Statistics for Biology and Health
Chapter 11 Multivariate Survival Analysis

Qi Guo

July 27,2019

THE UNIVERSITY OF TEXAS AT DALLAS
School of Natural Sciences and Mathematics

Contents

The accelerated failure time model

Introduction

- Let X denote the time to the event and Z a vector of fixed time explanatory covariates.
- The accelerated failure time model states that the survival function of an individual with covariate Z at time x is the same as the survival function of an individual with a baseline survival function at a time $x\exp(\theta' z)$, where $\theta' = (\theta_1, \cdots, \theta_p)$ is a vector of regression coefficients.
- The accelerated failure time model is defined by the relationship:

$$S(x|Z) = S_0[\exp(\theta' Z)x], \ \textit{for all } x$$

The accelerated failure time model

- The factor $\exp(\theta' Z)$ is called an accelerating factor telling the investigator how a change in covariate values changes the time scale from the baseline time scale.
- One application of this model is that the hazard rate for an individual with covariate Z is related to a baseline hazard by:

$$h(x|Z) = \exp(\theta' Z) h_0[\exp(\theta' Z) x], \; \textit{for all } x$$

- The second implication is that median time to the event with a covariate Z is the baseline median time to event divided by the acceleration factor.
- About the parametric regression model, we can look at the notes before.

The frailty model

The frailty model

- Before the analysis based on the assumption that the survival times of distinct individuals are independent of each other, but it is quite probable that there is some association within groups of survival times in some sample.
- A model for modeling association between individual survival times within subgroups is the use of a frailty model. A frailty is an unobservable random effect shared by subjects within a subgroup.
- The most common model for the frailty is a common random effect that acts multiplicatively on the hazard rates of all subgroup members. In this model, families with a large value of the frailty will experience the event at earlier times then families with small values of the random effect.

The frailty model

- The most common model for a frailty is the so-called shared frailty model extension of the proportional hazards regression model.
- Assume that the hazard rate for the jth subject in the ith subgroup, given the frailty, is of the form

$$h_{ij}(t) = h_0(t)\exp(\sigma\omega_i + \beta' Z_{ij}), i = 1, \cdots, G, j = 1, \cdots, n_i$$

where $h_0(t)$ is an arbitrary baseline hazard rate, Z_{ij} is the vector of covariates, β the vector of regression coefficients, and $\omega_1, \cdots, \omega_G$ the frailties.

The frailty model

- Here assume that ω's are an independent sample from some distribution with mean 0 and variance 1. In some applications, it's more convenient to write the model as:

$$h_{ij}(t) = h_0(t)\mu_i \exp(\beta' Z_{ij}), i = 1, \cdots, G, j = 1, \cdots, n_i$$

  where the μ_i's are an independent and identically distributed sample from a distribution with mean 1 and some unknown variance.
- When nature picks a value of μ_i greater than 1, it means individuals within a given family tend to fail at a rate faster than under an independence model where μ_i equal to 1 with probability 1.

The frailty model

- When we collect data, the ω_i's are not observable, so the joint distribution of the survival times of individuals within a group, found by taking the expectation of $\exp[-\sum_{j=1}^{n_i} H_{ij}(t)]$ with respect to μ_i is given by:

$$\begin{aligned} S(x_{i1}, \cdots, x_{in_i}) &= P[X_{i1} > x_{i1}, \cdots, X_{in_i} > x_{in_i}] \\ &= LP[\sum_{j=1}^{n_i} H_0(x_{ij}) \exp(\beta' Z_{ij})] \end{aligned}$$

Here, $LP(\nu) = E_U[\exp(-U\nu)]$ is the Laplace transform of the frailty U.

Score Test for Association

Score Test for Association

- To test for association between subgroups of individuals, a score test has been developed by Commenges and Andersen.The test can be used to test for association, after an adjustment for covariate effects has been made using Cox's proportional hazards model, or can be applied when there are no covariates present.
- Here we have G subgroups of subjects, and within the ith subgroup, we have n_i individuals.
- We are interested in testing the $H_0 : \sigma = 0$ and $H_A ; \sigma \neq 0$, no distributional assumptions are made about the distribution of the random effects ω_i.

Test statistics

- The test statistics consists of the on study time T_{ij}, the event indicator δ_{ij}, and the covariate vector Z_{ij} for the jth observation in the ith group, and the at risk indicator $Y_{ij}(t)$ at time t for an observation.
- Then, we fit the Cox proportional hazards model, $h(t|Z_{ij}) = h_0(t)\exp(\beta' Z_{ij})$, and we obtain the partial maximum likelihood estimates b of β and the Breslow estimate $\hat{H}_0(t)$ of the cumulative baseline hazard rate. Let

$$S^{(0)}(t) = \sum_{i=1}^{G}\sum_{j=1}^{n_i} Y_{ij}(t)\exp(b' Z_{ij})$$

Test statistics

- Then compute $M_{ij} = \delta_{ij} - \hat{H}_0(T_{ij})\exp(b'Z_{ij})$, the martingale residual for the ijth individual. The score statistic is given by:

$$T = \sum_{i=1}^{G} \left\{ \sum_{j=1}^{n_i} M_{ij} \right\}^2 - D + C$$

where D is the total number of deaths and:

$$C = \sum_{i=1}^{G} \sum_{j=1}^{n_i} \frac{\delta_{ij}}{S^{(0)}(T_{ij})^2} \sum_{h=1}^{G} \Big[\sum_{k=1}^{n_i} Y_{hk}(T_{ij}) \exp(b'Z_{hk}) \Big]^2$$

- The test statistic can also be written as:

$$T = \sum_{i=1}^{G} \sum_{j=1}^{n_i} \sum_{k=1, k \neq j}^{n_j} M_{ij} M_{ik} + \Big(\sum_{i=1}^{G} \sum_{j=1}^{n_i} M_{ij}^2 - N \Big) + C$$

where N is the total sample size.

