

文章编号:2095-6134(2016)04-0443-11

# 失效信息随机缺失时可加危险率模型的统计推断<sup>\*</sup>

陈菲菲, 孙志华<sup>†</sup>, 叶 雪

(中国科学院大学数学科学学院, 北京 100049; 中国科学院大数据挖掘与知识管理重点实验室, 北京 100190)

(2015 年 6 月 1 日收稿; 2015 年 11 月 20 日收修改稿)

Chen F F, Sun Z H, Ye X. Statistic inference of additive hazards model when censoring indicators are missing at random[J]. Journal of University of Chinese Academy of Sciences, 2016, 33(4): 443-453.

**摘 要** 对失效信息随机缺失时的可加危险率模型的估计进行研究. 充分利用失效信息和缺失信息的概率模型的信息, 通过构建估计方程, 得到回归参数和基准累积风险函数的 3 个估计. 证明了所提估计的渐近正态性, 并进行数值模拟研究其有限样本性质. 利用数值模拟研究比较所提估计与文献中的估计的有限样本性质, 并通过分析一个实际数据验证了本文方法的有效性.

**关键词** 删失信息; 随机缺失; 可加危险率模型; 加权估计方程; 插补估计

**中图分类号:** O212.1 **文献标志码:** A **doi:** 10. 7523/j. issn. 2095-6134. 2016. 04. 003

## Statistic inference of additive hazards model when censoring indicators are missing at random

CHEN Feifei, SUN Zhihua, YE Xue

(School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China;

Key Laboratory of Big Data Mining and Knowledge Management, Chinese Academy of Sciences, Beijing 100190, China)

**Abstract** In this work, we consider a semi-parametric additive hazards regression model for right-censored data with censoring indicators missing at random. By employing the information of the response and censoring probability models, we propose three estimators of the regression coefficient and the baseline cumulative hazard function. We prove that the proposed estimators are consistent and asymptotically normal. Simulation studies are conducted to evaluate the numerical performance of the proposed estimators in comparison with the existing estimators. A real data set is analyzed to validate the effectiveness of the proposed methods.

**Key words** censoring information; missing at random; additive hazards regression model; weighting estimating equation; imputation estimating

在很多研究领域, 数据经常出现右删失的情况, 比如研究个体的中途退出或失去跟踪都会导

<sup>\*</sup> 国家自然科学基金(10901162, U1430103, 11571340)、中国科学院大学校长基金、中国科学院大数据挖掘与知识管理重点实验室开放课题和安徽省振兴计划团队项目(统计学前沿问题及应用)资助

<sup>†</sup> 通信作者, E-mail: sunzh@amss.ac.cn

致生存时间的右删失. 令  $T$  和  $C$  分别表示分布函数为  $F$  和  $G$  的生存时间和删失时间变量, 当生存数据出现右删失时, 仅观察到  $X = \min(T, C)$ . 通常假定在给定  $p$  维协变量  $\mathbf{Z}$  时,  $T$  和  $C$  条件独立.

记删失指示变量  $\delta = I(T \leq C)$ ,  $I(A)$  表示  $A$  的示性函数. 删失指示变量  $\delta$  不是总能观测到, 有时会出现缺失的现象. 例如, 某些研究个体的医学记录丢失了; 另外, 当获取死亡信息需要进行解剖时, 昂贵的解剖代价往往使得许多个体的解剖没有进行, 从而使死亡原因无法收集到, 这时失效信息也缺失了. 令  $\xi$  表示失效信息缺失的指示变量, 当观测到  $\delta$  时  $\xi = 1$ , 否则  $\xi = 0$ .

若  $\delta$  缺失的概率既与  $\delta$  本身的值无关, 又与观测到的变量  $X$  和  $\mathbf{Z}$  的值无关, 即  $P\{\xi = 1 \mid \delta, X, \mathbf{Z}\} = P\{\xi = 1\}$ , 则称  $\delta$  的缺失为完全随机缺失 (MCAR); 若  $\delta$  缺失的概率仅仅与  $X$  和  $\mathbf{Z}$  的值相关, 但是与  $\delta$  本身的值无关, 即  $P\{\xi = 1 \mid \delta, X, \mathbf{Z}\} = P\{\xi = 1 \mid X, \mathbf{Z}\}$ , 则称  $\delta$  的缺失为随机缺失 (MAR); 若  $\delta$  的缺失与  $\delta$  本身有关, 则称  $\delta$  的缺失为非随机缺失 (MNAR). 通常假设数据的缺失是 MAR 的, 且在很多情形下这是合乎情理的<sup>[1]</sup>. 大量的统计文献在做统计推断时将数据的随机缺失假设作为基本准则, 例如文献[2]对失效原因随机缺失的竞争风险数据进行研究; 文献[3]对响应变量随机缺失的半参数回归模型进行研究; 文献[4]研究死亡原因随机缺失时生存函数的半参数估计. 本文考虑失效信息  $\delta$  是随机缺失的, 关于失效信息缺失的研究也有很多, 如文献[5-8]等.

本文中, 我们利用可加危险率模型研究协变量  $\mathbf{Z}$  对生存时间  $T$  的影响. 可加危险率模型是一种较为常见的半参数回归模型, 在可加危险率回归模型下, 条件风险率函数  $\lambda(t \mid \mathbf{Z}) = \lambda_0(t) + \beta'_0 \mathbf{Z}$ , 其中  $\lambda_0(t)$  是未知的基准风险函数,  $\beta_0$  是  $p$  维未知回归参数. 关于可加危险率模型的研究有很多, 可参考文献[9-11]等.

我们的目的是估计未知的回归参数  $\beta_0$  和基准风险函数  $\lambda_0(t)$ . 对可加危险率模型, 当失效信息随机缺失时, 文献[12]采用逆概率加权的思想, 构建关于未知参数  $\beta_0$  和  $\lambda_0(t)$  的简单加权估计方程组和完全增广加权估计方程组, 并证明了估计的渐近性质; 文献[13]亦对同种类型的数据进行研究, 通过一种基于核的插补方法 (Kernel-assisted imputation estimating method) 得到一组关

于  $\beta_0$  和  $\lambda_0(t)$  的估计方程组, 从而得到  $\beta_0$  和  $\lambda_0(t)$  的估计. 上述估计方法在构建估计方程组时, 均使用了条件概率  $P\{\delta = 1 \mid X, \mathbf{Z}\}$  的非参数核估计. 我们知道, 非参数核方法受到“维数祸根”的困扰, 从而在一定程度上限制了方法在协变量维数比较高时的使用.

我们发现, 有时是可以获得关于删失模型  $P\{\delta = 1 \mid X, \mathbf{Z}\}$  的信息的, 比如当生存时间和删失时间都服从威布尔分布时, 根据  $P\{\delta = 1 \mid X = x\} = \frac{\lambda_t(x)}{\lambda_t(x) + \lambda_c(x)} = \frac{1}{1 + \lambda_c(x)/\lambda_t(x)}$ , 其中  $\lambda_t(x)$  和  $\lambda_c(x)$  分别表示  $T$  和  $C$  的风险函数, 文献[14]表明  $\theta_1/(\theta_1 + x^{\theta_2})$ ,  $\theta_1 > 0$ ,  $\theta_2 \in R$  是一个可供选择的模型. 更多关于对  $P\{\delta = 1 \mid X, \mathbf{Z}\}$  进行参数模型假定的细节可参考文献[14]. 显然, 假定  $P\{\delta = 1 \mid X, \mathbf{Z}\}$  具有参数形式, 就可避免非参数光滑方法的使用, 从而避免“维数祸根”的问题. 即使协变量的维数比较高, 本文所提的方法也是有效的.

## 1 模型介绍

假设我们观测到一个独立同分布的样本  $\{X_i, \xi_i, \delta_i, \mathbf{Z}_i\}_{i=1}^n$ . 进一步假定参数概率模型:

$$\pi(\mathbf{V}_i, \boldsymbol{\alpha}) = P\{\xi_i = 1 \mid \mathbf{V}_i\},$$

$$\omega(\mathbf{V}_i, \boldsymbol{\gamma}) = P\{\delta_i = 1 \mid \mathbf{V}_i\},$$

其中,  $\pi(\cdot, \boldsymbol{\alpha})$  和  $\omega(\cdot, \boldsymbol{\gamma})$  表示形式已知的函数,  $\boldsymbol{\alpha}$  和  $\boldsymbol{\gamma}$  表示未知参数,  $\mathbf{V}_i = (X_i, \mathbf{Z}_i')'$ . 对可加危险率回归模型, 条件危险率函数有如下形式

$$\lambda(t \mid \mathbf{Z}) = \lambda_0(t) + \beta'_0 \mathbf{Z}, \quad (1)$$

其中  $\lambda_0(t)$  是未知的基准危险率函数,  $\beta_0$  是  $p$  维未知回归参数.

显然  $\pi(\mathbf{V}_i, \boldsymbol{\alpha}) = E\{\xi_i \mid \mathbf{V}_i\}$ ,  $\omega(\mathbf{V}_i, \boldsymbol{\gamma}) = E\{\delta_i \mid \mathbf{V}_i\}$  分别表示给定  $\mathbf{V}_i$  时  $\xi_i$  和  $\delta_i$  的条件期望, 因而可使用已有的二元数据分析的方法对缺失模型和删失模型进行分析, 可假定  $\pi(\mathbf{V}_i, \boldsymbol{\alpha})$  和  $\omega(\mathbf{V}_i, \boldsymbol{\gamma})$  满足常见的逻辑回归模型. 另外, 文献[14]给出删失参数模型需满足  $P\{\delta = 1 \mid X = x\}$

$= \frac{\lambda_t(x)}{\lambda_t(x) + \lambda_c(x)}$ . 该文同时还给出了一些具体可行的参数模型假设, 比如当  $\lambda_c(x)$  通过链接函数  $\Psi(x, \boldsymbol{\gamma})$  对  $\lambda_t(x)$  产生乘积效应, 即  $\lambda_t(x) = \Psi(x, \boldsymbol{\gamma})\lambda_c(x)$ , 其中  $\boldsymbol{\gamma}$  是未知参数, 那么可给出  $P\{\delta = 1 \mid X = x\} = \frac{\Psi(x, \boldsymbol{\gamma})}{1 + \Psi(x, \boldsymbol{\gamma})}$ . 更多关于对  $P\{\delta$

$= 1 \mid X\}$  进行参数模型假定的细节可参考文献 [14].

