An Additive Risk Model for Multi-type Recurrent Event Data

Eunyoung Choi

In this study, we consider the statistical modeling and analysis of the replicated multi-type point process data with covariates. Such data arises when heterogeneous subjects experience repeated events or failures which may be of several distinct types.

We propose a nonhomogeneous mixed Poisson process with random (subject) and fixed (covariate) effects with an additive intensity model. This method is applied to two examples, the first example involves 334 children who had at least three episodes of acute otitis media. The second example is a dataset of 661 homosexual men who had at least one episode of syphilis or gonorrhea.

This model is very useful for highlighting potential predictors of recurrent events. The model proposed in this paper allows one to examine a data set where related recurrent events are better analyzed as a single outcome, rather than individually. This model can provide useful information about potential predictors; however, it can be difficult to fit it into individual data sets.

◎ Key Words : recurrent event, mixed poisson process, random effect, additive risk

筆者: 本院 責任硏究員

▶ 원고를 검독하여 주신 張英植 副硏究委員과 金珍洙 責任硏究員께 감사드립니다.

I. Introduction

A variety of models are available for the analysis of time to event data where only a single outcome is of interest. These include the proportional hazards model(Cox, 1972), the additive risk model (Aalen, 1980), parametric models and extensions of these models. For this type of data, Cox's proportional hazards model has seen the widest use. This is due to the ease of interpretation of the covariate effects and the availability of software. While the proportional hazards model is useful in many settings, the additive risk model can be a useful alternative for when the data does not satisfy the proportionality assumption.

Another common type of data collection involves the recording of recurrent episodes, which may be of one or several types. Examples of this type of data include the occurrence of epileptic seizures (Albert, 1991), infection episodes, tumor recurrences (Gail et al., 1980, Freedman et al., 1989), sexually transmitted diseases, and bleeding incidents in medical studies. In engineering, the multivariate failure times may be the times to breakdown of a certain type of machinery, such as an electric computer or an automobile. Examples in sociology include studies of mobility, unemployment, fertility (Allison, 1984) and the experiences of different life events by each person. Examples in marketing research include the purchases of various products by each consumer.

Several methods have been proposed in the literature to deal with situations in which individuals experience repeated single type events. These methods impose specific structures of dependence among the

recurrences on each subject. Gail, Santner and Brown(1980) and Prentice, Williams and Peterson(1981) emphasize the interevent times and adopt the stratified proportional hazards model for use in this setting. Cox(1973) proposed the use of a modulated renewal process. Aalen and Husebye(1991) and Aalen, Bjertness and Sonju(1995) focus on renewal processes and consider the variation of interevent times within and between individuals.

Models based on modulated Poisson processes with group or subject random effects have also been developed (Lawless, 1987, Thall et al., 1988). Lawless and Neadeau(1995) developed a robust semiparametric generalization of these methods that remains valid without the assumption of an underlying Poisson or mixed Poisson process. These methods are based on the proportional intensity model.

A more complex problem is that of modeling repeated multitype events. Abu-Libdeh, Turnbull and Clark(1990) presented a non-homogeneous mixed Poisson process with a random (subject) and a fixed covariate effect and a proportional intensity for this type of data. Method has also been proposed to handle recurring and terminating events(Cook et al., 1997, Li et al., 1997).

We consider the statistical modeling and analysis of replicated multi-type point process data with covariates. We propose an extension of Abu-Libdeh, Turnbull and Clark's(1990) model incorporating an additive intensity. This model is developed in Section 2 and the application of this model to two examples is considered in Section 3. Final remarks are given in section 4.

