Advanced Engineering Forum *Online: 2012-09-26*
ISSN: 2234-991X, Vols. 6-7, pp 93-96
doi:10.4028/www.scientific.net/AEF.6-7.93

Additive-Accelerated Mean Regression Model for Multiple Type Recurrent Events

Huanbin Liu

College of Mathematics and Computer Sciences,

Huanggang Normal University, Hubei 438000, China

lhb@hgnu.edu.cn

Keywords: Recurrent events data; additive-accelerated mean regression model; parameter estimation

Abstract. Recurrent events data is often observed in applied research fields like biostatistics, clinical experiment, and so on. In this paper, an additive-accelerated mean regression model is established for multiple type recurrent events data, and the estimation methods of unknown parameter and non-parameter function based on the idea of estimating equation are given.

Introduction

Recurrent events data refers to the reoccurrence time sequence of interested events observed for individuals[1-4]. If only one type of resulted data is concerned, it is referred to as single type recurrent events data. Examples are the recurrent time sequence of acute coronary heart disease and machine faults. Recurrent event data is a large class of important incomplete data existed in survival analysis, biological medicine research, reliability life test and other practical problems, and the statistical analysis for them has been valued all over the world, especially by developed countries[5-6].

Complex statistical analysis of recurrent event data is focal point of research of modern statistics and the important part of the development of various disciplines. Analyzing complex data, establishing the corresponding statistical model, and revealing the internal laws of complex data are the important foundation of their relevant disciplines. Especially in the research on biology, medicine, ecology, demography, environmentology and economics and other disciplines, with the development of experimental techniques, testing methods and means of data analysis, the data obtained are more and more complex and precise in structure, and the information provided is more and more miscellaneous, which put forward higher requirements for the quantitative analysis of data[7-9]. How to make statistical modeling and statistical inference has become the frontier topic of biology, medicine, ecology, demography, environmentology and economics and other interdiscipline. In this research field, there still exist some problems to be solved by developing effective statistical methods[10-13].

This paper discusses the additive-accelerated mean regression model for multiple type recurrent events data, puts forward an estimation method of unknown parameter and non-parameter function based on the idea of estimating equation.

Model construction and estimation method

To describe the observation data of multiple type recurrent events, suppose there are n individuals to be observed during an observation period, each individual experiences k different types of recurrent events, and they are mutually independent. Let T_{ikj} represent the occurrence time of the kth-type, jth-time observed event of the ith individual after the experiment begins, and $j = 1,2,\cdots,l_k$, where $i = 1,2,\cdots,n$, $k = 1,2,\cdots,K$.

Let $V_{ik}^{*}(t)$ be the number of recurrence of the kth type event of the ith individual at time t, Definition (1) is given as follows:

$$N_{ik}^{*}(t)=\sum_{j=1}^{n} I(T_{ikj}\le t), \tag{1}$$

I(.) is an indicative function. At the same time, we assume the counting process $N_{ik}^{*}(t)$ is related to the covariant $X_{ik}, Z_{ik}, W_{ik}(t)$ with p1, p2 and p3 dimension respectively, and denote as:

$W_{ik}^{*}(t)=\int_0^t W_{ik}(s)ds$,

The following additive-accelerated mean regression model is suggested to be adopted here.

$$E[N_{ik}^{*}(t)\big| X_{ik}, Z_{ik}, W_{ik}(t)]=\mu_0(te^{\beta_0^{'}X_{ik}})g(\gamma_0^{'}Z_{ik})+\alpha_0^{'}W_{ik}^{*}(t). \tag{2}$$

where β_0, γ_0 and $\alpha_0^{'}$ are the unknown regression parameter vector of the p1, p2 and p3 dimension respectively, and $\mu_0(\cdot)$ is the unknown benchmark mean continuous function. Denote $\widetilde{T}_{ikj}=T_{ikj}e^{\beta^{'}X_{ik}}$, then the counting process is as follows:

$\widetilde{N}_{ik}^{*}(t;\beta)=\sum_{j=1}^{\infty} I(\widetilde{T}_{ikj}\le t)$

$=\sum_{j=1}^{\infty} I(\widetilde{T}_{ikj}\le te^{-\beta^{'}X_{ik}})$

$=\widetilde{N}_{ik}^{*}(te^{-\beta^{'}X_{ik}})$

and model (2-2) can be expressed as the following form:

$$E[N_{ik}^{*}(t;\beta_0])=\mu_0(t)g(\gamma_0^{'}Z_{ik})+\alpha_0^{'}W_{ik}^{*}(te^{-\beta_0^{'}X_{ik}}). \tag{3}$$

In many practical applications, individuals are always observed within a limited period, thus $N_{ik}^{*}(t)$ can not be observed completely.

