Chapter 4
Analysis of Survival Data Under an Assumed Copula

Abstract This chapter introduces statistical methods for analyzing survival data subject to dependent censoring. We review the copula-graphic estimator, parametric likelihood methods, and semi-parametric likelihood methods developed under a variety of copula models. All these approaches employ an *assumed copula*, a copula function that is completely specified including its parameter value to avoid the non-identifiability.

Keywords Burr distribution · Competing risk · Copula-graphic estimator
Maximum likelihood estimator · Spline · Weibull distribution

4.1 Introduction

The idea of an *assumed copula* was suggested by Zheng and Klein (1995) in their analysis of survival data subject to dependent censoring. They considered a bivariate distribution function of survival time and censoring time, where the form of the copula function is completely specified, including its parameter value. This strong assumption of the copula is imposed to make the model identifiable. Assuming the independence copula is equivalent to the assumption of independent censoring between survival time and censoring time.

Zheng and Klein (1995) view censoring as a *competing risk* of death and view death as a competing risk of censoring. This is the setting of *bivariate competing risks* where one can observe the first-occurring event time and the type of the observed event (death or censoring whichever comes first). With this view, survival data with dependent censoring are equivalent to bivariate competing risks data. In the context of competing risks, the independence among event times is rarely assumed since many medical and engineering applications yield event times that are positively associated. Hence, statistical methods for analyzing bivariate competing risks data can be applicable for analyzing survival data with dependent censoring.

Under an assumed copula, Zheng and Klein (1995) estimated the marginal survival function by the *copula-graphic* (CG) estimator. The survival function

estimated by the CG estimator is analogous to the one estimated by the Kaplan–Meier estimator. The CG estimator reduces to the Kaplan–Meier estimator under the independence copula. In real applications, the CG estimator is calculated by assuming one of Archimedean copulas. Rivest and Wells (2001) obtained a simple expression of the CG estimator when the assumed copula belongs to Archimedean copulas. Nowadays, the CG estimator is an indispensable tool for analyzing survival data with dependent censoring (Braekers and Veraverbeke 2005; Staplin 2012; de Uña-Álvarez and Veraverbeke 2013; 2017; Emura and Chen 2016; Emura and Michimae 2017; Moradian et al. 2017). Note, however, that the CG estimator cannot handle covariates. Likelihood-based approaches can naturally deal with covariates under an assumed copula.

   Throughout this chapter, we review the copula-graphic estimator, parametric likelihood methods, and semi-parametric likelihood methods developed under an assumed copula.

4.2   The Copula-Graphic (CG) Estimator

Analysis of survival data often begins by drawing the Kaplan–Meier survival curve which graphically summarizes survival experience of patients in the data. However, under dependent censoring, the Kaplan–Meier estimator may give biased information about survival. A survival curve calculated from the CG estimator provides unbiased information about survival if the copula function between death time and censoring time is correctly specified. Below, we shall introduce the CG estimator under an Archimedean copula as derived in Rivest and Wells (2001).

   Consider *random variables*, defined as

- T: survival time
- U: censoring time

   Consider an Archimedean copula model

$$\Pr(T > t, U > u) = \phi_\theta^{-1}[\phi_\theta\{S_T(t)\} + \phi_\theta\{S_U(u)\}], \tag{4.1}$$

where $\phi_\theta : [0, 1] \mapsto [0, \infty]$ is a generator function, which is continuous and strictly decreasing from $\phi_\theta(0) = \infty$ to $\phi_\theta(1) = 0$ (Chap. 3); $S_T(t) = \Pr(T > t)$ and $S_U(u) = \Pr(U > u)$ are the marginal survival functions.

   Let (t_i, δ_i), $i = 1, \ldots, n$, be survival data without covariates, where $t_i = \min\{T_i, U_i\}$, $\delta_i = \mathbf{I}(T_i \leq U_i)$, and $\mathbf{I}(\cdot)$ is the indicator function. Assume that all the observed times are distinct ($t_i \neq t_j$ whenever $i \neq j$). Based on the data, one can estimate the survival function by the following estimator:

The CG estimator is defined as

$$\hat{S}_T(t) = \phi_\theta^{-1}\left[\sum_{t_i \leq t, \delta_i = 1} \phi_\theta\left(\frac{n_i - 1}{n}\right) - \phi_\theta\left(\frac{n_i}{n}\right)\right], \quad 0 \leq t \leq \max_i(t_i)$$

where $n_i = \sum_{\ell=1}^{n} \mathbf{I}(t_\ell \geq t_i)$ is the *number at risk* at time t_i; $\hat{S}_T(t) = 1$ if no death occurs up to time t; $\hat{S}_T(t)$ is undefined for $t > \max_i(t_i)$.

The derivation of the CG estimator: Assume that $S_T(t)$ is a decreasing step function with jumps at death times. Thus, $\delta_i = 1$ implies $S_T(t_i) \neq S_T(t_i - dt)$ and $S_U(t_i) = S_U(t_i - dt)$. Setting $t = u = t_i$ in Eq. (4.1), we have

$$\phi_\theta\{\Pr(T > t_i, U > t_i)\} = \phi_\theta\{S_T(t_i)\} + \phi_\theta\{S_U(t_i)\}.$$

In the left-side of the preceding equation, we estimate $\Pr(T > t_i, U > t_i)$ by $(n_i - 1)/n$, where $n_i - 1 = \sum_{\ell=1}^{n} \mathbf{I}(t_\ell > t_i)$ is the number of survivors at time t_i. Accordingly,

$$\phi_\theta\left(\frac{n_i - 1}{n}\right) = \phi_\theta\{S_T(t_i)\} + \phi_\theta\{S_U(t_i)\}. \tag{4.2}$$

