Comparision of two estimators for the marginal proportional intensity model for clustered event history data

Franz Quehenberger
Department of Medical Informatics, Statistics, and Documentation, Graz

Abstract: The Cox partial likelihood estimator still remains consistent when the subjects are correlated due to clusters. Only the variance estimate has to be adapted. Its properties are similar to an estimate that is based on a U-statistic. Application is demonstrated on a dental data set.

Keywords: marginal hazard, proportional intensity model, correlated data, U-statistics.

1 Introduction

Multivariate survival data arise in medical studies when each subject may experience several events or when subjects are grouped. Examples are the onset of blindness in a study of diabetic retinopathy, the durability of dental crowns within a set of teeth, the onset of disease within a family, or the onset of smoking within the pupils of different classes.

The clustering of subjects can be handled by stratification, mixture models (which are called frailty models in event history data analysis), and by marginal models. In the last approach the covariate vector only contains information about the subject, ignoring information from other subjects in the cluster. This interpretation of covariate effect is natural if statements about the population are intended.

Here two estimators for the multiplicative intensity model will be compared. The risk of failure of an individual is factorised in a baseline hazard function $\lambda_0(t)$ and a function that models the action of the covariates, which is mostly chosen to be $\exp\{\beta_0^\top Z(t)\}$ (where β_0 is a vector of regression coefficients).

The Cox partial likelihood does not require the estimation of λ_0 and is therefore called semiparametric. It was pointed out that the estimator is consistent even if the data are correlated, however the covariance estimates for $\hat{\beta}$ must be of sandwich type (Lee, Wei, and Amato 1992) . The pseudolikelihood approach of Liang, Self, and Chang (1993) is even more nonparametric. The two approaches will be compared in a small simulation study. The application will be demonstrated on a dental data set.

2 The model

The data are represented as $(N_{ij}(u), Z_{ij}(u), Y_{ij}(u), 0 \leq u \leq \infty)$. The subscript i indexes the n clusters, j indicates the K subjects within a cluster. N_{ij} is the realization of a counting process, i.e. it counts the number of events that have occured up to time u. $Z_{ij}(u)$ is a time-dependent vector of covariates and $Y_{ij}(u)$ indicates if a subject is at risk. Unequal cluster sizes are handled by filling up with individuals with $Y_{ij}(\cdot) = 0$, i.e. the subject is never at risk. This is a straightforward generalization of the usual survival analysis that is characterized by the indicator function Y set to 0 strictly after an event has occured.

The event history of an individual is modeled as counting process $N_{ij}(u)$ with intensity with respect to the filtration that was induced by the information on individual ij (Andersen et al. 1993) given by

$$\lambda_{ij}(t) = Y_{ij}(t) \exp\{\beta_0^\top Z_{ij}(t)\} \lambda_0(t). \tag{1}$$

3 The semiparametric estimate

Ignoring the cluster structure and using the notation

$$S^{(d)}(\beta, t) = \sum_{i=1}^{n} \sum_{j=1}^{K} Y_{ij}(t) \exp\{\beta_0^\top Z_{ij}(t)\} Z_{ij}(t)^{\otimes d},$$

$d = 0, 1, 2$, $a^{\otimes d}$ is 1, a and $a^\top a$, $E(\beta, t) = S_1(\beta; t)/S_0(\beta; t)$ following Andersen et al (1993), p. 481 ff. the score vector, defined as the vector of partial derivatives of the log (partial) likelihood, is shown to be

$$U(\beta) = \sum_{i=1}^{n} \sum_{j=1}^{K} \int_0^\infty \left[Z_{ij}(u) - E(\beta, u)\right] \, dN_{ij}(u) \tag{2}$$

An estimate of β is obtained as the root of this function. Under certain regularity conditions the asymptotic distribution of $n^{1/2}\hat{\beta}$ is normal with mean β_0 and a variance matrix that can be estimated by $nA^{-1}BA^{-1}$. Here A is the matrix of second derivatives of the log partial likelihood and $B = \sum_{i=1}^{n} \hat{w}_i^{\otimes 2}$ with $\hat{w}_i = \sum_{j=1}^{K} \int_0^\infty \left\{ Z_{ij}(t) - E(\hat{\beta}, t) \right\} \left\{ dN_{ij}(t) - Y_{ij}(t) \exp\{\beta_0^\top Z_{ij}(t)\} \, d\hat{\Lambda}_0(t) \right\}$ whereby $\hat{\Lambda}_0(t) = n^{-1} \int_0^t S^{(0)}(u)^{-1} \, dN(u)$ and $N(u) = \sum_{i=1}^{n} \sum_{j=1}^{K} N_{ij}(u)$.