Ⅱ. Statistical method for recurrent events

Suppose that there are n subjects, each with an associated vector Z_i $(1 \leq i \leq n)$ of baseline covariate values. Subject i is observed over time period $[0, T_i]$, where time is measured from a defined starting point for that subject. Individuals experience repeated events or failures, each of which can be any of J different types. Suppose that $K_{ij} \geq 0$ events of type j $(1 \leq j \leq J)$ are observed to occur on individual I, $1 \leq i \leq n$, at times $0 = t_{ij0} < t_{ij1} < \ldots < t_{ijK_{ij}}$. We suppose that there are no ties among events of the same type in the same individual. Also, the gap times $\{g_{ijk}, 1 \leq i \leq n, 1 \leq j \leq J, 1 \leq k \leq K_{ij}\}$ are defined by $g_{ijk} = t_{ijk} - t_{ijk-1}$ for $1 \leq k \leq K_{ij}$. The final gap time $g_{ijK_{ij}+1} = T_i - t_{ijK_{ij}} \geq 0$ is considered censored and used as such in the analysis.

The Poisson process assumption implies that the intensity, as a function of time since the previous event, is the same for each period and that the periods are independent. Thus, the intensity for the occurence of the next event for type j in individual i is given by

$$\theta_i \xi_{ij} h_i(t, \beta), \qquad i=1,\ldots,n \quad \ldots\ldots\ldots\ldots\ldots\ldots (1)$$

where $h_i(t, \beta)$ is a fixed function which depends on the covariate vector, and the proportionality factors, θ_i and ξ_{ij}, are random variables. Abu-Libdeh et al.(1990) assumed that $h_i(t, \beta)$ follows the proportional intensity. We now consider estimation and inference for this model when $h_i(t, \beta)$ follows the additive intensity.

Let the event processes of type j in individual i be independent nonhomogeneous mixed Poisson processes with respective intensity functions

$$\lambda_{ij} = \theta_i \xi_{ij} (\lambda_0(t) + Z_i \beta) \quad \cdots\cdots (2)$$

where $\lambda_0(t) > 0$ is a baseline intensity function, β is a column p-vector of regression coefficients, time t is measured from the ith subject's starting point, $\sum_{j=1}^{J} \xi_{ij} = 1$ for all i and θ_i and ξ_{ij} are non negative parameters. The parameters $(\theta_i, \xi_{i1}, \xi_{i2}, \ldots, \xi_{ij})$ corresponding to individual i are considered as random effects. The values $\theta_1, \theta_2, \ldots, \theta_n$ are considered to be an independent identically distributed sample from the gamma distribution with scale parameter γ and shape parameter ν. The vectors $(\xi_i : 1 \leq i \leq n)$ are assumed to be an independent identically distributed sample from a Dirichlet distribution with parameter $\alpha = (\alpha_1, \ldots, \alpha_J)$, which is independent of the $\{\theta_i\}$.

Define the baseline cumulative intensity function by $\Lambda_0(t) = \int_0^t \lambda_0(u) du$. Then, given θ_i and ξ_i, the contribution of the ith subject to the conditional likelihood is

$$\prod_{j=1}^{J} \{ \prod_{k=1}^{K_{ij}} \theta_i \xi_{ij} (\lambda_0(t_{ijk}) + Z_i \beta) \} e^{-\theta_i \xi_{ij} (\Lambda_0(T_i) + Z_i \beta T_i)}$$

$$= \left\{ \prod_{j=1}^{J} \xi_{ij}^{K_{ij}} \right\} \theta_i^{K_{i.}} e^{-\theta_i (\Lambda_0(T_i) + Z_i \beta T_i)} \prod_{j=1}^{J} \prod_{k=1}^{K_{ij}} (\lambda_0(t_{ijk}) + Z_i \beta),$$

where $K_{i.} = \sum_{j=1}^{J} K_{ij}$.

