arXiv:2603.02723v1 [stat.ME] 3 Mar 2026

部分参数与部分非参数加性风险模型

2021年8月

Nils Lid Hjort[1] 和 Emil Aas Stoltenberg[2]

[1]奥斯陆大学数学系[2]奥斯陆 BI 挪威商学院

摘要

Aalen 的线性风险率回归模型是 Cox 乘法风险率模型的一种实用且日益受欢迎的替代方案。该模型假设个体的风险率函数为$h(s) = z_1\alpha_1(s) + \cdots + z_r\alpha_r(s)$，其中涉及其协变量值 z_1、……、z_r。这些通常是各种风险因子的水平，也可能随时间变化。风险因子函数$\alpha_j(s)$是模型的参数，并根据数据进行估计。传统上，这通常采用完全非参数的方式完成。本文提出了一套方法，用于估计风险因子函数，其中一部分采用参数模型，其他部分则不作具体设定。在这一部分参数、部分非参数的框架内，本文得到了大样本结果，这也使我们能够评估模型参数分量的拟合优度。此外，我们还利用这些结果指出，与标准非参数方法相比，采用参数-非参数模型能提高多少精度。文中还包括一项真实数据应用和一项简短的模拟研究。

关键词：计数过程、事件史、拟合优度过程、线性风险回归模型、半参数

1 引言与总结

假设个体 i 具有可观测的协变量值 $z_{i,1}$、……、$z_{i,r}$，并且认为这些值会影响其寿命 T_i的概率分布。最常见的建模方式是采用 Cox 风险率 $h_i(s)$回归模型，假设其形式为$h_0(s)\exp(\beta_1 z_{i,1} + \cdots + \beta_r z_{i,r})$，其中 β_1、……、β_r为参数。过去几十年来，Aalen 线性风险率回归模型已成为一种实用且广受欢迎的替代方案。该模型假设个体 i 具有如下风险率函数

$$h_i(s) = h(s \mid z_i) = z_{i,1}\alpha_1(s) + \cdots + z_{i,r}\alpha_r(s) = z_i^{\mathrm{t}}\alpha(s), \tag{1.1}$$

其中 $\alpha_j(s)$函数未知。观测数据由个体 $i = 1, \ldots\ldots, n$的三元组(T_i, δ_i, z_i)组成，其中 δ_i 是未删失的指示变量。参见Aalen（1980、1989、1993）以及经典专著Andersen、Borgan、Gill和Keiding（1993，第VII章）的相关章节以及

Aalen、Borgan 和 Gjessing（2008，第VI章），以了解（1.1）模型的一般性讨论、最常用的估计方法及其性质，以及对各种数据集的应用。下文我们将评述基本Aalen模型（1.1）的各种扩展与进一步发展。本文是对这一文献的又一贡献，将部分回归函数取为参数形式，其余取为非参数形式。

可以将 $z_j = z_{i,j}$ 视为该个体的第 j 个风险因子水平，而将$\alpha_j(s)$函数视为相应的风险因子函数，或回归函数。通常，第一个协变量是常数 1，其余协变量经过缩放：对于离散协变量，零为最小值；对于连续协变量，零为均值。在后一种情况下，通常还会将协变量乘以经验标准差的倒数。在这类情况下，方程 (1.1) 将风险率建模为共同的 $\alpha_1(s)$，再加上风险因子 z_2、.....、 z_r 所带来的额外贡献。协变量也可以随时间变化，只要其变化方式是预见性的或可预测的；在时间 s，时间 s 刚到之前，协变量值 $z_i(s)$ 应当已知。只需保证 $z_i(s)$ 是关于在 $[0, s]$ 上已观测内容的左连续函数；也就是说，它们不能依赖于 s 之后才获得的信息。

重要的是，Aalen 加性模型通常采用非参数方法进行估计，除对（1.1）中所有 z_i 属于协变量分布支撑集的情形，要求 $z_i^{\mathrm{t}}\alpha(s)$ 满足正性和连续性外，不作其他假设。对于典型的应用，会计算并展示累积风险因子函数的非参数估计量

$$A_j(t) = \int_0^t \alpha_j(s)\,\mathrm{d}s \quad \text{for } j = 1, \ldots, r \tag{1.2}$$

并辅以变异性估计。研究者据此推断不同协变量因子随时间变化的相对影响。也可以从这里的模型中读取给定个体的生存曲线；如果某个个体的协变量向量为 $z = (z_1, \ldots, z_r)^{\mathrm{t}}$，且不随时间变化，则其生存曲线为

$$S(t \mid z) = \exp\{-z^{\mathrm{t}} A(t)\} = \exp\{-z_1 A_1(t) - \cdots - z_r A_r(t)\}. \tag{1.3}$$

