Analysis of Bivariate Survival Data based on copulas with log Generalized Extreme Value marginals

Dooti Roy[*], Vivekananda Roy[**] and Dipak Kumar Dey[*]

[*]University of Connecticut, Storrs, CT, USA

[**]Iowa State University, Ames, Iowa, USA

April 1, 2015

Abstract

This chapter introduces a novel copula based methodology to analyze right censored bivariate survival data using flexible log-GEV marginals. Copulas are quite popular in high-dimensional data modeling as they allow modeling and estimating the parameters of the marginal distributions separately from the dependence parameter of the copula. The Clayton copula structure has been used to represent the dependency between the survival times for this chapter. We propose an empirical Bayes method for estimating the dependency parameter which is more efficient than the other existing Bayesian methodologies.

Keywords: Bivaraite survival, Clayton copula, DRS, Empirical Bayes, GEV, Importance Sampling, Metropolis Hastings

1 Introduction

During the last two decades, there have been a growing interest in modeling paired data or bivariate survival data which arises primarily from the field of medicine. It has been medically observed that incidence of one disease often increases the risk of another in a patient affected by Human Immunodeficiency Virus. Diabetic retinopathy is the leading cause of blindness in United States. It has been observed that once one eye gets affected, the chances of the other eye contracting the disease

increases many fold. Clayton (1978) and Oakes (1982) were the first to develop a fully parametric approach to model bivariate survival data. In Huster et al. (1989), the authors extended the fully parametric approach to include censored bivariate paired data with covariates. Since then between year 1985 to year 1999, several methods both parametric and semiparametric, have been introduced and developed by authors such as: Nelson (1986), Oakes (1986), Shih and Louis (1995), Genest and Rivest (1993) and Wang and Wells (2000) among others. Employing copula to model dependency between marginal survival data have been particularly popular since the univariate marginals do not depend on the choice of the dependency structure and can be consequently estimated separately from the dependency parameters.

Bayesian methods to model bivariate survival data started developing since early 20's. In Sahu and Dey (2000), the authors used a Bayesian frailty model to model bivariate survival data with co-variates. Chen et al. (2002) introduced a Bayesian model for right censored bivariate survival data with cure fraction. Another interesting work is by Shemyakin and Youn (2000), where the authors develop a Bayesian joint survival model for life insurance data. In Romeo and Tanaka (2006), the authors used a copula based Bayesian framework to model bivaraite survival data. Since choosing an appropriate prior for the dependency parameter of any copula is not intuitive, the authors put a β-prior on the Kendall's τ instead. Although these methods are fairly attractive since they are mostly identifiable and produces smooth survival functions, they are not time efficient. Running MCMC algorithm can be expensive especially when for each individual value of selected τ, MCMC chains for both the univariate marginal parameters have to be run. In this paper, we propose a novel em-pirical Bayes method of estimating the dependency parameter. Unlike previous works, our method requires only two runs of MCMC chain for the marginal parameters. Our goal is to develop a Bayesian framework for inference and estimation of the marginals and the dependence parameter following the approach outlined in Doss (2010) and Roy (2014) using Clayton copula structure and the flexible Generalized Extreme Value (GEV) distribution to model the marginal distributions. The chapter is organized as follows: Section 1 provides a brief introduction to the problem and current methods are outlined including their roadblocks and restrictions. In section 2, the concept of copula and different types of copulas have been discussed. In section 3, the model settings are described. Our proposed method is outlined in details in section 4. Section 5 provides background and results

of simulation studies. A real data analysis is provided in the section 6. The chapter concludes with future work directions and limitations.

2 A Brief Overview of Copula

Copula "couples" a joint cumulative density function to its marginal distributions and hence the name. Essentially a copula is a bivariate distribution with uniform marginals. The Sklar's theorem (Nelson (1999), pg:41) defines a copula structure in the following way:

Given a m-dimensional joint cdf F with marginal cdfs $F_1, F_2, \ldots, F_m$, there exists an m-copula C such that,

$$F(x_1, x_2, \ldots, x_m) = C(F_1(x_1), F_2(x_2), \ldots, F_m(x_m)) \tag{1}$$

for all $(x_1, x_2, \ldots, x_m) \in \mathbb{R}^m$. If the marginal distributions are continuous, then C is unique, otherwise C is uniquely determined on Range $F_1 \times \cdots \times$ Range F_m. Conversely, if C is an m-copula and $F_1, F_2, \ldots, F_m$ are cdfs, then F defined by (1) is an m -dimensional cdf with marginal cdfs $F_1, F_2, \ldots, F_m$.

