复发事件下一类加性乘性转移模型 *

杜彦斌
(湖南师范大学数学与统计学院，长沙 410006)
戴家佳 † 金 君
(贵州大学数学与统计学院，贵阳 550025)
(†E-mail: sci.jjdai@gzu.edu.cn)

摘 要 复发事件数据频繁的出现在纵向研究中, 本文基于生物医学中的单类型复发事件数据, 提出了一类加性乘性转移模型，该模型包含了一些重要的半参数模型. 同时，模型允许协变量具有加性和乘性的影响，且加性影响随时间而变化. 利用广义估计方程的思想，对模型中未知参数和非参数函数进行了估计，并证明了所得估计的相合性和渐近正态性. 最后，用数值模拟的方法验证了所提估计的可行性.

关键词 复发事件数据；加性乘性影响；转移模型；估计方程
MR(2000) 主题分类 62G05; 62N01
中图分类 O212. 7

1 前言

在生物学，医学，社会和经济学等研究领域中，研究的个体有时会不止一次的经历某一事件，例如：肿瘤切除手术后癌细胞的反复扩散；某校电梯故障的多次发生；某品牌手机的多次死机等. 对于这种一段时间内，某种特定的重复发生的事件我们称它为复发事件 (recurrent events), 在复发事件过程中产生的数据，我们称为复发事件数据.

根据研究对象种类的不同，复发事件数据一般分为两种类型：一种是单类型复发事件，即感兴趣的事件只有一种类型，并且不止一次发生，例如，某校校园网断网的多次发生、某街道十字路口交通事故的多次发生、某地区雨天天气的反复发生等. 然而，在许多生活应用中，经常会遇到多种不同类型的复发事件，即多类型复发事件. 例如，在研究人群高血压病症时，我们要同时考虑家族性、后天性因素的影响；在研究犯罪分子再犯案与种族关系时，我们要同时考虑各个种族的影响.

本文 2017 年 4 月 10 日收到. 2018 年 5 月 3 日收到修改稿.
* 国家自然科学基金 (11361015) 资助项目.

复发事件数据在生活中普遍存在，但是它与我们了解的时间序列数据、截面数据、纵向数据有所不同．时间序列数据反映的是某一事物、现象随时间变化状态，例如贵阳市某一天气温的变化．截面数据是不同个体在同一时间点 (时期) 的观测数据，例如 2017 年某村每户家庭收入．纵向数据是多个个体随着时间变化观测的一系列指标．例如：某高中高一学生一年中每月的生活费用．而复发事件数据更为复杂，除了具有时间序列数据时序性和纵向数据多个个体，多次测量的特点外，还具有以下特点：

(1) 观测到的数据一般情况下都存在删失；

(2) 研究事件本身存在一定的相依性；

(3) 删失时间与事件的复发时间可能具有一定的相依性；

(4) 不同类型的事件间也存在一定的相依性.

通常对复发事件数据的分析，研究者关心的是协变量对复发事件发生率的影响．对单类型复发事件，已有的文献中提出了几种估计方法，包括条件强度模型 [1,2] 和边际强度模型 [3]．这些方法都是基于风险函数 (hazards function) 来建立模型的 [4,5]．但是在处理复发事件数据时，由于事件的均值函数 (means function) 比风险函数具有更强的解释能力，因此，学者更倾向基于均值函数或比率函数 (rates function) 建立模型 [6]．随着研究的不断深入，很多学者对复发事件数据建立了半参数模型 [8], 它既满足我们对感兴趣变量的建模需求又使模型变得更加灵活、稳定．在半参数模型方面，已有的文献主要研究的是加性比率模型 [9−11] 和乘性比率模型 [12−14].

假设在一段时间内, 考察的个体共有 n 个, 且各个个体间是相互独立的. 设 $N_i^*(t)$ 为第 i 个个体在时刻 t 所经历事件的次数. 在大多数实际应用中, 总是在有限的时间内来考察个体, 因此 $N_i^*(t)$ 不可能完全观察, 记第 i 个体的删失时间为 C_i, 设 $N_i(t)=N_i^*(t\wedge C_i)$, $Y_i(t)=I(C_i\geq t)$, 其中 $\mathrm{a}\wedge b=\min(a,b), I(\cdot)$ 为示性函数．设 $Z_i(t)$ 表示 p 维协变量过程向量，并假设在给定 $Z_i(\cdot)$ 条件下，删失时间 C_i 与 $N_i^*(t)$ 相互独立，即

$$E[\mathrm{d}N_i^*(t)|Z_i(t),\ C_i>t]=E[\mathrm{d}N_i^*(t)|Z_i(t)],\qquad t\geq 0,$$

对于乘性比率模型 Lin D Y, Wei L J 等提出了复发事件的半参数乘性比率模型 [2]:

$$E\{\mathrm{d}N_i^*(t)|Z_i(t)\}=\exp\{\beta_0'Z_i(t)\}\lambda_0(t)\,\mathrm{d}t, \tag{1.1}$$

其中 β_0 是 p 维未知回归参数向量， $\lambda_0(\cdot)$ 为未知基本比率函数.