此后，研究者围绕基础 Aalen 模型（1.1）–（1.3）的各个方面开展了大量进一步研究，对其加以扩展和完善，参见 Martinussen 和 Scheike（2007, 2002a,b, 2009b,a）。McKeague & Sasieni（1994）研究了一种模型，其中部分 $\alpha_j(s)$ 函数设为常数，其余函数采用非参数形式；本文进一步拓展了这些思路和方法。Stoltenberg（2020）研究了存在治愈比例时的 Aalen 模型。Borgan 等人（2007）将该模型的某些特性扩展至复发事件数据，以反映个体间异质性和缺失数据。与本文同样相关的是，Jullum & Hjort（2017）提出了一般性的模型选择方法，用于在参数和非参数候选模型中进行选择；Jullum & Hjort（2019）则研究了在 Cox 回归模型中指定参数化基线风险可能带来的效率增益。

在应用中，研究人员可能对某些协变量的效应所采用的函数形式有明确的先验看法，而对其他协变量了解较少。这促成了这样一种框架：其中一部分风险因子函数（例如前 p 个）以参数形式指定，而其余 $q = r - p$ 个则继续保持未指定状态；除了基本要求（即对于所有预期的协变量值以及所有时间 s，(1.1) 中的量均为非负）之外，不再施加其他限制。记 $z_{i,(1)}$ 为前 p 个分量，记 $z_{i,(2)}$ 为 z_i 的其余 q 个分量，并将 $\alpha(s)$ 进行类似的分块划分为

$\alpha_{(1)}(s,\theta)$ 与 $\alpha_{(2)}(s)$，该模型变为

$$h_i(s) = z_{i,(1)}^{\mathrm{t}}\alpha_{(1)}(s,\theta) + z_{i,(2)}^{\mathrm{t}}\alpha_{(2)}(s) = \sum_{j=1}^{p} z_{i,j}\alpha_j(s,\theta) + \sum_{j=p+1}^{p+q} z_{i,j}\alpha_j(s) \tag{1.4}$$

对于 $i = 1,\ldots,n$。这里， θ 是用于描述前 p 个风险因子函数的参数集合，其通常取 $\alpha_j(s,\theta_j)$ 的形式，其中 $j = 1,\ldots,p$。

(1.4) 中的协变量不依赖于时间。正如关于 (1.1) 的 Aalen模型所讨论的那样，将其扩展到时变协变量只需对与可预测性以及所有时间点上协变量的线性无关性相关的理论作少量修改。为便于表述，我们坚持使用不随时间变化的协变量。

我们的目标有两个方面。首先，我们旨在为 (1.4) 模型中的未知量建立可靠的估计方法，并给出描述这些估计量行为的大样本理论。其次，将构建相应的拟合优度度量，以评估参数分量的适当性。为实现这些目标，本文按如下方式展开。

我们在第 2 节首先给出模型 (1.4) 的纯非参数版本和纯参数版本的自然方法与结果，然后在第 3 节讨论同时包含参数分量与非参数分量情形下我们所偏好的估计策略。在第 4 节中，我们推导参数部分与非参数部分所需的大样本正态性结果，从而进行统计推断。我们所提出的一类方法中的一个特例是渐近最优的；有关此类论断的细节具有独立意义，并在附录 A 中加以总结。在 (1.1) 模型中使用参数分量而非非参数分量的部分益处在于，它能提高精度，对参数分量和非参数分量均是如此；这一点在附录 B 中加以评估和说明。

随后，在第5节中，我们构造拟合优度监测过程，这些过程尤其导出各类卡方检验。在第6节中，通过模拟研究说明我们的估计和推断方法的有限样本行为。我们还给出一项实证应用，涉及 $n = 312$ 双盲随机研究中的原发性胆汁性肝硬化患者，并将我们的方法与完全非参数Aalen估计量相关的方法进行比较。这些应用说明了我们方法的实用性，并展示了采用部分参数部分非参数（而非完全非参数）所获得的效率提升。本文在第7节以一系列评注作结，其中一些指向进一步研究工作。

2 完全非参数与完全参数情形

~ ~在此我们建立一些记号，并简要描述完全非参数模型中通常使用的估计量 A_1，.....，A_r，见第2.1–2.2节。这些将作为后文拟合参数分量与非参数分量的基础。我们还在第 2.3 节中讨论(1.4)的特殊情形（所有分量均以参数形式指定）的自然估计方法。

我们首先逐一审视并评论一些方便起见而作出的假设，本文将始终采用这些假设。

假设 1. (1) 遍历性：随着 n 增长，所有平均值 $n^{-1}\sum_{i=1}^{n}\phi(z_i)$ 都收敛到适当的极限。这些极限可解释为关于协变量分布的均值。这一假设有助于数学推导，也便于准确陈述估计量极限分布等内容。不过，大样本理论是在