Denote C_{ik} be the censored time of the kth type event of the ith individual, and it is independent from the condition $N_{ik}^{*}(t)$, then according to Anderson and Gill (1982), the number of observed events $N_{ik}^{*}(t)$ within the observation period for the kth type event of the ith individual can be defined as follows:

$$N_{ik}(t)=\sum_{j=1}^{\infty} I(T_{ikj}\le t\wedge C_{ik} \tag{4}$$

where $a\wedge b=\min(a,b)$. Based on the above hypothesis model, the time scale model of transformation is:

$$\widetilde{N}_{ik}(t;\beta)=\sum_{j=1}^{\infty} I(\widetilde{T}_{ikj}\le t\wedge \widetilde{C}_{ik}(\beta)) \tag{5}$$

where $\widetilde{C}_{ik}(\beta)=C_{ik}e^{\beta^{'}X_{ik}}$. Meanwhile define

$Y_{ik}(t;\beta)=I(C_{ik}\ge te^{-\beta^{'}X_{ik}})$,

$\theta=(\beta^{'},\gamma^{'},\alpha^{'})$,

$\theta_0=(\beta_0^{'},\gamma_0^{'},\alpha_0^{'})$, and also define the following process:

$$M_{ik}(t;\theta) = N_{ik}(s;\beta) - \int_0^t Y_{ik}(s;\beta)[g(\gamma' Z_{ik})d\mu_0(s) + e^{-\beta' X_{ik}\alpha'} W_{ik}\alpha'(se^{-\beta' X_{ik}})ds]. \quad (6)$$

Obviously,

$$\widetilde{N}_{ik}(t;\beta) = \int_0^t Y_{ik}(s;\beta)d\widetilde{N}_{ik}(s;\beta),$$

$$M_{ik}(t;\theta) = \int_0^t Y_{ik}(s;\beta)d\widetilde{N}_{ik}(s;\beta) - \mu_0(s)g(\gamma' Z_{ik}) - \alpha' W_{ik}^*(se^{-\beta' X_{ik}})e^{-\beta' X_{ik}})\};$$

According to (3), $M_{ik}(t;\theta_0)$ $(i = 1,2,\cdots,n; k = 1,2,\cdots,K)$ is a stochastic process with zero mean value, therefore, for given $\theta = (\beta', \gamma', \alpha')$, a natural estimation of $\mu_0(t)$ is the solution of the following equation.

$$\sum_{i=1}^{n}\sum_{k=1}^{K} \widetilde{N}_{ik}(t;\beta) - \int_0^t Y_{ik}(s;\beta)[g(\gamma' Z_{ik})d\mu_0(s) + e^{-\beta' X_{ik}\alpha'} W_{ik}\alpha'(se^{-\beta' X_{ik}})ds]\} = 0 \quad (7)$$

where $0 \le t \le \tau$, τ is a given constant that makes $P(C_{ik} \ge \tau e^{-\beta_0' X_{ik}} > 0$.

Denote the estimation of $\mu_0(t)$ as $\hat{\mu}_0(t;\theta)$, then

$$\hat{\mu}_0(t;\theta) = \sum_{i=1}^{n}\int_0^t \frac{\sum_{k=1}^{K} d\widetilde{N}_{ik}(t;\beta) - \sum_{k=1}^{K} Y_{ik}(s;\beta)e^{-\beta' X_{ik}\alpha'} W_{ik}\alpha'(se^{-\beta' X_{ik}})ds}{\sum_{i=1}^{n}\sum_{k=1}^{K} Y_{jl}(s;\beta)g(\gamma' Z_{ik})} \quad (8)$$

Let $X_{ik}^*(t;\beta) = \{X_{ik}', Z_{ik}', W_{ik}'(te^{-\beta' X_{ik}})\}$, to estimate θ_0, based on the idea of constructing generalized estimation equation of partial likelihood scoring function proposed by Anderson and Gill(1982),we can get the following estimation equation of unknown regression coefficient θ_0.