对于加性比率模型， Schaubel 等研究了复发事件半参数可加比率模型 [15]:

$$E\{\mathrm{d}N_i^*(t)|Z_i(t)\}=\beta_0'Z_i(t)\,\mathrm{d}t+\lambda_0(t)\,\mathrm{d}t, \tag{1.2}$$

Liu Yu Tao, Sun Liu Quan, Zhou Yong 提出复发事件数据下可加转移模型 [16]:

$$E({N_i}^*(t)|Z)=u_0(t)+Q(t,{\beta_0}^TZ), \tag{1.3}$$

其中， $\mu_0(\cdot)$ 为未知基准均值函数， β_0 是 p 维未知回归参数向量， $Q(t,x)$ 为一个预先设定的二元非负连接函数，且满足：对于任意 x, $Q(0,x)=0$, 容易发现，当 $Q(t,x)=tx$,

且协变量 $Z(t)$ 与时间 t 无关时，模型 (1.3) 即为

$$E(N_i{}^*(t)|Z) = u_0(t) + \beta_0{}^T Zt. \tag{1.4}$$

我们可以发现模型 (1.3) 中直接把时间 t 作为一个变量，通过连接函数 Q 把协变量与时间 t 联系在一起，共同反映协变量对复发事件率的影响，这种做法更为直接，且可以根据已有的数据和想要得到参数的解释为依据去选择合理的连接函数. 同时我们用模型来表示协变量对复发事件的影响也具有更大的灵活性和广泛性. 但在模型 (1.3) 中， Liu 等只考虑了协变量的加性影响而没考虑到协变量的乘性影响. 而在研究应用中，研究的个体可能具有多个协变量的影响，其中有些协变量的影响是加性的，另外一些协变量的影响是乘性的，或者某些协变量的影响既是加性的又是乘性的，如戴家佳，孙六全，杨振海等提出了一类加性 - 乘性比率模型 [17]:

$$E[dN_i{}^*(t)|Z_i(t)] = g\{\beta_0' W_i(t)\}\,\mathrm{d}t + h\{\gamma_0' X_i(t)\}\lambda_0(t)\,\mathrm{d}t, \tag{1.5}$$

其中， β_0 和 γ_0 分别表示 p_1 和 p_2 维回归系数向量， $g(\cdot)$ 和 $h(\cdot)$ 为已知联系函数. 显然，协变量 $W_i(t)$ 对基准均值函数具有加性的影响，而协变量 $X_i(t)$ 具有乘性的影响.

所以，结合模型 (1.3) 和 (1.5) 的思想，我们提出复发事件数据下一类加性乘性转移模型. 模型不仅解决了协变量对复发事件的影响随时间变化问题，更同时考虑了协变量可能具有的加性和乘性影响，使模型变得更加广泛，包含了 Liu 等提出的可加转移模型.

2 模型和估计方法

设 $W_i(t)$ 和 $X_i(t)$ 分别表示 p_1 和 p_2 维协变量， $Z_i(t) = (W_i(t)', X_i(t)')'$ 为 p 维协变量过程向量，其中 $p = p_1 + p_2$. 我们提出的复发事件下半参数加性乘性转移模型具有以下形式：

$$E[N_i^*(t)|Z_i(t)] = u_0(t) g\{a_0' X_i(t)\} + Q(t, \beta_0' W_i(t)), \tag{2.1}$$

其中 β_0 和 α_0 分别表示 p_1 和 p_2 维未知回归系数向量， $\mu_0(t)$ 是未知基准均值函数， $g(\cdot)$ 是已知联系函数， $Q(t,x)$ 是预先设定的二元非负连接函数.

当 $g(x) = 1$ 时，模型 (2.1) 就是模型 (1.3); 当 $Q(t,x) = x$ 时，模型 (2.1) 就是模型 (1.5). 可见我们提出的模型包含了一些重要的半参数模型.

在复发事件数据下，可观测的数据是 $\{N_i(\cdot), Y_i(\cdot), W_i(\cdot), X_i(\cdot),\ i = 1, \cdots, n\}$, 定义如下过程：

$$M_i(t;\theta_0) = Y_i(t)\big[N_i(t) - \{\mu_0(t) g(\alpha_0' X_i(t)) + Q(t, \beta_0' W_i(t))\}\big],$$

易知， $M_i(t;\theta_0)$ 是均值为 $\mathbf{0}$ 的随机过程.