给定观测到的协变量值的条件下建立的，因此给定这些值后，所有随机性都来自 (T_i, δ_i)。(2) 有限时间窗：对个体进行随访的时间区间固定且有限，例如 $[0, \tau]$。这在实践中并不构成限制。在对删失机制作出适当假设的情况下，大多数结果都可以推广到 $\tau = \infty$；但我们仅讨论有限时间范围，因为在此范围内，鞅极限理论的推导更为顺畅，所需技术细节也更少。(3) 独立删失和有限方差：产生数据 (t_i, δ_i) 的删失机制与生成风险率的生存机制无关。此外，对于每个 $s \in [0, \tau]$，$r \times r$ 矩阵函数 $n^{-1} \sum_{i=1}^{n} I(T_i \geq s) z_i z_i^{\mathrm{t}}$ 依概率收敛到一个满秩矩阵 r。这尤其意味着删失分布的支撑集不会严格小于 $[0, \tau]$，并且随着 n增大，在时间 s的风险集中会有足够多个线性无关的协变量向量 z_i。(4) 平滑的参数分量：(1.4)中的$\alpha_j(s, \theta)$关于 θ是平滑的，其一阶导数 $\alpha_j^{*}(s, \theta)$和二阶导数 $\alpha_j^{**}(s, \theta)$连续；对于真实参数 θ_0附近的 θ，上述性质均成立。

■

2.1 一般积分加权最小二乘估计量

数据由每个 n 个体对应的三元组 (t_i, δ_i, z_i) 构成，其中 t_i 是寿命（可能为右删失），δ_i 是未删失指示变量，z_i 是上述 r 维协变量向量。令 $N_i(t) = I\{t_i \leq t, \delta_i = 1\}$ 和 $Y_i(t) = I\{t_i \geq t\}$ 分别表示个体 i 的计数过程和在险指示变量，并引入鞅 $M_i(t) = N_i(t) - \int_0^t Y_i(s) z_i^{\mathrm{t}} \alpha(s)\, \mathrm{d}s$。于是

$$\sum_{i=1}^{n} w_i(s) z_i \, \mathrm{d}N_i(s) = \sum_{i=1}^{n} Y_i(s) w_i(s) z_i z_i^{\mathrm{t}} \, \alpha(s) \, \mathrm{d}s + \sum_{i=1}^{n} w_i(s) z_i \, \mathrm{d}M_i(s), \tag{2.1}$$

此处第二项为均值为零的鞅噪声。这里我们允许使用某些权重函数 $w_i(s)$。这些函数取为可料函数（其在时刻 s 的取值在时刻 $s-$ 已知），最常用的选择是 $w_i(s) = 1$。方程 (2.1) 是以下内容的动机

$$\mathrm{d}\widetilde{A}(s) = G_n(s)^{-1} n^{-1} \sum_{i=1}^{n} w_i(s) z_i \, \mathrm{d}N_i(s), \quad \text{where} \quad G_n(s) = n^{-1} \sum_{i=1}^{n} Y_i(s) w_i(s) z_i z_i^{\mathrm{t}}, \tag{2.2}$$

并伴有相应的累积函数 $\widetilde{A}_j(t) = \int_0^t \mathrm{d}\, \widetilde{A}_j(s)$，适用于 $j = 1$、.....、 r。假设在时刻 s 处的风险集 z_i 中，至少有 r个线性独立，因此 $G_n(s)$ 满秩。

这些估计量的性质已得到充分研究，参见 Aalen、Borgan 和 Gjessing（2008，第 VI 章）。特别是，借助计数过程与鞅的微积分，可以得到大样本结果。我们在此简要回顾相关结果，并引入后续推导所需的符号。考虑

$$U_n(t) = n^{-1/2} \sum_{i=1}^{n} \int_0^t w_i(s) z_i \, \mathrm{d}M_i(s), \tag{2.3}$$

这是一个鞅，其方差过程为 $H_n(t) = n^{-1} \sum_{i=1}^{n} \int_0^t Y_i(s) w_i(s)^2 z_i z_i^{\mathrm{t}} \, z_i^{\mathrm{t}} \alpha(s) \, \mathrm{d}s$。根据假设 1 中描述的正则性条件，可得定义良好的依概率极限，

$$G_n(t) \to_{\mathrm{pr}} G(t) \quad \text{and} \quad H_n(t) \to_{\mathrm{pr}} H(t),$$

随着 n 增大，其中 G 和 H 是满秩 $r \times r$ 矩阵函数。可得

$$\sqrt{n}\{\mathrm{d}\widetilde{A}(s) - \mathrm{d}A(s)\} = G_n(s)^{-1} \, \mathrm{d}U_n(s) \to_d G(s)^{-1} \, \mathrm{d}U(s), \tag{2.4}$$

~其中 U是方差水平为Var d$U(s) = \mathrm{d}H(s)$的高斯鞅。特别地， $\sqrt{n}\{A(t) - A(t)\} \to_d \int_0^t G(s)^{-1}\,\mathrm{d}U(s)$的方差为 $\int_0^t G(s)^{-1}\,\mathrm{d}H(s)\ G(s)^{-1}$。该极限方差可由数据估计为 $\int_0^t G_n(s)^{-1}\,\mathrm{d}\,\widehat{H}_n(s)\,G_n(s)^{-1}$。对$\mathrm{d}H(s)$进行一致估计有若干选择，包括

$$\mathrm{d}\widehat{H}_n(s) = n^{-1}\sum_{i=1}^n Y_i(s)w_i(s)^2 z_i z_i^{\mathrm{t}}\ z_i^{\mathrm{t}}\mathrm{d}\widetilde{A}(s) \ \text{ and } \ \mathrm{d}\widehat{H}(s) = n^{-1}\sum_{i=1}^n w_i(s)^2 z_i z_i^{\mathrm{t}}\,\mathrm{d}N_i(s).$$