Our estimation process consists of two steps. First, we estimate the unknown parameters α, ν, γ, δ and β. Using these parameter estimates, we can then estimate the random subject effects, θ_i and ξ_i. Let $\lambda_0(t) = \delta t^{\delta-1}$, then the contribution of the ith subject to the marginal likelihood function is

$$\int_0^1 \cdots \int_0^1 \Gamma\Big(\sum_{j=1}^{J} \alpha_j\Big) \prod_{j=1}^{J} \Big[\frac{\xi_i^{K_{ij}+\alpha_j-1}}{\Gamma(\alpha_j)} d\xi_{ij}\Big] \times$$

$$\int_0^\infty \Big[\prod_{j=1}^{J} \prod_{k=1}^{K_{ij}} (\lambda_0(t_{ijk}) + Z_i\beta)\Big] \frac{\theta_i^{K_{i.}+\nu-1} e^{-\theta_i(T_i^\delta + Z_i\beta T_i + 1/\gamma)}}{\Gamma(\nu)\gamma^\nu} d\theta_i$$

$$= \Big[\frac{\Gamma\Big(\sum_{j=1}^{J} \alpha_j\Big)}{\Gamma\Big(K_{i.} + \sum_{j=1}^{J} \alpha_j\Big)} \prod_{j=1}^{J} \frac{\Gamma(K_{ij}+\alpha_j)}{\Gamma(\alpha_j)}\Big]$$

$$\times \Big[\frac{\Gamma(K_{i.}+\nu)}{\Gamma(\nu)} \frac{\gamma^{K_{i.}} \prod_{j=1}^{J} \prod_{k=1}^{K_{ij}} (\delta t_{ijk}^{\delta-1} + Z_i\beta)}{[\gamma T_i^\delta + \gamma Z_i\beta T_i + 1]^{K_{i.}+\nu}}\Big]$$

$$= L_{1i}(\alpha) L_{2i}(\nu, \gamma, \delta, \beta).$$

The total log likelihood is then given by

$$L(\alpha, \nu, \gamma, \delta, \beta) = \sum_{i=1}^{n} \{\log(L_{1i}(\alpha)) + \log(L_{2i}(\nu, \gamma, \delta, \beta))\}.$$

For inference on α, expressions for the score vector, $U(\alpha)$, and for the sample information matrix, $I(\alpha)$, are given in the Appendix. For inference on ν, γ, δ and β, the score vector and information matrix are also given in the Appendix. Maximum likelihood estimates (mle) of the parameters of interest, α, ν, γ, δ and β, are obtained by setting the components of the score vector equal to zero and

solving for the unknown values. This is done using a numerical iterative technique such as the Newton-Raphson method.

The mles of α, β, ν, δ and γ are now used to compute the empirical Bayes estimates of the random variables θ_i and ξ_{ij} for each individual. These estimates are the conditional means given the information on each individual and represent estimates of the relative risk level for each individual when compared with the average. The posterior distribution of θ_i has a gamma distribution with shape parameter $K_{i.}+\nu$ and mean $\frac{(K_{i.}+\nu)}{(T_i^{\delta}+Z_i\beta T_i+\gamma^{-1})}$, the posterior distribution of ξ_i has a Dirichlet distribution with parameters $(\alpha+K_i)$. Therefore, we estimate θ_i and ξ_{ij} by

$$\widehat{\theta_i}=\frac{K_{i.}+\widehat{\nu}}{T_i^{\widehat{\delta}}+Z_i\widehat{\beta}T_i+\widehat{\gamma}^{-1}} \quad \text{and} \quad \widehat{\xi_{ij}}=\frac{\widehat{\alpha_j}+K_{ij}}{K_{i.}+\sum_{s=1}^{J}\widehat{\alpha_s}},$$

respectively.

Ⅲ. Examples

We now consider an example of multitype recurrent events in an otitis media study. The purpose of the study is to evaluate the possible long term effects of early-life otitis media on speech, language, cognitive, and psychosocial development. Over time, more than 5000 children with sufficient ear diseases have been enrolled in the study and randomized to one of two groups. In one arm of the study the children receive ear tubes while the second arm delays the

insertion of the ear tubes. The expectation is that the group receiving tubes will have less ear disease on average, while the group not receiving tubes is expected to have a substantial percentage of children who do not have sufficient additional episodes of ear disease to justify tubes. To determine if a participant is eligible for randomization into the study, monthly assessments of ear diseases are conducted, so that the study also provides information on the natural history of ear disease.