Estimation for the Gamma Frailty Model

Introduction

- Now we estimate the risk coefficients, baseline hazard rate, and frailty parameter for a frailty model based on a gamma distributed frailty.
- Assume that for the jth individual in the ith group, the hazard rate given the unobservable frailty μ_i is of the from $h_{ij}(t) = h_0(t)\mu_i \exp(\beta' Z_{ij})$ with μ_i i.i.d of gamma distribution which is:

$$g(\mu) = \frac{\mu^{(1/\theta - 1)\exp(-\mu/\theta)}}{\Gamma[1/\theta]\theta^{1/\theta}}$$

  where the mean is 1 and the variance is θ, so that large values of θ reflect a greater degree of heterogeneity among groups and a stronger association within groups.

The estimation for the Gamma Frailty Model

- The joint survival function for the n_i individuals within the ith group is given by:

$$
\begin{aligned}
S\{x_{i1}, \cdots, x_{in_i}\} &= P[X_{i1} > x_{i1}, \cdots, X_{in_i} > x_{in_i}] \\
&= \left[1 + \theta \sum_{j=1}^{n_i} H_0(x_{ij}) \exp(\beta' Z_{ij})\right]^{-1/\theta}
\end{aligned}
$$

- The association between group members as measured by Kendall's τ is $\theta/(\theta+2)$, and $\theta = 0$ corresponds to the case of independence.

The estimation for the Gamma Frailty Model

- Estimation for this model is based on the log likelihood function, the usual triple $(T_{ij}, \delta_{ij}, Z_{ij}),\ i = 1, \cdots, G, j = 1, \cdots, n_i$. Let $D_i = \sum_{j=1}^{n_i} \delta_{ij}$ be the number of events in the ith group. Then the observable log likelihood is given by:

$$
\begin{aligned}
L(\theta, \beta) = \sum_{i=1}^{G} & D_i \ln \theta - \ln[\Gamma(1/\theta)] + \ln[\Gamma(1/\theta + D_i)] \\
& - (1/\theta + D_i) \ln \left[1 + \theta \sum_{j=1}^{n_i} H_0(T_{ij} \exp(\beta' Z_{ij}))\right] \\
& + \sum_{j=1}^{n_i} \delta_{ij} \{\beta' Z_{ij} + \ln[h_0(T_{ij})]\}
\end{aligned}
$$

- If one assumes a parametric form for $h_0()$, the m.l.e can be used.

EM algorithm

- The EM algorithm provides a means of maximizing complex likelihoods.
- In the E (or Estimation) step of the algorithm the expected value of L_{FULL} is computed, given the current estimates of the parameters and the observable data.
- In the M (or maximization) step of the algorithm, estimates of the parameters which maximize the expected value of L_{FULL} from the E step are obtained. The algorithm iterates between these two steps until convergence.

EM algorithm

- To apply the E-step to our problem, given the data and the current estimates of the parameters, the μ_i's are independent gamma random variables with shape parameters $A_i = [1/\theta + D_i]$ and scale parameters $C_i = [1/\theta + \sum_{i=1}^{G} H_0(T_{ij}\exp(\beta' Z_{ij}))]$.Thus

$$E[\mu_i|Data] = \frac{A_i}{C_i} \text{ and } E[\ln \mu_i] = [\psi(A_i) - \ln C_i]$$

where $\psi()$ is the digamma function. Substituting these values in $L_1(\theta)$ and $L_2(\beta, H_0)$ complete the E-step of the algorithm.

EM algorithm

- For the M-step, note that $E[L_2(\beta, H_0)|Data]$ is expressed as:

$$E[L_2(\beta, H_0)|Data] = \sum_{i=1}^{G} \sum_{j=1}^{n_i} \delta_{ij}[\beta' Z_{ij} + \ln h_0(T_{ij})] - \frac{A_i}{C_i} H_0(T_{ij}) \exp(\beta' Z_{ij})$$

  which depends on $h_0()$.
- If we let $t_{(k)}$ be the kth smallest death time, regardless of subgroup, and $m_{(k)}$ the number of deaths at time $t_{(k)}$, for $k = 1, \cdots, D$ and denote bt $\hat{\mu}_h$ and Z_h the expected value of frailty and the covariate vector for hth individual, the partial likelihood to be maximized in the M step is

$$L_3(\beta) = \sum_{k=1}^{D} \left\{ S_{(k)} - m_{(k)} \ln \left[\sum_{h \in R(T_{(k)})} \hat{\mu}_h \exp(\beta' Z_h) \right] \right\}$$

EM algorithm

- where $S_{(k)}$ is the sum of the covariates of individuals who died at time $t_{(k)}$.
- An estimate of the $H_0(t)$ from this step is given by:

$$\hat{H}_0(t) = \sum_{t_{(k)} \leq t} h_{k0}$$

where

$$h_{k0} = \frac{m_{(k)}}{\sum_{h \in R(t_{(k)})} \hat{\mu}_h \exp(\beta' Z_h)}$$

EM algorithm

Here we use a grid of possible values for the frailty parameter θ. For each fixed value of θ, the following EM algorithm is used to obtain an estimate of β_θ

- Step 0. Provide an initial estimate of β(and hence h_{k0})
- Step 1.(E step) Compute $\hat{\mu}_h, h = 1, \cdots, n$ based on the current values of the parameters.
- Step 2.(M step) Update the estimate of β_θ(and the h_{k0}) using the partial likelihood in $L_3(\beta)$.
- Step 3. Iterate between steps 1 and 2 until convergence.