在我们的实证研究中，我们采用了第二种方案。

2.2 最优非参数估计

例如，利用与加权最小二乘理论的类比，可以证明，理论上的最优权重（使 $G_n(s)^{-1}\mathrm{d}H_n(s)G_n(s)^{-1}$ 最小）为

$$w_i^0(s) = 1/\{z_i^{\mathrm{t}}\alpha(s)\} \quad \text{for } i = 1,\ldots,n. \tag{2.5}$$

由此得到的最小方差对应于 $F_n(s)^{-1}\mathrm{d}s$，其中

$$F_n(s) = n^{-1}\sum_{i=1}^n Y_i(s)\frac{z_i z_i^{\mathrm{t}}}{z_i^{\mathrm{t}}\alpha(s)}. \tag{2.6}$$

在实践中，需要对这些量进行估计，例如使用 $\widetilde{w}_i(s) = 1/\{z_i^{\mathrm{t}}\widetilde{\alpha}(s)\}$，从而得到

$$\breve{A}(t) = \int_0^t \Big\{n^{-1}\sum_{i=1}^n Y_i(s)\widetilde{w}_i(s)z_i z_i^{\mathrm{t}}\Big\}^{-1} n^{-1}\sum_{i=1}^n \widetilde{w}_i(z)z_i\,\mathrm{d}N_i(s).$$

可以证明，采用估计所得最优权重的 $\sqrt{n}(\breve{A} - A)$ 与采用最优权重的 $\sqrt{n}(\widetilde{A} - A)$ 具有相同的极限分布 $\int_0^{\cdot} F(s)^{-1}\,\mathrm{d}U(s)$，前提是 $\widetilde{\alpha}(s)$ 估计量满足某些一致收敛条件，见 Huffer & McKeague（1991）。$\widetilde{\alpha}(s)$ 的候选方法包括对普通 Aalen估计量进行核平滑（该方法使用 $w_i(s) = 1$），以及局部线性似然平滑。该最优 $\breve{A}$ 估计量的极限分布方差为 $\int_0^t F(s)^{-1}\,\mathrm{d}s$，这是所有 $\int_0^t G(s)\,\mathrm{d}H(s)\,G(s)^{-1}$ 中的最小值。这里， $F(s)$ 是（2.6）中 $F_n(s)$ 的依概率极限，并假设该极限存在。

虽然 $F(s)^{-1}\,\mathrm{d}s$ 的规模可能略小于最常使用的 $G(s)^{-1}\mathrm{d}H(s)G(s)^{-1}$（权重为 $w_i(s) = 1$），但该估计量还涉及额外的变异性贡献，因此在有限的 n 下，并不一定优于 Aalen 估计量。因此，在实证研究中，我们默认选择式 (2.2) 中的“普通权重” $w_i(s) = 1$。

2.3 完全参数模型

现在考虑完全参数模型，其中 $\alpha_j(s) = \alpha_j(s,\theta)$，适用于 $j = 1,..,r$。我们研究^最大似然估计量 θ，即最大化对数似然，其可写为

$$\ell_n(\theta) = \sum_{i=1}^n \int_0^\tau \big[\log\{z_i^{\mathrm{t}}\alpha(s,\theta)\}\,\mathrm{d}N_i(s) - Y_i(s)z_i^{\mathrm{t}}\alpha(s,\theta)\,\mathrm{d}s\big].$$

这里 τ 是观察期的一个上界，假定为有限，见假设 1。令$\alpha^*(s,\theta) = \partial\alpha(s,\theta)/\partial\theta$ 为偏导数 $\partial\alpha_j(s,\theta)/\partial\theta_k$的 $r \times m$ 矩阵，其中 m 是参数向量 θ的长度。假定模型成立，且 θ_0 为真实参数值，令$\Omega_0 = \int_0^\tau \alpha^*(s,\theta_0)^{\mathrm{t}}F(s)\alpha^*(s,\theta_0)\,\mathrm{d}s$，其中 $F(s)$是(2.6)中 $F_n(s)$的依概率极限。于是我们有如下结果。

命题 2.1。在标准正则性条件（包括假设 1 中所述条件）下，并假定模型对真实参数 θ_0成立，且该参数为参数空间的一个内点，则 $\sqrt{n}(\widehat{\theta}-\theta_0)$依分布趋于 $\mathrm{N}_m(0,\Omega_0^{-1})$。

