Bayesian modelling of temporal dependence for bivariate survival data using copulas

Jose Romeo & Renate Meyer

SHORE & Whariki Research Centre, Massey University – Department of Statistics, University of Auckland

e-mail: j.romeo@massey.ac.nz

Introduction

In multivariate survival analysis there may be a natural association because individuals share biological and/or environmental conditions.
Examples:

- Lifetimes of pairs of human organs (eg kidneys, eyes)
- Recurrent events: (i) asthma attacks for a subject, (ii) the first and second infection from the insertion of the catheter in subjects using a portable dialysis equipment
- Clustered failure times such as failure times of twins

➪ The assumption of independence among lifetimes can be unrealistic.
➪ It is of interest to estimate and quantify the dependence among the lifetimes and the effects of covariates under the dependence structure.

Diabetic Retinopathy Study: Follow up times for 197 diabetic patients. The main endpoint is severe visual loss in each eye. Treatment was randomly assigned to one eye of each patient.
T_1: the time up to visual loss for the treatment eye, (73% cen.), T_2: time up to visual loss for the control eye, (49% cen.), covariate age, 1: adult, 0: young (58% young).

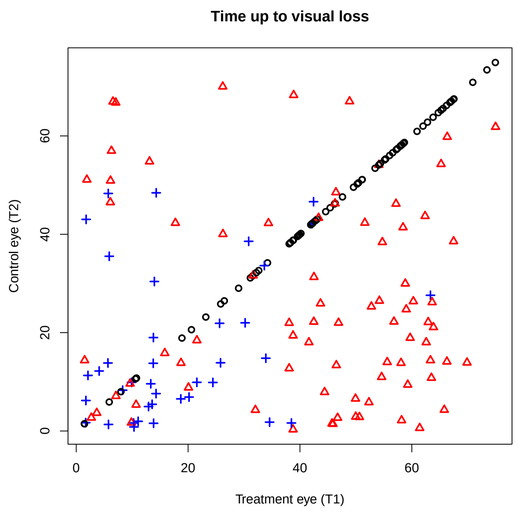

Figure 1: Scatterplot Diabetic retinopathy data

➪ It might be natural to assume the dependence between the lifetimes is not constant through time.

- Time-varying copula models allow parameter estimates to vary over time and therefore obtain a dynamic structure of the dependence between bivariate lifetimes.
- Copulas considering time varying dependence structure mainly analysing multivariate times series in finance, see e.g. Van den Goorbergh, Genest & Werker (2005) and the references of them. Also, see Abbara & Zevallos (2014). From a Bayesian approach, see Ausin & Lopes (2010).
- Dynamic frailty models: a Bayesian time-dependent frailty model in Manda & Meyer (2005), and Pennell & Dunson (2006).

➪ In this work we propose a bivariate model where the dependence structure is carry out temporally through a time-varying copula function, allowing to the dependence parameter to vary over time.
➪ We use the temporal factorization of the likelihood function following Gamerman (1991) and a two-steps estimation procedure.

Copulas

Let T_1, T_2 r.v.'s with $T_j \sim F_j$, $j = 1, 2$, the joint cdf can be written as $F(t_1, t_2) = C_\alpha^F(F_1(t_1), F_2(t_2)), \quad \alpha \in \mathcal{A}$, with joint pdf $f(t_1, t_2) = c_\alpha^F(F_1(t_1), F_2(t_2)) \prod_{j=1}^2 f_j(t_j)$. Also, the joint survival function is given by $S(t_1, t_2) = C_\alpha(S_1(t_1), S_2(t_2))$ where $S_j(\cdot) = 1 - F_j(\cdot)$.

