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

Dooti Roy, Vivekananda Roy

March 16, 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 a empirical Bayes method for estimating the dependency parameter which is more efficient than the other existing Bayesian methods. 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.

Keywords: Clayton copula, GEV, DRS, Empirical Bayes, Metropolis Hastings, Bivariate survival

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 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 semiparamtric, 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 be estimated separately than the dependency.

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 covariates. Chen et al. (2002) introduced a Bayesian model for right censored bivariate survival data with cure fraction. Another interesting work is 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 an 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 empirical 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 the Bayesian framework for inference and estimation of the marginals and the dependence parameter following the approach outlined in Roy (2014) using Clayton copula structure and the flexible GEV distribution to model the marginal distributions.

2 A Brief Overview of Copula

"Copula" "couples" a joint cdf to its marginal distributions and hence the name. Essentially a copula is a bivariate distribution with uniform marginals. The following theorem defines a copula structure.

Sklar's Theorem: 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 is an m-dimensional cdf with marginal cdfs $F_1, F_2, \ldots, F_m$.

Nelson in his book "Introduction to Copulas" defined the properties of a m-dimensional copula:

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 CDF. 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 τ and 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 modeling univariate right censored survival data with a cure fraction using Generalized Extreme Value distribution. The distribution characterized by three parameters is extremely flexible and also achieves proper posterior distribution with a variety of mild to diffused 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.

Minima Generalized Extreme Value distribution is defined following Roy et al. (2013). Going forward, it will be simply referred to as the GEV. 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 of GEV(μ, σ, ξ) distribution can be defined respectively as:

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

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

where $\mu \in R$, $\sigma \in R^+$, and $\xi \in R$ are the location, scale, and shape parameters respectively.

3.2 Bivariate Model

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, $\log T_1 \sim$ MGEV (μ_1, σ_1, ξ_1) and $\log T_2 \sim$ MGEV (μ_2, σ_2, ξ_2) for $i = 1, 2$, where MGEV is defined above following Roy et al. (2013). The joint survival function based on a copula $C_\phi, \phi \in G$ is given by

$$S(t_1, t_2) = C_\phi(S_1(t_1), S_2(t_2)),$$

where $S_1(t_1)$ and $S_2(t_2)$ are the marginal survival functions obtained from the previous section. 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

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

In this case $G = (-1, \infty) \setminus \{0\}$.

3.3 Model Settings

Let $T_{ij}(C_{ij})$ be the survival (censoring) time of the jth component for the ith subject, $j = 1, 2; i = 1, 2, ..., n.$. We assume (T_{i1}, T_{i2}) and (C_{i1}, C_{i2}) are independent and (T_{i1}, T_{i2}), $i = 1, 2, ..., n.$ be iid with common pdf $f(t_1, t_2)$ and survival time $S(t_1, t_2)$. So the observed data is $(\boldsymbol{y_1, y_2, \delta_1, \delta_2})$ where $\boldsymbol{y_i} = (y_{1i}, y_{2i}, ..., y_{ni})$, $\boldsymbol{\delta_i} = (\delta_{1i}, \delta_{2i}, ..., \delta_{ni})$. Let $y_{ij} = min(T_{ij}, (C_{ij}))$ and $\delta_{ij} = I(y_{ij} = T_{ij})$. Let $\boldsymbol{y} = (\boldsymbol{y_1, y_2})$ and $\boldsymbol{\delta} = (\boldsymbol{\delta_1, \delta_2})$. Here (S_1, S_2) and (f_1, f_2) are the marginal survival and density functions of logGEV respectively as defined before. θ_1 and θ_2 are the parameters associated with each of the marginals.