对于给定的 $\theta = \{\alpha', \beta'\}', \mu_0(t)$ 的一个自然估计是下列方程的解：

$$\sum_{\mathrm{i}=1}^{n} Y_i(t)\big[N_i(t) - \{\mu_0(t) g(\alpha' X_i(t)) + Q(t, \beta' W_i(t))\}\big] = 0, \qquad 0 \le t \le \tau, \tag{2.2}$$

其中 τ 是一个预先设定的常数使得 $p(C_i \geq \tau) > 0$.

求解方程 (2. 2) 得到：

$$\widehat{\mu}_0(t;\theta) = \frac{\sum\limits_{i=1}^{n} Y_i(t)[N_i(t) - Q(t, \beta' W_i(t))]}{\sum\limits_{i=1}^{n} Y_i(t) g\{\alpha' X_i(t)\}}. \tag{2.3}$$

为了估计 $\theta_0 = \{\alpha_0', \beta_0'\}'$, 应用估计方程的思想，我们可以得到下列估计方程：

$$\sum_{\mathrm{i}=1}^{n} \int_0^{\tau} Y_i(t) \big[N_i(t) - \{\mu_0(t) g(\alpha' X_i(t)) + Q(t, \beta' W_i(t))\} \big] Z_i(t) \, \mathrm{d}H_n(t) = 0_{p\times 1}, \tag{2.4}$$

其中，$\tau > 0$，$H_{\mathrm{n}}(t)$ 是 $[0, \tau]$ 上递增的权函数.

将式 (2.3) 代入式 (2.4) 中并做简单的代数运算可以得到：

$$\sum_{\mathrm{i}=1}^{n} \int_0^{\tau} \mathrm{Y_i}(t) \big[N_i(t) - Q(t, \beta' W_i(t)) \} \big] \{ Z_i(t) - \overline{Z}(t;\alpha) \} \, \mathrm{d}H_n(t) = 0_{p\times 1}, \tag{2.5}$$

其中

$$\overline{Z}(t;\alpha) = \frac{\sum\limits_{i=1}^{n} Y_i(t) g\{\alpha' X_i(t)\} Z_i(t)}{\sum\limits_{i=1}^{n} Y_i(t) g\{\alpha' X_i(t)\}}.$$

用 Newton-Raphson 迭代法可以求出方程 (2.5) 的解，记为 $\widehat{\theta} = (\widehat{\alpha}', \widehat{\beta}')'$, 则 $\mu_0(t)$ 的相应估计为 $\widehat{\mu}_0(t) \equiv \widehat{\mu}_0(t;\widehat{\theta})$.

3　估计量的渐近性质

为了研究 $\widehat{\theta}$ 和 $\widehat{\mu}_0(t)$ 的渐近性质，本文假设下列条件成立：

(C1)　$\{N_i(\cdot), Y_i(\cdot), Z_i(\cdot),\ i = 1, \cdots, n, \}$ 独立同分布，其中 $Z_i(\cdot) = (W_i(\cdot)', X_i(\cdot)')'$.

(C2)　$P\{C_i \geq \tau\} > 0$，$Z_i(t)$ 是一个有界变量.

(C3)　$\mathrm{g}(\cdot)$ 为二次连续可微函数，$\mathrm{g}\{\alpha_0' X_i(t)\}$ 不为 0 且有界.

(C4) $Q(t, x)$ 关于 t 单调递增，关于 x 严格单调递增且关于 x 的二阶偏导数连续.

(C5)　权函数 $H_n(t)$ 几乎一致的在 $[0, \tau]$ 上收敛于一个非随机的有界函数 $\widetilde{H}(t)$.

(C6)　A 是非奇异矩阵，其中

$$A = E\Big\{ \int_0^{\tau} Y_i(t)[Z_i(t) - \overline{z}(t)][\dot{g}\{\alpha_0' X_i(t)\} X_i(t)' \mu_0(t) \, \mathrm{d}\widetilde{H}(t), \dot{Q}\{t, \beta_0' W_i(t)\} W_i(t)' \, \mathrm{d}\widetilde{H}(t)] \Big\},$$

其中，E 表示数学期望，$\overline{z}(t)$ 是 $\overline{Z}(t)$ 的极限，$\dot{g}(t) = \mathrm{d}g(t)/\mathrm{d}t$，$\dot{Q}(t, x) = \partial Q(t, x)/\partial x$.

