{j=1}^{k} f(y_{i_j}|\xi) \exp\{u_j - (1 - S(y_{i_j}|\xi)) \exp(u_j)\} d\mathbf{u} d\xi \\
= &\int_{-1}^{1} \prod_{i:\delta_i=1, i\neq i_j, j=1,\ldots,k} g_i(\xi) \left\{ \prod_{j=1}^{k} f(y_{i_j}|\xi) \int_{R} \exp\{u_j - (1 - S(y_{i_j}|\xi)) \exp(u_j)\} du_j \right\} d\xi \\
= &\int_{-1}^{1} \prod_{i:\delta_i=1, i\neq i_j, j=1,\ldots,k} g_i(\xi) \prod_{j=1}^{k} \frac{f(y_{i_j}|\xi)}{1 - S(y_{i_j}|\xi)} d\xi,
\end{aligned} \tag{11}$$

The last equality holds, because we integrate out $\mathbf{u}$,$\int_R \exp\{u_j - (1 - S(y_{i_j}|\xi)) \exp(u_j)\} du_j = \frac{1}{1-S(y_{i_j}|\xi)}$.

Since $\frac{f(y_{i_j}|\xi)}{1-S(y_{i_j}|\xi)} = \frac{\frac{1}{y_{i_j}}(1+\xi \log y_{i_j})^{\frac{1}{\xi}-1} \exp[-(1+\xi \log y_{i_j})^{\frac{1}{\xi}}]}{1-\exp[-(1+\xi \log y_{i_j})^{\frac{1}{\xi}}]} \leq M_1 \cdot g_{i_j}(\xi)$,where $M_1 = g_0 \cdot \max_{i:\delta_i=1} y_i^{-1}$, ignoring the constant, (11) is less than or equal to$\int_{-1}^{1} \prod_{i:\delta_i=1} \frac{1}{1+\xi \log y_i} d\xi$.

Since $y_i > \exp(1)$,$(1+\xi\log y_i)^{\frac{1}{\xi}+1}$could not be 0 when $\xi \in [-1,1]$,$g(\xi) = \prod_{i:\delta_i=1} \frac{1}{(1+\xi\log y_i)^{\frac{1}{\xi}+1}}$ is bounded in $[-1,1]$. Thus $\int_{-1}^{1} \prod_{i:\delta_i=1} \frac{1}{1+\xi\log y_i} d\xi < \infty$

$\square$

References

Berkson, J. and Gage, R. P. (1952). Survival curve for cancer patients following treatment. *Journal of the ASA*, 47:501–515.

Chen, M.-H., brahim, J. G. I., and Sinha, D. (1999). A new bayesian model for survival data with a surviving fraction. *Journal of the ASA*, 94:909–919.

Coles, S. G. (2001). *An introduction to statistical modeling of extreme values*.

Hirose, H. (2007). Maximum likelihood estimation in the 3-parameter weibull distribution: a look through the generalized extreme-value distribution. *IEEE Trans.*, 14:257–260.

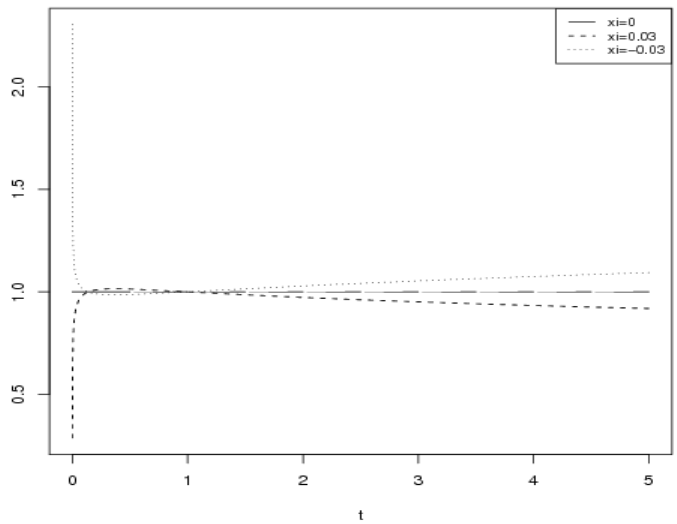

Figure 1: Plot of hazard functions corresponding to minima GEV distribution with $\mu = 0, \sigma = 1$ and different values of ξ.

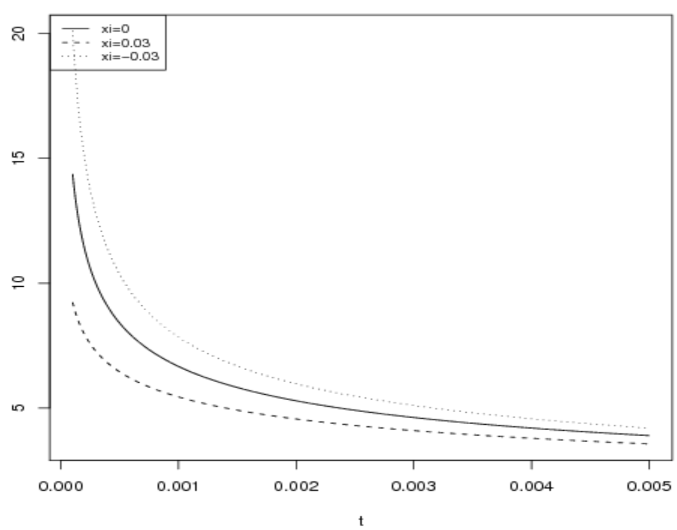

Figure 2: Plot of hazard functions corresponding to minima GEV distribution with $\mu = 0, \sigma = 1.5$ and different values of ξ.

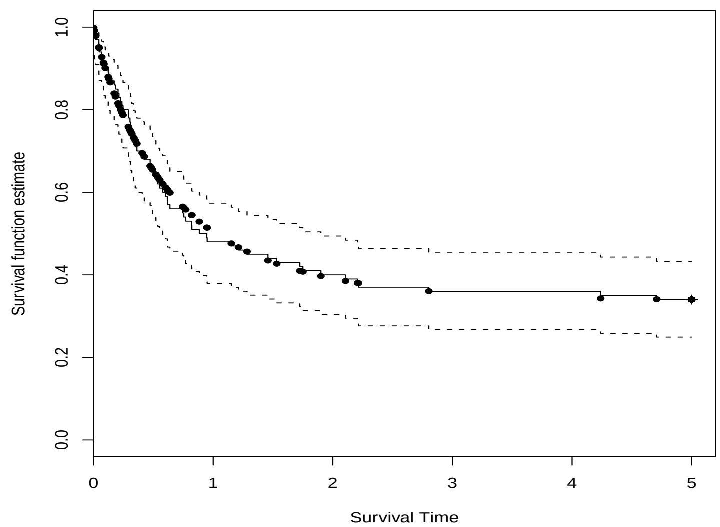

Figure 3: Estimated survival Curves for the simulated model Weibull(α=1.03,λ=1) by Kaplan-Meier method(solid line is the estimate, dashed lines are 95% confidence band for the survival function) and the fitting model $\log \text{GEV}(\mu = 0, \sigma = 1, \xi)$ (the dotted line).