Meanwhile, we set $t = u = t_i - dt$ in Eq. (4.1) and then estimate $\Pr(T > t_i - dt, U > t_i - dt)$ by n_i/n. Then,

$$\phi_\theta\left(\frac{n_i}{n}\right) = \phi_\theta\{S_T(t_i - dt)\} + \phi_\theta\{S_U(t_i)\}, \quad \delta_i = 1. \tag{4.3}$$

Equations (4.2) and (4.3) result in the system of difference equations

$$\phi_\theta\left(\frac{n_i - 1}{n}\right) - \phi_\theta\left(\frac{n_i}{n}\right) = \phi_\theta\{S_T(t_i)\} - \phi_\theta\{S_T(t_i - dt)\}, \quad \delta_i = 1.$$

We impose the usual constraint that $S_T(t_i - dt) = 1$ when t_i is the smallest death time. Then, the solution to the different equations is

$$\phi_\theta\{S_T(t)\} = \sum_{t_i \leq t, \delta_i = 1} [\phi_\theta\{S_T(t_i)\} - \phi_\theta\{S_T(t_i - dt)\}]$$

$$= \sum_{t_i \leq t, \delta_i = 1} \phi_\theta\left(\frac{n_i - 1}{n}\right) - \phi_\theta\left(\frac{n_i}{n}\right),$$

which is equivalent to the CG estimator. ∎

Under the independence copula, given by $\phi_\theta(t) = -\log(t)$, the CG estimator is equivalent to the Kaplan–Meier estimator. Under the Clayton copula, given by $\phi_\theta(t) = (t^{-\theta} - 1)/\theta$ for $\theta > 0$, the CG estimator is written as

$$\hat{S}_T(t) = \left[1 + \sum_{t_i \le t, \delta_i = 1} \left\{ \left(\frac{n_i - 1}{n} \right)^{-\theta} - \left(\frac{n_i}{n} \right)^{-\theta} \right\} \right]^{-1/\theta}.$$

This CG estimator can be computed by the *compound. Cox* R package (Emura et al. 2018).

The CG estimator provides a graphical summary of survival experience for patients in the same manner as the Kaplan–Meier estimator.

> *The survival curve is defined as the plot of $\hat{S}_T(t)$ against t, starting with t = 0 and ending with $t_{\max} = \max\limits_{i}(t_i)$. The curve is a step function that jumps only at points where a death occurs. On the curve, censoring times are often indicated as the mark "+".*

If $t_{\max} = \max\limits_{i}(t_i)$ corresponds to time-to-death of a patient, then $\hat{S}_T(t_{\max}) = \phi_\theta^{-1}(\infty) = 0$. This is because $\phi_\theta\left(\frac{n_i-1}{n}\right) = \phi_\theta(0) = \infty$ for some i in the definition of the CG estimator. If $t_{\max} = \max\limits_{i}(t_i)$ corresponds to censoring time of a patient, then $\hat{S}(t_{\max}) > 0$.

Additional remarks: The CG estimator can be modified to accommodate a variety of different censoring and truncation mechanisms. de Uña-Álvarez and Veraverbeke (2013) derived the CG estimator when survival time is subject to both dependent censoring and independent censoring. This estimator is convenient if the data provide the causes of censors for all patients. For instance, censoring caused by dropout may be dependent while censoring caused by the study termination is independent (see Chap. 14 of Collett (2015)). de Uña-Álvarez and Veraverbeke (2017) derived the CG estimator when survival time is subject to both dependent censoring and independent truncation. Chaieb et al. (2006) and Emura and Murotani (2015) derived the CG estimator when survival time is subject to independent censoring and *dependent truncation*.

4.3   Model and Likelihood

Throughout this chapter, we consider a bivariate survival function

$$\Pr(T > t, U > u|\mathbf{x}) = C_\theta\{S_T(t|\mathbf{x}), S_U(u|\mathbf{x})\},$$

where C_θ is a copula (Nelsen 2006) with a parameter θ; $S_T(t|\mathbf{x}) = \Pr(T > t|\mathbf{x})$ and $S_U(u|\mathbf{x}) = \Pr(U > u|\mathbf{x})$ are the marginal survival functions. The covariates are defined as $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2)$ such that $S_T(t|\mathbf{x}) = S_T(t|\mathbf{x}_1)$ and $S_U(u|\mathbf{x}) = S_U(t|\mathbf{x}_2)$. For instance, if $\mathbf{x}_1 = (\text{Age}, \text{gender})$ and $\mathbf{x}_2 = (\text{gender})$, the model does not consider the effect of age on censoring time.

Survival data consist of $(t_i, \delta_i, \mathbf{x}_i)$, $i = 1, \ldots, n$, where $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})'$ is a vector of covariates. The likelihood for the ith patient is expressed as

$$L_i = \Pr(T = t_i, U > t_i|\mathbf{x}_i)^{\delta_i} \Pr(T > t_i, U = t_i|\mathbf{x}_i)^{1-\delta_i} = f_T^{\#}(t_i|\mathbf{x}_i)^{\delta_i} f_U^{\#}(t_i|\mathbf{x}_i)^{1-\delta_i},$$

where

$$f_T^{\#}(t_i|\mathbf{x}_i) = -\frac{\partial}{\partial x}\Pr(T > x, U > t_i|\mathbf{x}_i)\Big|_{x=t_i},$$

$$f_U^{\#}(t_i|\mathbf{x}_i) = -\frac{\partial}{\partial y}\Pr(T > t_i, U > y|\mathbf{x}_i)\Big|_{y=t_i},$$

are called the *sub-density functions*. Therefore, the log-likelihood is defined as

$$\ell = \sum_{i=1}^{n} [\delta_i \log f_T^{\#}(t_i|\mathbf{x}_i) + (1 - \delta_i) \log f_U^{\#}(t_i|\mathbf{x}_i)]. \tag{4.4}$$