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

$$\begin{aligned} L(\theta_1, \theta_2, \phi | \boldsymbol{y, \delta}) = & \prod_{i=1}^{n} (f(y_{i1}, y_{i2}))^{\delta_{i1}\delta_{i2}} \Big(\frac{\partial S(y_{i1}, y_{i2})}{\partial y_{i1}} \Big)^{\delta_{i1}(1-\delta_{i2})} \Big(\frac{\partial S(y_{i1}, y_i 2)}{\partial y_{i2}} \Big)^{\delta_{i2}(1-\delta_{i1})} \\ & (S(y_{i1}, y_{i2}))^{(1-\delta_{i1})(1-\delta_{i2})} \\ = & \prod_{i=1}^{n} (c_\phi(S_{1\theta_1}(y_{i1}), S_{2\theta_2}(y_{i2}) f_{1\theta_1}(y_{i1}), f_{2\theta_2}(y_{i2}))^{\delta_{i1}\delta_{i2}} \\ & \Big(- \frac{\partial C_\phi(S_{1\theta_1}(y_{i1}), S_{2\theta_2}(y_{i2})}{\partial S_{1\theta_1}(y_{i1})} (-f_{1\theta_1}(y_{i1})) \Big)^{\delta_{i1}(1-\delta_{i2})} \\ & \Big(- \frac{\partial C_\phi(S_{1\theta_1}(y_{i1}), S_{2\theta_2}(y_{i2})}{\partial S_{2\theta_2}(y_{i1})} (-f_{2\theta_2}(y_{i2})) \Big)^{\delta_{i2}(1-\delta_{i1})} \\ & (C_\phi(S_{1\theta_1}(y_{i1}), S_{2\theta_2}(y_{i2}))^{(1-\delta_{i1})(1-\delta_{i2})} \end{aligned} \tag{3}$$

4 Proposed Methodology

4.1 Stage I

There is mainly two developed approach to solve the problem. Shih and Louis (1995) proposed a two step procedure. First assume independence between failure times and estimate the marginals and then estimate 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 unnatural assumption of independence to begin with. In the second case finding appropriate prior of ϕ is difficult as it

depends on which copula family is being used.

We start by putting appropriate priors on θ_1 and θ_2. For fixed ϕ, the joint posterior density of θ_1 and θ_2 is given by

$$\pi_\phi(\theta_1, \theta_2|\boldsymbol{y}, \boldsymbol{\delta}) = \frac{L(\theta_1, \theta_2, \phi|\boldsymbol{y}, \boldsymbol{\delta})\pi(\theta_1)\pi(\theta_2)}{m_\phi(\boldsymbol{y}, \boldsymbol{\delta})}, \quad (\theta_1, \theta_2) \in \Theta, \tag{4}$$

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 $m_\phi(\boldsymbol{y}, \boldsymbol{\delta})$. Instead of maximizing $m_\phi(\boldsymbol{y}, \boldsymbol{\delta})$, we choose to maximize, $a \times m_\phi(\boldsymbol{y}, \boldsymbol{\delta})$ as it often easier. We choose "a" as $m_{\phi_1}(\boldsymbol{y}, \boldsymbol{\delta})$ for a pre fixed value ϕ_1. Then by ergodic theorem we have a simple consistent estimator of B_{ϕ,ϕ_1},

$$\frac{1}{N}\sum_{l=1}^{N} \frac{L(\theta_1^{(l)}, \theta_2^{(l)}, \phi|\boldsymbol{y}, \boldsymbol{\delta})}{L(\theta_1^{(l)}, \theta_2^{(l)}, \phi_1|\boldsymbol{y}, \boldsymbol{\delta})} \overset{a.s.}{\to} \int_\Theta \frac{L(\theta_1, \theta_2, \phi|\boldsymbol{y}, \boldsymbol{\delta})}{L(\theta_1, \theta_2, \phi|\boldsymbol{y}, \boldsymbol{\delta})}\pi_{\phi_1}(\theta_1, \theta_2|\boldsymbol{y}, \boldsymbol{\delta})d\theta_1 d\theta_2 = \frac{m_\phi(\boldsymbol{y}, \boldsymbol{\delta})}{m_{\phi_1}(\boldsymbol{y}, \boldsymbol{\delta})}, \tag{5}$$

as $N \to \infty$ where $\{\theta_1^{(l)}, \theta_2^{(l)}\}_{l=1}^N$ is a single Harris ergodic Markov chain with stationary density $\pi_{\phi_1}(\theta_1, \theta_2|\boldsymbol{y}, \boldsymbol{\delta})$. Once the association parameter is determined, Gibb's sampler was used to determine the marginal parameter estimates.

4.2 Stage II

Roy (2014) mentioned that the above estimate of ϕ although simple is often unstable. To remove the instability introduced due to an arbitrary choice of ϕ_1, Roy (2014) proposed a revised method. Let $\phi_1, \phi_2, \ldots, \phi_m \in G$ be m appropriately chosen 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 $j = 1 \ldots, k$. 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 maximizing

$$\hat{B}_{\phi,\phi_1} = \sum_{j=1}^{m}\sum_{l=1}^{N_j} \frac{L(\theta_1^{(j;l)}, \theta_2^{(j;l)}, \phi|\boldsymbol{y}, \boldsymbol{\delta})}{\sum_{k=1}^{m} N_k L(\theta_1^{(j;l)}, \theta_2^{(j;l)}, \phi_k|\boldsymbol{y}, \boldsymbol{\delta})/\hat{r}_k}, \tag{6}$$

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).