对可加危险率模型,若失效信息没有缺失,令  $N_i(t) = I(X_i \leq t, \delta_i = 1)$ ,  $N_i^u(t) = I(X_i \leq t)$ ,  $Y_i(t) = I(X_i \geq t)$ , 那么

$$\begin{aligned} M_i(t) &= N_i(t) - \int_0^t Y_i(s) \{ \lambda_0(s) + \beta'_0 Z_i \} ds \\ &= \delta_i N_i^u(t) - \int_0^t Y_i(s) \{ \lambda_0(s) + \beta'_0 Z_i \} ds, \\ (i &= 1, 2, \dots, n). \end{aligned}$$

是零均值的鞅过程. 根据鞅的性质, 可得到关于未知回归参数  $\beta_0$  的估计方程<sup>[9]</sup>

$$\sum_{i=1}^n \int_0^\tau \{ Z_i - \bar{Z}(t) \} [dN_i(t) - Y_i(t) \beta'_0 Z_i dt] = 0, \quad (2)$$

其中  $\bar{Z}(t) = \frac{\sum_{i=1}^n Z_i Y_i(t)}{\sum_{i=1}^n Y_i(t)}$ ,  $\tau$  是给定的满足  $P(X_i \geq \tau) > 0$  且  $\Lambda_0(\tau) < \infty$  的正常数, 一般取  $\tau = \max\{X_i\}$ , 其中  $\Lambda_0(t) = \int_0^t \lambda_0(s) ds$  表示基准累积风险函数.

## 2 参数概率模型的估计

这一节考虑参数概率模型中未知参数  $\alpha$  和  $\gamma$  的估计. 因为  $\xi$  和  $\delta$  均服从条件伯努利分布, 可以构建如下的似然函数:

$$L_n(\alpha) = \prod_{i=1}^n \{ \pi(V_i, \alpha) \}^{\xi_i} \{ 1 - \pi(V_i, \alpha) \}^{1-\xi_i}, \quad (3)$$

$$L_n(\gamma) = \prod_{i=1}^n \{ \omega(V_i, \gamma) \}^{\xi_i \delta_i} \{ 1 - \omega(V_i, \gamma) \}^{\xi_i (1-\delta_i)}. \quad (4)$$

求解使上面似然函数达到最大的  $\alpha$  和  $\gamma$ , 分别记为  $\hat{\alpha}_n$  和  $\hat{\gamma}_n$ . 在估计  $\hat{\gamma}_n$  时, 我们只用到对应  $\xi_i = 1$  的观测数据, 也即只使用了失效信息完全观察到的数据. 在随机缺失条件下, 由上面似然函数得到的估计方程是相合的. 下面引理给出估计的渐近正态性.

**引理 2.1** 若附录中的条件 (C1) — (C3) 成立, 则有

$$\begin{aligned} \sqrt{n}(\hat{\alpha}_n - \alpha) &= \frac{V_\alpha^{-1}}{\sqrt{n}} \sum_{i=1}^n \frac{\pi'(V_i, \alpha) \xi_i - \pi(V_i, \alpha)}{\pi(V_i, \alpha) (1 - \pi(V_i, \alpha))} + o_p(1), \\ \sqrt{n}(\hat{\gamma}_n - \gamma) &= \frac{V_\gamma^{-1}}{\sqrt{n}} \sum_{i=1}^n \frac{\omega'(V_i, \gamma) \delta_i - \omega(V_i, \gamma)}{\omega(V_i, \gamma) (1 - \omega(V_i, \gamma))} \xi_i + o_p(1), \end{aligned}$$

这里

$$V_\alpha = E \left[ \frac{\pi'(V_i, \alpha)}{\pi(V_i, \alpha)} \frac{\pi'(V_i, \alpha)^T}{1 - \pi(V_i, \alpha)} \right],$$

$$V_\gamma = E \left[ \frac{\omega'(V_i, \gamma)}{\omega(V_i, \gamma)} \frac{\omega'(V_i, \gamma)^T}{1 - \omega(V_i, \gamma)} \pi(V_i, \alpha) \right],$$

$\pi'(V_i, \alpha)$  和  $\omega'(V_i, \gamma)$  分别为  $\pi(V_i, \alpha)$  和  $\omega(V_i, \gamma)$  关于  $\alpha$  和  $\gamma$  的一阶导数.

## 3 逆概率加权估计

当失效信息随机缺失时, 因为有些  $\delta_i$  是缺失的, 因此估计方程 (2) 不能直接使用. 利用逆概率加权方法, 可以构建零均值的随机过程

$$M_i^{[1]}(t) = \frac{\xi_i}{\pi(V_i, \alpha)} \left[ N_i(t) - \int_0^t Y_i(s) \{ \lambda_0(s) + \beta'_0 Z_i \} ds \right],$$

进一步可构建下面的估计方程组:

$$\sum_{i=1}^n \int_0^\tau \frac{\xi_i}{\pi(V_i, \hat{\alpha}_n)} Z_i [dN_i(t) - Y_i(t) \beta'_0 Z_i dt - Y_i(t) d\Lambda_0(t)] = 0, \quad (5)$$

$$\sum_{i=1}^n \frac{\xi_i}{\pi(V_i, \hat{\alpha}_n)} [dN_i(t) - Y_i(t) \beta'_0 Z_i dt - Y_i(t) d\Lambda_0(t)] = 0, \quad (6)$$

对任意的向量  $a$ , 记  $a^{\otimes 2} = aa'$ . 解上述方程组, 可得  $\beta_0$  和  $\Lambda_0$  的逆概率加权估计有如下形式:

$$\hat{\beta}_{1n} = \left[ \sum_{i=1}^n \int_0^\tau \frac{\xi_i}{\pi(V_i, \hat{\alpha}_n)} Y_i(t) \{ Z_i - \bar{Z}^*(t) \}^{\otimes 2} dt \right]^{-1}$$

$$\sum_{i=1}^n \int_0^\tau \frac{\xi_i}{\pi(V_i, \hat{\alpha}_n)} \{ Z_i - \bar{Z}^*(t) \} dN_i(t), \quad (7)$$

$$\hat{\Lambda}_{1n}(t) = \int_0^t \frac{\sum_{i=1}^n \frac{\xi_i}{\pi(V_i, \hat{\alpha}_n)} [dN_i(s) - Y_i(s) \hat{\beta}_{1n}' Z_i ds]}{\sum_{i=1}^n \frac{\xi_i}{\pi(V_i, \hat{\alpha}_n)} Y_i(s)}, \quad (8)$$

其中

$$\bar{Z}^*(t) = \frac{\sum_{i=1}^n \frac{\xi_i}{\pi(V_i, \hat{\alpha}_n)} Z_i Y_i(t)}{\sum_{i=1}^n \frac{\xi_i}{\pi(V_i, \hat{\alpha}_n)} Y_i(t)}.$$

首先给出  $\hat{\beta}_{1n}$  的渐近正态性.

**定理 3.1** 若附录中条件 (C1) — (C5) 成立, 那么

$$\sqrt{n}(\hat{\beta}_{1n} - \beta_0) = A^{-1} (n^{-1/2} \sum_{i=1}^n \phi_{1i}) + o_p(1),$$

其中

$$\begin{aligned} \boldsymbol{A} &= E\left[\int_0^\tau Y_i(t)\{\boldsymbol{Z}_i - \bar{\boldsymbol{z}}(t)\}^{\otimes 2} dt\right], \\ \bar{\boldsymbol{z}}(t) &= E[\boldsymbol{Z}_i Y_i(t)]/E[Y_i(t)], \\ \boldsymbol{\phi}_{1i} &= \int_0^\tau \frac{\xi_i}{\pi(\boldsymbol{V}_i, \boldsymbol{\alpha})} \{\boldsymbol{Z}_i - \bar{\boldsymbol{z}}(t)\} dM_i(t) + \boldsymbol{B}_{1\alpha} \boldsymbol{V}_\alpha^{-1} \boldsymbol{S}_{\alpha i}, \\ \boldsymbol{B}_{1\alpha} &= E\left[-\int_0^\tau \{\boldsymbol{Z}_i - \bar{\boldsymbol{z}}(t)\} \frac{\xi_i \boldsymbol{\pi}'(\boldsymbol{V}_i, \boldsymbol{\alpha})^T}{\pi^2(\boldsymbol{V}_i, \boldsymbol{\alpha})} dM_i(t)\right], \\ \boldsymbol{S}_{\alpha i} &= \frac{\boldsymbol{\pi}'(\boldsymbol{V}_i, \boldsymbol{\alpha}) \xi_i - \pi(\boldsymbol{V}_i, \boldsymbol{\alpha})}{\pi(\boldsymbol{V}_i, \boldsymbol{\alpha}) 1 - \pi(\boldsymbol{V}_i, \boldsymbol{\alpha})}. \end{aligned}$$

根据定理 3.1, 由中心极限定理可知  $\sqrt{n}(\hat{\boldsymbol{\beta}}_{1n} - \boldsymbol{\beta}_0)$  依分布收敛于均值为  $\boldsymbol{0}$ 、协方差矩阵为  $\boldsymbol{A}^{-1}E(\boldsymbol{\phi}_{1i}\boldsymbol{\phi}_{1i}^T)(\boldsymbol{A}^{-1})^T$  的正态分布.