- **Clayton**: $C_\alpha(S_1(t_1), S(t_2)) = \left(S_1(t_1)^{-\alpha} + S_2(t_2)^{-\alpha} - 1\right)^{1/\alpha}, \quad \alpha \in \mathbb{R}^+$
- **Positive stable**: $C_\alpha(S_1(t_1), S_2(t_2)) = \exp\left\{-\left[(-\log(S_1(t_1)))^{1/\alpha} + (-\log(S_2(t_2)))^{1/\alpha}\right]^\alpha\right\}, \quad \alpha \in [0, 1]$

Estimation - Two steps

Let (T_1, T_2) bivariate lifetime r.v.'s with survival functions (S_1, S_2) and pdf (f_1, f_2). Let (C_1, C_2) bivariate censoring times. For $i, \ldots, n$, suppose (T_{i1}, T_{i2}) and (C_{i1}, C_{i2}) independents. For each i we observe $Z_{ij} = \min(T_{ij}, C_{ij})$ and $\delta_{ij} = I[Z_{ij} = T_{ij}]$, $j = 1, 2$.

➪ **Step 1**: Estimation of Marginals. Piecewise exponential model (PE), $T_j|\boldsymbol{\lambda}^j \sim \text{PE}(\boldsymbol{\lambda}^j)$, $j = 1, 2$.
Suppose $T_1|\boldsymbol{\lambda}^1 \perp\!\!\!\perp T_2|\boldsymbol{\lambda}^2$ and the partition $0 = a_0 < a_1 < \cdots < a_{K-1} < a_K = \infty$, so that on the interval $I_k = (a_{k-1}, a_k]$ the hazard function is $h(t_j) = \lambda_k^j, \quad a_{k-1} \le t_j < a_k$.
We consider $\boldsymbol{\lambda}^j \sim \text{AR(1)Gamma}$. Thus, the joint prior density of $\boldsymbol{\lambda}^j$ is

$$\pi(\boldsymbol{\lambda}^j) = \pi(\lambda_k^j|\lambda_{k-1}^j) \cdot \pi(\lambda_{k-1}^j|\lambda_{k-2}^j) \cdots \pi(\lambda_2^j|\lambda_1^j) \cdot \pi(\lambda_1^j)$$

with $\lambda_k^j|\lambda_{k-1}^j \sim \text{Gamma}\left(c, \frac{c}{\lambda_{k-1}^j}\right), \ k = 2, \ldots, K$, and $\lambda_1^j \sim \text{Gamma}(c, c)$

Then we obtain $\widehat{\boldsymbol{\lambda}}^j$ from the posterior $\pi(\boldsymbol{\lambda}^j|\boldsymbol{z}_j, \boldsymbol{\delta}_j) \propto L(\boldsymbol{\lambda}^j|\boldsymbol{z}_j, \boldsymbol{\delta}_j)\pi(\boldsymbol{\lambda}^j)$. The likelihood function is expressed in terms of a temporal factorization (Gamerman, 1991)

$$L(\boldsymbol{\lambda}^j|\boldsymbol{z}_j, \boldsymbol{\delta}_j) = \prod_{k=1}^K L_k\left((\boldsymbol{z}_j, \boldsymbol{\delta}_j)^{(k)}|\boldsymbol{\lambda}^j, D_{k-1}^j\right),$$

where D_k^j is the information set with the observed information of each individual until a_{k-1} and

$$L_k\left((\boldsymbol{z}_j, \boldsymbol{\delta}_j)^{(k)}|\boldsymbol{\lambda}^j, D_{k-1}^j\right) = \prod_{i=1}^n f_j(z_{ij}|\boldsymbol{\lambda}^j, D_{k-1}^j)^{\delta_{ij}} S_j(z_{ij}|\boldsymbol{\lambda}^j, D_{k-1}^j)^{1-\delta_{ij}}$$

➪ **Step 2**: Estimation of temporal dependence. Interested in estimating $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K)$ into each interval. Given $\widehat{\boldsymbol{\lambda}}^j$, the pseudo-likelihood function is

$$L(\boldsymbol{\alpha}|\boldsymbol{Z}, \boldsymbol{\delta}, \widehat{\boldsymbol{\lambda}}^1, \widehat{\boldsymbol{\lambda}}^2) = \prod_{k=1}^K L_k\left((\boldsymbol{Z}, \boldsymbol{\delta})^{(k)}|\boldsymbol{\alpha}, D_{k-1}\right)$$