$$U_n(\theta) = \sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^\tau X_{ik}^*(t;\beta)[d\widetilde{N}_{ik}(t;\beta) - Y_{ik}(t;\beta)\{g(\gamma' Z_{ik})d\hat{\mu}_0(t;\theta) + e^{-\beta' X_{ik}\alpha'} W_{ik}\alpha'(te^{-\beta' X_{ik}})dt\}]$$

Then plug (2-8) into the above equation, and we get

$$U_n(\theta) = \sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^\tau \{X_{ik}^*(t;\beta) - \overline{X}^*(t;\beta;\gamma)[d\widetilde{N}_{ik}(t;\beta) - Y_{ik}(t;\beta)e^{-\beta' X_{ik}\alpha'} W_{ik}\alpha'(te^{-\beta' X_{ik}})dt\}],$$

Denote

$$\overline{X}^*(t;\beta;\gamma) = \frac{\sum_{i=1}^{n}\sum_{k=1}^{K} Y_{ik}(t;\beta)g(\gamma' Z_{ik})X_{ik}^*(t;\beta)}{\sum_{i=1}^{n}\sum_{k=1}^{K} Y_{ik}(t;\beta)g(\gamma' Z_{ik})}$$

As $U_n(\theta)$ is a section function of the unknown regression coefficient θ, we can define θ as zero solution of $U_n(\theta)$ or minimum value of $\|U_n(\theta)\|$. When low dimensional covariant exists, grid search method can be used to solve for θ.

When high dimensional covariant exists, simulated annealing algorithm(Lin and Geyer,1992) can be more effective to solve for θ. The solution $\hat{\theta} = (\hat{\beta}', \hat{\gamma}', \hat{\alpha}')'$ is taken as the estimation of θ_0.

Substitute $\hat{\theta}$ into (8), and then the estimation of the benchmark mean function $\hat{\mu}_0(t)$ is

$$\hat{\mu}_0(t) = \hat{\mu}_0(t;\hat{\theta}) = \sum_{i=1}^{n}\int_0^t \frac{\sum_{k=1}^{K} d\widetilde{N}_{ik}(s;\hat{\beta}) - \sum_{k=1}^{K} Y_{ik}(s;\hat{\beta})e^{-\beta' X_{ik}}\hat{\alpha} W_{ik}(se^{-\beta' X_{ik}})ds}{\sum_{i=1}^{n}\sum_{k=1}^{K} Y_{jl}(s;\hat{\beta})g(\gamma' Z_{jl})}.$$

Conclusions

This paper mainly introduces an additive-accelerated mean regression model for multiple type recurrent events. For multiple type recurrent events data, based on the idea of generalized estimating equation, this paper respectively estimates the unknown regression coefficient in the model, as well as the benchmark mean function. In our next work, we will give their asymptotic property under the case of large-scale samples, and also makes interval estimation.

Acknowledgments

This work is supported by the Natural Science Foundation of Hubei Province, China (No. 2011CDB167), the Major Research Program of Hubei Provincial Department of Education, China (No. Z20092701), the Ph.D. Fund of Huanggang Normal University to H. B. Liu, and the Innovative Group Project of Hubei Provincial Department of Education (No. 03BA85).

References

[1] J. Hyde, Biometrika 64 (1977) 225-230.

[2] D.G. Clayton, J. Cuzick, J. R. Statist. Soc. A 148 (1985) 82-117.

[3] U. Uzunogullari, J.L. Wang, Biometrika 79 (1992) 297-310.

[4] T.L. Lai, Z. Ying, Ann. Statist. 19 (1991) 531-556.

[5] Y. Li, X. Lin, Biometrika 87 (2000) 849-866.

[6] D.G. Clayton, Biometrics 47 (1991) 467-485.

[7] T.L. Lai, Z. Ying, Statistica Sinica 2 (1992) 17-46.

[8] H. Aslanidou, D.K. Dey, D. Sinha, Can. J. Statist. 26 (1998) 33-48.

[9] K.Y. Liang, S.G. Self, Y. Chang, J. Roy. Statist. Soc. Ser. B 55 (1993) 441-453.

[10] Y.D. Lin, Z. Ying, Biometrika 81 (1994) 61-71.

[11] D.Y. Lin, Z. Ying, J. Amer. Statist. Assoc. 88 (1993) 1341-1349.

[12] J.P. Klein, Biometrics 48 (1992) 795-806.

[13] I.D. Ha, Y. Lee, J.K. Song, Biometrika 88 (2001) 233-243.