采用传统的 plug-in 方法, 可得到渐近方差  $\boldsymbol{A}^{-1}E(\boldsymbol{\phi}_{1i}\boldsymbol{\phi}_{1i}^T)(\boldsymbol{A}^{-1})^T$  的相合估计为

$$\hat{\boldsymbol{A}}^{-1}\left(\frac{1}{n}\sum_{i=1}^n \hat{\boldsymbol{\phi}}_{1i}\hat{\boldsymbol{\phi}}_{1i}^T\right)(\hat{\boldsymbol{A}}^{-1})^T, \text{ 其中}$$

$$\begin{aligned} \hat{\boldsymbol{A}} &= \frac{1}{n}\int_0^\tau Y_i(t)\{\boldsymbol{Z}_i - \bar{\boldsymbol{Z}}(t)\}^{\otimes 2} dt, \\ \hat{\boldsymbol{\phi}}_{1i} &= \int_0^\tau \frac{\xi_i}{\pi(\boldsymbol{V}_i, \hat{\boldsymbol{\alpha}}_n)} \{\boldsymbol{Z}_i - \bar{\boldsymbol{Z}}(t)\} [dN_i(t) - Y_i(t) \\ &\quad (d\hat{\Lambda}_{1n}(t) + \hat{\boldsymbol{\beta}}'_{1n}\boldsymbol{Z}_i dt)] + \hat{\boldsymbol{B}}_{1\alpha} \hat{\boldsymbol{V}}_\alpha^{-1} \hat{\boldsymbol{S}}_{\alpha i}, \\ \hat{\boldsymbol{B}}_{1\alpha} &= \frac{1}{n}\sum_{i=1}^n \left[-\int_0^\tau \{\boldsymbol{Z}_i - \bar{\boldsymbol{Z}}(t)\} \frac{\xi_i \boldsymbol{\pi}'(\boldsymbol{V}_i, \hat{\boldsymbol{\alpha}}_n)^T}{\pi^2(\boldsymbol{V}_i, \hat{\boldsymbol{\alpha}}_n)} \right. \\ &\quad \left. \{dN_i(t) - Y_i(t)(d\hat{\Lambda}_{1n}(t) + \hat{\boldsymbol{\beta}}'_{1n}\boldsymbol{Z}_i dt)\}\right], \\ \hat{\boldsymbol{V}}_\alpha &= \frac{1}{n}\sum_{i=1}^n \left[\frac{\boldsymbol{\pi}'(\boldsymbol{V}_i, \hat{\boldsymbol{\alpha}}_n)}{\pi(\boldsymbol{V}_i, \hat{\boldsymbol{\alpha}}_n)} \frac{\boldsymbol{\pi}'(\boldsymbol{V}_i, \hat{\boldsymbol{\alpha}}_n)^T}{1 - \pi(\boldsymbol{V}_i, \hat{\boldsymbol{\alpha}}_n)}\right], \\ \hat{\boldsymbol{S}}_{\alpha i} &= \frac{\boldsymbol{\pi}'(\boldsymbol{V}_i, \hat{\boldsymbol{\alpha}}_n) \xi_i - \pi(\boldsymbol{V}_i, \hat{\boldsymbol{\alpha}}_n)}{\pi(\boldsymbol{V}_i, \hat{\boldsymbol{\alpha}}_n) 1 - \pi(\boldsymbol{V}_i, \hat{\boldsymbol{\alpha}}_n)}. \end{aligned}$$

下面考察  $\hat{\Lambda}_{1n}(t)$  的渐近正态性.

**定理 3.2** 在定理 3.1 成立的条件下, 有

$$\sqrt{n}(\hat{\Lambda}_{1n}(t) - \Lambda_0(t)) = n^{-1/2} \sum_{i=1}^n \psi_{1i}(t) + o_p(1),$$

其中

$$\begin{aligned} \psi_{1i}(t) &= \int_0^t \frac{\xi_i}{\pi(\boldsymbol{V}_i, \boldsymbol{\alpha})} dM_i(s) - \boldsymbol{C}(t)^T \boldsymbol{A}^{-1} \boldsymbol{\phi}_{1i} - \boldsymbol{D}_{1\alpha}(t) \boldsymbol{V}_\alpha^{-1} \boldsymbol{S}_{\alpha i}, \\ \boldsymbol{\mu}(t) &= E[Y_i(t)], \boldsymbol{C}(t) = \int_0^t \bar{\boldsymbol{z}}(s) ds, \\ \boldsymbol{D}_{1\alpha}(t) &= E\left[\int_0^t \frac{\xi_i \boldsymbol{\pi}'(\boldsymbol{V}_i, \boldsymbol{\alpha})^T}{\pi^2(\boldsymbol{V}_i, \boldsymbol{\alpha})} dM_i(s)\right]. \end{aligned}$$

根据定理 3.2,  $\sqrt{n}(\hat{\Lambda}_{1n}(t) - \Lambda_0(t))$  依分布收敛于一个均值为 0 的高斯过程, 且在  $(s, t)$  的渐近协方差函数为  $E(\psi_{1i}(s)\psi_{1i}(t))$ . 易得

$$E(\psi_{1i}(s)\psi_{1i}(t)) \text{ 的相合估计为 } \frac{1}{n} \sum_{i=1}^n (\hat{\psi}_{1i}(s)\hat{\psi}_{1i}(t)),$$

其中

$$\begin{aligned} \hat{\psi}_{1i}(t) &= \int_0^t \frac{\xi_i}{\pi(\boldsymbol{V}_i, \hat{\boldsymbol{\alpha}}_n)} [dN_i(s) - Y_i(s)\{d\hat{\Lambda}_{1n}(s) + \hat{\boldsymbol{\beta}}'_{1n}\boldsymbol{Z}_i ds\}] \\ &\quad - \frac{1}{n} \sum_{i=1}^n Y_i(s) \\ &\quad \hat{\boldsymbol{C}}(t)^T \hat{\boldsymbol{A}}^{-1} \hat{\boldsymbol{\phi}}_{1i} - \hat{\boldsymbol{D}}_{1\alpha}(t) \hat{\boldsymbol{V}}_\alpha^{-1} \hat{\boldsymbol{S}}_{\alpha i}, \\ \hat{\boldsymbol{C}}(t) &= \int_0^t \bar{\boldsymbol{Z}}(s) ds, \hat{\boldsymbol{\mu}}(t) = \frac{1}{n} \sum_{i=1}^n Y_i(t), \\ \hat{\boldsymbol{D}}_{1\alpha}(t) &= \frac{1}{n} \sum_{i=1}^n \left[\int_0^t \frac{\xi_i \boldsymbol{\pi}'(\boldsymbol{V}_i, \hat{\boldsymbol{\alpha}}_n)^T}{\pi^2(\boldsymbol{V}_i, \hat{\boldsymbol{\alpha}}_n)} dN_i(s) - \right. \\ &\quad \left. Y_i(s)(d\hat{\Lambda}_{1n}(s) + \hat{\boldsymbol{\beta}}'_{1n}\boldsymbol{Z}_i ds)\right]. \end{aligned}$$

### 4 增广加权估计

上面构建的逆概率加权估计只用到了完全观测样本(对应  $\xi_i = 1$  的个体)的信息. 忽略不完全观测样本( $\xi_i = 0$  的个体)可能会使结果的有效性降低. 尤其当缺失比率较高时, 这种现象更为严重. 注意到

$$\begin{aligned} M_i^{[2]}(t) &= \frac{\xi_i}{\pi(\boldsymbol{V}_i, \boldsymbol{\alpha})} N_i(t) + \\ &\quad \left(1 - \frac{\xi_i}{\pi(\boldsymbol{V}_i, \boldsymbol{\alpha})}\right) \omega(\boldsymbol{V}_i, \boldsymbol{\gamma}) N_i^u(t) - \\ &\quad \int_0^t Y_i(s) \{\lambda_0(s) + \boldsymbol{\beta}'_0 \boldsymbol{Z}_i\} ds, \end{aligned}$$

是零均值的随机过程.

考虑下面的关于  $\boldsymbol{\beta}_0$  和  $\Lambda_0$  的增广加权估计方程:

$$\begin{aligned} \sum_{i=1}^n \int_0^\tau Z_i \left[ \frac{\xi_i}{\pi(\boldsymbol{V}_i, \hat{\boldsymbol{\alpha}}_n)} dN_i(t) + \left(1 - \frac{\xi_i}{\pi(\boldsymbol{V}_i, \hat{\boldsymbol{\alpha}}_n)}\right) \omega(\boldsymbol{V}_i, \hat{\boldsymbol{\gamma}}_n) \right. \\ \left. dN_i^u(t) - Y_i(t) \boldsymbol{\beta}'_0 \boldsymbol{Z}_i dt - Y_i(t) d\Lambda_0(t) \right] = 0, \end{aligned} \tag{9}$$

$$\begin{aligned} \sum_{i=1}^n \left[ \frac{\xi_i}{\pi(\boldsymbol{V}_i, \hat{\boldsymbol{\alpha}}_n)} dN_i(t) + \left(1 - \frac{\xi_i}{\pi(\boldsymbol{V}_i, \hat{\boldsymbol{\alpha}}_n)}\right) \omega(\boldsymbol{V}_i, \hat{\boldsymbol{\gamma}}_n) \right. \\ \left. dN_i^u(t) - Y_i(t) \boldsymbol{\beta}'_0 \boldsymbol{Z}_i dt - Y_i(t) d\Lambda_0(t) \right] = 0, \end{aligned} \tag{10}$$

解上述方程组,可得  $\beta_0$  和  $\Lambda_0$  的增广加权估计有如下形式:

$$\hat{\beta}_{2n} = \left[ \sum_{i=1}^n \int_0^\tau Y_i(t) \{Z_i - \bar{Z}(t)\}^{\otimes 2} dt \right]^{-1} \sum_{i=1}^n \int_0^\tau \{Z_i - \bar{Z}(t)\} \left[ \frac{\xi_i}{\pi(V_i, \hat{\alpha}_n)} dN_i(t) + \left(1 - \frac{\xi_i}{\pi(V_i, \hat{\alpha}_n)}\right) \omega(V_i, \hat{\gamma}_n) dN_i^u(t) \right], \quad (11)$$

$$\hat{\Lambda}_{2n}(t) = \int_0^t \sum_{i=1}^n \left\{ \frac{\xi_i}{\pi(V_i, \hat{\alpha}_n)} dN_i(s) + \left(1 - \frac{\xi_i}{\pi(V_i, \hat{\alpha}_n)}\right) \omega(V_i, \hat{\gamma}_n) dN_i^u(s) - Y_i(s) \hat{\beta}_{2n}' Z_i ds \right\} / \sum_{i=1}^n Y_i(s). \quad (12)$$

我们给出记号:

$$S_{\gamma i} = \frac{\omega'(V_i, \gamma) \delta_i - \omega(V_i, \gamma)}{\omega(V_i, \gamma) 1 - \omega(V_i, \gamma)} \xi_i,$$

$$\phi_{2i} = \int_0^\tau \{Z_i - \bar{z}(t)\} dM_i^{[2]}(t) + B_{2\alpha} V_\alpha^{-1} S_{\alpha i} + B_{2\gamma} V_\gamma^{-1} S_{\gamma i},$$

$$B_{2\alpha} = E \left[ \int_0^\tau \{Z_i - \bar{z}(t)\} \frac{\xi_i \pi'(V_i, \alpha)^T}{\pi^2(V_i, \alpha)} (\omega(V_i, \gamma) - \delta_i) dN_i^u(t) \right],$$

$$B_{2\gamma} = E \left[ \int_0^\tau \{Z_i - \bar{z}(t)\} \left(1 - \frac{\xi_i}{\pi(V_i, \alpha)}\right) \omega'(V_i, \gamma)^T dN_i^u(t) \right].$$

下面定理给出  $\hat{\beta}_{2n}$  的渐近正态性.