Nelson in his book "Introduction to Copulas (1999)" defined the properties of a m-dimensional copula on page 40:

1. For every $u = (u_1, u_2, \ldots, u_m) \in [0,1]^m$, $C(u) = 0$ if at least one element $u_i = 0$.

2. If all coordinates of u are 1, except u_i, then $C(u) = u_i$.

3. For every $a = (a_1, a_2, \ldots, a_m), b = (b_1, b_2, \ldots, b_m) \in [0,1]^m$, such that $a_i \leq b_i$ for all $i = 1, 2, \ldots, m$,

$$\Delta_{a_m}^{b_m} \Delta_{a_{m-1}}^{b_{m-1}} \cdots \Delta_{a_1}^{b_1} C(u) \geq 0, \tag{2}$$

where $\Delta_{a_k}^{b_k}$ defines the first order differences as

$$\Delta_{a_k}^{b_k} C(u) = C(u_1, \ldots, u_{k-1}, b_k, u_{k+1}, \ldots, u_m) - C(u_1, \ldots, u_{k-1}, a_k, u_{k+1}, \ldots, u_m).$$

The first and third property are satisfied if $C(u)$ is a distribution function. The second property is satisfied if the marginal distributions of C are uniform. Using copulas to model dependent bivariate data is advantageous due to many reasons. Copula structure being extremely flexible, allows non-linear dependence between the two associated marginal distributions, or to measure dependence for

heavy tailed distributions. It can be used under fully parametric, semi parametric or non parametric setting and also allows for faster computations. Several families of copula have been studied in details, for example: Gaussian copulas, Clayton copulas, Frank copula, Gumbel copula etc. Copula's are chosen depending on the tail concentrations. In this chapter, Clayton copula is chosen to develop the framework and model the Diabetes Retinopathy Study data, since the copula has a nice relationship with the Kendall's τÂ· Further Romeo and Tanaka (2006) have shown that Clayton copula performs reasonably well on the same real data.

3 Marginal Distributions and Bivariate Survival Model

3.1 Generalized Extreme Value Distribution as Marginal

Roy et al. (2013) introduced introduced modeling univariate right censored survival data with a cure fraction using Generalized Extreme Value (GEV) distribution (See also Roy and Dey (2014)). The GEV distribution characterized by three parameters is extremely flexible and also achieves proper posterior distribution with a variety of mild to diffuse priors. Commonly used lifetime distributions such as the Exponential, the Rayleigh and the Weibull are all special cases of the minima Generalized Extreme Value distribution.

If we assume a GEV distribution for $\log T$, where T is a marginal survival time, i.e., $\log T \sim GEV(\mu, \sigma, \xi)$, then the corresponding density and survival functions can be defined respectively as:

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

$$S(t|\xi) = \begin{cases} \exp[-(1 + \xi\frac{(\log t - \mu)}{\sigma})_+^{\frac{1}{\xi}}] & \text{if } \xi \neq 0 \\ \exp[-\exp(\frac{\log t - \mu}{\sigma})] & \text{if } \xi = 0, \end{cases}$$

where $\mu \in R$, $\sigma \in R^+$ and $\xi \in R$ are the location, scale and shape parameters respectively. The above distribution was called mGEV in Roy et al. (2013).

3.2 Bivariate Model

Let (T_1, T_2) denote failure times of two events for each subject or failure times of members of each group. Marginally we assume that

$$\log T_i \sim \text{GEV}(\mu_i, \sigma_i, \xi_i) \tag{3}$$

for $i = 1, 2$. The joint survival function based on a copula $C_\phi, \phi \in G$ is given by

$$S(\log t_1, \log t_2) = C_\phi(S_1(\log t_1), S_2(\log t_2)), \tag{4}$$

where $S_1(\log t_1)$, and $S_2(\log t_2)$ are the marginal survival functions obtained from (3). The parameter ϕ measures the "intensity" of dependence between the individual failure times. We use the popular bivariate copula from the Clayton family which is given by

$$C_\phi(u_1, u_2) = \max\{(u_1^{-\phi} + u_2^{-\phi} - 1)^{-1/\phi}, 0\}. \tag{5}$$