证明。该证明沿用 Borgan (1984) 与 Hjort (1986, 1992) 的思路。我们需要 $z_i^{\mathrm{t}}\alpha(s,\theta)$ 的一阶导数和二阶导数，并分别将其记为维数为 $1\times m$ 的 $\alpha^*(s,\theta)^{\mathrm{t}}z_i$，以及$\sum_{j=1}^{r} z_{i,j}\alpha_j^{**}(s,\theta)$，其中 $\alpha_j^{**}(s,\theta)$ 是 $\alpha_j(s,\theta)$ 的二阶导数的 $m\times m$矩阵。 ℓ_n 的一阶导数为

$$u_n(\theta)=\sum_{i=1}^{n}\int_0^{\tau}\Big\{\frac{\alpha^*(s,\theta)^{\mathrm{t}}z_i}{\alpha(s,\theta)^{\mathrm{t}}z_i}\,\mathrm{d}N_i(s)-Y_i(s)\alpha^*(s,\theta)^{\mathrm{t}}z_i\,\mathrm{d}s\Big\}.$$

利用鞅 $M_i(t)=N_i(t)-\int_0^t Y_i(s)\alpha(s,\theta_0)^{\mathrm{t}}z_i\,\mathrm{d}s$，我们看到

$$n^{-1/2}u_n(\theta_0)=n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\frac{\alpha^*(s,\theta_0)^{\mathrm{t}}z_i}{\alpha(s,\theta_0)^{\mathrm{t}}z_i}\,\mathrm{d}M_i(s),$$

这是一个在 τ 处取值的鞅，其方差过程为

$$\begin{aligned}J_n&=n^{-1}\sum_{i=1}^{n}\int_0^{\tau}\Big(\frac{\alpha^*(s,\theta_0)^{\mathrm{t}}z_i}{\alpha(s,\theta_0)^{\mathrm{t}}z_i}\Big)\Big(\frac{\alpha^*(s,\theta_0)^{\mathrm{t}}z_i}{\alpha(s,\theta_0)^{\mathrm{t}}z_i}\Big)^{\mathrm{t}}Y_i(s)\alpha(s,\theta_0)^{\mathrm{t}}z_i\,\mathrm{d}s\\&=\int_0^{\tau}\alpha^*(s,\theta_0)^{\mathrm{t}}F_n(s)\alpha^*(s,\theta_0)\,\mathrm{d}s.\end{aligned}$$

由此可知，在模型条件下， $n^{-1/2}u_n(\theta_0)$ 趋于一个 $\mathrm{N}_m(0,\Omega_0)$ 随机变量 s.

接下来，我们需要处理 ℓ_n 的二阶导数 $i_n(\theta)$，以证明在该模型下 $-n^{-1}i_n(\theta)=J_n+o_{\mathrm{pr}}(1)$。我们得到

$$i_n(\theta)=\sum_{i=1}^{n}\int_0^{\tau}\Big[\frac{\sum_{j=1}^{r}z_{i,j}\alpha_j^{**}(s,\theta)\alpha(s,\theta)^{\mathrm{t}}z_i-\{\alpha^*(s,\theta)^{\mathrm{t}}z_i\}^2}{\{\alpha(s,\theta)^{\mathrm{t}}z_i\}^2}\,\mathrm{d}N_i(s)-Y_i(s)\sum_{j=1}^{r}z_{i,j}\alpha_j^{**}(s,\theta)\,\mathrm{d}s\Big].$$

再次利用鞅并化简，可得

$$-n^{-1}i_n(\theta)=n^{-1}\sum_{i=1}^{n}\int_0^{\tau}\frac{((\alpha^*)^{\mathrm{t}}z_i)^2}{\alpha^{\mathrm{t}}z_i}Y_i\,\mathrm{d}s+n^{-1}\sum_{i=1}^{n}\int_0^{\tau}\Big[\frac{((\alpha^*)^{\mathrm{t}}z_i)^2}{(\alpha^{\mathrm{t}}z_i)^2}-\frac{\sum_{j=1}^{r}z_{i,j}\alpha_j^{**}}{\alpha^{\mathrm{t}}z_i}\Big]\,\mathrm{d}M_i(s).$$

在真值 θ_0 处，第一项等于 $\int_0^{\tau}(\alpha^*)^{\mathrm{t}}F_n\alpha^*\,\mathrm{d}s=J_n$，而根据 Lenglart 不等式（例如参见 Andersen 等人（1993，第 86 页）），第二项依概率趋于零。进一步分析（与 Hjort（1992，第 2–3 节）的内容类似）最终表明，$\sqrt{n}(\widehat{\theta}-\theta_0)$ 与 $J_n^{-1}n^{-1/2}u_n(\theta_0)$ 至多相差 $o_{\mathrm{pr}}(1)$，而后者服从极限分布 $\mathrm{N}_m(0,\Omega_0^{-1})$。

□

3 参数与非参数模型中的估计

本节介绍参数-非参数模型 (1.4) 的估计方法。这些方法包括步骤（a），用于估计参数部分 $A_{(1)}(t,\theta)$；估计所得结果还会用于步骤（b），以估计非参数部分。具体而言，我们对这些$A_{(2)}(t)$所用的估计量利用了 $A_{(1)}(t,\theta)$的参数结构，并不等同于直接的 Aalen$\tilde{}$估计量 $A_{(2)}(t)$；这样做的目的是利用参数知识，提高精度。