4.3 Estimating $\hat{r_k}$

Following Geyer (1994), $\hat{r_i}$ are estimated by maximizing the quasi-likelihood function,

$$l_n(r) = \sum_{j=1}^{m} \sum_{l=1}^{N} \log p_j(X_{lj}, r) \tag{7}$$

where $X_{ij} = \{\theta_1^{(l;j)}, \theta_2^{(l;j)}\}$, $r = (r_1, r_2, ..., r_m)$, N is the number of MCMC samples generated for each skeletal point $\phi_j,\ j = 1, .., m$. To avoid indentifiability issues, $\hat{r_1}$ is assumed to be 1. The function $p_j(.)$ is defined as

$$p_j(X_{ij}, r) = \frac{L(\theta_1^{(j;l)}, \theta_2^{(j;l)}, \phi_j | \boldsymbol{y}, \boldsymbol{\delta}) e^{r_j}}{\sum_{k=1}^{m} L(\theta_1^{(j;l)}, \theta_2^{(j;l)}, \phi_k | \boldsymbol{y}, \boldsymbol{\delta}) e^{r_k}} \tag{8}$$

where ϕ_k is the $k'th$ skeletal point and $\theta_1^{(j;l)} = (\mu_1^{(j;l)}, \sigma_1^{(j;l)}, xi_1^{(j;l)})$ is the $l'th$ MCMC sample of the chain generated by choosing skeletal point $\phi_j; j = 1, ..., k$ and $l = 1, ..., N$ assuming the length of the MCMC chains for each skeletal point were equal.

5 Simulation

5.1 Stage I

R package "Copula" was used to generate bivariate $(u_1, u_2) \sim$ U(0,1). The data has dependency according to previously mentioned Clayton Copula structure. Next we used "Inverse Survival Function" approach to get survival times generated from $mGEV(0, 1, \xi)$.

- Case I : $\xi \neq 0$.

$$\begin{aligned} 1 - u =& S(y) \\ =& \exp\left[-(1 + \xi \log(y))^{\frac{1}{\xi}+}\right] \\ \Rightarrow y =& \exp\left[\frac{-1 + (-\log(u))^{\xi}}{\xi}\right] \end{aligned} \tag{9}$$

- Case II: $\xi = 0$.

$$\begin{aligned} 1 - u =& S(y) \\ =& \exp(-y) \\ \Rightarrow y =& -\log(1 - u) \end{aligned} \tag{10}$$

Select $\xi_1 = \xi_2 = 0.3$ and $\phi = 5$. Censoring percentages in both Y_1 and Y_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.

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)

5.2 Stage II

Four skeletal points (2, 4, 6, 8) were selected. 1000 data points were simulated following similar method as in Stage I. This time 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 $\hat{r_1} = 1$ due to model identifiability issues) and then using them, estimate of ϕ was obtained. Then using ϕ, ξ_1 and ξ_2 were estimated. Table 2 provides the details.

$Param$	Estimate[S.E.]	95% HPD
$\hat{\phi}$	5.212	-
ξ_1	0.308[0.008]	(0.290,0.321)
ξ_2	0.311[0.008]	(0.298,0.326)

5.3 Important Note on Simulation and Data Analysis

Since the extreme value distributions have inter dependent parameters, i.e., the criterion: $1 + \xi\frac{\log y - \mu}{\sigma} > 0$ 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 X for simplicity. We introduce the covariate through μ parameter as $\mu = \beta_0 + \beta_1 X$. So overall we will have 8 parameters in the likelihood function: $\sigma_1, \xi_1, \sigma 2, \xi_2, \beta_{01}$(intercept for survival time y_1), β_{11} (slope for survival time y_1), β_{02} (intercept for survival time y_2), β_{12} (slope for 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 $\sigma > 0$.(ξ_1 and ξ_2 will have similar bounds.) Then,

$$max(-1, max[\frac{\sigma}{\nu_i}\mathbb{1}_{\nu_i>0}]) < \xi < min(-1, min[\frac{\sigma}{\nu_i}\mathbb{1}_{\nu_i>0}]) \tag{11}$$

For σ, define $\zeta_i = \xi(\log(y_i) - \mu_i), i = 1, ..., n$. We assume n is the sample size and $\sigma > 0$.(σ_1 and σ_2 will have similar bounds.) Then,

$$\sigma > max(0, max(-\zeta_i)) \tag{12}$$

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

For the slope β_1, define $B_i = \frac{\log(y_i)+\frac{\sigma}{\xi}-\beta_0}{X_i}$. For $\xi > 0$, $\beta_1 \in (-\infty, minB_i)$. For $\xi < 0$, $\beta_1 \in (maxB_i, \infty)$.