The data set used for this example is a sample of 334 children who were followed for 2 years and experienced 3 to 10 episodes of Acute Otitis Media(AOM). The events of interest are the successive incidence times of right or left ear AOM. We denote an infection in the right ear as type I and an infection in the left ear as type II. Table 1 summarizes the distribution of patients according to the number of failures of each type. Analyses were restricted to the first five events in each ear. The covariates used in this analysis were SEX, an indicator which is 1 for females and RACE, an indicator which is 1 for caucasian.

Table 1. Frequency distribution of the number of AOM episodes in each ear

Number of episodes	Right ear	Left ear
0	6	8
1	25	25
2	50	62
3	89	87
4	61	57
5	103	95
Total failures	1,151	1,113

Table 2. Mixed Poisson Process model. Endpoint is time to AOM

Parameter	Estimate	S.E.	p-value
$\alpha 1$	22.74	2.44	< .001
$\alpha 2$	21.96	2.35	< .001
ν 1)	3.40	0.507	< .001
γ 1)	0.138	0.039	< .001
δ 1)	1.218	0.023	< .001
SEX	-0.431	0.087	< .001
RACE	-1.148	0.043	< .001

Note: 1) the null hypothesis is H0: ν =1, H0 : γ =1 and H0 : δ =1.

The results obtained from fitting the model in equation (2) to this data are presented in table 2. Recall that the parameters ν and γ are the shape and scale parameter, respectively, of the gamma distribution representing the random effects. The parameter δ is associated with the baseline hazard function. Covariate effects are interpreted in the standard fashion.

Table 2 contains the results of this analysis. The shape parameter ν and the scale parameter γ are statistically significant. This implies that the gamma distribution representing the random effect. The baseline hazard estimate, δ, is highly significant. Similarly, both of the covariates included in the model are also significant. The two covariates, SEX and RACE, have negative coefficients, indicating that females and caucasians tend to develop less AOM.

The second example to be considered is based on data from the Multicenter AIDS Cohort Study(MACS). This study is a major ongoing prospective study of the natural history of HIV infection and AIDS in U. S. homosexual and bisexual men. Information on

sexually transmitted diseases is collected semi-annually in this group. The data set used for this example is a sample of 661 homosexual and bisexual men who had experienced at least one syphilis(type I) episode or one gonorrhea(type II) episode. Table 3 summarizes the distribution of patients according to the number of failures of each type. The covariates considered for this analysis were AIDS, an indicator variable which is 1 for subjects with AIDS, DRUG, an indicator variable which is 1 for drug use and age at baseline. The mixed Poisson process model given by equation (2) was fit to this data. The covariates AIDS, DRUG and age at baseline were included in the intensity function.

Table 3. Frequency distribution of the number episodes of syphilis and gonorrhea.

Number of episodes	Syphilis	Gonorrhea
0	459	152
1	167	404
2	26	86
3	5	13
4	3	5
5	0	1
6	1	0
Total failures	252	640

Table 4 contains the results of this analysis. The shape parameter ν is not significant and the scale parameter γ is statistically significant. This implies that in the gamma distribution representing the random effect. The baseline hazard estimate, δ, is highly significant. Similarly, all three of the covariates included in the model are also significant. Both AIDS and DRUG have negative

coefficients inidicating a protective effect for the development of either gonorrhea or syphilis. The coefficient for AGE is positive and statistically significant; however, the magnitude of this coefficient is smaller than that of the coefficients for both AIDS and DRUG.