为了方便推导，我们给出以下标记：

$$\widehat{M}_i(t;\widehat{\theta}) = Y_i(t)\big[N_i(t) - \{\widehat{\mu}_0(t)g(\widehat{\alpha}'X_i(t)) + Q(t,\widehat{\beta}'W_i(t))\}\big],$$

$$\widehat{A} = n^{-1}\sum_{i=1}^{n}\int_0^{\tau} Y_i(t)\big[Z_i(t) - \overline{z}(t)\big]\big[\dot{g}\{\widehat{\alpha}'X_i(t)\}X_i(t)'\widehat{\mu}_0(t)\,\mathrm{d}\widetilde{H}(t), \dot{Q}\{t,\widehat{\beta}'W_i(t)\}W_i(t)'\,\mathrm{d}\widetilde{H}(t)\big],$$

$$U(\theta) = \sum_{i=1}^{n}\int_0^{\tau} Y_i(t)\big[N_i(t) - Q(t,\beta'W_i(t))\}\big]\,\{Z_i(t) - \overline{Z}(t;\alpha)\}\,\mathrm{d}H_n(t).$$

定理 3.1 在条件 (C1)–(C6) 下， $\widehat{\theta}$ 和 $\widehat{\mu}_0(t)$ 是存在且唯一的，并且 $\widehat{\theta}$ 是 θ_0 的一个强相合估计， $\widehat{\mu}_0(t)$ 在 $[0,\tau]$ 上几乎一致收敛于 $\mu_0(t)$.

定理 3.2 在条件 (C1–C6) 下， $n^{1/2}(\widehat{\theta}-\theta_0)$ 渐近服从均值为 **0** 且协方差矩阵为 $A^{-1}\sum A^{-1}$ 的正态分布，协方差阵的相合估计为 $\widehat{A}^{-1}\widehat{\sum}\widehat{A}^{-1}$, 其中 A 定义于条件 (C6),

$$\widehat{\sum} = n^{-1}\sum_{i=1}^{n}\Big[\int_0^{\tau}\{Z_i(t) - \overline{Z}_i(t)\}\widehat{M}_i(t;\widehat{\theta})\,\mathrm{d}H_n(t)\Big]^{\otimes 2},$$

对于任意向量 ν, $\nu^{\otimes 2} = \nu\nu'$.

定理 3.3 在条件 (C1)–(C6) 下，$n^{1/2}\{\widehat{\mu}_0(t)-\mu_0(t)\}$ 弱收敛于一个均值为 **0** 的 Gauss 过程，且在 (s,t) 处的协方差函数为

$$\Gamma(s,t) = E\{\psi_i(s)\psi_i(t)\},$$

用估计量代替未知量，可以得到协方差函数的一个相合估计为

$$\widehat{\Gamma}(s,t) = \frac{1}{n}\sum_{i=1}^{n}\widehat{\psi}_i(s)\widehat{\psi}_i(t).$$

其中

$$\widehat{\psi}_i(t) = \frac{\widehat{M}_i(t;\widehat{\theta})}{n^{-1}\sum\limits_{i=1}^{n}Y_i(t)g\{\alpha'X_i(t)\}} - \widehat{\mathrm{B}}(t;\widehat{\theta})\widehat{A}^{-1}\int_0^{\tau}\{Z_i(t) - \overline{Z}_i(t)\}\widehat{M}_i(t;\widehat{\theta})\,\mathrm{d}H_n(t),$$

$$\widehat{\mathrm{B}}(t;\widehat{\theta}) = \frac{\frac{1}{n}\sum\limits_{i=1}^{n}Y_i(t)[\dot{g}\{\alpha'X_i(t)\}X_i(t)'\widehat{\mu}_0(t), \dot{Q}\{t,\beta'W_i(t)\}W_i(t)']}{\frac{1}{n}\sum\limits_{i=1}^{n}Y_i(t)g\{\alpha'X_i(t)\}}.$$

4 模拟研究

为了更好地了解模型的优劣，我们通过随机模拟的方法去检验有限样本下模型中估计量的性质．在研究中，我们取 $W_i \sim B(n,0.5)$， $X_i \sim U(0,1)$. 删失时间 $C \sim U(1,6)$, $\tau = 3$ 为复发事件的最长跟踪时间．对于给定的协变量 $W(t), X(t)$ 复发事件的发生次数在模型 (2.1) 下通过泊松过程产生，其中，取 $g(x) = \exp(x)$， $\mu_0(t) = t$ 和 $1.5\log(1+t)$,

连接函数 $Q(t,x)$ 考虑两种情况 $Q(t,x)=tx$ 和 $Q(t,x)=te^x$. 选取样本容量为 200 个，重复模拟 1000 次，权函数取 $H_n(t)=t$.