An equivalent expression is

$$\ell = \sum_{i=1}^{n} [\delta_i \log h_T^{\#}(t_i|\mathbf{x}_i) + (1 - \delta_i) \log h_U^{\#}(t_i|\mathbf{x}_i) - \Phi(t_i, t_i|\mathbf{x}_i)], \tag{4.5}$$

where

$$h_T^{\#}(t_i|\mathbf{x}_i) = \frac{f_T^{\#}(t_i|\mathbf{x}_i)}{\Pr(T > t_i, U > t_i|\mathbf{x}_i)}, \quad h_U^{\#}(t_i|\mathbf{x}_i) = \frac{f_U^{\#}(t_i|\mathbf{x}_i)}{\Pr(T > t_i, U > t_i|\mathbf{x}_i)},$$

are the *cause-specific hazard* functions, and

$$\Phi(t_i, t_i|\mathbf{x}_i) = -\log \Pr(T > t_i, U > t_i|\mathbf{x}_i) = -\log \Pr(\min\{T, U\} > t_i|\mathbf{x}_i)$$

is the cumulative hazard function for $\min\{T, U\}$.

With appropriate models on C_θ, $S_T(\cdot|\mathbf{x})$ and $S_U(\cdot|\mathbf{x})$, one can obtain the maximum likelihood estimator (MLE) with Eqs. (4.4) or (4.5).

4.4 Parametric Models

4.4.1 The Burr Model

Escarela and Carrière (2003) considered a copula model with the Burr distribution defined as

$$S_T(t|\mathbf{x}_{1i}) = \left\{1 + \gamma_1(\lambda_{1i}t)^{v_1}\right\}^{-1/\gamma_1}, \quad t \geq 0;$$
$$S_U(u|\mathbf{x}_{2i}) = \left\{1 + \gamma_2(\lambda_{2i}u)^{v_2}\right\}^{-1/\gamma_2}, \quad u \geq 0,$$

where $v_j > 0$, $\gamma_j > 0$, and $\lambda_{ji} = \exp(\beta_{j0} + \boldsymbol{\beta}_j'\mathbf{x}_{ji})$ for $j = 1$ and 2. The Burr distribution includes many distributions as special cases; $v_j = 1$ gives the Pareto distribution, $\gamma_j = 1$ gives the log-logistic distribution, and $\gamma_j \to 0$ gives the Weibull distribution. For the copula, Escarela and Carrière (2003) considered the Frank copula.

$$C_\theta(u, v) = -\frac{1}{\theta}\log\left\{1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1}\right\}, \quad \theta \neq 0.$$

Their motivation to use the Frank model is that they wish to consider both positive dependence ($\theta > 0$) and negative dependence ($\theta < 0$) between two variables.

4.4.2 The Weibull Model

Likelihood-based analyses of Escarela and Carrière (2003) focused on the Weibull model

$$S_T(t|\mathbf{x}_{1i}) = \exp\{-(\lambda_{1i}t)^{v_1}\}, \quad t \geq 0; \quad S_U(u|\mathbf{x}_{2i}) = \exp\{-(\lambda_{2i}u)^{v_2}\}, \quad u \geq 0.$$

With the Frank copula model, they maximize the log-likelihood of Eq. (4.4) with respect to $(\beta_{10}, \boldsymbol{\beta}_1, v_1, \beta_{20}, \boldsymbol{\beta}_2, v_2)$ given the value θ. This leads to the profile likelihood

$$\ell^*(\theta) = \max_{(\beta_{10}, \boldsymbol{\beta}_1, v_1, \beta_{20}, \boldsymbol{\beta}_2, v_2)} \ell(\beta_{10}, \boldsymbol{\beta}_1, v_1, \beta_{20}, \boldsymbol{\beta}_2, v_2|\theta).$$

The MLE of $(\beta_{10}, \boldsymbol{\beta}_1, v_1, \beta_{20}, \boldsymbol{\beta}_2, v_2)$ is obtained at a given value $\hat{\theta} = \arg\max_\theta \ell^*(\theta)$.

The data analysis of Escarela and Carrière (2003) revealed that the estimator $\hat{\theta}$ had a wide confidence interval (CI) if no covariate enters the model. This phenomenon is related to the non-identifiability of the model. The CI of $\hat{\theta}$ was shrunken if many covariates enter the model. Heckman and Honoré (1989) showed that the non-identifiability is resolved by adding covariates into the marginal models. Unfortunately, there are no papers that give the conditions (e.g., how many covariates or how many samples) required to give reasonable precision of $\hat{\theta}$ for estimating the true value θ.

In this context, we suggest regarding the approach of Escarela and Carrière (2003) as a two-step fashion. The first stage *selects* (not *estimates*) θ via the profile likelihood. With the selected value $\hat{\theta}$, the second stage estimates the remaining parameters $(\beta_{10}, \boldsymbol{\beta}_1, \nu_1, \beta_{20}, \boldsymbol{\beta}_2, \nu_2)$ by the MLE. The SEs of $(\beta_{10}, \boldsymbol{\beta}_1, \nu_1, \beta_{20}, \boldsymbol{\beta}_2, \nu_2)$ may not account for the variation of $\hat{\theta}$ following the approaches of an assumed copula.

