Copulas and Competing Risks: Applications for Long-term Mixture Survival Models

Ronny Westerman[1], Ulrich Mueller[1]
[1]Institute of Medical Sociology and Social Medicine, Medical School,
Philipps-University of Marburg Karl-von-Frisch-Str. 4, 35043 Marburg
Correspondence address: westerm2@staff.uni-marburg.de

Abstract

In terms of competing risks Long-term Mixture Survival Models are widely used for the analysis of individuals who may never experience the considered type of failure. If we add the possibility of a lasting therapy success. some individuals have to be treated as immune to a specific cause of failure or to be defined as long-term survivors. In case of multi- or bivariate cause-specific survival data different dependence structures can be modeled with different copula functions. There are two main methodical goals for modeling marginal distributions to be considered: First flexibility and second masked causes. We propose a Bivariate Mixture Long-term Survival model based on the Farlie-Gumbel-Morgenstern (FGM) copula. Data simulations will be provided with SEER Breast Cancer Data.

Key Words: Competing Risks, Copula, Long-term Survival, masked causes

1. Background

Competing Risks Models are popular in medical and public health studies. Still a challenge, however, is the application of cause-specific survival models in case of missing data or misclassification of cause of death. Masking is present if the considered cause of failure or cause of death is not or only partially known (Flehinger et al. 2002; Craiu and Lee, 2005; Lu and Liang, 2008; Sen et al. 2010, Roman et al. 2012).
Long-term Survival Mixture Models have been applied for the analysis of individuals who may never experience the considered type of failure. Under the condition of an unobserved prognostic factor, some individuals will be treated as immune to a certain cause of failure or be defined as long-term survivors (Maller and Zhou, 1996, Roman et al. 2012; Louzada et al. 2012).

The also applied parametric or semi-parametric versions of the Proportional Hazard Model (PH) or the Mixed Proportional Hazard Model (MPH) will be also used in advance on practice. The structure of the properties for these models seems easy to explore but their application to real data is mostly tedious. The assumption on the marginal distributions of the latent variables and their dependence structure is mostly restrictive, under some conditions inadequate.

To prevent such conditions researchers have to assume independence of the latent variables in these models.
An alternative to assuming independence would be a access the joint dependence by the means of a copula function (e.g. Escalera and Carrière, 2003; Lo and Wilke, 2009; Wienke, 2011). With different families of copulas available, the model allows for flexible specification of the dependence structure between competing random variables (Nelsen,2006).

.

2. The Model

Following Maller and Zhou (1996), here we consider a model based on two components, one component representing the failure or the survival time of individuals susceptible to a certain risk and the other component representing the not susceptible individuals, the so called immune ones (see also Francisco et al. 2012; Roman et al., 2012)

$$S_{popj}(t_j) = p_j S_j(t_j) + (1 - p_j)S_0(t_j)$$

with S_j as the Survival function for the non-susceptible (or cured) individuals, S_0 as the Survival function for susceptible (or non-cured) individuals and p_j the probability (or cured fraction) of an individual to belong to the non-susceptible group.

$$if\ S_j(t_j) = P(T > t) = 1, \forall t \geq 0, then\ S_{popj}(t_j)\ can\ be\ rewritten\ as:$$

$$S_{popj}(t_j) = p_j + (1 - p_j)S_0(t_j)$$

Follow the bivariate Archimedean copula with a single parameter

$$C\left(S_{pop1}(t_1), S_{pop2}(t_2)\right) = S_{pop1}(t_1) + S_{pop2}(t_2) - 1 + \tilde{C}\begin{pmatrix}1 - S_{pop1}(t_1),\\ 1 - S_{pop2}(t_2)\end{pmatrix}$$

can be also applied for Clayton (1978), Ali-Mikhail-Haq (1978), or Frank copula (1979).

In comparison to the alternate mixture approach the bivariate long-term survival model with Farlie-Gumble-Morgenstein distribution (FGM) Copula Model (Conway, 1983) will be defined with the joint survival function of the copula C_φ with the density function c_φ $[0,1]^2$ for $\varphi \in R$.