Table 4. Mixed Poisson Process model. MACS data set Endpoint is time to gonorrhea and/or syphilis

Parameter	Estimate	S.E.	p-value
$\alpha 1$	0.826	0.307	< 0.001
$\alpha 2$	2.203	0.809	< 0.001
ν 1)	0.771	0.144	0.112
γ 1)	0.553	0.0217	0.039
δ 1)	1.173	0.021	< 0.001
AIDS	-1.580	0.0954	< 0.001
DRUG	-0.310	0.039	< 0.001
AGE	0.001	0.0005	0.001

Note: 1) the null hypothesis is H0: ν =1, H0: γ =1 and H0: δ =1.

Ⅳ. Discussion

This model is very useful for highlighting potential predictors of recurrent events. Many methods that are currently available for this problem are based either on a conditional approach or a complex model. The model proposed in this paper allows one to examine a data set where related recurrent events are better analyzed as a single outcome, rather than individually. This model can provide useful information about potential predictors; however, they can be difficult to fit into individual data sets.

The model proposed here is based on an intensity comprising

linear hazards and random effects. The interpretation of the parameters obtained from these models is straightforward, since the parameters can be partitioned into random effects, baseline hazard and regression piece. Thus, for the regression aspect of this approach the interpretation is the same as that of any hazard regression model.

Appendix

Components of Score vector and Information Matrix for Likelihood based on Random Effects Poisson process Model of Section 2.

For inference on α, the score vector $U(\alpha)$, and for the sample information matrix, $I(\alpha)$, are given by

$$U_{\alpha_j}(\alpha)=\sum_{i=1}^{n}\{\sum_{k=1}^{K_{ij}}(\alpha_j+k-1)^{-1}-\sum_{s=1}^{K_i}(\sum_{j=1}^{J}\alpha_j+s-1)^{-1}\}.$$

$$I_{\alpha_j\alpha_k}(\alpha)=\sum_{i=1}^{n}\{\sum_{k=1}^{K_{ij}}-\delta_{jk}(\alpha_j+k-1)^{-2}+\sum_{s=1}^{K_i}(\sum_{j=1}^{J}\alpha_j+s-1)^{-2}\}.$$

For inference on ν, γ, δ and β, the score vector and information matrix are given by

$$U_{\nu}(\nu,\gamma,\delta,\beta)=\sum_{i=1}^{n}\{\sum_{s=1}^{K_i}(\nu+s-1)^{-1}-\log[1+\gamma T_i^{\delta}+\gamma Z_i\beta T_i]\},$$

$$U_{\gamma}(\nu,\gamma,\delta,\beta)=\sum_{i=1}^{n}\{\frac{K_{i.}}{\gamma}-\frac{(K_{i.}+\nu)(T_i^{\delta}+Z_i\beta T_i)}{1+\gamma T_i^{\delta}+\gamma Z_i\beta T_i}\},$$

$$U_{\delta}(\nu,\gamma,\delta,\beta)=\sum_{i=1}^{n}\{\sum_{j=1}^{J}\sum_{k=1}^{K_{ij}}\frac{t_{ijk}^{\delta-1}[1+\delta\log(t_{ijk})]}{\delta t_{ijk}^{\delta-1}+Z_i\beta}$$

$$-\frac{(K_{i.}+\nu)\gamma T_i^{\delta}\log(T_i)}{1+\gamma T_i^{\delta}+\gamma Z_i\beta T_i}\},$$

$$U_{\beta_l}(\nu,\gamma,\delta,\beta)=\sum_{i=1}^{n}\{\sum_{j=1}^{J}\sum_{k=1}^{K_{ij}}\frac{Z_{il}}{\delta t_{ijk}^{\delta-1}+Z_i\beta}-\frac{(K_{i.}+\nu)\gamma Z_{il}T_i}{1+\gamma T_i^{\delta}+\gamma Z_i\beta T_i}\},$$

$$I_{\nu\nu}(\nu,\gamma,\delta,\beta)=\sum_{i=1}^{n}\{\sum_{s=1}^{K_i}(\nu+s-1)^{-2}\},$$