4.4.3 The Pareto Model

In the absence of covariates, Shih et al. (2018) considered the Pareto marginal models

$$S_T(t) = (1 + \alpha_1 t)^{-\gamma_1}, \quad t \geq 0; \quad S_U(u) = (1 + \alpha_2 u)^{-\gamma_2}, \quad u \geq 0,$$

where $\alpha_j > 0$ and $\gamma_j > 0$ are re-parameterized from the Burr models. The marginal hazard functions are $h_T(t) = \alpha_1 \gamma_1 / (1 + \alpha_1 t)$ and $h_U(u) = \alpha_2 \gamma_2 / (1 + \alpha_2 u)$ and the marginal density functions are $f_T(t) = h_T(t) S_T(t)$ and $f_U(u) = h_U(u) S_U(u)$. Applying the Frank copula to Eq. (4.4), the log-likelihood can be written as

$$\ell(\alpha_1, \alpha_2, \gamma_1, \gamma_2 | \theta) = \sum_{i=1}^{n} \delta_i \{ \log f_T(t_i) - \theta S_T(t_i) + \log(e^{-\theta S_T(t_i)} - 1) - \log(e^{-\theta} - 1) + \theta S(t_i) \}$$

$$+ \sum_{i=1}^{n} (1 - \delta_i) \{ \log f_U(t_i) - \theta S_U(t_i) + \log(e^{-\theta S_U(t_i)} - 1) - \log(e^{-\theta} - 1) + \theta S(t_i) \},$$

where $S(t) = C_\theta \{ S_T(t), S_U(t) \}$. The MLE is obtained by maximizing the preceding equation.

They developed a Newton–Raphson algorithm to obtain the MLE of $(\alpha_1, \alpha_2, \gamma_1, \gamma_2)$ given the value θ. The *Bivariate.Pareto* R package (Shih and Lee 2018) can be used to compute the MLE and the SE for the parameters. Hence, this model uses an assumed copula. Their Newton–Raphson algorithm employs a *randomization* scheme to reduce the sensitivity of the convergence results against the initial values, which is termed the *randomized Newton–Raphson* algorithm (Hu and Emura 2015). When θ is unknown, the profile likelihood estimate was suggested, namely $\hat{\theta} = \arg\max_\theta \ell^*(\theta)$, where $\ell^*(\theta) = \max_{(\alpha_1, \alpha_2, \gamma_1, \gamma_2)} \ell(\alpha_1, \alpha_2, \gamma_1, \gamma_2 | \theta)$.

However, they reported that the profile likelihood occasionally does not have a peak and $\hat{\theta}$ has a large sampling variation. These problems are related to the non-identifiability of competing risks data (Tsiatis 1975).

Due to the difficulty of estimating θ, Shih et al. (2018) considered a restricted model $S_T(t) = S_U(t) = (1 + \alpha t)^{-\gamma}$. The model makes a strong assumption that the two marginal distributions are the same. Under the Frank copula, they developed the randomized Newton–Raphson algorithm to obtain the MLE of (α, γ, θ). While the peak of the likelihood always exists under this restricted model, the variation of estimating θ remains large. Including covariates into the marginal Pareto models may improve the precision of $\hat{\theta}$. Alternatively, a sensitivity analysis may be considered under a few selected values of θ.

4.4.4 The Burr III Model

In the absence of covariates, Shih and Emura (2018) considered the Burr III marginal distributions

$$S_T(t) = 1 - (1 + t^{-\gamma})^{-\alpha}, \quad t > 0; \quad S_U(u) = 1 - (1 + u^{-\gamma})^{-\beta}, \ u > 0,$$

where (α, β, γ) are positive parameters. They considered the generalized FGM copula with a copula parameter θ. In their model, the copula is imposed on a bivariate distribution function rather than a bivariate survival function. More details about this copula, such as the range of θ and the expressions of Kendall's tau, are referred to Amini et al. (2011), Domma and Giordano (2013) and Shih and Emura (2016, 2018).

Shih and Emura (2018) used the randomized Newton–Raphson algorithm to obtain the MLE of (α, β, γ) given the value of θ. When the value of θ is unknown, they suggested making inference for (α, β, γ), followed by the profile likelihood estimate $\hat{\theta} = \arg\max_\theta \ell^*(\theta)$, where $\ell^*(\theta) = \max_{(\alpha, \beta, \gamma)} \ell(\alpha, \beta, \gamma | \theta)$. They also proposed a goodness-of-fit method to test the validity of the generalized FGM copula and the Burr III marginal models. The estimation and goodness-of-fit algorithms are implemented in the *GFGM.copula* R package (Shih 2018). Their method is developed for bivariate competing risks data, where dependent censoring is a competing risk of death, and death is a competing risk of dependent censoring.

4.4.5 The Piecewise Exponential Model

The piecewise exponential model has been considered to fit survival data with dependent censoring (Staplin et al. 2015; Emura and Michimae 2017). Let $0 = \alpha_0 < a_1 < \cdots < a_m$ be a knot sequence, where m is the number of knots. Assume

that the hazard function for T in an interval $(a_{j-1}, a_j]$ is a constant e^{θ_j} for $j = 1, \ldots, m$, such that $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)$ are parameters without restriction to their ranges. The survival function is

$$S_T(t; \boldsymbol{\theta}) = \exp\left\{ -e^{\theta_j}(t - a_{j-1}) - \sum_{k=1}^{j-1} e^{\theta_k}(a_k - a_{k-1}) \right\}, \qquad t \in (a_{j-1}, a_j],$$

where $\sum_{k=1}^{0} (\cdot) \equiv 0$. In a similar fashion, define the survival function $S_U(u; \boldsymbol{\gamma})$ for the censoring time U, where $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_m)$.