where the likelihood in the I_k interval in terms of the copula is

$$L_k\Big((\boldsymbol{Z}, \boldsymbol{\delta})^{(k)}|\boldsymbol{\alpha}, D_{k-1}\Big) = \prod_{i=1}^n [c_{\alpha_k}(\hat{S}_1(z_{i1k}), \hat{S}_2(z_{i2k}))]^{\delta_{i1k}\delta_{i2k}} \cdot \left[\frac{\partial C_{\alpha_k}(\hat{S}_1(z_{i1k}), \hat{S}_2(z_{i2k}))}{\partial S_1(z_{i1k})}\right]^{\delta_{i1k}(1-\delta_{i2k})}$$
$$\cdot \left[\frac{\partial C_{\alpha_k}(\hat{S}_1(z_{i1k}), \hat{S}_2(z_{i2k}))}{\partial S_2(z_{i2k})}\right]^{(1-\delta_{i1k})\delta_{i2k}} C_{\alpha_k}(\hat{S}_1(z_{i1k}), \hat{S}_2(z_{i2k}))^{(1-\delta_{i1k})(1-\delta_{i2k})},$$

where $\hat{S}_j(\cdot) = S_j(\cdot|\widehat{\boldsymbol{\lambda}}^j)$ is the estimate survival function from Step 1.

The estimation of $\boldsymbol{\alpha}$ is based on the 'posterior distribution'

$$\pi(\boldsymbol{\alpha}|\boldsymbol{Z}, \boldsymbol{\delta}, \hat{\boldsymbol{\lambda}}^1, \hat{\boldsymbol{\lambda}}^2) \propto \prod_{k=1}^K L_k\left((\boldsymbol{Z}, \boldsymbol{\delta})^{(k)}|\boldsymbol{\alpha}, D_{k-1}\right)\pi(\boldsymbol{\alpha})$$

Application - Diabetic Retinopathy Study

Step 1: $K = 8$ and same number of events in each I_k.

- $T_1|\boldsymbol{\lambda}^1 \sim \text{PE}(\boldsymbol{\lambda}^1), \quad \lambda_k^1|\lambda_{k-1}^1 \sim \text{AR(1)Gamma}(3.4, \frac{3.4}{\lambda_{k-1}^1}), \quad \lambda_1^1 \sim \text{Gamma}(3.4, 3.4)$
- $T_2|\boldsymbol{\lambda}^2 \sim \text{PE}(\boldsymbol{\lambda}^2), \quad \lambda_k^2|\lambda_{k-1}^2 \sim \text{AR(1)Gamma}(3.5, \frac{3.5}{\lambda_{k-1}^2}), \quad \lambda_1^2 \sim \text{Gamma}(3.5, 3.5)$

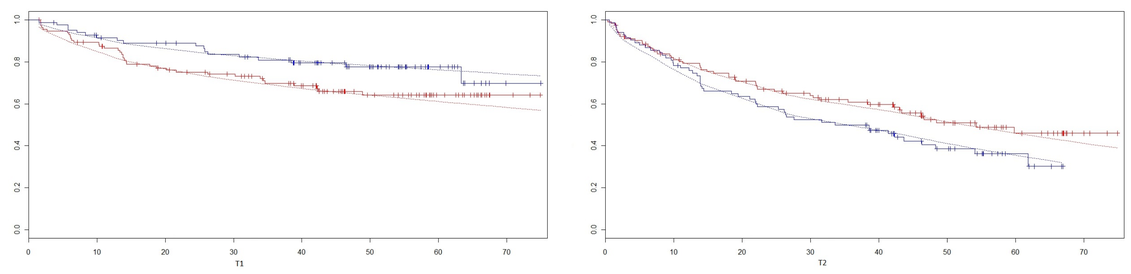

Figure 2: Estimated survival function for T_1 and T_2, PE model

Step 2: The estimates from Step 1 are replaced into the temporal pseudo-likelihood of the dynamic copula model to obtain the posterior distribution of $\boldsymbol{\alpha}$

$$\pi(\boldsymbol{\alpha}|\boldsymbol{Z}, \boldsymbol{\delta}, \hat{\boldsymbol{\lambda}}^1, \hat{\boldsymbol{\lambda}}^2) \propto \prod_{k=1}^K L_k\left((\boldsymbol{Z}, \boldsymbol{\delta})^{(k)}|\boldsymbol{\alpha}, D_{k-1}\right)\pi(\boldsymbol{\alpha})$$