**定理 4.1** 在定理 3.1 成立的条件下,

$$\sqrt{n}(\hat{\beta}_{2n} - \beta_0) = A^{-1} \left( n^{-1/2} \sum_{i=1}^n \phi_{2i} \right) + o_p(1).$$

由定理 4.1 可得  $\sqrt{n}(\hat{\beta}_{2n} - \beta_0)$  依分布收敛于均值为 0、协方差矩阵为  $A^{-1} E(\phi_{2i} \phi_{2i}^T) (A^{-1})^T$  的正态分布. 我们可以给出渐近方差  $A^{-1} E(\phi_{2i} \phi_{2i}^T) (A^{-1})^T$  的相合估计:

$$\hat{A}^{-1} \left( \frac{1}{n} \sum_{i=1}^n \hat{\phi}_{2i} \hat{\phi}_{2i}^T \right) (\hat{A}^{-1})^T, \text{ 其中}$$

$$\hat{\phi}_{2i} = \int_0^\tau \{Z_i - \bar{Z}(t)\} \left[ \frac{\xi_i}{\pi(V_i, \hat{\alpha}_n)} dN_i(t) + \left(1 - \frac{\xi_i}{\pi(V_i, \hat{\alpha}_n)}\right) \omega(V_i, \hat{\gamma}_n) dN_i^u(t) - Y_i(t) (\hat{\Lambda}_{2n}(t) + \hat{\beta}_{2n}' Z_i dt) \right] +$$

$$\hat{B}_{2\alpha} \hat{V}_\alpha^{-1} \hat{S}_{\alpha i} + \hat{B}_{2\gamma} \hat{V}_\gamma^{-1} \hat{S}_{\gamma i},$$

$$\hat{B}_{2\alpha} = \frac{1}{n} \sum_{i=1}^n \left[ \int_0^\tau \{Z_i - \bar{Z}(t)\} \frac{\xi_i \pi'(V_i, \hat{\alpha}_n)^T}{\pi^2(V_i, \hat{\alpha}_n)} (\omega(V_i, \hat{\gamma}_n) - \delta_i) dN_i^u(t) \right],$$

$$\hat{B}_{2\gamma} = \frac{1}{n} \sum_{i=1}^n \left[ \int_0^\tau \{Z_i - \bar{Z}(t)\} \left(1 - \frac{\xi_i}{\pi(V_i, \hat{\alpha}_n)}\right) \omega'(V_i, \hat{\gamma}_n)^T dN_i^u(t) \right].$$

$$\omega'(V_i, \hat{\gamma}_n)^T dN_i^u(t) \Big],$$

$$\hat{V}_\gamma = \frac{1}{n} \sum_{i=1}^n \left[ \frac{\omega'(V_i, \hat{\gamma}_n)}{\omega(V_i, \hat{\gamma}_n)} \frac{\omega'(V_i, \hat{\gamma}_n)^T}{1 - \omega(V_i, \hat{\gamma}_n)} \pi(V_i, \hat{\alpha}_n) \right],$$

$$\hat{S}_{\gamma i} = \frac{\omega'(V_i, \hat{\gamma}_n)}{\omega(V_i, \hat{\gamma}_n)} \frac{\delta_i - \omega(V_i, \hat{\gamma}_n)}{1 - \omega(V_i, \hat{\gamma}_n)} \xi_i.$$

下面的定理 4.2 给出  $\hat{\Lambda}_{2n}(t)$  的渐近性质.

**定理 4.2** 在定理 3.1 成立的条件下,有

$$\sqrt{n}(\hat{\Lambda}_{2n}(t) - \Lambda_0(t)) = n^{-1/2} \sum_{i=1}^n \psi_{2i}(t) + o_p(1),$$

其中

$$\psi_{2i}(t) = \int_0^t \frac{dM_i^{[2]}(s)}{\mu(s)} - C(t)^T A^{-1} \phi_{2i} + D_{2\alpha}(t) V_\alpha^{-1} S_{\alpha i} + D_{2\gamma}(t) V_\gamma^{-1} S_{\gamma i},$$

$$D_{2\alpha}(t) = E \left[ \int_0^t \frac{\xi_i \pi'(V_i, \alpha)^T}{\pi^2(V_i, \alpha) \mu(s)} (\omega(V_i, \gamma) - \delta_i) dN_i^u(s) \right],$$

$$D_{2\gamma}(t) = E \left[ \int_0^t \left(1 - \frac{\xi_i}{\pi(V_i, \alpha)}\right) \frac{\omega'(V_i, \gamma)^T}{\mu(s)} dN_i^u(s) \right].$$

由定理 4.2, 进一步可得到  $\sqrt{n}(\hat{\Lambda}_{2n}(t) - \Lambda_0(t))$  依分布收敛于一个均值为 0 的高斯过程, 且在  $(s, t)$  的渐近协方差函数为  $E(\psi_{2i}(s) \psi_{2i}(t))$ . 我们给出  $E(\psi_{2i}(s) \psi_{2i}(t))$  的相合估计  $\frac{1}{n} \sum_{i=1}^n (\hat{\psi}_{2i}(s) \hat{\psi}_{2i}(t))$ , 其中

$$\hat{\psi}_{2i}(t) = \int_0^t \left\{ \frac{\xi_i}{\pi(V_i, \hat{\alpha}_n)} dN_i(s) + \left(1 - \frac{\xi_i}{\pi(V_i, \hat{\alpha}_n)}\right) \omega(V_i, \hat{\gamma}_n) dN_i^u(s) - Y_i(s) (\hat{\Lambda}_{2n}(s) + \hat{\beta}_{2n}' Z_i ds) \right\} -$$

$$\hat{C}(t)^T \hat{A}^{-1} \hat{\phi}_{2i} + \hat{D}_{2\alpha}(t) \hat{V}_\alpha^{-1} \hat{S}_{\alpha i} + \hat{D}_{2\gamma}(t) \hat{V}_\gamma^{-1} \hat{S}_{\gamma i},$$

$$\hat{D}_{2\alpha}(t) = \frac{1}{n} \sum_{i=1}^n \left[ \int_0^t \frac{\xi_i \pi'(V_i, \hat{\alpha}_n)^T}{\pi^2(V_i, \hat{\alpha}_n) \hat{\mu}(s)} (\omega(V_i, \hat{\gamma}_n) - \delta_i) dN_i^u(s) \right],$$

$$\hat{D}_{2\gamma}(t) = \frac{1}{n} \sum_{i=1}^n \left[ \int_0^t \left(1 - \frac{\xi_i}{\pi(V_i, \hat{\alpha}_n)}\right) \frac{\omega'(V_i, \hat{\gamma}_n)^T}{\hat{\mu}(s)} dN_i^u(s) \right].$$

## 5 基于模型的插补估计

在处理缺失数据时,插补是一种常用的方法. 下面考虑一种基于模型的插补方法. 当  $\xi_i = 1$ ,  $I(X_i \leq t, \delta_i = 1) = \delta_i I(X_i \leq t)$  是可以观察到的. 但是  $\xi_i = 0$  时, 因为  $\delta_i$  的值缺失了, 因此  $I(X_i \leq t, \delta_i = 1)$  无法由观察数据得到. 此时, 对  $\delta_i$  用其条

件均值  $\omega(\mathbf{V}_i, \boldsymbol{\gamma})$  替代, 以此得到下面零均值的随机过程

$$M_i^{[3]}(t) = \xi_i N_i(t) + (1 - \xi_i) \omega(\mathbf{V}_i, \boldsymbol{\gamma}) N_i^u(t) - \int_0^t Y_i(s) \{ \lambda_0(s) + \boldsymbol{\beta}'_0 \mathbf{Z}_i \} ds,$$

那么可得如下关于  $\boldsymbol{\beta}_0$  和  $\Lambda_0$  的估计方程:

$$\sum_{i=1}^n \int_0^\tau \mathbf{Z}_i [\xi_i dN_i(t) + (1 - \xi_i) \omega(\mathbf{V}_i, \hat{\boldsymbol{\gamma}}_n) dN_i^u(t) - Y_i(t) \boldsymbol{\beta}'_0 \mathbf{Z}_i dt - Y_i(t) d\Lambda_0(t)] = 0, \quad (13)$$

$$\sum_{i=1}^n [\xi_i dN_i(t) + (1 - \xi_i) \omega(\mathbf{V}_i, \hat{\boldsymbol{\gamma}}_n) dN_i^u(t) - Y_i(t) \boldsymbol{\beta}'_0 \mathbf{Z}_i dt - Y_i(t) d\Lambda_0(t)] = 0, \quad (14)$$

解上述方程组, 可得  $\boldsymbol{\beta}_0$  和  $\Lambda_0$  的基于模型的插补估计有如下形式:

$$\hat{\boldsymbol{\beta}}_{3n} = \left[ \sum_{i=1}^n \int_0^\tau Y_i(t) \{ \mathbf{Z}_i - \bar{\mathbf{Z}}(t) \}^{\otimes 2} dt \right]^{-1} \sum_{i=1}^n \int_0^\tau \{ \mathbf{Z}_i - \bar{\mathbf{Z}}(t) \} [\xi_i dN_i(t) + (1 - \xi_i) \omega(\mathbf{V}_i, \hat{\boldsymbol{\gamma}}_n) dN_i^u(t)], \quad (15)$$

$$\hat{\Lambda}_{3n}(t) = \int_0^t \left\{ \sum_{i=1}^n \xi_i dN_i(s) + (1 - \xi_i) \omega(\mathbf{V}_i, \hat{\boldsymbol{\gamma}}_n) dN_i^u(s) - Y_i(s) \hat{\boldsymbol{\beta}}'_{3n} \mathbf{Z}_i ds \right\} / \sum_{i=1}^n Y_i(s). \quad (16)$$

与定理 4.1 和定理 4.2 相似,  $\hat{\boldsymbol{\beta}}_{3n}$  和  $\hat{\Lambda}_{3n}(t)$  的渐近性质由下面的定理给出.