Emura and Michimae (2017) considered a copula model

$$\Pr(T > t, U > u) = C_\theta\{S_T(t; \boldsymbol{\theta}), S_U(u; \boldsymbol{\gamma})\}, \quad \boldsymbol{\theta} = (\theta_1, \ldots, \theta_m), \boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_m),$$

where $S_T(t; \boldsymbol{\theta})$ and $S_U(u; \boldsymbol{\gamma})$ follow the piecewise exponential models. The Clayton copula and the Joe copula were chosen for their numerical studies. They developed inference procedures based on the likelihood in Eq. (4.4) given the value θ. Hence, they applied an assumed copula. They did not use the profile likelihood for selecting θ since it may not work with many parameters in the marginal distributions. Alternatively, they suggested a sensitivity analysis to examine the result under a few different values of θ.

Staplin et al. (2015) originally proposed the piecewise exponential models for dependent censoring, but did not use copulas. Consequently, the sub-density functions in their likelihood function require some numerical integrations of the joint density of T and U.

4.5 Semi-parametric Models

4.5.1 The Transformation Model

Chen (2010) considered a semi-parametric transformation model defined as

$$S_T(t|\mathbf{x}_{1i}) = \exp[-G_1\{\Lambda_0(t)e^{\boldsymbol{\beta}_1' \mathbf{x}_{1i}}\}], \quad S_U(u|\mathbf{x}_{2i}) = \exp[-G_2\{\Gamma_0(u)e^{\boldsymbol{\beta}_2' \mathbf{x}_{2i}}\}],$$

where $\boldsymbol{\beta}_j$ are regression coefficients, and $G_j(\cdot)$ is a known and nonnegative increasing function such that $G_j(0) = 0$, $G_j(\infty) = \infty$, and $g_j(t) \equiv dG_j(t)/dt > 0$ for $j = 1$ and 2; Λ_0 and Γ_0 are unknown increasing functions. No distributional assumptions are imposed on Λ_0 and Γ_0. The linear transformation $G_j(t) = t$ corresponds to the Cox model.

Under the semi-parametric transformation model, the cause-specific hazard functions are

$$h_T^{\#}(t|\mathbf{x}_i) = \lambda_0(t)e^{\boldsymbol{\beta}_1'\mathbf{x}_{1i}}\eta_{1i}(t; \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \Lambda_0, \Gamma_0|\theta), \quad h_U^{\#}(t|\mathbf{x}_i) = \gamma_0(t)e^{\boldsymbol{\beta}_2'\mathbf{x}_{2i}}\eta_{2i}(t; \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \Lambda_0, \Gamma_0|\theta),$$

where $\lambda_0(t) = d\Lambda_0(t)/dt$, $\gamma_0(t) = d\Gamma_0(t)/dt$,

$$\eta_{1i}(t; \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \Lambda_0, \Gamma_0|\theta) = g_1\{\Lambda_0(t)e^{\boldsymbol{\beta}_1'\mathbf{x}_{1i}}\}S_T(t|\mathbf{x}_{1i})D_{\theta,1}[S_T(t|\mathbf{x}_{1i}), S_U(t|\mathbf{x}_{2i})],$$

$$\eta_{2i}(t; \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \Lambda_0, \Gamma_0|\theta) = g_2\{\Gamma_0(t)e^{\boldsymbol{\beta}_2'\mathbf{x}_{2i}}\}S_U(t|\mathbf{x}_{2i})D_{\theta,2}[S_T(t|\mathbf{x}_{1i}), S_U(t|\mathbf{x}_{2i})],$$

$$D_{\theta,1}(u, v) = \frac{\partial C_\theta(u, v)/\partial u}{C_\theta(u, v)}, \quad D_{\theta,2}(u, v) = \frac{\partial C_\theta(u, v)/\partial v}{C_\theta(u, v)}.$$

Under the independence copula $C_\theta(u, v) = uv$, the cause-specific hazard functions are equal to the marginal hazards:

$$h_T^{\#}(t|\mathbf{x}_i) = \lambda_0(t)e^{\boldsymbol{\beta}_1'\mathbf{x}_{1i}}g_1\{\Lambda_0(t)e^{\boldsymbol{\beta}_1'\mathbf{x}_{1i}}\}, \quad h_U^{\#}(t|\mathbf{x}_i) = \gamma_0(t)e^{\boldsymbol{\beta}_2'\mathbf{x}_{2i}}g_2\{\Gamma_0(t)e^{\boldsymbol{\beta}_2'\mathbf{x}_{2i}}\}.$$

To obtain the MLE of $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \Lambda_0, \Gamma_0)$, we treat Λ_0 and Γ_0 as increasing step functions that have jumps sizes $d\Lambda_0(t_i) = \Lambda_0(t_i) - \Lambda_0(t_i-)$ for $\delta_i = 1$ and $d\Gamma_0(t_i) = \Gamma_0(t_i) - \Gamma_0(t_i-)$ for $\delta_i = 0$. Putting the cause-specific hazard functions into Eq. (4.5) and replacing $\lambda_0(t_i)$ by $d\Lambda_0(t_i)$ and $\gamma_0(t_i)$ by $d\Gamma_0(t_i)$, we obtain the log-likelihood

$$\begin{aligned}
\ell(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \Lambda_0, \Gamma_0|\theta) = &\sum_i \delta_i[\boldsymbol{\beta}_1'\mathbf{x}_{1i} + \log \eta_{1i}(t_i; \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \Lambda_0, \Gamma_0|\theta) + \log d\Lambda_0(t_i)] \\
&+ \sum_i (1 - \delta_i)[\boldsymbol{\beta}_2'\mathbf{x}_{2i} + \log \eta_{2i}(t_i; \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \Lambda_0, \Gamma_0|\theta) + \log d\Gamma_0(t_i)] \\
&- \sum_i \Phi_\theta[S_T(t_i|\mathbf{x}_{1i}), S_U(t_i|\mathbf{x}_{2i})],
\end{aligned}$$

where $\Phi_\theta(u, v) = -\log C_\theta(u, v)$. Since the marginal distributions have a number of parameters to be estimated, the profile likelihood may not properly identify a suitable value of θ. Chen (2010) suggested a sensitivity analysis to examine the result under a few different values of θ, possibly selected by prior knowledge and expert opinion.

