Abstract

Correlated Frailty Models in the Analysis of Dependent Failure Times

Anatoli I. Yashin
Max Planck Institute for Demographic Research
18057 Rostock, Germany
yashin@demogr.mpg.de

The multivariate survival distributions are used for description of dependent failure times (life spans) (Clayton and Cuzik 1985; Hougaard 1995, Yashin et al. 1995). The convenient representation of such distributions involves semi-parametric (copula) structure of multivariate survival function. Such structure is important in applications where the cause for dependence between life spans is the subject of the study (Yashin and Iachine 1994, Yashin and Iachine 1995a). It turns out that the wide class of frailty models allows for such a representation (Yashin et al. 1999). In genetic-epidemiological applications these models can be used to evaluate the presence and the magnitude of genetic influence on susceptibility to death as well as its variability in the population (Yashin and Iachine 1995b; 1997). In particular the correlated gamma frailty model applied to survival data on monozygotic (MZ) and dizygotic (DZ) twins can be used for this purpose. This model is characterized by the bivariate survival function $S(x_1, x_2) = P(T_1 > x_1, T_2 > x_2)$

$$S(x_1, x_2) = S(x_1)^{1-\frac{\sigma_1}{\sigma_2}\rho} S(x_1)^{1-\frac{\sigma_2}{\sigma_1}\rho} \left[S(x_1)^{-\sigma_1^2} + S(x_2)^{-\sigma_2^2} - 1\right]^{-\frac{\rho}{\sigma_1\sigma_2}} \tag{1}$$

where σ^2 and ρ are the variances and correlation coefficient of bivariate frailty distribution, $T_i, i = 1,2$ are two life spans of related individuals and $S(x)=P(T_i >x)$ is the univariate survival function. Statistically significant difference in correlation coefficients of frailty ρ_{MZ} and ρ_{DZ}, estimated for MZ and DZ twins, may indicate the presence of genetic influence on susceptibility to death (Yashin and Iachine 1995a,bc). The values of genetic parameter, such as heritability can be also estimated from this data. It is unclear, however, how the results of such analyses depend on the type of the frailty distribution used in the correlated frailty model. In this paper we derive a model of correlated frailty which includes gamma- and inverse Gaussian frailty models as particular cases. This model is based on a bivariate extension of "the three-parameter" distribution of frailty $P(z;\alpha,\delta,\theta)$ (here α,δ,θ are parameters) which was introduced in univariate survival analysis by Hougaard (1984).

Let $\mu_i(t,Z_i)=Z_i\mu_{0i}(x)$, T_i i=1,2, be conditional hazards and respective life spans for two related individuals where Z_i, i =1,2 are frailty variables and $\mu_{0i}(t)$ are respective underlying hazards. Assume that T_1 and T_2 are conditionally independent given Z, Z_2. Let Y_i i = 0,1,2 be three $P(z;\alpha,\delta_i,\theta)$-distributed independent random variables, i = 0,1,2. This distribution is characterized by the Laplace transform

$$L(s) = \exp\left\{-\frac{\beta}{\alpha}\left[(\theta+s)^\alpha - \theta^\alpha\right]\right\} \tag{2}$$

Let $Z_1 = Y_0 + Y_1$, $Z_2 = \beta(Y_0 + Y_2)$. Denote $Corr(Z_1,Z_2)=\rho$ and assume that $Var(Z_1)= Var(Z_2)= \sigma^2$. $E(Z_1)= E(Z_2)=1$. By construction

$$S(x_1, x_2 \mid Y_0, Y_1, Y_2) = \exp(-Y_0(H(x_1) + H(x_2)) - Y_1 H(x_1) - Y_2 H(x_2)) \tag{3}$$

where $H(x) = \int_0^x \mu_0(u)du$. The averaging of (3) with respect to Y_0, Y_1, Y_2 after some transformations yields the marginal bivariate survival function:

$$S\left(x_1,x_2\right)=S_1\left(x_1\right)^{1-\rho\frac{\sigma 1}{\sigma 2}}S_2\left(x_2\right)^{1-\rho\frac{\sigma 2}{\sigma 1}}$$

$$\times\exp\left\{\frac{(1-\alpha)}{\alpha}\frac{\rho}{\sigma_1\sigma_2}\left[1-\left(\left(1-\frac{\alpha}{1-\alpha}\sigma_1^2\ln S_1\left(x_1\right)\right)^{\frac{1}{\alpha}}+\left(1-\frac{\alpha}{1-\alpha}\sigma_2^2\ln S_2\left(x_2\right)\right)^{\frac{1}{\alpha}}-1\right)^{\alpha}\right]\right\} \tag{4}$$

where ρ≤min(σ₁/σ₂,σ₂/σ₁). The correlated gamma-frailty model (1) may be obtained from (4) as limiting case when α↓0. The correlated inverse Gaussian model corresponds to (4) with α=0.5. In this case

$$S(x_1,x_2)=S(x_1)^{1-\frac{\sigma_1}{\sigma_2}\rho}S(x_2)^{1-\frac{\sigma_2}{\sigma_1}\rho}\exp\left\{\frac{\rho}{\sigma_1\sigma_2}(1-((1-\sigma_1{}^2\ln S(x_1))^2+(1-\sigma_2{}^2\ln S(x_2))^2-1)^{1/2}\ )\right\} \tag{5}$$

For twins of the same sex it is convenient to assume $\sigma_1^2=\sigma_2^2=\sigma^2$, and $\mu_{01}(x)=\mu_{02}(x)=\mu_0(x)$. When $\rho=1$ (1) transforms to

$$S(x_1,x_2)=S(x_1)^{1-\frac{\sigma_1}{\sigma_2}}S(x_1)^{1-\frac{\sigma_2}{\sigma_1}}\left[S(x_1)^{-\sigma_1^2}+S(x_2)^{-\sigma_2^2}-1\right]^{-\frac{1}{\sigma_1\sigma_2}} \tag{6}$$

This is the new version of the shared gamma frailty model. When $\sigma_1^2=\sigma_2^2=\sigma^2$, (6) becomes

$$S(x_1,x_2)=\left[S(x_1)^{-\sigma^2}+S(x_2)^{-\sigma^2}-1\right]^{-\frac{1}{\sigma^2}} \tag{7}$$

This is a traditional version of the shared gamma frailty model (Oakes 1989). When $\sigma_1^2=\sigma_2^2=\sigma^2\rightarrow\infty$ and $\rho=1$, (4) reduces to the positive stable shared frailty model discussed by Hougaard (1984; 1995).

$$S(x_1,x_2)=exp\left\{-\left((-\ln S_1(x_1))^{\frac{1}{\alpha}}+(-\ln S_2(x_2))^{\frac{1}{\alpha}}\right)^{\alpha}\right\} \tag{8}$$

Similar representations may be derived in the multivariate case. For example, the general model (4) can be extended to

$$S\left(x_1,x_2,\ldots x_n\right)=\prod_{i=1}^{n}S_i\left(x_i\right)^{1-\beta\sigma_i{}^2}\exp\left\{\frac{(1-\alpha)}{\alpha}\beta\left[1-\left(\sum_{i=1}^{n}\left(1-\frac{\alpha}{1-\alpha}\sigma_i^2\ln S_i\left(x_i\right)\right)^{\frac{1}{\alpha}}-n+1\right)^{\alpha}\right]\right\} \tag{9}$$

Here β= $\dfrac{\rho_{ij}}{\sigma_i\sigma_j}$. In case of gamma-frailty the multivariate survival function *S(x₁,x₂,…,xₙ)* is:

$$S(x_1,x_2,\ldots x_n)=\prod_{i=1}^{n}S(x_i)^{1-\rho\frac{\sigma_i}{\sigma_j}}\left[\sum_{i=1}^{n}S(x_i)^{-\sigma_i^2}+n-1\right]^{-\frac{\rho_{ij}}{\sigma_i\sigma_j}} \tag{9}$$

Here ρ_{ij} are the correlation coefficients between Z_i and Z_j, and σ_i^2 are the variances of Z_i, $i,j=1,2,\ldots n;\ i\neq j$ and $\dfrac{\rho_{ij}}{\sigma_i\sigma_j}$ is, in fact, a constant not depending on i,j.