$$I_{\nu\gamma}(\nu,\gamma,\delta,\beta)=\sum_{i=1}^{n}\{\frac{(T_i^{\delta}+Z_i\beta T_i)}{1+\gamma T_i^{\delta}+\gamma Z_i\beta T_i}\},$$

$$I_{\nu\delta}(\nu,\gamma,\delta,\beta)=\sum_{i=1}^{n}\{\frac{\gamma T_i^{\delta}\log(T_i)}{1+\gamma T_i^{\delta}+\gamma Z_i\beta T_i}\},$$

$$I_{\nu\beta_i}(\nu,\gamma,\delta,\beta)=\sum_{i=1}^{n}\{\frac{\gamma Z_{ij}T_i}{1+\gamma T_i^{\delta}+\gamma Z_i\beta T_i}\},$$

$$I_{\gamma\gamma}(\nu,\gamma,\delta,\beta)=\sum_{i=1}^{n}\{\frac{K_{i.}}{\gamma^2}-\frac{(K_{i.}+\nu)(T_i^{\delta}+Z_i\beta T_i)^2}{(1+\gamma T_i^{\delta}+\gamma Z_i\beta T_i)^2}\},$$

$$I_{\gamma\delta}(\nu,\gamma,\delta,\beta)=\sum_{i=1}^{n}\{\frac{(K_{i.}+\nu)T_i^{\delta}\log(T_i)}{(1+\gamma T_i^{\delta}+\gamma Z_i\beta T_i)^2}\},$$

$$I_{\gamma\beta_i}(\nu,\gamma,\delta,\beta)=\sum_{i=1}^{n}\{\frac{(K_{i.}+\nu)Z_{ij}T_i}{(1+\gamma T_i^{\delta}+\gamma Z_i\beta T_i)^2}\},$$

$$I_{\delta\delta}(\nu,\gamma,\delta,\beta)=\sum_{i=1}^{n}\{\sum_{j=1}^{J}\sum_{k=1}^{K_{ij}}\frac{t_{ijk}^{(\delta-1)}(t_{ijk}^{(\delta-1)}-\log(t_{ijk})Z_i\beta[2+\delta\log(t_{ijk})])}{(\delta t_{ijk}^{\delta-1}+Z_i\beta)^2}$$

$$+\frac{(K_{i.}+\nu)\gamma T_i^{\delta}[\log(T_i)]^2(1+\gamma Z_i\beta T_i)}{(1+\gamma T_i^{\delta}+\gamma Z_i\beta T_i)^2}\},$$

$$I_{\delta\beta_i}(\nu,\gamma,\delta,\beta)=\sum_{i=1}^{n}\{\sum_{j=1}^{J}\sum_{k=1}^{K_{ij}}\frac{t_{ijk}^{(\delta-1)}(1+\delta\log(t_{ijk}))Z_{il}}{(\delta t_{ijk}^{\delta-1}+Z_i\beta)^2}$$

$$-\frac{(K_{i.}+\nu)\gamma^2 T_i^{\delta+1}\log(T_i)Z_{il}}{(1+\gamma T_i^{\delta}+\gamma Z_i\beta T_i)^2}\},$$

$$I_{\beta_i\beta_m}(\nu,\gamma,\delta,\beta)=\sum_{i=1}^{n}\{\sum_{j=1}^{J}\sum_{k=1}^{K_{ij}}\frac{Z_{il}Z_{im}}{(\delta t_{ijk}^{\delta-1}+Z_i\beta)^2}$$

$$-\frac{Z_{il}Z_{im}(K_{i.}+\nu)\gamma^2 T_i^2}{(1+\gamma T_i^{\delta}+\gamma Z_i\beta T_i)^2}\}.$$

Reference

Aalen, O. O., "A model for non-parametric regression analysis of counting processes", Lecture Notes in Statistics, Vol.2, Springer Verlag: New york, 1980, pp.1~25.