The approach of Chen (2010) reduces to Cox's partial likelihood approach (Cox 1972) under the independence copula and the linear transformation. Under these assumptions, the MLE $(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\Lambda}_0, \hat{\Gamma}_0)$ is obtained by maximizing two functions

$$\ell_1(\boldsymbol{\beta}_1, \Lambda_0) = \sum_i \delta_i[\boldsymbol{\beta}_1'\mathbf{x}_{1i} + \log d\Lambda_0(t_i)] + \sum_i \log S_T(t_i|\mathbf{x}_{1i}),$$

$$\ell_2(\boldsymbol{\beta}_2, \Gamma_0) = \sum_i (1 - \delta_i)[\boldsymbol{\beta}_2'\mathbf{x}_{2i} + \log d\Gamma_0(t_i)] + \sum_i \log S_U(t_i|\mathbf{x}_{2i}),$$

since $\ell(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \Lambda_0, \Gamma_0) = \ell_1(\boldsymbol{\beta}_1, \Lambda_0) + \ell_2(\boldsymbol{\beta}_2, \Gamma_0)$. Then, the MLE $(\hat{\boldsymbol{\beta}}_1, \hat{\Lambda}_0)$ for $(\boldsymbol{\beta}_1, \Lambda_0)$ is the partial likelihood estimator $\hat{\boldsymbol{\beta}}_1$ and the Breslow estimator $\hat{\Lambda}_0$ (Chap. 2).

4.5.2   The Spline Model

Emura et al. (2017) considered a spline-based model defined as

$$S_T(t|\mathbf{x}_{1i}) = \exp\{-\Lambda_0(t)e^{\boldsymbol{\beta}_1' \mathbf{x}_{1i}}\}, \quad S_U(u|\mathbf{x}_{2i}) = \exp\{-\Gamma_0(u)e^{\boldsymbol{\beta}_2' \mathbf{x}_{2i}}\},$$

where $\boldsymbol{\beta}_j$ are regression coefficients, and the baseline hazard functions are modeled by

$$\frac{d}{dt}\Lambda_0(t) = \lambda_0(t) = \sum\nolimits_{\ell=1}^{5} g_\ell M_\ell(t) = \mathbf{g}'\mathbf{M}(t), \quad \frac{d}{dt}\Gamma_0(t) = \gamma_0(t) = \sum\nolimits_{\ell=1}^{5} h_\ell M_\ell(t) = \mathbf{h}'\mathbf{M}(t),$$

where $\mathbf{M}(t) = (M_1(t), \ldots, M_5(t))'$ are the cubic M-spline basis functions (Ramsay 1988). Here, $\mathbf{g}' = (g_1, \ldots, g_5)$ and $\mathbf{h}' = (h_1, \ldots, h_5)$ are unknown positive parameters. These five-parameter approximations give a good flexibility in estimation for real applications (Ramsay 1988) and are one of reasonable choices (Commenges and Jacqmin-Gadda 2015). Since the spline bases are easy to integrate, the baseline cumulative hazard functions are computed as $\Lambda_0(t) = \sum_{\ell=1}^{5} g_\ell I_\ell(t)$ and $\Gamma_0(t) = \sum_{\ell=1}^{5} h_\ell I_\ell(t)$, where $I_\ell(t)$ is the integration of $M_\ell(t)$, called the *I-spline* basis (Ramsay 1988).

The *joint.Cox* package (Emura 2018) offers functions *M.spline ()* for computing $M_\ell(t)$ and *I.spline ()* for $I_\ell(t)$. To compute these spline bases, one needs to specify the range of t. The package uses the range $t \in [\xi_1, \xi_3]$ for the equally spaced knots $\xi_1 < \xi_2 < \xi_3$, where $\xi_2 = (\xi_1 + \xi_3)/2$. A possible choice is $\xi_1 = \min_i(t_i)$ and $\xi_3 = \max_i(t_i)$. The expressions of $M_\ell(t)$ and $I_\ell(t)$ are given in Appendix A. Figure 4.1 displays the M- and I-spline basis functions with the knots $\xi_1 = 1$, $\xi_2 = 2$, and $\xi_3 = 3$.
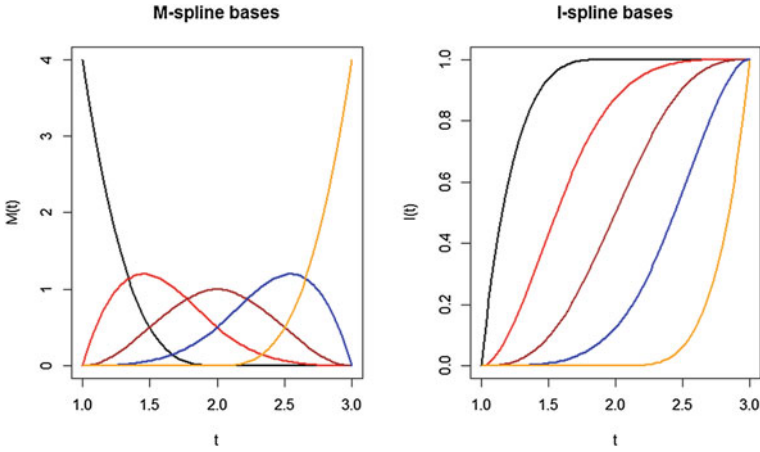
Fig. 4.1 M-spline basis functions (left-panel) and I-spline basis functions (right-panel) with knots $\xi_1 = 1$, $\xi_2 = 2$, and $\xi_3 = 3$