In this case $G = (-1, \infty) \setminus \{0\}$. One of the major goals while studying bivariate data is to gauge the direction and extent of the association between the two variables. When linear correlation coefficient is not a valid measure of association due to nonlinearity of the association, Kendal's τ and Spearman's ρ are two most popular measures of pairwise concordance (Nelson, 1999, chapter 5). Certain families of copula have very nice relationship with Kendal's τ. For example, in case of Clayton copula family, the Kendal's τ measure is given by

$$\tau_\phi = \frac{\phi}{\phi + 2} \in (-1, 1).$$

Due to the easy computation and interpretiblity, Clayton copula remains one of the widely used copula families. For proof of the above relationship, see Nelson (1999), page 162–163.

3.3 Model Settings

Let T_{ij} (C_{ij}) be the survival (censoring) time of the jth event for the ith subject, $j = 1, 2; i = 1, 2, \ldots, n$. We assume (T_{i1}, T_{i2}) and (C_{i1}, C_{i2}) are independent for $i = 1, 2, \ldots, n$. We also assume that (T_{i1}, T_{i2}), are iid with common pdf $f(t_1, t_2)$ and survival time $S(t_1, t_2)$, $i = 1, 2, \ldots, n$. The observed data is $(\boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{\delta}_1, \boldsymbol{\delta}_2)$ where $\boldsymbol{y}_1 = (y_{11}, y_{21}, \ldots, y_{n1}), \boldsymbol{y}_2 = (y_{12}, y_{22}, \ldots, y_{n2}), \boldsymbol{\delta}_1 = (\delta_{11}, \delta_{21}, \ldots, \delta_{n1})$, and $\boldsymbol{\delta}_2 = (\delta_{12}, \delta_{22}, \ldots, \delta_{n2})$, with $y_{ij} = \log(min(T_{ij}, C_{ij}))$ and $\delta_{ij} = I(y_{ij} = $

$\log T_{ij}$) for $j = 1, 2; i = 1, 2, \ldots, n$. Let $\boldsymbol{y} = (\boldsymbol{y}_1, \boldsymbol{y}_2)$, and $\boldsymbol{\delta} = (\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)$. Let (S_1, S_2) and (f_1, f_2) be the marginal survival and density functions of GEV distribution as given in Section 3. Let θ_1 and θ_2 be the parameters associated with each of the marginal distribution. So, $\theta_i = (\mu_i, \sigma_i, \xi_i)$, $i = 1, 2$. where μ, σ, and ξ are location, scale and shape parameters of the GEV distribution. Then the complete data likelihood function of the parameters $(\theta_1, \theta_2, \phi)$ can be written as (see also Chen et al. (2012)):

$$
\begin{aligned}
L(\theta_1, \theta_2, \phi | \boldsymbol{y}, \boldsymbol{\delta}) &= \prod_{i=1}^{n} (f(y_{i1}, y_{i2}))^{\delta_{i1}\delta_{i2}} \Big(\frac{\partial S(y_{i1}, y_{i2})}{\partial y_{i1}}\Big)^{\delta_{i1}(1-\delta_{i2})} \Big(\frac{\partial S(y_{i1}, y_{i2})}{\partial y_{i2}}\Big)^{(1-\delta_{i1})\delta_{i2}} \\
&\quad (S(y_{i1}, y_{i2}))^{(1-\delta_{i1})(1-\delta_{i2})} \\
&= \prod_{i=1}^{n} (c_\phi(S_{1\theta_1}(y_{i1}), S_{2\theta_2}(y_{i2})) f_{1\theta_1}(y_{i1}), f_{2\theta_2}(y_{i2}))^{\delta_{i1}\delta_{i2}} \\
&\quad \Big(-\frac{\partial C_\phi(S_{1\theta_1}(y_{i1}), S_{2\theta_2}(y_{i2}))}{\partial S_{1\theta_1}(y_{i1})}(-f_{1\theta_1}(y_{i1}))\Big)^{\delta_{i1}(1-\delta_{i2})} \\
&\quad \Big(-\frac{\partial C_\phi(S_{1\theta_1}(y_{i1}), S_{2\theta_2}(y_{i2}))}{\partial S_{2\theta_2}(y_{i2})}(-f_{2\theta_2}(y_{i2}))\Big)^{(1-\delta_{i1})\delta_{i2}} \\
&\quad (C_\phi(S_{1\theta_1}(y_{i1}), S_{2\theta_2}(y_{i2})))^{(1-\delta_{i1})(1-\delta_{i2})}.
\end{aligned}
\tag{6}
$$

where $c_\phi(., .)$ is the second derivative of the copula function $C_\phi(., .)$ defined in equation 5.

