Joint Modeling of Repeated Measures and Survival Time Data

Il Do Ha  
*Kyungsan University, Department of Statistics, 712-240, Kyungsan, Korea*  
idha@kyungsan.ac.kr

Taesung Park and Youngjo Lee  
*Seoul National University, Department of Statistics, 151-742, Seoul, Korea*  
tspark@stats.snu.ac.kr, youngjo@plaza.snu.ac.kr

1. Introduction

Recently, many long-term clinical trials collect both repeated measures data at a number of time points and survival time data at which times to terminating events are recorded: see Hogan and Laird (1997). For example, after a kidney transplantation clinicians usually monitor the patient’s serum creatinine level, an indicator of kidney function, at many time points. However, a single terminating event may occur, for instance, failure of the kidney graft or death of the patient: see Sung et al. (1998). Thus, a patient has both repeated measures data and survival time data, so that these data may induce a dependence due to a shared random effect such as the patient’s genetic or environmental effect.

In this presentation we introduce a random effect model for the joint analysis of combined repeated and survival time data of this form, and then propose an inferential method using hierarchical-likelihood (Lee and Nelder, 1996; Ha et al., 2001). Finally, our method is illustrated with kidney transplantation data presented by Sung et al. (1998).

2. The model

We consider the situation that for the $i$th ($i = 1, \cdots, q$) patient there are both repeated measures responses $y_{ij}$ ($j = 1, \cdots, n_i$) and survival time $T_i$. Let $y_i = (y_{i1}, \cdots, y_{in_i})^T$ denote the $n_i \times 1$ vector of the $i$th responses among $y_{ij}$’s, $X_i^{(1)} = (x_{i1}^{(1)}, \cdots, x_{in_i}^{(1)})^T$ the $n_i \times p_1$ matrix of covariates corresponding to $y_i$, where $x_{ij}^{(1)} = (x_{ij1}^{(1)}, \cdots, x_{ijp_1}^{(1)})^T$ is the $p_1 \times 1$ covariate vector corresponding to $y_{ij}$, and $x_i^{(2)} = (x_{i1}^{(2)}, \cdots, x_{i}^{(2)})^T$ the $p_2 \times 1$ covariate vector corresponding to $T_i$. Let $C_i$ be the censoring time corresponding to $T_i$, and $t_i^* = \min(T_i, C_i)$ and $\delta_i = I(T_i \leq C_i)$, where $I(\cdot)$ is the indicator function. Denote by $v_i$ a shared random effect for the $i$th patient. In this paper, we assume that $y_i$ and $T_i$ are conditionally independent given $v_i$, and that $T_i$ and $C_i$ are also conditionally independent given $v_i$. 
The $y_i$ and $T_i$ ($i = 1, \cdots, q$) are assumed to have the following models:

(i) $y_i | v_i \sim N(\mu_i^{(1)}, \phi_{n_i \times n_i})$ with $\mu_i^{(1)} = X_i^{(1)} \beta^{(1)} + z_i \gamma_1 v_i$,
where $I_{n_i \times n_i}$ is the $n_i \times n_i$ identity matrix, $\beta^{(1)}$ is a $p_1 \times 1$ vector of regression parameters, $z_i$ is a $n_i \times 1$ vector with one, and $\gamma_1$ is the real-valued parameter which describes the random effect, say $v_{i1} = \gamma_1 v_i$, of $y_i$.

(ii) $T_i | v_i$ has the conditional hazard function

$$\lambda_i(t | v_i) = \lambda_0(t) \exp(x_i^{(2)T} \beta^{(2)} + \gamma_2 v_i),$$

where $\lambda_0(\cdot)$ is a parametric (e.g. Weibull) or nonparametric baseline hazard function, $\beta^{(2)}$ is a $p_2 \times 1$ vector of regression parameters, and $\gamma_2$ is the real-valued parameter which describes the random effect, say $v_{i2} = \gamma_2 v_i$, of $T_i$.

(iii) $v_i \sim N(0, 1)$.

The above model is the joint of mixed linear model and frailty model. The random effect $v_i$ models the dependence between $y_i$ and $T_i$, and $\gamma_1$ and $\gamma_2$ are parameters describing the dependence. That is, if the two parameters take the same sign, then $y_i$ and $T_i$ have a positive correlation; otherwise they have a negative correlation.

REFERENCES