Under the spline model, the cause-specific hazard functions are

$$h_T^{\#}(t|\mathbf{x}_i) = \lambda_0(t)e^{\boldsymbol{\beta}_1'\mathbf{x}_{1i}}\eta_{1i}(t; \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \Lambda_0, \Gamma_0|\theta), \quad h_U^{\#}(t|\mathbf{x}_i) = \gamma_0(t)e^{\boldsymbol{\beta}_2'\mathbf{x}_{2i}}\eta_{2i}(t; \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \Lambda_0, \Gamma_0|\theta),$$

where

$$\eta_{1i}(t; \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \Lambda_0, \Gamma_0|\theta) = S_T(t|\mathbf{x}_{1i})D_{\theta,1}[S_T(t|\mathbf{x}_{1i}), S_U(t|\mathbf{x}_{2i})],$$
$$\eta_{2i}(t; \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \Lambda_0, \Gamma_0|\theta) = S_U(t|\mathbf{x}_{2i})D_{\theta,2}[S_T(t|\mathbf{x}_{1i}), S_U(t|\mathbf{x}_{2i})].$$

Putting these formulas into Eq. (4.5), we obtain the log-likelihood

$$\ell(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \mathbf{g}, \mathbf{h}|\theta) = \sum_i \delta_i[\boldsymbol{\beta}_1'\mathbf{x}_{1i} + \log \eta_{1i}(t_i; \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \Lambda_0, \Gamma_0|\theta) + \log \lambda_0(t_i)]$$
$$+ \sum_i (1 - \delta_i)[\boldsymbol{\beta}_2'\mathbf{x}_{2i} + \log \eta_{2i}(t_i; \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \Lambda_0, \Gamma_0|\theta) + \log \gamma_0(t_i)]$$
$$- \sum_i \Phi_\theta[S_T(t_i|\mathbf{x}_{1i}), S_U(t_i|\mathbf{x}_{2i})].$$

The estimator of $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \mathbf{g}, \mathbf{h})$ is obtained by maximizing the penalized log-likelihood

$$\ell(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \mathbf{g}, \mathbf{h}|\theta) - \kappa_1 \int \ddot{\lambda}_0(t)^2 dt - \kappa_2 \int \ddot{\gamma}_0(t)^2 dt,$$

where $\ddot{f}(t) = d^2f(t)/dt^2$, and (κ_1, κ_2) are given nonnegative values. The parameters (κ_1, κ_2) are called *smoothing parameters*, which control the degrees of penalties on the roughness of the two baseline hazard functions. It is shown in Appendix A that

$$\int_{\xi_1}^{\xi_3} \ddot{\lambda}_0(t)^2 dt = \mathbf{g}'\Omega\mathbf{g}, \quad \int_{\xi_1}^{\xi_3} \ddot{\gamma}_0(t)^2 dt = \mathbf{h}'\Omega\mathbf{h},$$

$$\Omega = \frac{1}{\Delta^5} \begin{bmatrix} 192 & -132 & 24 & 12 & 0 \\ -132 & 96 & -24 & -12 & 12 \\ 24 & -24 & 24 & -24 & 24 \\ 12 & -12 & -24 & 96 & -132 \\ 0 & 12 & 24 & -132 & 192 \end{bmatrix},$$

where $\Delta = \xi_2 - \xi_1 = \xi_3 - \xi_2$. A naïve approach is to set $\kappa_1 = \kappa_2 = 0$ as in Shih and Emura (2018).

A more sophisticated approach is to choose (κ_1, κ_2) by optimizing a likelihood cross-validation (LCV) criterion (O' Sullivan 1988). Under the independence copula, the penalized log-likelihood is written as the sum of two marginal penalized log-likelihoods,

$$\left[\ell_1(\boldsymbol{\beta}_1, \Lambda_0) - \kappa_1 \int \ddot{\lambda}_0(t)^2 dt \right] + \left[\ell_2(\boldsymbol{\beta}_2, \Gamma_0) - \kappa_2 \int \ddot{\gamma}_0(t)^2 dt \right],$$

where

$$\ell_1(\boldsymbol{\beta}_1, \Lambda_0) = \sum_i \delta_i [\boldsymbol{\beta}_1' \mathbf{x}_{1i} + \log \lambda_0(t_i)] - \sum_i \Lambda_0(t_i) \exp(\boldsymbol{\beta}_1' \mathbf{x}_{1i}),$$

$$\ell_2(\boldsymbol{\beta}_2, \Gamma_0) = \sum_i (1 - \delta_i)[\boldsymbol{\beta}_2' \mathbf{x}_{2i} + \log \gamma_0(t_i)] - \sum_i \Gamma_0(t_i) \exp(\boldsymbol{\beta}_2' \mathbf{x}_{2i}).$$