Bivariate survival functions can always be represented in the form

$$S(x_1,x_2)=S_1(x_1)S_2(x_2)e^{A(x_1,x_2)} \tag{10}$$

with $A\left(x_1,x_2\right)=\ln\dfrac{S\left(x_1,x_2\right)}{S_1\left(x_1\right)S_2\left(x_2\right)}$. If a bivariate density distribution function for *T₁*, *T₂* exists, then there is a function *φ(u,v)* such that

$$A(x_1, x_2) = \int_0^{x_1}\int_0^{x_2} \varphi(u, v)dudv \tag{11}$$

Representation (11) is called the *exponential representation* of a bivariate survival function. It turns out that in the case of random hazards, the function $\varphi(u,v)$ in (11) can be calculated using the bivariate conditional distribution of these hazards. Let $Z_i,\ i = 1{,}2$ be two random variables. We assume that the survival chances of the i^{th} individual depend on Z_i, i.e., the conditional survival function $S_i(x_i|Z_i) = P(T_i > x_i|Z_i)$ is

$$S_i(\,x_i \,|\, Z_i\,) = e^{-\int_0^x \mu_i(Z_i,u)du} \tag{12}$$

with individual hazard $\mu_i(\,Z_i\,,x),\ i = 1{,}2.$ We assume that given Z_1,Z_2 the random variables T_1,T_2 are conditionally independent. Yashin and Iachine (1995c) show that in this case

$$\begin{aligned}\varphi(x_1, x_2) &= Cov(\mu_1(Z_1, x_1), \mu_2(Z_2, x_2) \,|\, T_1 > x_1, T_2 > x_2) \\ &= E(\mu_1(Z_1, x_1)\mu_2(Z_2, x_2) - \overline{\mu}_1(x_1, x_2)\overline{\mu}_2(x_1, x_2) \,|\, T_1 > x_1, T_2 > x_2)\end{aligned} \tag{13}$$

Here $\overline{\mu}_i(u,v)$ can be represented as (14).

Remark 1. Representation (13) holds when random variables Z_1, Z_2 in the individual hazards are replaced by stochastic processes, $Z_{1t}, Z_{2t}, t \geq 0$, such that conditional mathematical expectations in (13) and (15) exist.

Remark 2. The hazards $\overline{\mu}_i(u, v), i = 1{,}2$ in (13), which are associated with the bivariate survival function $S(x_1, x_2)$ as

$$\overline{\mu}_i(x_1, x_2) = \frac{\partial}{\partial x_i}\ln S(x_1, x_2)\,,\ i{=}1{,}2 \tag{14}$$

are different from the hazards $\overline{\mu}_i(u), i = 1{,}2$

$$\overline{\mu}_i(u) = -\frac{d}{du}\ln S_i(u) = E(\mu_i(Z_i, u)|T_i > u) = \overline{\mu}_i(u{,}0),\ i = 1{,}2 \tag{15}$$

which are associated with univariate survival function $S_i(u)$. Both hazards (14) and (15) are conditional means of the same random hazard, but the respective mathematical expectations are calculated under different conditions. These survival functions have been successfully used in the statistical analysis of data on related life spans (Pickles and Crouchley 1995; Yashin and Iachine 1997; Iachine et al. 1999; Begun et al. 2000; Wienke et al. 2000). The methods of quadratic hazards are promising alternative to the multivariate frailty modeling (Aalen 1987; Yashin and Iachine 1996).

Acknowledgments

This research was partly supported by NIH/NIA grant PO1 AG08791-01.

References

Aalen, O. O. (1987). Mixed Distributions on a Markov Chain. *Scandinavian Journal of Statistics* 14, 281-289.