Aalen, O. O., Husebye, E., "Statistical analysis of repeated events forming renewal processes", Statistics in Medicine 10, 1991, pp.1227~1240.

Aalen, O. O., Bjertness, E. and Sonju, T., "Analysis of dependent survival data applied to lifetimes of amalgam fillings", Statistics in Medicine 14, 1995, pp.1819~1829.

Abu-Libdeh, H., Turnbull, B. W. and Clark, L. C., "Analysis of multi-type recurrence events in longitudinal studies; application to a skin cancer prevention trial" Biometrics 46, 1990, pp.1017~1034.

Albert, P. S., "A two-state Markov mixture model for a time series of epileptic seizure counts", Biometrics 47, 1991, pp.1371~1381.

Allison, P. D., Event History Analysis: Regression For Longitudinal Data, Beverly Hills, California: Sage Publications, 1984.

Cook, R. J., Lawless, J. F., "Marginal Analysis of Recurrent Events and a Terminating Event", Statistics in Medicine 16(8), 1997, pp.911~924.

Cox, D. R., "Regression models and life-tables", Journal of Royal Statistics Society B 34, 1972, pp.87~220.

Cox, D. R., "The statistical analysis of dependencies in point processes", In Symposium on Point Processes, Ed. P. A. W. Lewis, New York: Wiley, 1973, pp. 55~66.

Freedman, L., Sylvester, R. and Byar, D. P., "Using permutation tests and boostrap confidence limits to analyze repeated events data from clinical trials", Controlled Clinical Trials 10, 1989, pp.129~141.

Gail, M. H., Santner, T. J. and Brown, C. C., "Analysis of comparative carcinogenesis experiments based on multiple times to tumor" Biometrics 36, 1980, pp.255~266.

Lawless, J. F., "Regression methods for Poisson process data" Journal of the American Statistical Association 82, 1987, pp. 808~815.

Lawless, J. F. and Neadeau, J. C., "Nonparametric estimation of cumulative mean functions for recurrent events", Technometrics 37, 1995, pp.158~168.

Li, Q. H. and Lagakos, S. W., "Use of the Wei-Lin-Weissfeld Method for the Analysis of a Recurring and a Terminating Event" Statistics in Medicine 16(8), 1997, pp.925~940.

Prentice, R. L., Williams, J. and Peterson, A. V., "On the regression analysis of multivariate failure time data", Biometrika 68, 1981, pp.373~379.

Thall, P. F. and Lachin, J. M., "Analysis of recurrent events: nonparametric methods for random-interval count data", Journal of the American Statistical Association 83, 1988, pp.339~347.

요 약

多形態 再發事件 資料分析을 위한 加法 危險函數 模型에 관한 硏究

崔 恩 英

공변량을 가진 반복되는 다형태 점과정(Multi-type point process) 자료를 분석하기 위해 가법 위험함수 모형(Additive risk model)을 적용하였다. 이러한 자료는 이질적 사람들이 다형태 재발사건이나 여러 가지 종류의 반복되는 실패(failure)를 경험할 때 발생한다.

이 자료를 분석하기 위한 방법으로 가법 강도 모형을 가정한 비동질 포아송 과정을 제안한다. 이 과정에서 랜덤한 개인효과와 고정된 공변량 효과를 고려했다. 다음 두 예제에 이 방법을 적용하여 실제 이 방법이 어떻게 이용될 수 있는 지를 보였다. 첫 번째 예는 세 번 이상 중이염이 재발한 어린이 자료이고 두 번째 예는 매독이나 임질에 걸린 경험이 있는 661명의 남자 동성연애자에 관한 자료이다.

이 모형은 재발사건에 영향을 주는 공변량을 찾는데 매우 유익하고 단순 실패를 한 자료 분석보다 다형태 재발사건에 관련된 자료를 분석하기에 유익하다. 이 모형은 영향력 있는 변수에 대한 정보를 제공할 수 있다.