Prior distribution for $\boldsymbol{\alpha}$ in the Clayton copula:

$$\alpha_k|\alpha_{k-1} \sim \text{AR(1)Gamma}\left(b, \frac{b}{\alpha_{k-1}}\right), \quad k = 2, \ldots, 8, \quad \alpha_1|b \sim \text{Gamma}(d, d), \quad d \sim \text{Gamma}(0.01, 0.01)$$

Prior distribution for $\boldsymbol{\alpha}$ in the Positive stable copula:

$$\alpha_k|\alpha_{k-1} \sim \text{AR(1)Beta}(b \cdot \alpha_{k-1}, b \cdot (1 - \alpha_{k-1})), \quad k = 2, \ldots, 8, \quad \alpha_1|b \sim \text{Beta}(d, d), \quad d \sim \text{Gamma}(0.01, 0.01)$$

Model selection criteria:

Copula	AIC	BIC	DIC
Clayton	11160.7	11186.9	11150.4
Positive stable	11873.0	11899.3	11863.2

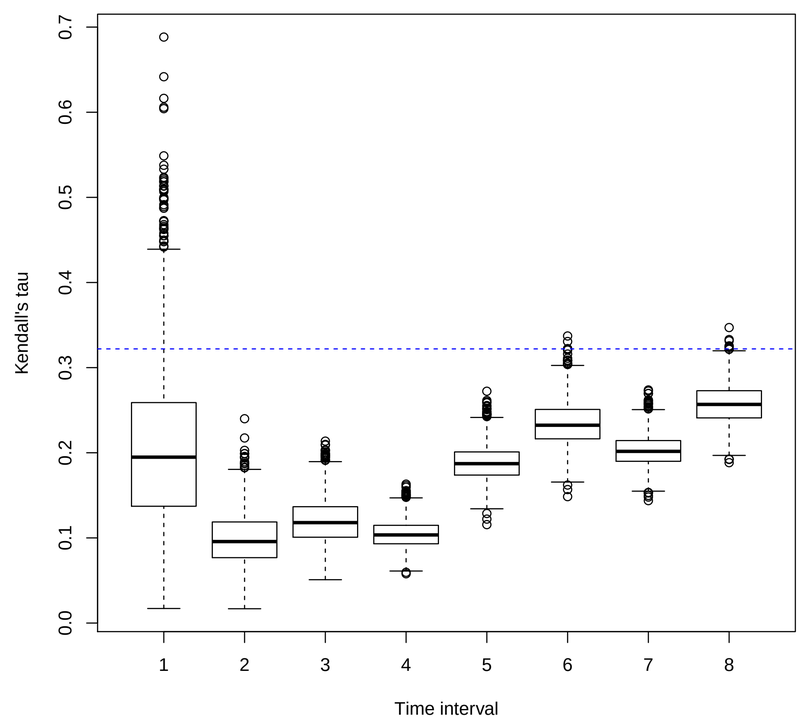

Figure 3: Temporal dependence τ_{α_k} - Clayton copula

➪ **Interpretation**: On the first months failure times are moderately dependent, then treatment becomes effective, and by the end of study (6 years) the disease returns due to abnormal blood vessels growth.

Discussion and future work

- Flexibility of models based on copulas: dependence structure and marginals distributions
- Dependence can vary over time in real situations
- Future work: Simulation study, joint estimation procedure, ...

Main References

Ausin, M. and Lopes, H. (2010). Time-varying joint distribution through copulas. *Computational Statistics & Data Analysis*, **54**, 2383-2399.

Gamerman, D. (1991). Dynamic Bayesian models for survival data. *Applied Statistics*, **40**, 63-79.

Hougaard, P. (2000). *Analysis of Multivariate Survival Data.* Springer, New York.

Romeo, J.S., Tanaka, N.I. and Pedroso de Lima, A.C. (2006). Bivariate survival modeling: A Bayesian approach based on Copulas. *Lifetime Data Analysis*, **12**, 205–222.

van den Goorbergh, R.W., Genest, C. and Werker, B.J. (2005). Bivariate option pricing using dynamic copula models. *Insurance: Mathematics and Economics*, **37**, 101-114.