4 Proposed Methodology

There are mainly two developed approaches to solve the estimation problem. Shih and Louis (1995) proposed a two step procedure. They first assume independence between failure times and estimate the marginals and then estimate the dependency parameter ϕ assuming the estimated marginals are fixed. The Bayesian approach simultaneously estimates $(\theta_1, \theta_2, \phi)$ from joint posterior. The issue with the first approach is the unnatural assumption of independence to begin with. In the second case finding appropriate prior of ϕ is difficult because of complicated range depending on which copula family is being used. The dependency parameter for Clayton copula, for example, has the range $(-1, \infty) \setminus \{0\}$.

We start by putting appropriate priors on θ_1 and θ_2. Let $\pi(\theta_1)$ and $\pi(\theta_2)$ be the priors on θ_1 and θ_2. Then for fixed ϕ, the joint posterior density of θ_1 and θ_2 is given by,

$$
\pi_\phi(\theta_1, \theta_2 | \boldsymbol{y}, \boldsymbol{\delta}) = \frac{L(\theta_1, \theta_2, \phi | \boldsymbol{y}, \boldsymbol{\delta})\pi(\theta_1)\pi(\theta_2)}{m_\phi(\boldsymbol{y}, \boldsymbol{\delta})}, \quad (\theta_1, \theta_2) \in \Theta
\tag{7}
$$

where $m_\phi(\boldsymbol{y}, \boldsymbol{\delta})$ is the normalizing constant given by

$$m_\phi(\boldsymbol{y}, \boldsymbol{\delta}) = \int_\Theta L(\theta_1, \theta_2, \phi | \boldsymbol{y}, \boldsymbol{\delta}) \pi(\theta_1) \pi(\theta_2) d\theta_1 d\theta_2.$$

Following Roy (2014), we select that value of $\phi \in G$ which maximizes the marginal likelihood of the data $m_\phi(\boldsymbol{y}, \boldsymbol{\delta})$. Note that $m_\phi(\boldsymbol{y}, \boldsymbol{\delta})$ is not available in closed form and needs to be estimated. Instead of maximizing $m_\phi(\boldsymbol{y}, \boldsymbol{\delta})$, we choose to maximize $a \times m_\phi(\boldsymbol{y}, \boldsymbol{\delta})$ as it is often easier to estimate. We choose "a" as $1/m_{\phi_1}(\boldsymbol{y}, \boldsymbol{\delta})$ for a pre fixed value ϕ_1. Then by ergodic theorem we have a simple importance sampling consistent estimator of $B_{\phi,\phi_1} := \frac{m_\phi(\boldsymbol{y},\boldsymbol{\delta})}{m_{\phi_1}(\boldsymbol{y},\boldsymbol{\delta})}$.

$$\frac{1}{N} \sum_{l=1}^N \frac{L(\theta_1^{(l)}, \theta_2^{(l)}, \phi | \boldsymbol{y}, \boldsymbol{\delta})}{L(\theta_1^{(l)}, \theta_2^{(l)}, \phi_1 | \boldsymbol{y}, \boldsymbol{\delta})} \xrightarrow{a.s.} \int_\Theta \frac{L(\theta_1, \theta_2, \phi | \boldsymbol{y}, \boldsymbol{\delta})}{L(\theta_1, \theta_2, \phi | \boldsymbol{y}, \boldsymbol{\delta})} \pi_{\phi_1}(\theta_1, \theta_2 | \boldsymbol{y}, \boldsymbol{\delta}) d\theta_1 d\theta_2 = \frac{m_\phi(\boldsymbol{y}, \boldsymbol{\delta})}{m_{\phi_1}(\boldsymbol{y}, \boldsymbol{\delta})}, \quad (8)$$