**定理 5.1** 在定理 3.1 成立的条件下, 我们有

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{3n} - \boldsymbol{\beta}_0) = \mathbf{A}^{-1} \left( n^{-1/2} \sum_{i=1}^n \boldsymbol{\phi}_{3i} \right) + o_p(1),$$

$$\sqrt{n}(\hat{\Lambda}_{3n}(t) - \Lambda_0(t)) = n^{-1/2} \sum_{i=1}^n \psi_{3i}(t) + o_p(1),$$

其中

$$\begin{aligned} \boldsymbol{\phi}_{3i} &= \int_0^\tau \{ \mathbf{Z}_i - \bar{\mathbf{z}}(t) \} dM_i^{[3]}(t) + \mathbf{B}_{3\gamma} \mathbf{V}_\gamma^{-1} \mathbf{S}_{\gamma i}, \\ \mathbf{B}_{3\gamma} &= E \left[ \int_0^\tau \{ \mathbf{Z}_i - \bar{\mathbf{z}}(t) \} (1 - \xi_i) \omega'(\mathbf{V}_i, \boldsymbol{\gamma})^T dN_i^u(t) \right], \\ \psi_{3i}(t) &= \int_0^t \frac{dM_i^{[3]}(s)}{\mu(s)} - \mathbf{C}(t)^T \mathbf{A}^{-1} \boldsymbol{\phi}_{3i} + \mathbf{D}_{3\gamma}(t) \mathbf{V}_\gamma^{-1} \mathbf{S}_{\gamma i}, \\ \mathbf{D}_{3\gamma}(t) &= E \left[ \int_0^t (1 - \xi_i) \frac{\omega'(\mathbf{V}_i, \boldsymbol{\gamma})^T}{\mu(s)} dN_i^u(s) \right]. \end{aligned}$$

由定理 5.1 可得  $\sqrt{n}(\hat{\boldsymbol{\beta}}_{3n} - \boldsymbol{\beta}_0)$  依分布收敛于均值为  $\boldsymbol{\theta}$ 、协方差矩阵为  $\mathbf{A}^{-1} E(\boldsymbol{\phi}_{3i} \boldsymbol{\phi}_{3i}^T) (\mathbf{A}^{-1})^T$  的正态分布,  $\sqrt{n}(\hat{\Lambda}_{3n}(t) - \Lambda_0(t))$  依分布收敛于均值为 0、在  $(s, t)$  的渐近协方差函数为  $E(\psi_{3i}(s) \psi_{3i}(t))$

的高斯过程, 与前面类似的讨论即可得到其渐近方差和渐近协方差函数的相合估计.

## 6 数值模拟和实例分析

### 6.1 数值模拟

本节通过数值模拟对回归参数  $\boldsymbol{\beta}_0$  的 3 个估计结果  $\hat{\boldsymbol{\beta}}_{1n}$ 、 $\hat{\boldsymbol{\beta}}_{2n}$  和  $\hat{\boldsymbol{\beta}}_{3n}$  的有限样本性质进行研究. 对考虑的可加危险率模型  $\lambda(t | \mathbf{Z}) = \lambda_0(t) + \boldsymbol{\beta}'_0 \mathbf{Z}$ , 基准风险函数取为  $\lambda_0(t) \equiv 0.5$ , 回归系数  $\boldsymbol{\beta}_0 = 1$ . 协变量考虑 2 种情形:  $Z \sim \text{Uniform}(0, 2)$  和  $Z \sim \text{Bernoulli}(0.5)$ . 由可加危险率模型  $\lambda(t | \mathbf{Z}) = \lambda_0(t) + \boldsymbol{\beta}_0 Z$ , 根据风险函数和密度函数的关系, 可以生成生存时间变量  $T_i, i = 1, 2, \dots, n$ . 假定删失时间变量  $C$  的风险函数为  $(\lambda_0(t) + \boldsymbol{\beta}_0 Z) / \exp(\gamma_1 Z + \gamma_2)$ , 并据此生成删失时间变量  $C_i, i = 1, 2, \dots, n$ . 样本量  $n$  取为 100 和 200.

设定响应概率模型  $P\{\xi_i = 1 | \mathbf{V}_i\} = \pi(\mathbf{V}_i, \boldsymbol{\alpha})$  的参数形式为

$$\pi(\mathbf{V}_i, \boldsymbol{\alpha}) = \frac{\exp(\boldsymbol{\alpha}'_1 \mathbf{V}_i + \alpha_2)}{1 + \exp(\boldsymbol{\alpha}'_1 \mathbf{V}_i + \alpha_2)},$$

考虑到  $\omega(v, \gamma) = \frac{\lambda_t}{\lambda_t + \lambda_c}$ , 其中  $\lambda_t$  和  $\lambda_c$  分别表示  $T$  和  $C$  的风险函数, 因此可求得

$$P\{\delta_i = 1 | X_i, Z_i\} = \omega(\mathbf{V}_i, \boldsymbol{\gamma}) = \frac{\exp(\gamma_1 Z_i + \gamma_2)}{1 + \exp(\gamma_1 Z_i + \gamma_2)},$$

考虑不同的删失率和缺失率, 这通过取不同的参数的值达到. 参数不同的取值以及相应的删失率和缺失率详见表 1.

基于 1 000 次重复模拟的结果, 计算所提的回归参数  $\boldsymbol{\beta}$  的 3 种估计的偏差 (BIAS), 样本标准差 (SSE), 估计的标准差 (ESE) 和基于正态近似的 95% 的覆盖率 (CP). 我们也给出文献 [12-13] 所提估计的偏差, 样本标准差, 估计的标准差和基于正态近似的 95% 的覆盖率, 具体结果列于表 2 和表 3.

从模拟结果可以看出, 本文所提的 3 个估计都有比较小的偏差, 样本标准差和估计标准差; 样本标准差和估计标准差比较接近. 当缺失概率和删失概率不是很大时, 所提估计的覆盖率在 95% 附近. 因此, 所提的估计有比较好的有限样本性质. 进一步比较所提的 3 个估计. 显然, 增广加权估计和基于模型的插补估计, 相比逆概率加权估计, 具有更小的偏差, 样本标准差和估计标准差及

表 1 模型参数设置

Table 1 Parametric model settings

| 删失率                            | 缺失率              |                  |                  |                  |
|--------------------------------|------------------|------------------|------------------|------------------|
|                                | 20%              |                  | 40%              |                  |
|                                | $\alpha$         | $\gamma$         | $\alpha$         | $\gamma$         |
| $Z \sim \text{Uniform}(0,2)$   |                  |                  |                  |                  |
| 20%                            | [ 1.50, -0.71 ]' | [ 1.00,0.48 ]'   | [ 0.50, -0.38 ]' | [ 1.00,0.48 ]'   |
| 40%                            | [ 0.50, 0.70 ]'  | [ 1.00, -0.56 ]' | [ 0.50, -0.30 ]' | [ 1.00, -0.56 ]' |
| $Z \sim \text{Bernoulli}(0.5)$ |                  |                  |                  |                  |
| 20%                            | [ 1.00, 0.15 ]'  | [ 2.00, 0.68 ]'  | [ 0.50, -0.25 ]' | [ 2.00, 0.68 ]'  |
| 40%                            | [ 0.50, 0.90 ]'  | [ 2.00, -0.49 ]' | [ 0.50, -0.16 ]' | [ 2.00, -0.49 ]' |