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 Stage II, 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 considering maximizing:

$$l_n(r) = \sum_{j=1}^{m}\sum_{l=1}^{N} \log p_j(X_{lj}, r) \tag{13}$$

where,

$$
\begin{aligned}
p_j(X_{ij}, r) &= \frac{1}{\frac{\sum_{k=1}^{m} e^{L^*(\theta_1^{(j;l)}, \theta_2^{(j;l)}, \phi_k | \boldsymbol{y}, \boldsymbol{\delta})} e^{r_k}}{e^{L^*(\theta_1^{(j;l)}, \theta_2^{(j;l)}, \phi_j | \boldsymbol{y}, \boldsymbol{\delta})} e^{r_j}}} \\
&= \frac{1}{\sum_{k=1}^{m} \frac{e^{L^*(\theta_1^{(j;l)}, \theta_2^{(j;l)}, \phi_k | \boldsymbol{y}, \boldsymbol{\delta})} e^{r_k}}{e^{L^*(\theta_1^{(j;l)}, \theta_2^{(j;l)}, \phi_j | \boldsymbol{y}, \boldsymbol{\delta})} e^{r_j}}} \\
&= \frac{1}{\sum_{k=1}^{m} e^{L_k^* - L_j^*} e^{r_k - r_j}}
\end{aligned}
\tag{14}
$$

Where $L_k^* = L^*(\theta_1^{(j;l)}, \theta_2^{(j;l)}, \phi_k | \boldsymbol{y}, \boldsymbol{\delta})$. Depending on the converged MCMC chains, the quantity $L_k^* - L_j^*$ should not be overly large and hence the maximization can be carried out. Similar technique can be applied to maximization of function $\hat{B}_{\phi,\phi_1}$ in the next part of stage II. where $L^*()$ is the log likelihood function.

6 Application to Diabetes Retinopathy Study Data

We consider the well known Diabetes Retinopathy Study (DRS) Data for our analyses. This data was first introduced by Huster et al. (1989) in their seminal paper on modeling bivariate data with covariates. The Diabetes retinopathy is the leading cause of blindness in Unites States under 60 years accounting for approximately 12% of all new cases. It has been shown that 90% of patients who have been diabetic for more than a decade will eventually develop retinopathy, or an ocular manifestation of blindness. Huster et al. (1989) analyzed the data using a frequentist approach with the Clayton copula family and Weibull marginal distributions. Therneau and Grambsch (2000) considered frailty model and Sahu and Dey (2000) considered exponential and Weibull bivariate distributions with a Bayesian approach. Romeo and Tanaka (2006) considered copulas to model the dependency of the data and follow the same two step procedures outlined by Shih and Louis (1995). The Diabetes Retinopathy Study (DRS) data consists of 197 patients with severe retinopathy affecting the eye. For each patient, the data records a treated eye and a control, untreated eye. Time from detection till blindness is the response variable. One or both eye may be censored. The censoring happens in approximately 79% of the treated eyes and 49% of the untreated eyes. Age at the onset of diabetes is considered as a covariate.

Let us assume survival time, $y_j \sim GEV(\mu_j, \sigma_j, \xi_j)$, j = 1,2. The covariate is included through $\mu_j = \beta_{0j} + \beta_{1j}$, j = 1,2. For simplicity, independent priors are considered on μ, σ, ξ and β. Since we know from 5.3, that the values of the marginal parameters μ, σ, ξ are all interdependent, we adjust for the dependency by restricting the proposed parameter space within the Monte Carlo simulations. Also, we take $\sigma^2 \sim IG(a_\sigma, b_\sigma)$ and $\xi \sim Uniform(-a, a)$. We pick $a = 1$, $a_\sigma = 0.01$, $b_\sigma = 2$ .We put diffuse priors for β_0s and β_1s as N(0, 100). The choices of the hyper parameters are such that the posterior distribution is proper but the prior is diffuse.