Then, let $(T_{1,}T_2)$ denoted as the paired failure S_{popj} and f_{popj} denote the marginal long-term survival functions and the marginal long-term density function of $T_{j,j} = 1,2$ (see also Maller and Zhou, 1996, Roman et al. 2012, Louzada et al. 2012).

$$S_{pop}(t_{1,}t_{2,}) = C_\varphi(S_{pop1}(t_1), S_{pop2}(t_2)), t_{1,}t_{2,} > 0$$

$$f_{pop}(t_{1,}t_{2,}) = c_\varphi(S_{pop1}(t_1), S_{pop2}(t_2)),\ f_{pop1}(t_1) f_{pop2}(t_2) t_{1,}t_{2,} > 0$$

The Farlie-Gumble-Morgenstern copula (FGM) was first considered by Conway (1983) as the application to a Bayesian approach estimates the effect of three copula structures by modeling the dependence effect on prevalence and performance test parameters (Bairamov and Kotz, 2002; Fisher and Klein, 2007;Amblard and Girard, 2008; Tovar Cuevas and Archor, 2011).

$$C\varphi(u, v) = uv[1 + \varphi(1 - u)(1 - v)]$$

where $0 \le u, v \le 1\ and - 1 \le \varphi \le 1$, for $\varphi > 0$ if dependence structure for $u\ and\ v$ is positive and $\varphi < 0$ if depence structure for $u\ and\ v$ is negative

Consider $(T_{1,}T_{2,})$ for FMG copula the joint long-term survival of $(T_{1,}T_{2,})$ will be given with

$$S_{pop}(t_{1,}t_{2,}) = S_{pop1}(t_1), S_{pop2}(t_2)\left[1 + \varphi\left(1 - S_{pop1}(t_1)\right)\left(1 - S_{pop2}(t_2)\right)\right]$$

Then φ parameter measures the intensity of the dependence between the lifetimes.

If $\varphi = 0,\ S_{pop1}(t_1) = S_{pop2}(t_2)$ is valid then the random variables T_1 and T_2 are independent

Copula functions rely on sophisticated methodical advances because their focus will not be on correlation coefficients but more over on scale invariant measures of association. These measures of association are functions of a measure of dependence between marginals. Then the association parameter can be defined with different values specified on the copula. In comparison to that measures of associations like the Pearson's correlation coefficient are bounded.
Modeling copulas will be arranged with the Gibbs Sampler belonging to the class of the Markov Chain Monte Carlo (MCMC) methodology.

For T_j we assume a Weibull mixture distribution with the parameters α_j and γ_j and

$$p_j = exp(\beta_{0j} + \beta_{1jx})/\left(1 + exp(\beta_{0j} + \beta_{1jx})\right)$$

with $\gamma_j \sim Gamma\ (a_j, b_j)$ and $\alpha_j \sim Gamma\ (c_j, d_j)$

3. Data

The long-term survival mixture approach will be applied for Breast Cancer Data provided by the SEER Cancer Statistic Data Base National Cancer Institute, DCCPS, Surveillance Research Program, and Cancer Statistics Branch that were released in April 2013.
The data set contains information on the incidence by race, gender and age for different period of time. We use cause specific mortality data including all types of cancer. The SEER public use dataset includes the vital status of breast cancer patients from 1992-2010 (n=69,990 in Situ).
The mixture cure model defines: S_0 will be susceptible (non-cured) individuals, identified as breast cancer case and with S_j will be non-breast cancer including all masking cases.

Simulation will be performed with OpenBUGS. (For more details in the program code: see Spiegelhalter et al. 2007)

The results from the parameter estimates are presented in table1.

4. Results

Table 1: SEER Breast Cancer Data,
Summary results from the posterior distribution, mean,
standard deviation (SD) and HPD (95%) interval for the FGM copula

	Parameter	Mean	SD	HPD (95%)
Time 1	α_1	1.457	0.158	(1.1089; 1.589)
	λ_1	0.052	0.018	(0.031; 0.073)
	β_{01}	-2.134	0.993	(-4.534; -0.675)
	β_{11}	0.754	0.976	(-1.452; 2.871)
Time 2	α_2	1.564	0.176	(1.286; 1.834)
	λ_2	0.052	0.019	(0.030; 0.074)
	β_{02}	-0.781	0.511	(-1.547; 1.034)
	β_{12}	0.843	0.574	(0.641; 0.984)
Copula	φ	0.673	0.345	(0.031; 0.978)

5. Conclusion

The major gains of this approach yield on the high flexibility to account for different dependence structure. The eventual computation problems can be neglected.
The estimates should be realized on a hierarchical two-step procedure:
First the marginals have to be estimated, then in a second step the copula to perform the joint distribution. On the other hand the identification problem for the joint distribution is still present.
Misclassification in cause of failure should be account for because the bias has serious effects on the estimates and determine lower statistical power type of misclassification also drives the bias:
Non differential misclassifications have less impact on the estimation bias than the systematic misclassifications (Sarfati et al. 2010).