These results were established by (Lee, Wei, and Amato 1992) . They are based on the fact that the $\hat{w}_i$ fulfill the conditions of a central limit theorem. The martingale structure on which the asymptotic results of the proportional intensity model rely is destroyed by the cluster structure of the data. But it can still be exploited approximately by a generalization of the Slutsky lemma to stochastic processes.

4 The pseudolikelihood estimate

The individuals of different clusters are independent. Liang, Self, and Chang (1993) suggested an estimating equation that combines pairwise likelihood elements of Cox type of individuals from different clusters that result in pairwise scores of type (2). From this so called pseudolikelihood the following score function is derived

$$S_{PL}(\beta) = \sum_{\substack{i,j=1 \\ i \neq j}}^{n} \sum_{l,k=1}^{K} \int_0^{\infty} W_i^{-1}(u) \left[Z_{ij}(u) - E_{ij;lk}(\beta, u) \right] \, dN_{ij}(u) \tag{3}$$

The weight function $W_i(u)$ should account for an undue number of nonzero contributions of early events. The authors suggested the number of individuals in the risk set without the individuals in the i-th cluster. $E_{ij;lk}(\beta, u)$ corresponds tho the Function $E(\beta, u)$ in (2)

In (3) the sum of individuals from cluster i and j can be interpreted as the kernel of a U-statistic of degree 2. As it is a sum of unbiased equations, the estimating equation is unbiased . Its distribution is asymptotically normal by the central limit theorem for U-statistics. Its variance can be estimated from the jack-knife estimate of variance for U-statistics. The asymptotic variance of the parameter estimate $\hat{\beta}$ is again of sandwich type.

Although the kernel of the U-statistic representation of $S_{PL}(\beta)$ still depends on the entire sample through the weight function the asymptotic results are still valid.

Liang et al. went on to develop an asymptotically equivalent estimating equation with a kernel of degree 3 (method *Liang3*). It can be shown that it is algebraically identical to a U-statistic with kernel of degree 2 (method *Liang2*). The latter kernel is the same as in (3) except for the weight function that is only asymptotically equivalent. The fact that the somewhat cumbersome estimating equation can be simplified is attractive for computing, however the jackknife estimates of variance still differ. It was also pointed out by Liang et al. that for the jackknife estimator of variance of U-statistics with kernel of degree greater than 2 a finite sample correction should be applied.

5 Simulation study

Although both Lee, Wei, and Amato (1992) and Liang, Self, and Chang (1993) made simulations to check the finite sample performance of their estimates, these estimates were not compared with each other. Furthermore there was an error in the simulations of the latter.

One of the simulations that were performed had exactly the same design as in Liang, Self, and Chang (1993) with the exception of the random vector of covariates. 50 pairs of data from a gamma frailty model were generated. Here higher values of the frailty parameter θ correspond to more correlation, no correlation means

		$\theta = 1$				$\theta = 3$			
Design	Method	$\hat{\beta}$	Var $\hat{\beta}$	$\widehat{\text{Var}}\, \beta$	Cov. 95%	$\hat{\beta}$	Var $\hat{\beta}$	$\widehat{\text{Var}}\, \beta$	Cov. 95%
indep.	Cox72	0.011	0.138	0.156	4.2	0.010	0.162	0.156	5.2
	Lee93	0.011	0.138	0.152	4.7	0.010	0.162	0.154	6.3
	PL	0.008	0.144	0.162	4.0	0.006	0.167	0.168	5.6
	Liang3	0.014	0.148	0.169	4.7	0.016	0.165	0.177	6.2
	Liang2	0.014	0.147	0.169	4.6	0.016	0.165	0.176	6.1
dep.	Cox72	0.002	0.155	0.141	6.1	0.005	0.236	0.142	12.1
	Lee93	0.002	0.155	0.133	8.3	0.005	0.236	0.238	6.1
	PL	0.004	0.153	0.139	6.7	0.007	0.238	0.255	5.1
	Liang3	-0.008	0.151	0.151	7.0	0.008	0.239	0.274	6.0
	Liang2	-0.008	0.153	0.151	6.8	0.008	0.242	0.272	5.7
inverse	Cox72	-0.001	0.146	0.138	5.4	0.010	0.034	0.138	0.0
	Lee93	-0.001	0.146	0.133	6.1	0.010	0.034	0.035	4.9
	PL	-0.002	0.155	0.146	5.7	0.010	0.037	0.040	4.7
	Liang3	-0.002	0.141	0.145	6.0	0.012	0.033	0.035	5.3
	Liang2	-0.002	0.142	0.145	6.0	0.012	0.032	0.037	4.3