表 2  $Z \sim \text{Uniform}(0,2)$  时的模拟结果

Table 2 Simulation results when  $Z \sim \text{Uniform}(0,2)$

| 删失率                   | 估计                    | 缺失率             |         |         |         |          |         |         |         |
|-----------------------|-----------------------|-----------------|---------|---------|---------|----------|---------|---------|---------|
|                       |                       | 20%             |         |         |         | 40%      |         |         |         |
|                       |                       | BIAS            | SSE     | ESE     | CP      | BIAS     | SSE     | ESE     | CP      |
| $n = 100$             |                       |                 |         |         |         |          |         |         |         |
| 20%                   | $\hat{\beta}_1$       | 0.015 2         | 0.325 2 | 0.348 9 | 0.962 0 | 0.025 0  | 0.374 0 | 0.488 0 | 0.977 0 |
|                       | $\hat{\beta}_2$       | 0.008 6         | 0.304 8 | 0.299 2 | 0.946 0 | 0.021 9  | 0.317 7 | 0.302 3 | 0.935 0 |
|                       | $\hat{\beta}_3$       | 0.012 5         | 0.286 3 | 0.292 4 | 0.952 0 | 0.021 9  | 0.289 9 | 0.298 9 | 0.948 0 |
|                       | $\hat{\beta}_1^{[S]}$ | 0.053 3         | 0.313 4 | 0.381 7 | 0.983 0 | 0.053 3  | 0.365 0 | 0.438 8 | 0.979 0 |
|                       | $\hat{\beta}_2^{[S]}$ | 0.060 9         | 0.305 7 | 0.381 7 | 0.987 0 | 0.007 1  | 0.312 6 | 0.438 8 | 0.983 0 |
|                       | $\hat{\beta}_3^{[Q]}$ | 0.148 1         | 0.289 8 | 0.375 1 | 0.988 0 | -0.065 0 | 0.276 6 | 0.374 7 | 0.977 0 |
|                       | 40%                   | $\hat{\beta}_1$ | 0.044 0 | 0.383 3 | 0.445 6 | 0.975 0  | 0.035 1 | 0.441 6 | 0.618 4 |
| $\hat{\beta}_2$       |                       | 0.040 6         | 0.367 3 | 0.341 9 | 0.938 0 | 0.026 3  | 0.383 5 | 0.360 9 | 0.979 0 |
| $\hat{\beta}_3$       |                       | 0.039 9         | 0.340 5 | 0.340 3 | 0.950 0 | 0.024 8  | 0.319 1 | 0.358 7 | 0.967 0 |
| $\hat{\beta}_1^{[S]}$ |                       | 0.053 6         | 0.383 0 | 0.465 9 | 0.982 0 | 0.060 6  | 0.438 6 | 0.519 6 | 0.973 0 |
| $\hat{\beta}_2^{[S]}$ |                       | 0.031 9         | 0.363 8 | 0.465 9 | 0.987 0 | 0.000 3  | 0.370 6 | 0.519 6 | 0.984 0 |
| $\hat{\beta}_3^{[Q]}$ |                       | 0.022 9         | 0.351 2 | 0.438 3 | 0.984 0 | 0.062 1  | 0.329 4 | 0.435 1 | 0.975 0 |
| $n = 200$             |                       |                 |         |         |         |          |         |         |         |
| 20%                   | $\hat{\beta}_1$       | 0.007 7         | 0.228 3 | 0.241 8 | 0.950 0 | 0.004 5  | 0.253 6 | 0.335 5 | 0.988 0 |
|                       | $\hat{\beta}_2$       | 0.004 2         | 0.216 0 | 0.211 8 | 0.940 0 | 0.003 7  | 0.219 1 | 0.214 1 | 0.942 0 |
|                       | $\hat{\beta}_3$       | 0.005 0         | 0.200 0 | 0.206 8 | 0.953 0 | 0.003 9  | 0.198 3 | 0.211 5 | 0.964 0 |
|                       | $\hat{\beta}_1^{[S]}$ | 0.030 2         | 0.217 2 | 0.261 2 | 0.983 0 | 0.029 2  | 0.241 5 | 0.294 1 | 0.980 0 |
|                       | $\hat{\beta}_2^{[S]}$ | 0.034 7         | 0.213 7 | 0.261 2 | 0.983 0 | -0.011 7 | 0.213 2 | 0.294 1 | 0.987 0 |
|                       | $\hat{\beta}_3^{[Q]}$ | 0.131 4         | 0.205 2 | 0.259 2 | 0.978 0 | -0.069 3 | 0.188 8 | 0.259 6 | 0.979 0 |
|                       | 40%                   | $\hat{\beta}_1$ | 0.009 0 | 0.254 7 | 0.312 4 | 0.977 0  | 0.014 2 | 0.293 7 | 0.424 2 |
| $\hat{\beta}_2$       |                       | 0.008 3         | 0.242 2 | 0.241 7 | 0.943 0 | 0.015 2  | 0.254 9 | 0.253 7 | 0.953 0 |
| $\hat{\beta}_3$       |                       | 0.006 8         | 0.223 6 | 0.240 5 | 0.964 0 | 0.011 8  | 0.214 1 | 0.251 1 | 0.972 0 |
| $\hat{\beta}_1^{[S]}$ |                       | 0.016 7         | 0.254 8 | 0.315 1 | 0.980 0 | 0.035 6  | 0.286 4 | 0.347 8 | 0.981 0 |
| $\hat{\beta}_2^{[S]}$ |                       | 0.000 6         | 0.240 9 | 0.315 1 | 0.986 0 | -0.004 8 | 0.249 4 | 0.347 8 | 0.988 0 |
| $\hat{\beta}_3^{[Q]}$ |                       | -0.005 8        | 0.230 5 | 0.300 6 | 0.984 0 | -0.058 7 | 0.222 0 | 0.302 4 | 0.983 0 |

注： $\hat{\beta}_1$ ：本文的逆概率加权估计； $\hat{\beta}_1^{[S]}$ ：文献[12]的简单加权估计； $\hat{\beta}_2$ ：本文的增广加权估计； $\hat{\beta}_2^{[S]}$ ：文献[12]的增广加权估计； $\hat{\beta}_3$ ：本文的基于模型的插补估计； $\hat{\beta}_3^{[Q]}$ ：文献[13]的插补估计；BIAS：偏差；SSE：样本标准差；ESE：估计标准差；CP：基于正态近似的 95% 的覆盖率。

表 3  $Z \sim \text{Bernoulli}(0.5)$  时的模拟结果  
Table 3 Simulation results when  $Z \sim \text{Bernoulli}(0.5)$

| 删失率                   | 估计                    | 缺失率             |         |         |         |          |         |         |         |
|-----------------------|-----------------------|-----------------|---------|---------|---------|----------|---------|---------|---------|
|                       |                       | 20%             |         |         |         | 40%      |         |         |         |
|                       |                       | BIAS            | SSE     | ESE     | CP      | BIAS     | SSE     | ESE     | CP      |
| $n = 100$             |                       |                 |         |         |         |          |         |         |         |
| 20%                   | $\hat{\beta}_1$       | 0.020 3         | 0.270 7 | 0.280 6 | 0.958 0 | 0.016 0  | 0.311 1 | 0.360 6 | 0.966 0 |
|                       | $\hat{\beta}_2$       | 0.019 3         | 0.258 9 | 0.243 2 | 0.926 0 | 0.018 2  | 0.261 2 | 0.249 8 | 0.943 0 |
|                       | $\hat{\beta}_3$       | 0.018 9         | 0.250 5 | 0.240 8 | 0.934 0 | 0.020 5  | 0.246 7 | 0.247 2 | 0.951 0 |
|                       | $\hat{\beta}_1^{[S]}$ | 0.022 8         | 0.260 8 | 0.291 4 | 0.971 0 | 0.026 8  | 0.275 8 | 0.314 5 | 0.983 0 |
|                       | $\hat{\beta}_2^{[S]}$ | 0.021 8         | 0.260 2 | 0.291 4 | 0.969 0 | 0.014 7  | 0.260 2 | 0.314 5 | 0.987 0 |
|                       | $\hat{\beta}_3^{[Q]}$ | 0.030 6         | 0.247 6 | 0.287 0 | 0.981 0 | -0.084 8 | 0.230 6 | 0.288 8 | 0.958 0 |
|                       | 40%                   | $\hat{\beta}_1$ | 0.036 1 | 0.302 4 | 0.341 4 | 0.970 0  | 0.035 0 | 0.353 5 | 0.458 8 |
| $\hat{\beta}_2$       |                       | 0.034 0         | 0.286 8 | 0.275 0 | 0.946 0 | 0.026 1  | 0.304 4 | 0.289 3 | 0.933 0 |
| $\hat{\beta}_3$       |                       | 0.031 9         | 0.269 1 | 0.273 6 | 0.962 0 | 0.025 2  | 0.259 8 | 0.287 0 | 0.973 0 |
| $\hat{\beta}_1^{[S]}$ |                       | 0.038 9         | 0.293 5 | 0.336 7 | 0.983 0 | 0.040 4  | 0.327 9 | 0.362 5 | 0.978 0 |
| $\hat{\beta}_2^{[S]}$ |                       | 0.031 1         | 0.286 1 | 0.336 7 | 0.984 0 | 0.016 9  | 0.301 6 | 0.362 5 | 0.983 0 |
| $\hat{\beta}_3^{[Q]}$ |                       | 0.013 6         | 0.274 8 | 0.327 1 | 0.983 0 | -0.072 8 | 0.264 8 | 0.327 9 | 0.969 0 |
| $n = 200$             |                       |                 |         |         |         |          |         |         |         |
| 20%                   | $\hat{\beta}_1$       | 0.005 7         | 0.190 5 | 0.197 4 | 0.955 0 | 0.009 6  | 0.216 7 | 0.250 5 | 0.970 0 |
|                       | $\hat{\beta}_2$       | 0.004 0         | 0.180 0 | 0.173 6 | 0.943 0 | 0.004 0  | 0.182 5 | 0.177 5 | 0.948 0 |
|                       | $\hat{\beta}_3$       | 0.005 2         | 0.174 6 | 0.172 1 | 0.948 0 | 0.006 0  | 0.172 2 | 0.175 2 | 0.957 0 |
|                       | $\hat{\beta}_1^{[S]}$ | 0.006 6         | 0.181 9 | 0.203 2 | 0.971 0 | 0.013 5  | 0.189 7 | 0.217 1 | 0.972 0 |
|                       | $\hat{\beta}_2^{[S]}$ | 0.005 9         | 0.180 7 | 0.203 2 | 0.974 0 | 0.001 5  | 0.182 8 | 0.217 1 | 0.982 0 |
|                       | $\hat{\beta}_3^{[Q]}$ | 0.018 3         | 0.174 4 | 0.202 5 | 0.977 0 | -0.091 4 | 0.161 2 | 0.201 1 | 0.952 0 |
|                       | 40%                   | $\hat{\beta}_1$ | 0.002 3 | 0.210 9 | 0.237 1 | 0.972 0  | 0.004 9 | 0.246 0 | 0.314 2 |
| $\hat{\beta}_2$       |                       | 0.003 8         | 0.199 0 | 0.192 6 | 0.954 0 | 0.000 9  | 0.213 7 | 0.203 6 | 0.937 0 |
| $\hat{\beta}_3$       |                       | 0.004 7         | 0.186 9 | 0.191 8 | 0.947 0 | 0.002 8  | 0.182 8 | 0.202 0 | 0.966 0 |
| $\hat{\beta}_1^{[S]}$ |                       | 0.003 2         | 0.203 4 | 0.230 8 | 0.972 0 | 0.011 3  | 0.225 6 | 0.246 9 | 0.969 0 |
| $\hat{\beta}_2^{[S]}$ |                       | -0.002 4        | 0.200 6 | 0.230 8 | 0.976 0 | -0.004 9 | 0.213 8 | 0.246 9 | 0.979 0 |
| $\hat{\beta}_3^{[Q]}$ |                       | -0.017 1        | 0.193 4 | 0.225 8 | 0.977 0 | -0.089 9 | 0.185 1 | 0.227 3 | 0.962 0 |

注： $\hat{\beta}_1$ ：本文的逆概率加权估计； $\hat{\beta}_1^{[S]}$ ：文献[12]的简单加权估计； $\hat{\beta}_2$ ：本文的增广加权估计； $\hat{\beta}_2^{[S]}$ ：文献[12]的增广加权估计； $\hat{\beta}_3$ ：本文的基于模型的插补估计； $\hat{\beta}_3^{[Q]}$ ：文献[13]的插补估计；BIAS：偏差；SSE：样本标准差；ESE：估计标准差；CP：基于正态近似的 95% 的覆盖率。

更加准确的覆盖率. 比较增广加权估计和基于模型的插补估计,后者具有更好的有限样本性质. 前文提出回归系数和基准累积风险函数的增广加权估计时,就已指出,逆概率加权估计只用到了完全观测样本(对应  $\xi_i = 1$  的个体)的信息,忽略了不完全观测样本( $\xi_i = 0$  的个体)的信息,可能会使结果的有效性降低. 增广加权估计中由于也采用了逆概率的方法,有时会导致结果稍逊于插补估计.