Begun, A. S., I. A. Iachine and A. I. Yashin (2000). Genetic Nature of Individual Frailty: Comparison of Two Approaches. *Twin Research* 3, 51-57.

Clayton, D. G. and J. Cuzick. (1985). Multivariate generalizations of the proportional hazards model (with discussion). *Journal of Royal Statistical Society* Ser. A 148, 82-117.

Hougaard, P. (1984). Life table methods for heterogeneous populations: distributions describing the heterogeneity. *Biometrika* 71, 75-83.

Hougaard, P. (1995). Frailty Models for Survival Data. *Lifetime Data Analysis* 1, 255-273.

Iachine, I. A., A. Z. Begun, M. K. Iachina, N. V. Holm, J. R. Harris, M. Latinen, J. Kaprio and A. I. Yashin (1998). How Heritable is Individual Susceptibility to Death? The Results of an Analysis of Survival Data on Danish, Swedish and Finnish Twins. *Twin Research* 1, 196-205.

Oakes, D. (1989). Bivariate Survival Models Induced by Frailties. *Journal of American Statistical Association.* 84, pp. 487-493.

Pickles, A. and R. Crouchley (1995). A Comparison of Frailty Models for Multivariate Survival Data. *Statistics in Medicine* 14, 1447-1461.

Wienke, A., K. Christensen, N. V. Holm, and A. I. Yashin (2000). Heritability of Death from Respiratory Diseases: An Analysis of Danish Twin Survival Data Using a Correlated Frailty Model. In *"Medical Infobahn for Europe".* Proceedings of MIE2000 and GMDS2000. A. Hasman, B. Blobel, J. Dudeck, G. Gell, H.-U.Prockosch (Eds), 407-411, IOS Press, 2000, Amsterdam.

Yashin, A. I. and I. A. Iachine (1994). Mortality Models with Application to Twin Survival Data. CISS - First Joint Conference of International Simulation Societies Proceedings. Edited by Jürgen Halin, Walter Karplus and Rainer Rimane. August 22-25 (1994), ETH, Zurich, Switzerland (1994), pp.567-571.

Yashin, A. I., J. W. Vaupel, and I. A. Iachine (1995). Correlated Individual Frailty: An Advantageous Approach to Survival Analysis of Bivariate Data. *Mathematical Population Studies* 5, 145-159.

Yashin, A. I. and I. A. Iachine (1995a). How long can humans live? Lower bound for biological limit of human longevity calculated from Danish twin data using correlated frailty model. *Mechanisms of Ageing and Development* 80, 147-169.

Yashin, A. I. and I. A. Iachine (1995b). Genetic analysis of durations: Correlated frailty model applied to survival of Danish Twins. *Genetic Epidemiology* 12, 529-538.

Yashin, A. I. and I. A. Iachine (1995c). Survival of related individuals: an extension of some fundamental results of heterogeneity analysis. *Mathematical Population Studies* 5, 321-339.

Yashin, A. I. and I. A. Iachine (1996a). Random effect models of bivariate survival: Quadratic hazard as a new alternative. In *Symposium in Anvendt Statistik* (Kristensen G, ed.). Odense, Denmark: Odense University, pp. 87-101.

Yashin, A. I. and I. A. Iachine (1997). How Frailty Models Can be Used in the Analysis of Mortality And Longevity Limits. *Demography* 34, 31-48.

Yashin, A. I., A. S. Begun I. A. Iachine (1999). Genetic Factors in Susceptibility to Death: Comparative Analysis of Bivariate Survival Models. *Journal of Epidemiology and Biostatistics* 4(1), 53-60.

Yashin, I. A. and I. A. Iachine (1999a). What Difference Does the Dependence Between Durations Make? *Life Time Data Analysis* 5, 5-22.

Yashin, A. I. and I. A. Iachine (1999b). Dependent Hazards in the Problem of Multivariate Survival. *Journal of Multivariate Analysis* 71, 241-261.