References

Ali, M.M., Mikhail, N.N., Haq, M.S. (1978) A class of bivariate distributions including the bivariate logistic. J. Multivariate Anal. 8, 405-412.
Amblard, C.; Girard, S. (2008) A new extension of bivariate FGM copulas. Metrika, Secaucus, v.70, p.1-17.
Bairamov, I., Kotz, S. Dependence structure and symmetry of Huang-Kotz FGM distributions and their extensions. (2002) Metrika, Secaucus,56,(1), p.55-72.
Clayton, D.G. (1978). A model for association in bivariate life-tables and its application in epidemiological studies of family tendency in chronic disease incidence. Biometrika 65, 141-151.
Conway, D.A. (1983). Farlie-Gumbel-Morgenstern distributions. In Encyclopedia of Statistical Sciences (Edited by Kotz and N.L. Johnson), Volume 3, 28-31. Wiley, New York
Craiu R.V. and Lee T.C.M. (2005). Model Selection for the Competing-Risks Model with and without masking. Lifetime Data Anal (2006) 12:21–33
Flehinger, B. J., Reiser, B. and Yashchin, E. (2002). Parametric modeling for survival with competing risks and masked failure causes. Lifetime Data Anal. 8, 177-203.
Escarela, G. and Carrière, J.F. (2003). Fitting competing risks with an assumed copula. Statistical Methods in Medical Research 12, 333-349.
Fischer, M., Klein, I. (2007) Constructing generalized FGM copulas by means of certain univariate distributions. Metrika, Secaucus, 65(2), p.243-260.
Frank, M.J. (1979) on the simultaneous associativity of F(x,y) and x+y-F(x,y). Aequationes Mathematicae 19, 194-226.
Hsieh, J.J. and Huang Y.T. (2012). Regression analysis based on conditional likelihood approach under semi-competing risks data. Lifetime Data Analysis 18, 302-320.

Maller, R.A. and Zhou, X. (1996). Survival Analysis with Long-term Survivors. Wiley, NewYork.
Li, Y., Tiwari, R.C. and Guha, S. (2007): Mixture cure survival models with dependent censoring. Journal of the Royal Statistical Society 69, 285-306
Louzada, F., Suzuki, A.K., Cancho, V.G., Prince F.L. and Pereira, G.A. (2012) The Longterm Bivariate. Survival FGM Copula Model: An application to a Brazilian HIV Data. Journal of Data Science 10, 511-535.
Lo, S.M.S. and Wilke, R.A.(2009). A copula model for dependent competing risks. Discussion Papers in Economics No.09/01 University of Nottingham.
Lu W. and Liang Y. (2008): Analysis of competing risks data with missing cause of failure under additive hazard model. Statistica Sinica 18, 219-234.
Nelsen (2006) An introduction to Copulas, 2nd Edition, Springer, New York
National Cancer Institute DCCPS Surveillance Research Programme, Surveillance, Epidemiology and End Results (SEER) Programme (www.cancer.seer.gov) Research Data (1973-2010) (Released April 2013)
Rodrigues, J., Cancho, V.G., de Castro, M. and Louzada-Neto, F. (2009). On the unification of long-term survival models. Statistics and Probailities Letters 79, 753-759.
Roman, M., Louzada, F., Cancho, V.G. and Leite, J.G. (2012). A New Long-term Survival Distribution for cancer Data. Journal of Data Science 10, 241-258.
Sarfati D, Blakely T, Pearce N.(2010) Measuring cancer survival in populations: relative survival vs cancer-specific survival Int J Epidemiol. 39(2):598-610
Spiegelhalter, D.J., Thomas, A., Best, N. and Lunn, D. (2007) Open BUGS: User Manual, version 3.0.2 MRC Biostatistics Unit, Cambridge.
http://mathstat.helsinki.fi/openbugs/
Wienke , A. (2011). Frailty Models in Survival Analysis. Chapman and Hall, Boca Raton
Zhang, S., Zhang, Y., Chaloner,K. and Stapleton, J.T.(2010). A copula model for bivariatehybrid censored survival data with application to the MACS study. Lifetime Data Analysis 16, 231-249.