as $N \to \infty$ where $\{\theta_1^{(l)}, \theta_2^{(l)}\}_{l=1}^N$ is a single Harris ergodic Markov chain with stationary density $\pi_{\phi_1}(\theta_1, \theta_2 | \boldsymbol{y}, \boldsymbol{\delta})$. So one *single* MCMC sample $\{\theta_1^{(l)}, \theta_2^{(l)}\}_{l=1}^N$ from $\pi_{\phi_1}(\theta_1, \theta_2 | \boldsymbol{y}, \boldsymbol{\delta})$ is used to estimate B_{ϕ,ϕ_1} for all $\phi \in G$. Once the association parameter is estimated, $\boldsymbol{\theta_1}$ and $\boldsymbol{\theta_2}$ are estimated using Markov Chain (Gibb's sampler) samples from $\pi_{\hat{\phi}}(\theta_1, \theta_2 | \boldsymbol{y}, \boldsymbol{\delta})$.

4.1 A two-stage procedure

Roy (2014) mentioned that the above estimate of B_{ϕ,ϕ_1} although simple is often unstable. To remove this instability introduced due to an arbitrary choice of ϕ_1, following Doss (2010), Roy (2014) used generalized importance sampling method. Let $\phi_1, \phi_2, \ldots, \phi_k \in G$ be k appropriately chosen skeleton points. See Roy et al. (2014) for a discussion on how to choose the skeleton points. Let $\{\theta_1^{(j;l)}, \theta_2^{(j;l)}\}_{l=1}^{N_j}$ be a Markov chain with stationary density $\pi_{\phi_j}(\theta_1, \theta_2 | \boldsymbol{y}, \boldsymbol{\delta})$ for $j = 1 \ldots, m$. Define $r_k = m_{\phi_k}(\boldsymbol{y}, \boldsymbol{\delta}) / m_{\phi_1}(\boldsymbol{y}, \boldsymbol{\delta})$ for $k = 2, 3, \ldots, m$, with $r_1 = 1$. Then B_{ϕ,ϕ_1} is consistently estimated by

$$\hat{B}_{\phi,\phi_1} = \sum_{j=1}^m \sum_{l=1}^{N_j} \frac{L(\theta_1^{(j;l)}, \theta_2^{(j;l)}, \phi | \boldsymbol{y}, \boldsymbol{\delta})}{\sum_{k=1}^m N_k L(\theta_1^{(j;l)}, \theta_2^{(j;l)}, \phi_k | \boldsymbol{y}, \boldsymbol{\delta}) / \hat{r}_k}, \quad (9)$$

where $\hat{r}_1 = 1$, $\hat{r}_k$, $k = 2, 3, \ldots, m$ are consistent estimator of r_k's obtained by the "reverse logistic regression" method proposed by Geyer (1994) described in details in the next section.

4.1.1 Estimation of $\hat{r}_k$

Following Geyer (1994), $\hat{r}_k$ are estimated by maximizing the quasi-likelihood fucntion,

$$l_n(r) = \sum_{j=1}^m \sum_{l=1}^{N_j} \log p_j(X_{jl}, r) \quad (10)$$

where $X_{jl} = (\theta_1^{(j;l)}, \theta_2^{(j;l)})$, $r = (r_1, r_2, \ldots, r_m)$, N_j is the number of MCMC samples with stationary density $\pi_{\phi_j}(\theta_1, \theta_2 | \boldsymbol{y}, \boldsymbol{\delta})$ corresponding to the skeletal point ϕ_j, $j = 1, 2, ..., m$ and the function $p_j(.,.)$ is defined below. To aviod identifiability issues, $\hat{r}_1$ is assumed to be 1. The function $p_j(.,.)$ is defined as,

$$p_j(X_{jl}, r) = \frac{L(\theta_1^{(jl)}, \theta_2^{(jl)}, \phi_j | \boldsymbol{y}, \boldsymbol{\delta}) e^{r_j}}{\sum\limits_{k=1}^{m} L(\theta_1^{(jl)}, \theta_2^{(jl)}, \phi_k | \boldsymbol{y}, \boldsymbol{\delta}) e^{r_k}} \tag{11}$$

where ϕ_k is the kth skeletal point and $\theta_i^{(jl)} = (\mu_i^{(jl)}, \sigma_i^{(jl)}, \xi_i^{(jl)})$, for $i = 1, 2$ is the lth MCMC sample of the chain with stationary density $\pi_{\phi_j}(\theta_1, \theta_2 | \boldsymbol{y}, \boldsymbol{\delta})$ corresponding to the skeletal point ϕ_j, $j = 1, 2, ..., m$ and $l = 1, ..., N_j$.