Table 1: Simulation results

$\theta = 1$. 15% censoring occured at fixed times. The one and only covariate was time-independent and came from a uniform distribution. Z_{i2} was either independent from Z_{i1}, equal to Z_{i1} (dependent design) or $1 - Z_{i1}$ (inversely dependent design). 1000 data sets were generated.

The bias of the estimates is satisfactory except in the inversely independent design, where the bias is more than 30% of the variance. The bias of the variance estimate is considerable for the Cox72 estimate when correlation is present. The Liang 2 and the Liang3 estimate are almost identical.

6 Example

In a retrospective cohort study the durability of 652 telescopic crowns within 196 patients of the local dental clinic was investigated. The prognostic factors in consideration were jaw (upper=1, lower=0), kind of tooth (molar, premolar, canine, incisive), kind of prostheses (casting=1, plastic=2), crowns per individual, age and indicators of dental care such as usage of tooth paste, tooth brush or hydrogen peroxide. Only 52 crowns had failed with a median observation time of 4 years. There were up to 10 crowns per dentition with median 3 and mode 2. The 5-year and 8-year duration rate were 0.952 and 0.915 (Breslow estimate of final model).

Teeth were assumed to be individuals that are clustered within a dentition. A marginal proportional hazards model was fitted for time to failure.

The molars had no failure. In such cases the estimating equations have no

	Cox72			Lee93			PL			Liang3		
jaw	1.58	±	0.43	1.58	±	0.41	1.62	±	0.54	1.62	±	0.51
prostheses	-0.73	±	0.31	-0.73	±	0.38	-0.78	±	0.42	-0.83	±	0.39
incisive	1.46	±	0.40	1.46	±	0.35	1.73	±	0.44	1.73	±	0.41
H_2O_2	-2.80	±	1.01	-2.80	±	1.01	-2.73	±	1.04	-2.74	±	1.03

Table 2: Estimates from dental example

roots (Jacobsen 1989) and therefore molars were pooled with the premolars. A reasonable set of covariates was selected by backward selection (generalized Wald test). Cluster size was not related to failure.

It is expected that the standard errors of covariates that are constant within clusters are overestimated in the Cox model and covariates that vary between clusters are underestimated. Covariates that are only partly constant within clusters should remain more or less unchanged. Jaw is of the last type, because some patients had got crowns on both jaws and others had not. The type of prostheses did not change within a patient and the bias of the standard error in the Cox model is in the expected direction. The estimate for incisive behaves similar, but the other way round. The standard error of hydrogen peroxide remains unchanged in contradiction to theory. All estimates of standard error of pseudolikelihood estimate were larger than in the marginal Cox model. There was also a tendency towards larger parameter estimates in the former.

References

Andersen, P. K., Borgan, Ø., Gill, R. D., and Keiding, N. (1993) . *Statistical models based on counting processes.* New York: Springer.

Jacobsen, M. (1989). Existence and unicity of MLEs in discrete exponential family distributions. *Scandinavian Journal of Statistics* **16**, 335–349.

Liang, K. Y., Self, S. G., and Chang, Y. C.(1993). Modeling marginal hazards in multivariate failure time data. *Journal of the Royal Statistical Society Series B* **55**, 441–453.

Lee, E. W., Wei, L. J., and Amato, D. A. (1992). Cox-type tegression analysis for large numbers of small groups of correlated failure time observations. In: *Survival Analysis: State of the Art* (J. P. Klein, P. K. Goel, Eds.), 237–247. Dordrecht: Kluwer Academic Publishers.