For the first step, following the method outlined in stage I, an arbitrary $\phi_1 = 3$ value was chosen. Several initial values of ϕ_1 were tested, however not much difference was observed in the final estimates. Next, using this ϕ_1, we use the Gibbs sampler with Metropolis-Hastings steps to get the

posterior samples for parameters. For this data set, we have 10,000 MCMC iterations. Convergence was checked using the trace plots, ergodic mean plots and the autocorrelation plots for all the parameters. And we find that 1,000 iterations are adequate as burn-in. Using 9,000 posterior samples, we then estimate the final *phi* value by maximizing $B_{\phi,\phi 1}$. Final estimates and HPD intervals for all 8 parameters were calculated based on the final estimated ϕ. The following table provides us the estimates for the parameters.

Param	Estimate[S.E.]	95% HPD
$\hat{\phi}$	0.782	-
σ_1	0.765[0.127]	(0.536,1.028)
σ_2	1.102 [0.168]	(0.714,1.354)
ξ_1	-0.019[0.145]	(-0.304,0.267)
ξ_2	-0.004[0.167]	(-0.329,0.299)
β_{01}	2.904[0.236]	(2.457,3.363)
β_{11}	0.009[0.009]	(-0.01,0.025)
β_{02}	3.033 [0.297]	(2.424,3.612)
β_{12}	-0.005[0.012]	(-0.028,0.017)

Next following the method outlined in Stage II, four skeletal points were chosen as $\phi_1 = 0.2,\ \phi_2 = 0.5,\ \phi_3 = 1,\ \phi_4 = 3$. 10,000 MCMC samples were generated for each skeletal points and 1000 were considered as burn-in. The maximum likelihood estimates of the marginal parameters given the value of skeletal point ϕ were estimated by maximizing the log likelihood function and were considered as initial starting value for the MCMC simulations. Once all the four MCMC chains were obtained, the function (7) was maximized to obtain estimated $\hat{r_j}$s. The estimated values of $\hat{r_j}$s were quite robust to the change in initial start values. The following tables provides the values of $\hat{r_j}$s.

$\mathbf{r}$	r_1	r_2	r_3	r_4
$\hat{r}$	1	1.071	2.844	12.218

Using the estimated $\hat{r_j}$s, $\hat{B}_{\phi,\phi_1}$ is maximized as a function of ϕ and final $\hat{\phi}$ was obtained. Lastly using the final estimate of the association parameter $\hat{\phi}$, the marginal parameters were determined using Gibb's sampling algorithm as before. The following table provides the all the final estimates.

Param	Estimate[S.E.]	95% HPD
$\hat{\phi}$	1.718	-
σ_1	0.878[0.154]	(0.601,1.203)
σ_2	1.181 [0.220]	(0.754,1.619)
ξ_1	0.019[0.158]	(-0.301,0.311)
ξ_2	0.055[0.178]	(-0.323,0.344)
β_{01}	3.109[0.239]	(2.621,3.549)
β_{11}	0.005[0.01]	(-0.014,0.023)
β_{02}	3.237 [0.348]	(2.578,3.962)
β_{12}	-0.008[0.014]	(-0.039,0.018)

7 Discussion

In Romeo and Tanaka (2006), the authors found the estimated ϕ to be approximately 1.061. Our estimated ϕ is higher but in the same direction.It can be concluded that the marginal time to blindness for the patients eyes are indeed associated. The estimates of slopes of predictor Age is not significant for both the marginals. This finding agrees with Romeo and Tanaka (2006). Our method will work on any bivariate data not just survival and using any of the popular copula families. This makes our method much generalized and widely applicable. A future direction of work includes extending the model for multivariate response. We believe that the efficiency gain due to our method will be even more substantial under the multivariate setting.

References

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

Clayton, D. (1978), "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.

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

Geyer, C. J. (1994), "Estimating Normalizing Constants and reweighting Mixtures in Markov Chain Monte Carlo," *Technical Report*.

Huster, W., Brookmeyer, R., and Self, S. (1989), "Modelling Paired Survival Data with Covariates," *Biometrics*, 45, 145–156.

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

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

— (1986), "Semiparametric inference in a model for association in bivanate survival data," *Biometrika*, 73, 353–361.

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

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

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

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

Shemyakin, A. and Youn, H. (2000), "Bayesian Estimation of Joint Survival Functions in Life Insurance," *ISBA Proceedings*.

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

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