3.1 参数部分的估计

我们首选的步骤（a）版本如下。理想情况下，应找到使积分加权二次型

$$\int_0^\tau \{\alpha_{(1)}(s,\theta)-\alpha_{(1)}(s)\}^{\mathrm{t}} V_n(s)\{\alpha_{(1)}(s,\theta)-\alpha_{(1)}(s)\}\,\mathrm{d}s$$

尽可能小。这里，τ 是一个上限时间点，可根据当前应用的便利性来选择，而 $V_n(s)$是一个满秩对称 $p\times p$ 矩阵权重函数。由于二次型取决于未知函数，因此无法直接实现这一最小化。不过，展开后并略去不涉及参数的那一项，经验版本

$$C_n(\theta)=\int_0^\tau \alpha_{(1)}(s,\theta)^{\mathrm{t}} V_n(s)\alpha_{(1)}(s,\theta)\,\mathrm{d}s-2\int_0^\tau \alpha_{(1)}(s,\theta)^{\mathrm{t}} V_n(s)\,\mathrm{d}\widetilde{A}_{(1)}(s), \tag{3.1}$$

~ ~便由此得到。这里，$\mathrm{d}A_{(1)}(s)$包含式（2.2）中非参数$\mathrm{d}A(s)$的前 p 个分量，并且^我们令 θ 为准则函数 $C_n(\theta)$的最小化者。

注意，$V_n(s)$完全可以依赖数据。我们通常考虑这样的情形：对于一个合适的极限矩阵函数，$V_n(s)\ \to_{\mathrm{pr}}\ V(s)$；见下一节，其中我们还将给出 $V_n(s)$的一种特定选择，该选择在 n较大时可带来最优性能。不过，这涉及非平凡的估计量 $1/\{z_i^{\mathrm{t}}\widetilde{\alpha}(s)\}$，其讨论见式（2.5）–（2.6）；而且对于较小和中等的 n，这些估计量往往过于不稳定。我们的默认选择更为简单：

$$V_n(s)=n^{-1}\sum_{i=1}^n Y_i(s)z_{i,(1)}z_{i,(1)}^{\mathrm{t}}, \tag{3.2}$$

$p\times p$($n^{-1}\sum_{i=1}^n Y_i)sz_iz_i^{\mathrm{t}}$的左上角区块。根据假设1，它有一个定义良好的依概率极限函数$V(s)$。对于参数化风险函数分量为常数这一最简单的情形，令$\alpha_{(1,j)}(s,\theta)\ =\theta_j$，其中 $j=1,\ldots,p$，则上式可得

$$\widehat{\theta}=\Big\{\int_0^\tau V_n(s)\,\mathrm{d}s\Big\}^{-1}\int_0^\tau V_n(s)\,\mathrm{d}\widetilde{A}_{(1)}(s).$$

^ ~这些是最佳常数，可视为对非参数 $A_{(1,j)}(t)$的近似 $\theta_j t$，适用于 $t\in[0,\tau]$和 $j=1,\ldots,p$；权重矩阵 $V_n(s)$的选取也决定了这一点。

^采用式（3.2）中的默认权重函数时，估计量 θ 与 McKeague & Sasieni（1994，式（2.4），第503页）提出的估计量相似，但并不完全相同。为了得到他们的估计量，McKeague & Sasieni 对得分函数作适当修改，由此得到一个关于 θ 的线性估计方程（即他们的 β）。类似的技术也可用于更一般的参数化风险函数，因此，我们在此使用的准则函数 $C_n(\theta)$或许可以替换为略有不同的准则函数。

3.2 通过回代重新估计非参数部分

^现在，在步骤（a）得出$j=1$、.....、p的参数估计量 $\alpha_j(s,\ \theta)$之后，我们来描述步骤（b）的一种形式。考虑方程（2.1）的非参数部分，即

$$\sum_{i=1}^n w_i(s)z_{i,(2)}\,\mathrm{d}N_i(s)\ =\sum_{i=1}^n Y_i(s)w_i(s)z_{i,(2)}\{z_{i,(1)}^{\mathrm{t}}\alpha_{(1)}(s,\theta)+z_{i,(2)}^{\mathrm{t}}\alpha_{(2)}(s)\}\,\mathrm{d}s$$
$$+\sum_{i=1}^n w_i(s)z_{i,(2)}\,\mathrm{d}M_i(s).$$

现在，既然工作是在（1.4）框架内进行的，其中涉及的鞅可以更精确地定义为

$$M_i(t) = N_i(t) - \int_0^t Y_i(s)\{z_{i,(1)}^{\mathrm{t}}\alpha_{(1)}(s,\theta_0) + z_{i,(2)}^{\mathrm{t}}\alpha_{(2)}(s)\}\,\mathrm{d}s, \tag{3.3}$$