表 1 给出了 $\theta_0=(\alpha_0,\beta_0)'$ 与估计量 $\widehat{\theta}=(\widehat{\alpha},\widehat{\beta})'$ 平均值的偏差 (Bias), 样本的标准方差 (SSE), 估计 $\widehat{\theta}$ 的渐近方差的平均值 (SEE) 和基于正态近似的 95% 的经验覆盖率 CP. 从表中可以看出估计几乎是无偏的，经验覆盖率 CP 也非常接近置信度为 95% 的置信水平. 特别是当 $Q(t,x)=tx$, $\mu_0(t)=1.5\log(1+t)$ 时，真值 α_0,β_0 与估计值的偏差都非常小，经验覆盖率 CP 也更接近 95% 的置信度，效果非常的好.

表 1 真值 θ_0 估计量 $\widehat{\theta}$ 的模拟结果

		真值 $\alpha_0=0.1$				真值 $\beta_0=0.5$			
$Q(t,x)$	$\mu_0(t)$	Bias	SSE	SEE	CP	Bias	SSE	SEE	CP
tx	t	0.0102	0.0718	0.0710	0.954	0.0700	0.1884	0.1364	0.941
	$1.5\log(1+t)$	0.0103	0.0750	0.0743	0.952	0.0728	0.1862	0.1359	0.949
te^x	t	0.0225	0.0765	0.0731	0.952	0.1161	0.1486	0.1226	0.942
	$1.5\log(1+t)$	0.0174	0.0729	0.0708	0.950	0.1129	0.1546	0.1277	0.954

总体的看，对参数 α_0 的估计结果要比 β_0 好些，偏差相对更小，经验覆盖率 CP 也更接近 95% 的置信度. 总的来说，我们提出模型的估计方法比较合理，估计效果比较好.

当取特殊的联系函数 $g(x)=1$, $Q(t,x)=x$ 时，模型就退化为简单的半参数可加均值模型：

$$E[N_i^*(t)|Z_i(t)]=u_0(t)+{\beta_0}'Z_i(t), \tag{4.1}$$

同样，我们取协变量 $Z_{i1}\sim B(n,0.5)$, $Z_{i2}\sim U(0,1)$. 删失时间 $C\sim U(1,6)$, $\tau=3$ 为复发事件的最长跟踪时间. 复发事件的发生次数在模型 (4.1) 下通过泊松过程产生，其中，取 $\mu_0(t)=t$, $t^{\frac{3}{2}}$ 和 $1.5\log(1+t)$, 选取样本容量为 200 个，重复模拟 1000 次，权函数取 $H_n(t)=t$.

表 2 真值 β_0 估计量 $\widehat{\beta}$ 的模拟结果

		真值 $\beta_{01}=0$				真值 $\beta_{02}=0.2$			
	$\mu_0(t)$	Bias	SSE	SEE	CP	Bias	SSE	SEE	CP
t	-0.0007	0.0762	0.0762	0.949	0.0232	0.2176	0.1269	0.955	
$t^{\frac{3}{2}}$	0.0011	0.0728	0.0728	0.952	0.0183	0.2267	0.1356	0.945	
$1.5\log(1+t)$	0.0003	0.0734	0.0734	0.950	0.0230	0.2171	0.1258	0.958	

从表 2 中同样可以看出估计值与真值的 Bias 比较小，说明估计量几乎是无偏的，且经验覆盖率 CP 也非常接近置信度为 95% 的置信水平. 所以，我们提出模型的估计方法比较合理.

5 结论

在本文中，我们对复发事件数据提出了一类半参数转移模型，模型不仅考虑了协变量的加性和乘性影响，同时假定加性影响是时间的函数．利用广义估计方程的思想，我们对模型中参数进行了估计，并证明了估计量的大样本性质．最后，通过数值模拟我们发现真值与估计值的偏差 (Bias) 非常小，几乎是无偏的，且基于正态近似的经验覆盖率 CP 也非常接近 95% 的置信度，说明我们所提的估计方法是可行的．

6 附录：渐近性质的证明

定理 3.1 的证明 为了证明 $\widehat{\theta}$ 和 $\widehat{\mu}_0(t)$ 的存在性和唯一性，只需证明 $U(\theta)=0$ 存在唯一解就可以了．定义

$$\widehat{A}(\theta)=-\frac{1}{\mathrm{n}}\partial U(\theta)/\partial\theta',$$

由于

$$\sum_{i=1}^{n}\int_0^\tau \mathrm{Y_i}(t)\mu_0(t)g(\alpha' X_i(t))\{\mathrm{Z_i}(t)-\overline{Z}(t;\alpha)\}\,\mathrm{d}H_n(t)=0_{p\times 1},$$

所以

$$U(\theta)=\sum_{i=1}^{n}\int_0^\tau\{Z_i(t)-\overline{Z}(t)\}M_i(t;\theta)\,\mathrm{d}H_n(t),$$