We suggest choosing κ_1 and κ_2 based on the two marginal LCVs defined as

$$LCV_1 = \hat{\ell}_1 - \text{tr}\{\hat{H}_{PL1}^{-1}\hat{H}_1\}, \quad LCV_2 = \hat{\ell}_2 - \text{tr}\{\hat{H}_{PL2}^{-1}\hat{H}_2\},$$

where $\hat{\ell}_1$ and $\hat{\ell}_2$ are the log-likelihood values evaluated at their marginal penalized likelihood estimates, and $\hat{H}_{PL1}$ and $\hat{H}_{PL2}$ are the converged Hessian matrices for the marginal penalized likelihood estimations, $\hat{H}_1$ and $\hat{H}_2$ are the converged Hessian matrices for the marginal log-likelihoods such that

$$\hat{H}_1 = \hat{H}_{PL1} + 2\kappa_1 \begin{bmatrix} O_{p_1 \times p_1} & O_{p_1 \times 5} \\ O_{5 \times p_1} & \Omega \end{bmatrix}, \quad \hat{H}_2 = \hat{H}_{PL2} + 2\kappa_2 \begin{bmatrix} O_{p_2 \times p_2} & O_{p_2 \times 5} \\ O_{5 \times p_2} & \Omega \end{bmatrix},$$

where O is a zero matrix and p_j is the dimension of $\boldsymbol{\beta}_j$ for $j = 1$ and 2. The values of (κ_1, κ_2) are obtained by maximizing LCV_1 for κ_1 and LCV_2 for κ_2, separately. One may apply the R function *splineCox.reg* in the *joint.Cox* R package to find the optimal value of κ_1 (or κ_2).

References

Amini M, Jabbari H, Mohtashami Borzadaran GR (2011) Aspects of dependence in generalized Farlie-Gumbel-Morgenstern distributions. Commun Stat Simul Comput 40(8):1192–1205

Braekers R, Veraverbeke N (2005) A copula-graphic estimator for the conditional survival function under dependent censoring. Can J Stat 33:429–447

Chaieb LL, Rivest LP, Abdous B (2006) Estimating survival under a dependent truncation. Biometrika 93(3):655–669

Chen YH (2010) Semiparametric marginal regression analysis for dependent competing risks under an assumed copula. J R Stat Soc Ser B Stat Methodol 72:235–251

Collett D (2015) Modelling survival data in medical research, 3rd edn. CRC Press, London

Commenges D, Jacqmin-Gadda H (2015) Dynamical biostatistical models. CRC Press, London

Cox DR (1972) Regression models and life-tables (with discussion). J R Stat Soc Ser B Stat Methodol 34:187–220

de Uña-Álvarez J, Veraverbeke N (2013) Generalized copula-graphic estimator. TEST 22(2):343–360

de Uña-Álvarez J, Veraverbeke N (2017) Copula-graphic estimation with left-truncated and right-censored data. Statistics 51(2):387–403

Domma F, Giordano S (2013) A copula-based approach to account for dependence in stress-strength models. Stat Pap 54(3):807–826

Emura T, Chen YH (2016) Gene selection for survival data under dependent censoring, a copula-based approach. Stat Methods Med Res 25(6):2840–2857

Emura T, Murotani K (2015) An algorithm for estimating survival under a copula-based dependent truncation model. TEST 24(4):734–751

Emura T, Michimae H (2017) A copula-based inference to piecewise exponential models under dependent censoring, with application to time to metamorphosis of salamander larvae. Environ Ecol Stat 24(1):151–173

Emura T, Nakatochi M, Murotani K, Rondeau V (2017) A joint frailty-copula model between tumour progression and death for meta-analysis. Stat Methods Med Res 26(6):2649–2666

Emura T (2018) joint.Cox: penalized likelihood estimation and dynamic prediction under the joint frailty-copula models between tumour progression and death for meta-analysis, CRAN

Emura T, Chen HY, Matsui S, Chen YH (2018) compound.Cox: univariate feature selection and compound covariate for predicting survival, CRAN

Escarela G, Carrière JF (2003) Fitting competing risks with an assumed copula. Stat Methods Med Res 12(4):333–349

Heckman JJ, Honore BE (1989) The identifiability of the competing risks models. Biometrika 76:325–330

Hu YH, Emura T (2015) Maximum likelihood estimation for a special exponential family under random double-truncation. Comput Stat 30(4):1199–1229

Moradian H, Denis Larocque D, Bellavance F (2017). Survival forests for data with dependent censoring. Stat Methods Med Res, https://doi.org/10.1177/0962280217727314

Nelsen RB (2006) An introduction to copulas, 2nd edn. Springer, New York

O' Sullivan F (1988) Fast computation of fully automated log-density and log-hazard estimation. SIAM J Sci Stat Comput 9:363–379

Ramsay J (1988) Monotone regression spline in action. Stat Sci 3:425–461

Rivest LP, Wells MT (2001) A martingale approach to the copula-graphic estimator for the survival function under dependent censoring. J Multivar Anal 79:138–155

Shih JH, Emura T (2016) Bivariate dependence measures and bivariate competing risks models under the generalized FGM copula. Stat Pap, https://doi.org/10.1007/s00362-016-0865-5

Shih JH, Lee W, Sun LH, Emura T (2018) Fitting competing risks data to bivariate Pareto models. Commun Stat Theory, https://doi.org/10.1080/03610926.2018.1425450

Shih JH, Emura T (2018) Likelihood-based inference for bivariate latent failure time models with competing risks under the generalized FGM copula. Comput Stat, https://doi.org/10.1007/s00180-018-0804-0

Shih JH (2018) GFGM.copula: generalized Farlie-Gumbel-Morgenstern copula, CRAN

Shih JH and Lee W (2018) Bivariate.Pareto: bivariate Pareto models, CRAN

Staplin ND (2012) Informative censoring in transplantation statistics. Doctoral Thesis, University of Southampton, School of Mathematics

Staplin ND, Kimber AC, Collett D, Roderick PJ (2015) Dependent censoring in piecewise exponential survival models. Stat Methods Med Res 24(3):325–341

Tsiatis A (1975) A nonidentifiability aspect of the problem of competing risks. Proc Natl Acad Sci 72(1):20–22

Zheng M, Klein JP (1995) Estimates of marginal survival for dependent competing risks based on an assumed copula. Biometrika 82(1):127–138