Hence the two-stage procedure for estimating B_{ϕ, ϕ_1} and hence $\hat{\phi}$ is as follows. In the first stage, we run MCMC chains $\{\theta_1^{(j;l)}, \theta_2^{(j;l)}\}_{l=1}^{N_j}$ with invariant density $\pi_{\phi_j}(\theta_1, \theta_2 | \boldsymbol{y}, \boldsymbol{\delta})$, $j = 1, 2, ..., m$. We estimate $\hat{r}$ using these first stage posterior samples using the reverse logistic method as described in section 4.1.1. In the second stage, independent of the first stage samples, we obtain *new* samples $\{\theta_1^{(j;l)}, \theta_2^{(j;l)}\}_{l=1}^{N_j}$ from $\pi_{\phi_j}(\theta_1, \theta_2 | \boldsymbol{y}, \boldsymbol{\delta})$, $j = 1, 2, ..., m$ to estimate B_{ϕ, ϕ_1} using equation (9). Finally, the estimate $\hat{\phi}$ where $\hat{B}_{\phi, \phi_1}$ attains maximum, can be obtained by an optimization procedure or simply looking at the plot of $\hat{B}_{\phi, \phi_1}$ against ϕ.

5 Simulation Study

5.1 Generating data

R package "Copula" was used to generate samples from bivariate distribution with Uniform (0, 1) marginals. The data has dependency according to previously mentioned Clayton copula structure. Next we used "Inverse Survival Function" approach to get survival times generated from bivariate distribution with $GEV(0, 1, \xi)$ marginals. Let $(u_1, u_2) \sim U(0, 1)$. Let t denote the generated survival time.

Case I: $\xi \neq 0$

$$1 - u_i = S(t_i)$$
$$= \exp[-(1 + \xi \log t_i)_+^{\frac{1}{\xi}}] \tag{12}$$
$$\Rightarrow t_i = \exp[\frac{-1 + (-\log(1 - u_i))^{\xi}}{\xi}] \tag{13}$$

8

Case II: $\xi = 0$

$$1 - u_i = S(t_i) \tag{14}$$

$$= \exp(-t_i)$$

$$\Rightarrow t_i = -\log(1 - u_i) \tag{15}$$

Table 1: Naive Importance Sampling Results for simulated data

ϕ_1	$Param$	Estimate[S.E.]	95% HPD
1.00	$\hat{\phi}$	5.2114	-
	ξ_1	0.308[0.009]	(0.293, 0.323)
	ξ_2	0.312[0.009]	(0.299, 0.326)
2.00	$\hat{\phi}$	5.2116	-
	ξ_1	0.308[0.008]	(0.292,0.322)
	ξ_2	0.312[0.006]	(0.300,0.324)
4.00	$\hat{\phi}$	5.2116	-
	ξ_1	0.307[0.008]	(0.291,0.321)
	ξ_2	0.312[0.007]	(0.300,0.321)
6.00	$\hat{\phi}$	5.2115	-
	ξ_1	0.308[0.008]	(0.291,0.322)
	ξ_2	0.310[0.009]	(0.294,0.322)
8.00	$\hat{\phi}$	5.2117	-
	ξ_1	0.308[0.008]	(0.294,0.322)
	ξ_2	0.310[0.008]	(0.295,0.326)

A random sample of 1000 observations were generated following the above method. We selected true $\xi_1 = \xi_2 = 0.3$ and true $\phi = 5$. Censoring percentages in both t_1 and t_2 were taken to be about 12%. 3000 MCMC samples were generated and the first 1000 were discarded as burnin. A uniform prior of (-0.7, 0.7) was considered for ξ. Several initial values of ϕ_1 were tried. The result of the simulation is displayed in Table 1. Although, the simulation results of the naive importnace

9

sampling are quite stable, it is not the case always and there is a need to find a stable estimation procedure. This will be demonstrated by the real data analysis in the coming section.