直接计算则可以得到

$$\begin{aligned}\widehat{A}(\theta)=&\frac{1}{n}\sum_{i=1}^{n}\int_0^\tau\{Z_i(t)-\overline{Z}(t)\}Y_i(t)\\&\cdot\left[\dot{g}\{\alpha' X_i(t)\}X_i(t)'\mu_0(t)\,\mathrm{d}H_n(t),\dot{Q}\{t,\beta' W_i(t)\}W_i(t)'\mathrm{d}H_n(t)\right]\\&+\frac{1}{\mathrm{n}}\sum_{i=1}^{n}\int_0^\tau\frac{\partial\overline{Z}(t;\beta)}{\partial\theta'}M_i(t;\theta)\,\mathrm{d}H_n(t).\end{aligned}\tag{6.1}$$

由一致强大数定律可知， $\widehat{A}(\theta)$ 关于 θ 几乎处处一致收敛于某非随机函数 $A(\theta)$. 易知， $A(\theta_0)=A$, 其中 A 由 (C6) 给出．这样，由 $\widehat{A}(\theta)$ 的一致收敛性， $A(\theta)$ 的连续性和 A 的正定性可以得到：当 n 足够大时，在 θ_0 的一个小邻域里存在唯一的 $\widehat{\theta}$ 使得 $U(\theta)=0$ 成立．由于 θ_0 的邻域可以任意小，所以同样可以得到 $\widehat{\theta}$ 的强相合性．在 [17] 中定理 2 也对 $\widehat{\theta}$ 的唯一性和强相合性进行了证明．

由 $\widehat{\theta}$ 的存在性和唯一性，可得 $\widehat{\mu}_0(t)$ 的存在性和唯一性. 类似 [16] 中定理 3.1 的证明可知 $\widehat{\mu}_0(t)$ 的相合性．由于 $Y_i(t)\{N_i(t)-Q(t,\beta' W_i(t))\}$ 可以被写成和的形式或者可以看成是由 t 的单调函数组成的，他们都是可控的，且 β 的所有成分也是可控的，则由强大数定律，在 $t\in(0,\tau)$, 可以得到 $\frac{1}{n}\sum\limits_{i=1}^{n}Y_i(t)g(\alpha' X_i(t))$ 和 $\frac{1}{n}\sum\limits_{i=1}^{n}Y_i(t)\{N_i(t)-Q(t,\beta' W_i(t))\}$

几乎处处分别收敛于 $E[Y_i(t)g(\alpha' X_i(t))]$ 和 $E[Y_i(t)\{N_i(t)-Q(t,\beta' W_i(t))\}]$. 这样由式 (2.3) 可以得到，当 $t\in(0,\tau)$ 时， $\widehat{\mu}_0(t;\theta)$ 几乎处处收敛到 $\mu_0(t;\theta)$, 其中，

$$\mu_0(t;\theta)=\frac{E[Y_i(t)\{N_i(t)-Q(t,\beta' W_i(t))\}]}{E[Y_i(t)g(\alpha' X_i(t))]},$$

这样， $\widehat{\mu}_0(t;\widehat{\theta})\equiv\widehat{\mu}_0(t)$ 几乎处处收敛于 $\mu_0(t;\theta)\equiv\mu_0(t)$. 证毕.

定理 3.2 的证明 由泰勒展开式可得：

$$U(t;\widehat{\theta})-U(t;\theta_0)=\frac{\partial U(t;\theta^*)}{\partial\theta'}(\widehat{\theta}-\theta_0)=-n\widehat{A}(t;\theta^*)(\widehat{\theta}-\theta_0),$$

其中 $\theta^*=\theta_0+\mathrm{m}(\widehat{\theta}-\theta_0)$, $\mathrm{m}\in(0,1)$.

再由 $\widehat{A}(\theta)$ 的一致收敛性， $\widehat{\theta}$ 的相合性以及 A 的正定性可得：

$$n^{\frac{1}{2}}(\widehat{\theta}-\theta_0)=-n^{-\frac{1}{2}}\widehat{A}^{-1}(t;\theta^*)[U(t;\widehat{\theta})-U(t;\theta_0)]=n^{-\frac{1}{2}}A^{-1}U(t;\theta_0)+o_p(1). \tag{6.2}$$

由 [17] 中定理 1 可知， $n^{-1/2}U(t;\theta_0)$ 弱收敛于零均值的 Gauss 过程，且协方差阵为

$$\sum=E\Big[\int_0^\tau\{Z_i(t)-\overline{Z}_i(t)\}\widehat{M_i}(t;\widehat{\theta})\,\mathrm{d}H_n(t)\Big]^{\otimes 2}.$$

详细证明过程类似参考 [3] 中附录 6.2 的证明.