我们对所提的 3 个估计和文献中的估计:文献[12]的简单逆概率加权估计和增广加权估计,以及文献[13]的插补估计的有限样本性质进行比较. 发现我们所提的增广加权估计和基于模型的插补估计具有比文献中 3 个估计更好的有限样本性质:具有更小的偏差,样本标准差和估计标准差和更加准确的覆盖率. 因为本文所提的增广加权估计和基于模型的插补估计充分地利用了参数模型的信息,因此相比采用非参数方法的文献



中的 3 个估计更为准确,这是非常合理的. 这些说明我们所提的增广加权估计和基于模型的插补估计在参数模型假定正确时,的确更为有效. 这种优势在样本量小( $n = 100$ )时尤其明显. 这也是合理的,当样本量比较大时,非参数方法的估计也比较准确,这时基于参数概率模型的方法和非参数概率模型的方法的差异就变小了.

6.2 实例分析

本节我们通过对乳腺癌数据<sup>[15]</sup>的分析验证所提出的估计方法. 该数据来自 ECOG(东部肿瘤协作组)的 E1178 临床试验,有 169 名患有第 II 型乳腺癌的年长女性的相关信息. 我们感兴趣的是注射它莫西芬病人的腋窝淋巴结数量对其生存时间的影响,该批病人中,共有 86 名病人接受了它莫西芬注射,其中 21 人死于乳腺癌( $\xi = 1, \delta = 1$ ),55 人死于其他原因( $\xi = 1, \delta = 0$ ),而剩余 10 人死亡原因未知( $\xi = 0$ ). 我们建立可加危险率模型,利用本文所提的方法,采用与前面的模拟设置相同的参数模型结构对数据进行分析,具体结果见表 4.

表 4 乳腺癌数据研究

| Table 4 Aanalysis of breast cancer data |                           |                   |         |
|---|---------------------------|-------------------|---------|
| 方法                                      | $\hat{\beta} \times 10^3$ | ESE $\times 10^4$ | P-value |
| IPW                                     | 0.129 4                   | 0.599 5           | 0.030 9 |
| FW                                      | 0.126 8                   | 0.581 4           | 0.029 2 |
| IM                                      | 0.123 4                   | 0.572 4           | 0.031 2 |
| Song1                                   | 0.140 2                   | 0.699 2           | 0.044 9 |
| Song2                                   | 0.138 5                   | 0.699 2           | 0.047 6 |
| Qiu                                     | 0.132 9                   | 0.683 5           | 0.051 9 |
| CC                                      | 0.120 6                   | 0.633 3           | 0.056 9 |

注:IPW,FW 和 IM 分别表示本文的逆概率加权,增广加权和基于模型的插补方法;Song1 和 Song2 分别表示文献[12]的简单加权和增广加权方法;Qiu 表示文献[13]的插补方法;CC 表示 Complete-Case 方法.

我们也利用文献[12-13]中的方法,以及 Complete - Case 方法分析数据. 文献中的方法用到删失概率函数和缺失概率函数的非参数估计,我们使用大拇指法则选择窗宽,所得结果见表 4. 从表 4 的结果可以看出,乳腺癌病人的腋窝淋巴结数量是显著的,并且可以看出本文所提出方法的估计标准差明显小于文献中的方法,且检验的  $P$  值也要小一些.

7 结束语

本文给出在失效信息随机缺失时可加危险率

模型的回归参数和基准累积风险函数的逆概率加权估计、增广加权估计和基于模型的插补估计. 本文结果表明 所得的估计计算简单并且具有良好的理论性质,其渐近协方差矩阵的也可由传统的 plug - in 方法得到. 通过数值模拟研究,我们发现当参数模型的假定正确时,本文所提的增广加权估计和基于模型的插补估计具有优于文献[12-13]的估计的有限样本性质. 乳腺癌数据的对比分析结果也证明了本文方法的良好效果.

参考文献

[ 1 ] Little R J A, Rubin D B. Statistical analysis with missing data [M]. 2nd ed. New York: John Wiley & Sons, 2002.

[ 2 ] Lu K, Tsiatis A A. Multiple imputation methods for estimating regression coefficients in the competing risks model with missing cause of failure[J]. Biometrics,2001, 57(4): 1 191-1 197.

[ 3 ] Wang Q, Linton O, Härdle W. Semiparametric regression analysis with missing response at random[J]. Journal of the American Statistical Association, 2004, 99(466): 334-345.

[ 4 ] Sun Z, Wang Q. Semiparametric estimation of survival function with cause of death data missing at random [J]. Journal of Applied Probability and Statistics, 2007, 2: 189-209.

[ 5 ] Zhou X, Sun L. Additive hazards regression with missing censoring information[J]. Statistica Sinica, 2003, 13(4): 1 237-1 257.

[ 6 ] Wang Q, Ng K W. Asymptotically efficient product-limit estimators with censoring indicators missing at random [J]. Statistica Sinica, 2008, 18(2):749-768.

[ 7 ] Wang Q, Dinse G E. Linear regression analysis of survival data with missing censoring indicators [J]. Lifetime Data Analysis, 2011, 17(2):256-279.

[ 8 ] Sun Z, Xie T, Liang H. Statistical inference for right-censored data with nonignorable missing censoring indicators [J]. Science China Mathematics, 2013, 56(6): 1 263-1 278.

[ 9 ] Lin D Y, Ying Z. Semiparametric analysis of the additive risk model[J]. Biometrika, 1994, 81(1): 61-71.

[10] Yin G, Cai J. Additive hazards model with multivariate failure time data[J]. Biometrika, 2004, 91(4): 801-818.

[11] Lu W, Liang Y. Analysis of competing risks data with missing cause of failure under additive hazards model[J]. Statistica Sinica, 2008, 18(1): 219.

[12] Song X, Sun L, Mu X, et al. Additive hazards regression with censoring indicators missing at random [J]. Canadian Journal of Statistics, 2010, 38(3): 333-351.

[13] Qiu Z, Chen X, Zhou Y. A kernel-assisted imputation estimating method for the additive hazards model with missing

censoring indicator [J]. Statistics & Probability Letters, 2015, 98: 89-97.

[14] Dikta G. On semiparametric random censorship models[J]. Journal of Statistical Planning and Inference, 1998, 66(2): 253-279.

[15] Goetghebeur E, Ryan L. Analysis of competing risks survival data when some failure types are missing[J]. Biometrika, 1995, 82(4): 821-833.

[16] Pollard D. Empirical processes: theory and applications[C]// NSF-CBMS regional conference series in probability and statistics. Institute of Mathematical Statistics and the American Statistical Association, 1990: 1-86.

## 附录 引理和定理的证明

下面列出定理成立所需的正则条件:

(C1):  $\omega(v, \gamma)$  和  $\pi(v, \alpha)$  关于  $\gamma$  和  $\alpha$  分别有逐点连续的二阶偏导数;

(C2): 对所有的  $v \geq 0, 0 < \omega(v, \gamma) < 1$  且  $0 < \pi(v, \alpha) < 1$ ;

(C3):  $L_n(\alpha)$  和  $L_n(\gamma)$  关于  $\alpha$  和  $\gamma$  分别有唯一的极大值解;

(C4):  $\Lambda_0(\tau) < \infty, P(X \geq \tau) > 0, \mathbf{Z}$  有界, 并且  $\int_0^\tau \lambda_0(t) + \beta' \mathbf{Z} dt > 0$  a. e.;

(C5):  $\mathbf{A} = E \left[ \int_0^\tau Y_i(t) \{ \mathbf{Z}_i - \bar{\mathbf{Z}}(t) \}^{\otimes 2} dt \right]$  是非奇异的.

**引理 2.1 的证明** 由(3), 对似然函数  $L_n(\alpha)$  取对数并对  $\alpha$  求导, 令其为 0, 可得估计方程如下:

$$\sum_{i=1}^n \frac{\pi'(V_i, \hat{\alpha}_n)}{\pi(V_i, \hat{\alpha}_n)} \frac{\xi_i - \pi(V_i, \hat{\alpha}_n)}{1 - \pi(V_i, \hat{\alpha}_n)} = 0,$$

记上式左边为  $\mathbf{A}_n$ , 则有

$$\begin{aligned} \mathbf{A}_n &= \sum_{i=1}^n \frac{\pi'(V_i, \hat{\alpha}_n)}{\pi(V_i, \hat{\alpha}_n)} \frac{\xi_i - \pi(V_i, \alpha)}{1 - \pi(V_i, \hat{\alpha}_n)} + \\ &\quad \sum_{i=1}^n \frac{\pi'(V_i, \hat{\alpha}_n)}{\pi(V_i, \hat{\alpha}_n)} \frac{\pi(V_i, \alpha) - \pi(V_i, \hat{\alpha}_n)}{1 - \pi(V_i, \hat{\alpha}_n)} \\ &:= \mathbf{A}_{n1} + \mathbf{A}_{n2}, \end{aligned}$$

容易证得

$$\mathbf{A}_{n1} = \sum_{i=1}^n \frac{\pi'(V_i, \alpha)}{\pi(V_i, \alpha)} \frac{\xi_i - \pi(V_i, \alpha)}{1 - \pi(V_i, \alpha)} + o_p(\sqrt{n}),$$

利用泰勒展开, 可得

$$\begin{aligned} \mathbf{A}_{n2} &= - \sum_{i=1}^n \frac{\pi'(V_i, \alpha)}{\pi(V_i, \alpha)} \frac{\pi'(V_i, \alpha)^T}{1 - \pi(V_i, \alpha)} (\hat{\alpha}_n - \alpha) + \\ &\quad o_p(\sqrt{n}), \end{aligned}$$

由大数定律, 有

$$\frac{1}{n} \sum_{i=1}^n \frac{\pi'(V_i, \alpha)}{\pi(V_i, \alpha)} \frac{\pi'(V_i, \alpha)^T}{1 - \pi(V_i, \alpha)} = \mathbf{V}_\alpha + o_p(1),$$

从而

$$\begin{aligned} \sqrt{n}(\hat{\alpha}_n - \alpha) &= \frac{\mathbf{V}_\alpha^{-1}}{\sqrt{n}} \sum_{i=1}^n \frac{\pi'(V_i, \alpha)}{\pi(V_i, \alpha)} \frac{\xi_i - \pi(V_i, \alpha)}{1 - \pi(V_i, \alpha)} + \\ &\quad o_p(1), \end{aligned}$$

类似的方法可得

$$\begin{aligned} \sqrt{n}(\hat{\gamma}_n - \gamma) &= \frac{\mathbf{V}_\gamma^{-1}}{\sqrt{n}} \sum_{i=1}^n \frac{\omega'(V_i, \gamma)}{\omega(V_i, \gamma)} \frac{\delta_i - \omega(V_i, \gamma)}{1 - \omega(V_i, \gamma)} \xi_i + \\ &\quad o_p(1), \end{aligned}$$

引理得证.