其中 θ_0 为真实参数。为利用参数知识，从而提高非参数分量的估计精度，促使我们采用

$$\sum_{i=1}^n w_i(s)z_{i,(2)}\{\mathrm{d}N_i(s) - Y_i(s)z_{i,(1)}^{\mathrm{t}}\alpha_{(1)}(s,\widehat{\theta})\,\mathrm{d}s\} = \sum_{i=1}^n Y_i(s)w_i(s)z_{i,(2)}z_{i,(2)}^{\mathrm{t}}\,\mathrm{d}\alpha_{(2)}(s) + \text{noise}$$

来构造

$$\mathrm{d}\widehat{A}_{(2)}(s) = G_{n,22}(s)^{-1}n^{-1}\sum_{i=1}^n w_i(s)z_{i,(2)}\{\mathrm{d}N_i(s) - Y_i(s)z_{i,(1)}^{\mathrm{t}}\alpha_{(1)}(s,\widehat{\theta})\,\mathrm{d}s\}. \tag{3.4}$$

这定义了针对 $j = p+1,\ldots,p+q$ 的修正估计量 $A_j(t)$。这里 $G_{n,\,22}(s)$ 是 $G_n(s)$的下方 $q\times q$子矩阵。

需要注意的是，这里概述的方法实际上是一类程序，因为在 (3.4) 中可以使用不同的权重方案，而且在最小化 $C_n(\theta)$函数以获得 θ 估计量时，也可以使用不同的权重函数 V_n 。例如，在 (3.4) 中，我们可以使用普通权重 $w_i(s) = 1$，也可以使用第 2.2 节中更复杂的 $\widetilde{w}_i(s)$。下一节将给出一个渐近最优方案。

4 大样本行为与最优性

这里我们证明第 3 节中估计量的渐近正态性，即最小化(3.1)中的 $C_n(\theta)$ 和 (3.4) 中的 $A_{(2)}(t)$，并给出极限分布的方差与协方差的精确公式。结果是在模型条件 (1.4) 下推导出的，其中 θ_0 表示参数部分的真实参数 $\alpha_{(1),\,j}(s,\theta)$ 对应于 $j = 1,\ldots,p$。令 $\alpha_{(1)}^*(s,\theta)$ 为分量函数的一阶导数矩阵 $p\times m$，其元素为 $\alpha_j^*(s,\theta) = \partial\alpha_j(s,\theta)/\partial\theta$，其中 m 是完整θ 向量的长度。

4.1 参数部分的大样本理论

为了研究我们的估计量，我们还需要函数 $Q(s)$，其定义如下：

$$Q(s)\,\mathrm{d}s = [G(s)^{-1}\,\mathrm{d}H(s)G(s)^{-1}]_{11}, \tag{4.1}$$

也就是说，与 Aalen 估计量前 p 个分量的方差（即 $A_{(1)}$；见 (2.4)）相关的完整 $G(s)^{-1}\mathrm{d}H(s)G(s)^{-1}$ 矩阵的左上角 $p\times p$ 块矩阵。

命题 4.1. 假设假设 1 中列出的正则性条件均成立，且 $V_n(s) \to_{\mathrm{pr}} V(s)$ 在 $s\in[0,\tau]$ 上一致成立。那么，在参数模型的条件下，$\Lambda_n = \sqrt{n}(\theta-\theta_0)$ 收敛于 $\mathrm{N}_m(0,\Gamma^{-1}\Omega\Gamma^{-1})$，其中

$$\Gamma = \int_0^\tau \alpha_{(1)}^*(s,\theta_0)^{\mathrm{t}}V(s)\alpha_{(1)}^*(s,\theta_0)\,\mathrm{d}s \quad\text{and}\quad \Omega = \int_0^\tau \alpha_{(1)}^*(s,\theta_0)^{\mathrm{t}}V(s)Q(s)V(s)\alpha_{(1)}^*(s,\theta_0)\,\mathrm{d}s.$$

证明。令准则函数 (3.1) 的导数等于零，得到估计方程 $S_n(\theta) = 0$，其中

$$S_n(\theta) = \int_0^\tau \alpha_{(1)}^*(s,\theta)^{\mathrm{t}}V_n(s)\{\mathrm{d}\widetilde{A}_{(1)}(s) - \alpha_{(1)}(s,\theta)\,\mathrm{d}s\}.$$

^这在适当条件下将 θ 重新定义为一种 M 型估计量；参见 Hjort（1985年，第 4 节）、Hjort（1992年，第5节）。注意，

$$\sqrt{n}S_n(\theta_0) \to_d \int_0^\tau \alpha^*_{(1)}(s,\theta_0)^{\mathrm{t}} V(s)[G(s)^{-1}\,\mathrm{d}U(s)]_{(1)} = S,$$

在真实的 θ_0 下，它服从方差矩阵为 Ω 的零均值正态分布。再稍作推导，可得 $\theta(\ S_n)$ 的 Γ_n 矩阵的表达式，其中包含 $\theta(\ m\ )$ 对 参数的导数的负值，即

$$\Gamma_n(\theta) = \int_0^\tau \alpha^*_{(1)}(s,\theta)^{\mathrm{t}} V_n(s)\alpha^*_{(1)}(s,\theta)\,\mathrm{d}s + E_n(\theta).$$