那么，由式 (6.2) 可知， $n^{1/2}(\widehat{\theta}-\theta_0)$ 渐近服从均值为 **0** 的正态分布，且协方差阵为 $A^{-1}\sum A^{-1}$, 它可以被 $\widehat{A}^{-1}\widehat{\sum}\widehat{A}^{-1}$ 相合估计. 证毕.

定理 3.3 的证明 首先

$$\widehat{\mu}_0(t)-\mu_0(t)=\{\widehat{\mu}_0(t;\widehat{\theta})-\widehat{\mu}_0(t;\theta_0)\}+\{\widehat{\mu}_0(t;\theta_0)-\mu_0(t)\}, \tag{6.3}$$

运用泰勒展开式，式 (6.3) 右端第一项可以写为

$$n^{1/2}\{\widehat{\mu}_0(t;\widehat{\theta})-\widehat{\mu}_0(t;\theta_0)\}=\frac{\partial\widehat{\mu}_0(t;\theta_0)}{\partial\theta'}n^{1/2}(\widehat{\theta}-\theta_0), \tag{6.4}$$

由于

$$\widehat{\mu}_0(t;\theta)=\frac{\sum\limits_{i=1}^n Y_i(t)[N_i(t)-Q(t,\beta' W_i(t))]}{\sum\limits_{i=1}^n Y_i(t)g\{\alpha' X_i(t)\}},$$

直接计算可得

$$\frac{\partial\widehat{\mu}_0(t;\theta_0)}{\partial\theta'}=-\mathrm{B}(t;\theta). \tag{6.5}$$

再由式 (6.2) 可得

$$\begin{aligned}n^{\frac{1}{2}}(\widehat{\theta}-\theta_0)=&n^{-\frac{1}{2}}A^{-1}U(t;\theta_0)+o_p(1)\\=&A^{-1}n^{-\frac{1}{2}}\sum_{\mathrm{i}=1}^n\int_0^\tau Y_i(t)\{Z_i(t)-\overline{Z}(t;\alpha_0)\}M_i(t;\theta_0)\,\mathrm{d}H_n(t)+o_p(1),\end{aligned} \tag{6.6}$$

因此，由式 (6.4)–(6.6) 可得

$$\begin{aligned}&n^{1/2}\{\widehat{\mu}_0(t;\widehat{\theta})-\widehat{\mu}_0(t;\theta_0)\}\\=&-\mathrm{B}(t;\theta_0)A^{-1}n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^{\tau}Y_i(t)\{Z_{\mathrm{i}}(t)-\overline{Z}(t;\alpha_0)\}M_i(t;\theta_0)\,\mathrm{d}H_n(t)+o_p(1),\end{aligned}\tag{6.7}$$

另一方面可推导

$$n^{1/2}\{\widehat{\mu}_0(t;\theta_0)-\mu_0(t)\}=n^{1/2}\sum_{i=1}^{n}\frac{M_i(t;\theta_0)}{n^{-1}\sum\limits_{i=1}^{n}Y_i(t)g\{\alpha'_0X_i(t)\}}+o_p(1).\tag{6.8}$$

最后由式 (6.3)–(6.8) 可得

$$n^{1/2}\{\widehat{\mu}_0(t;\widehat{\theta})-\widehat{\mu}_0(t)\}=n^{-\frac{1}{2}}\sum_{i=1}^{n}\psi_i(t)+o_p(1).\tag{6.9}$$

对于每一个的 t, $\psi_i(t)$ 均为独立的以 0 为均值的随机向量, 则由式 (6.9) 可知 $n^{1/2}\{\widehat{\mu}_0(t;\widehat{\theta})-\mu_0(t)\}$ 渐近于均值为 0 的独立同分布的随机变量之和. 再由多元中心极限定理可知， $n^{1/2}\{\widehat{\mu}_0(t;\widehat{\theta})-\mu_0(t)\}$ 依有限维分布收敛到一个均值为 0 的 Gauss 过程. 由于 $\psi_i(t)$ 是紧的, 这样 $n^{1/2}\{\widehat{\mu}_0(t;\widehat{\theta})-\mu_0(t)\}$ 也是紧的, 且弱收敛到一个均值为 0 的 Gauss 过程, 它在 (s,t) 处的协方差函数可以写为

$$\Gamma(s,t)=E\{\psi_i(s)\psi_i(t)\}.$$

它的一个相合估计由定理 3.3 给出的 $\widehat{\Gamma}(s,t)$.