下面给出定理 4.1 和定理 4.2 的证明, 使用相似的方法即可证明定理 3.1, 定理 3.2 和定理 5.1.

**定理 4.1 的证明** 利用式(9)和式(10), 消去  $d\Lambda_0(t)$ , 可得关于  $\beta$  的方程

$$\begin{aligned} \sum_{i=1}^n \int_0^\tau \{ \mathbf{Z}_i - \bar{\mathbf{Z}}(t) \} \left[ \frac{\xi_i}{\pi(V_i, \hat{\alpha}_n)} dN_i(t) + \left( 1 - \frac{\xi_i}{\pi(V_i, \hat{\alpha}_n)} \right) \right. \\ \left. \omega(V_i, \hat{\gamma}_n) dN_i^u(t) - Y_i(t) \beta' \mathbf{Z}_i dt \right] = 0. \end{aligned}$$

考虑到  $\sum_{i=1}^n \int_0^\tau \{ \mathbf{Z}_i - \bar{\mathbf{Z}}(t) \} Y_i(t) \lambda_0(t) dt = 0$ , 上式左边可写成:

$$\begin{aligned} \sum_{i=1}^n \int_0^\tau \{ \mathbf{Z}_i - \bar{\mathbf{Z}}(t) \} \left[ \frac{\xi_i}{\pi(V_i, \hat{\alpha}_n)} dN_i(t) + \left( 1 - \frac{\xi_i}{\pi(V_i, \hat{\alpha}_n)} \right) \right. \\ \left. \omega(V_i, \hat{\gamma}_n) dN_i^u(t) - Y_i(t) (\lambda_0(t) + \beta' \mathbf{Z}_i) dt \right] = \mathbf{U}(\beta). \end{aligned}$$

将  $\mathbf{U}(\hat{\beta}_{2n})$  关于  $\beta$  一阶泰勒展开:

$$0 = \mathbf{U}(\hat{\beta}_{2n}) = \mathbf{U}(\beta_0) + \frac{\partial \mathbf{U}(\beta)}{\partial \beta} \Big|_{\beta=\beta^*} (\hat{\beta}_{2n} - \beta_0),$$

其中  $\beta^*$  介于  $\hat{\beta}_{2n}$  与  $\beta_0$  之间, 于是

$$\sqrt{n}(\hat{\beta}_{2n} - \beta_0) = \left[ - \frac{1}{n} \frac{\partial \mathbf{U}(\beta)}{\partial \beta} \Big|_{\beta=\beta^*} \right]^{-1} \cdot \frac{1}{\sqrt{n}} \mathbf{U}(\beta_0).$$

根据大数定律可得

$$\begin{aligned} - \frac{1}{n} \frac{\partial \mathbf{U}(\beta)}{\partial \beta} &= \frac{1}{n} \sum_{i=1}^n \int_0^\tau Y_i(t) \{ \mathbf{Z}_i - \bar{\mathbf{Z}}(t) \}^{\otimes 2} dt \\ &= \mathbf{A} + o_p(1), \end{aligned}$$

其中  $\mathbf{A}$  在定理 3.1 中给出. 接下来, 将  $\mathbf{U}(\beta_0)$  看做关于  $\alpha$  和  $\gamma$  的二元函数, 进行泰勒展开可得

$$\mathbf{U}(\beta_0) = \mathbf{U}_1(\beta_0) + \mathbf{U}_2(\beta_0) + \mathbf{U}_3(\beta_0),$$

其中

$$\mathbf{U}_1(\beta_0) = \sum_{i=1}^n \int_0^\tau \{ \mathbf{Z}_i - \bar{\mathbf{Z}}(t) \} dM_i^{[2]}(t),$$

$$\mathbf{U}_2(\beta_0) = \sum_{i=1}^n \int_0^\tau \{ \mathbf{Z}_i - \bar{\mathbf{Z}}(t) \} \frac{\xi_i \pi'(V_i, \alpha)^T}{\pi^2(V_i, \alpha)} (\omega(V_i, \gamma) - \delta_i)$$

$$dN_i^u(t)(\hat{\alpha}_n - \alpha),$$

$$U_3(\beta_0) = \sum_{i=1}^n \int_0^\tau \{Z_i - \bar{Z}(t)\} \left(1 - \frac{\xi_i}{\pi(V_i, \alpha)}\right) \omega'(V_i, \gamma)^T$$

$$dN_i^u(t)(\hat{\gamma}_n - \gamma).$$

注意到  $U_1(\beta_0)$  是鞅积分, 因此

$$U_1(\beta_0) = \sum_{i=1}^n \int_0^\tau \{Z_i - \bar{z}(t)\} dM_i^{[2]}(t) + o_p(\sqrt{n}), \quad (17)$$

对  $U_2(\beta_0)$ , 有

$$\begin{aligned} U_2(\beta_0) &= \sum_{i=1}^n \int_0^\tau \{Z_i - \bar{z}(t)\} \frac{\xi_i \pi'(V_i, \alpha)^T}{\pi^2(V_i, \alpha)} (\omega(V_i, \gamma) - \\ &\quad \delta_i) dN_i^u(t)(\hat{\alpha}_n - \alpha) + \sum_{i=1}^n \int_0^\tau \{\bar{z}(t) - \bar{Z}(t)\} \\ &\quad \frac{\xi_i \pi'(V_i, \alpha)^T}{\pi^2(V_i, \alpha)} (\omega(V_i, \gamma) - \delta_i) dN_i^u(t)(\hat{\alpha}_n - \alpha) \\ &:= U_{21}(\beta_0) + U_{22}(\beta_0). \end{aligned}$$

显然

$$U_{21}(\beta_0) = \sum_{i=1}^n B_{\alpha i}(\hat{\alpha}_n - \alpha), \quad (18)$$

由大数定律和泛函中心极限定理<sup>[16]</sup>, 得

$$\begin{aligned} \bar{z}(t) - \bar{Z}(t) &= o_p(1), \\ \sum_{i=1}^n \frac{\xi_i \pi'(V_i, \alpha)^T}{\pi^2(V_i, \alpha)} (\omega(V_i, \gamma) - \delta_i) N_i^u(t)(\hat{\alpha}_n - \alpha) \\ &= O_p(\sqrt{n}), \end{aligned}$$

从而

$$U_{22}(\beta_0) = o_p(\sqrt{n}), \quad (19)$$

根据式(18)和式(19), 可得

$$U_2(\beta_0) = \sum_{i=1}^n B_{\alpha i}(\hat{\alpha}_n - \alpha) + o_p(\sqrt{n}),$$

再根据引理 2.1, 有

$$\frac{1}{\sqrt{n}} U_2(\beta_0) = B_{2\alpha} \frac{V_\alpha^{-1}}{\sqrt{n}} \sum_{i=1}^n S_{\alpha i} + o_p(1). \quad (20)$$

同样可证得

$$\frac{1}{\sqrt{n}} U_3(\beta_0) = B_{2\gamma} \frac{V_\gamma^{-1}}{\sqrt{n}} \sum_{i=1}^n S_{\gamma i} + o_p(1). \quad (21)$$

综上, 可得

$$\frac{1}{\sqrt{n}} U(\beta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \phi_{2i} + o_p(1). \quad (22)$$

定理得证.

**定理 4.2 的证明** 根据  $\hat{\Lambda}_{2n}(t)$  的定义以及  $\Lambda_0(t)$

$$= \int_0^t \lambda_0(s) ds, \text{ 经计算可得}$$

$$\begin{aligned} \sqrt{n}(\hat{\Lambda}_{2n}(t) - \Lambda_0(t)) &= \sqrt{n} \sum_{i=1}^n \int_0^t \left\{ \frac{\xi_i}{\pi(V_i, \hat{\alpha}_n)} dN_i(s) + \right. \\ &\quad \left. \left(1 - \frac{\xi_i}{\pi(V_i, \hat{\alpha}_n)}\right) \omega(V_i, \hat{\gamma}_n) dN_i^u(s) - Y_i(s) \right\} \\ &\quad \left\{ d\hat{\Lambda}_2(s) + \hat{\beta}'_{2n} Z_i ds \right\} \Big/ \left\{ \frac{1}{n} \sum_{i=1}^n Y_i(s) \right\}. \end{aligned}$$

将  $\sqrt{n}(\hat{\Lambda}_{2n}(t) - \Lambda_0(t))$  看做关于  $\beta, \alpha, \gamma$  的三元函数, 根据多元函数的一阶泰勒展开, 以及  $\beta_{2n}, \hat{\alpha}_n$  和  $\hat{\gamma}_n$  的性质, 采用与定理 4.1 相同的讨论, 即可得

$$\begin{aligned} \sqrt{n}(\hat{\Lambda}_{2n}(t) - \Lambda_0(t)) &= n^{-1/2} \sum_{i=1}^n \int_0^t \frac{dM_i^{[2]}(s)}{\mu(s)} - C(t)^T \\ &\quad \sqrt{n}(\hat{\beta}_{2n} - \beta_0) + D_{2\alpha}(t) V_\alpha^{-1} S_{\alpha i} + D_{2\gamma}(t) V_\gamma^{-1} S_{\gamma i} \\ &= n^{-1/2} \sum_{i=1}^n \psi_{2i}(t) + o_p(1). \end{aligned}$$

定理得证.