5.2 Estimation of Parameters

Table 2: Estimation results for Generalized Importance Sampling

$Param$	Estimate[S.E.]	95% HPD
$\hat{\phi}$	5.212	-
$\hat{\tau}$	0.723	-
ξ_1	0.[0.018]	(0.502,0.569)
ξ_2	0.540[0.015]	(0.510,0.562)

In order to use our multi chain importance sampling method, four skeletal points (2, 4, 6, 8) were selected. We simulated 1000 data points following the method described in Section 5. The true value of ϕ was 5 and $\xi_1 = \xi_2 = 0.3$. For each of the four ϕ, 3000 MCMC samples were generated. Using the MCMC samples, following Geyer (1994), first $\hat{r}_k$, $k = 2, 3, 4$ were estimated (we assume $hatr_1 = 1$ due to model identifiability issues) and then using them n second stage, estimate of ϕ was obtained. Finally fixing ϕ as $\hat{\phi}$, ξ_1 and ξ_2 were estimated. Table 2 provides the details.

5.3 Note on Monte Carlo Simulation

Since the extreme value distributions have inter dependent parameters, i.e., the criterion: $1 + \xi \frac{\log t - \mu}{\sigma} > 0$ must always hold, when performing Monte Carlo simulations, there is a need to specify bounds for each parameter contained in the likelihood function. When conducting a real data analysis we assume that both the marginal GEV's have three unknown parameters:μ, σ and ξ. Also more often than not there is one or more associated covariates. Let us assume that there is univariate Z for simplicity. We introduce the covariate through μ parameter as $\mu = \beta_0 + \beta_1 Z$. So overall we will have 8 parameters in the likelihood function: $\sigma_1, \xi_1, \beta_{01}$ (intercept for log of survival time y_1), β_{11} (slope for log of survival time y_1), β_{02} (intercept for log of survival time y_2), β_{12} (slope for log of survival time y_2).

We define the bounds for each of the parameters in the following way:

For ξ, define $\nu_i = \log y_i - \mu_i$, $i = 1, .., n$. We assume n is the sample size and the standard

assumption that $\sigma > 0$. (The parameters ξ_1 and ξ_2 will have similar bounds.) Then,

$$max(-1, max_{1 \leq i \leq n}[-\frac{\sigma}{\nu_i}\mathbb{1}_{\nu_i > 0}]) < \xi < min(1, min_{1 \leq i \leq n}[-\frac{\sigma}{\nu_i}\mathbb{1}_{\nu_i < 0}])$$

For σ, define $\zeta_i = \xi(\log y_i - \mu_i), \; i = 1, ..., n$. As before, parameters σ_1 and σ_2 will have similar bounds. Then,

$$\sigma > max(0, max_{1 \leq i \leq n}(-\zeta_i))$$

For the intercept β_0, we define $A_i = \log y_i + \frac{\sigma}{\xi} - \beta_1 Z_i$. For $\xi > 0$, $\beta_0 \in (-\infty, minA_i)$. For $\xi < 0$, $\beta_0 \in (maxA_i, \infty)$.

For the slope β_1, define $B_i = \frac{\log y_i + \frac{\sigma}{\xi} - \beta_0}{Z_i}$. For $\xi > 0$, $\beta_1 \in (-\infty, minB_i)$. For $\xi < 0$, $\beta_1 \in (maxB_i, \infty)$. Note that these bounds hold for postive $Z_i, \; i = 1, ..., n$. Further adjustments are needed in case the covariate can take negative values.

5.4 Note on Implementation of the Algorithm

In practice, often log likelihood function is used instead of likelihood for scaling issues and also to obtain a simplified form. While estimating $\hat{r}_k$ or $\hat{B}_{\phi,\phi_1}$ function in the estimation procedure, using likelihood can be a deterrent due to the large magnitude of the evaluated likelihood. One way around this problem is to deal with log likelihood and consider maximizing:

$$l_n(r) = \sum_{j=1}^{m} \sum_{l=1}^{N_j} \log p_j(X_{jl}, r) \tag{16}$$

where,

$$
\begin{aligned}
p_j(X_{jl}, r) &= \frac{1}{\frac{\sum_{k=1}^{m} e^{L^*(\theta_1^{(jl)}, \theta_2^{(jl)}, \phi_k | \boldsymbol{y}, \boldsymbol{\delta})} e^{r_k}}{e^{L^*(\theta_1^{(jl)}, \theta_2^{(jl)}, \phi_j | \boldsymbol{y}, \boldsymbol{\delta})} e^{r_j}}} \\
&= \frac{1}{\sum_{k=1}^{m} \frac{e^{L^*(\theta_1^{(jl)}, \theta_2^{(jl)}, \phi_k | \boldsymbol{y}, \boldsymbol{\delta})} e^{r_k}}{e^{L^*(\theta_1^{(jl)}, \theta_2^{(jl)}, \phi_j | \boldsymbol{y}, \boldsymbol{\delta})} e^{r_j}}} \\
&= [\sum_{k=1}^{m} e^{L_k^* - L_j^*} e^{r_k - r_j}]^{-1}
\end{aligned}
\tag{17}
$$

Romeo and Tanaka (2006)

References

M.H. Chen, J. Ibrahim, and D. Sinha. Bayesian Inference for Multivariate Survival Data with a Cure Fraction. *Journal of Multivariate Analysis*, 80:101–126, 2002.

Z. Chen, D. Oakes, and O. Hyrien. A Flexible Copula Model for Bivariate Survival Data. *Dissertation Thesis, Department of Biostatistics and Computational Biology, School of Medicine and Dentistry, University of Rochester*, pages 30–31, 2012.

D.G. Clayton. A Model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease coincidence. *Biometrika*, 65:141–151, 1978.

H. Doss. Estimation of Large Families of Bayes Factors from Markov Chains output. *Statistica Sinica*, 20:537 –560, 2010.

C. Genest and L.P. Rivest. Statistical Inference Procedures for Bivariate Archimedean Copulas. *Journal of the American Statistical Association*, 88:1034 – 1043, 1993.

C.J. Geyer. Estimating Normalizing Constants and reweighting Mixtures in Markov Chain Monte Carlo. *Technical Report, School of Statistics, University of Minnesota*, 1994.

W. J. Huster, R. Brookmeyer, and S.G. Self. Modelling Paired Survival Data with Covariates. *Biometrics*, 45:145–156, 1989.

R.B. Nelson. Properties of a one-parameter family of bivariate distributions with specified marginals. *Communications in Statistics, Part A*, 153:3277–3285, 1986.

R.B. Nelson. *An Introduction to Copulas*. Springer, 1999.

D. Oakes. A model for association in bivariate survival data. *Journal of Royal Statistical Society B*, 44:414–428, 1982.

D. Oakes. Semiparametric inference in a model for association in bivariate survival data. *Biometrika*, 73:353–361, 1986.

J. Romeo and N. Tanaka. Bivariate survival modeling: a Bayesian approach based on copulas. *Lifetime Data Analysis*, 12:205 – 222, 2006.

D. Roy, V. Roy, and D. Dey. Analysis of survival data with a cure fraction under generalized extreme value distribution. *Tech. Report 49, Department of Statistics, University of Connecticut*, 2013.

V. Roy. Efficient Estimation of the link function parameter in a robust Bayesian binary regression model. *Computational Statistics and Data Analysis*, 73:87–102, 2014.

V. Roy and D. Dey. Propriety of Posterior Distributions arising in Catgorical and Survival Models under generalized extreme value distribution. *Statistica Sinica*, 2:699–722, 2014.

V. Roy, E. Evangelou, and Z. Zhu. Efficient estimation and prediction for the Bayesian binary spatial model with flexible link functions. *Techical Report, Iowa State University*, 2014.

S.K. Sahu and D. Dey. A Comparison of Frailty and Other Models for Bivariate Survival Data. *Lifetime Data Analysis*, 6:207–228, 2000.

A. Shemyakin and H. Youn. Bayesian Estimation of Joint Survival Functions in Life Insurance. *ISBA Proceedings*, pages 463– 471, 2000.

J.H. Shih and T.A. Louis. Inferences on the Association Parameter in Copula Models for Bivariate Survival Data. *Biometrics*, 51:1384–1399, 1995.

W. Wang and M. Wells. Model Selection and Semiparametric Inference for Bivariate Failure-Time Data. *Journal of the American Statistical Association*, 95:62– 72, 2000.