参 考 文 献

[1] Robins J, Rotnitzky. Additive Hazards Model with Multivariate Failure Time Data. *Biometrika*, 2004, 91(4): 801–818

[2] Lin D Y, Ying Z. Semiparametric Analysis of General Additive-Multiplicative Hazard Models for Counting Processes. 1995, 23(5): 1712–1734

[3] Lin D Y, Wei L J, Yang I. Semiparametric Regression for the Mean and Rate Functions of Recurrent Events. *Journal of the Royal Statistical Society*, 2000, 62(4): 711–730

[4] Huang C Y, Wang M C. Joint Modeling and Estimation for Recurrent Event Processes and Failure Time Data. *Journal of the American Statistical Association*, 2004, 99(468): 1153–1165

[5] Zeng D, Yin G, Ibrahim J G. Inference for a Class of Transformed Hazards Models. *Journal of the American Statistical Association*, 2005, 100(471): 1000–1008

[6] Chen X, Wang Q, Cai J. Semiparametric Additive Marginal Regression Models for Multiple Type Recurrent Events. *Lifetime data analysis*, 2012, 18(4): 504–527

[7] Lawless J F, Nadeau C. Some Simple Robust Methods for the Analysis of Recurrent Events. *Technometrics*, 1995, 37(2): 158–168

[8] Cook R J, Lawless J F. The Statistical Analysis of Recurrent Events. Springer Science & Business Media, 2007

[9] Ye P, Zhao X Q, Sun L Q. A Semiparametric Additive Rates Model for Multivariate Recurrent Events with Missing Event Categories. *Computational Statistics & Data Analysis*, 2015, 89(C): 39–50

[10] Zeng D, Cai J. A Semiparametric Additive Rate Model for Recurrent Events with an Informative Terminal Event. *Biometrika*, 2010, 97(3): 699–712

[11] He S, Wang F, Sun L Q. A Semiparametric Additive Rates Model for Clustered Recurrent Event Data. *Acta Mathematicae Applicatae Sinica, English Series*, 2013, 29(1): 55–62

[12] Liu Y, Wu Y, Cai J. Additive-multiplicative Rates Model for Recurrent Events. *Lifetime Data Analysis*, 2010, 16(3): 353–373

[13] Dai J J, Sun L Q, Yang Z H. A General Additive-multiplicative Rates Model for Recurrent Event Data. *Science in China Series A: Mathematics*, 2009, 52(10): 2257–2265

[14] Sun L Q, Kang F Y. An Additive-multiplicative Rates Model for Recurrent Event Data with Informative Terminal Event. *Lifetime Data Analysis*, 2013, 19(1): 117–137

[15] Schaubel D E, Zeng D, Cai J. A semiparametric Additive Rates Model for Recurrent Event Data. *Lifetime Data Analysis*, 2006, 12(4): 389–406

[16] Liu Y T, Sun L Q, Zhou Y. Additive Transformation Models for Recurrent Events. *Communications in Statistics-theory and Methods*, 2013, 42(22): 4043–4055

[17] 戴家佳, 孙六全, 杨振海. 复发事件下加性乘积比率回归模型. 中国科学 (A 辑: 数学), 2009, 39(05): 605–613

(Dai J J, Sun L Q, Yang Z H. A General Additive-Multiplicative Rates Model for Recurrent Event Data. *Science in China Series A: Mathematics*, 2009, 39(05): 605–613)

[18] Zeng D, Schaubel D E, Cai J. Semiparametric Transformation Rate Model for Recurrent Event Data. *Statistics in Biosciences*, 2011, 3(2): 187–207

[19] Hu Z, Yang Q L, Qu L Q. A class of Transformation Rate Models for Recurrent Event Data. *Science China*, 2016, 59(11): 1–18

A Class of Additive-Multiplicative Transformation Model for Recurrent Events Data

DU Yanbin

(College of Mathematics and Statistics, Hunan Normal University, Changsha 410006, China)

DAI Jiajia† JIN Jun

(College of Mathematics and Statistics, Guizhou University, Guiyang 550025)

(†E-mail: sci.jjdai@gzu.edu.cn)

Abstract Recurrent event data appear frequently in longitudinal studies. Based on the data of single types of recurrence events in biomedicine, a semi-parametric transformation model is proposed, which contains some important semiparametric models. At the same time, the model allows covariates to have additive and multiplicative effects, and additive effects change over time. By using the idea of generalized estimating equations, the unknown and nonparametric functions in the model are estimated and the consistency and asymptotic normality of estimators are proved. Finally, the numerical simulation method is used to verify the quality of the estimation. The result indicates that the model we proposed is feasible.

Key words recurrent event data; additive-multiplicative effects; transformation model; estimating equation

MR(2000) Subject Classification 62G05; 62N015

Chinese Library Classification O212. 7