这里，第二个矩阵的各分量是如下各项的线性组合： θ 的光滑有界 ~ 函数，乘以 $\mathrm{d}A_{(1)}(s) - \alpha_{(1)}(s,\theta)\,\mathrm{d}s$ 的各分量，并对 $[0,\tau]$ 积分。关键在于，在模型条件下，$\theta(\ E_n)$ 的所有项在 θ_0 时都依概率趋于零，因此 $\Gamma_n(\theta_0) \to_{\mathrm{pr}} \Gamma$。由此得到 $\Lambda_n \to_d \Gamma^{-1}S$，从而证明了该结论。

□

在模型条件下追求最佳性能，至少对于较大的 n 而言，等同于选择 $p\times p$ 矩阵函数 V，使 $\Gamma^{-1}\Omega\Gamma^{-1}$ 最小。假设 $Q(s)$ 在整个范围 $[0,\tau]$ 上均为满秩 $p\times p$，则当 $V(s)$ 与 $Q(s)^{-1}$ 成正比时，可达到这一目标。于是 $\Gamma = \Omega\ = \Omega_0,$，记作 Ω，此时最小方差矩阵等于

$$\Omega_0^{-1} = \Big\{\int_0^\tau \alpha^*_{(1)}(s,\theta_0)^{\mathrm{t}} Q(s)^{-1}\alpha^*_{(1)}(s,\theta_0)\,\mathrm{d}s\Big\}^{-1}. \tag{4.2}$$

为证明这是维数最小的矩阵，设 $Z(t)$ 是一个高斯鞅，其增量方差为 $\mathrm{Var}\,\mathrm{d}Z(s) = Q(s)\,\mathrm{d}s$，并考虑随机向量 $X = \int_0^\tau \alpha^*_{(1)}V\mathrm{d}Z\,\mathrm{and}Y = \int_0^\tau \alpha^*_{(1)}Q^{-1}\,\mathrm{d}Z$。它们的联合方差矩阵为

$$\Sigma = \begin{pmatrix} \int_0^\tau (\alpha^*_{(1)})^{\mathrm{t}} VQV\alpha^*_{(1)}\,\mathrm{d}s & \int_0^\tau (\alpha^*_{(1)})^{\mathrm{t}} V\alpha^*_{(1)}\,\mathrm{d}s \\ \int_0^\tau (\alpha^*_{(1)})^{\mathrm{t}} V\alpha^*_{(1)}\,\mathrm{d}s & \int_0^\tau (\alpha^*_{(1)})^{\mathrm{t}} Q^{-1}\alpha^*_{(1)}\,\mathrm{d}s \end{pmatrix}.$$

采用通常的分块记法， $\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ 必须是半正定的。这等价于所述的最小化结论。

接下来的问题是如何使 Ω_0^{-1} 尽可能小。但这等同于对 $Q(s)\,\mathrm{d}s = [G(s)^{-1}\,\mathrm{d}H(s)\,G(s)^{-1}]_{11}$ 进行最小化；我们已经看到，这在采用最优权重 (2.5) 时能够实现，此时有 $Q(s) = [F(s)^{-1}]_{11} = F^{11}(s)$，记作如此。因此，渐近最优方法是将一个在可能情况下依概率收敛至 $V(s) = F^{11}(s)^{-1}$ 的矩阵函数用作 $V_n(s)$。而这可通过以下方式实现

$$V_n(s) = \widetilde{F}_n^{11}(s)^{-1} = \widetilde{F}_{n,11}(s) - \widetilde{F}_{n,12}(s)\widetilde{F}_{n,22}(s)^{-1}\widetilde{F}_{n,21}(s),$$

~其中 F_n 与（2.6）中的 F_n 相同，但插入了权重 $z_i^{\mathrm{t}}\widetilde{\alpha}(s)$。我们可以得出结论：该方法在 n较大时给出最优性能，其极限方差矩阵

$$\Big\{\int_0^\tau (\alpha^*_{(1)})^{\mathrm{t}}(F^{11})^{-1}\alpha^*_{(1)}\,\mathrm{d}s\Big\}^{-1} = \Big\{\int_0^\tau (\alpha^*_{(1)})^{\mathrm{t}}(F_{11} - F_{12}F_{22}^{-1}F_{21})\alpha^*_{(1)}\,\mathrm{d}s\Big\}^{-1}. \tag{4.3}$$

事实上，采用任何其他估计方法都不可能对此加以改进。附录 A 详细说明了这一点确实如此。

4.2 非参数部分的大样本理论

$\widehat{\quad}$ $\widehat{\quad}$为了研究 A_{p+1}、.....、 A_{p+q} 的行为，我们需要 $q \times m$ 函数

$$\phi_n(s) = n^{-1} \sum_{i=1}^{n} Y_i(s) w_i(s) z_{i,(2)} z_{i,(1)}^{\mathrm{t}} \alpha_{(1)}^*(s, \theta_0),$$

在先前所述的温和一般条件下，其依概率极限函数为 $\phi(s)$。

命题 4.2。假设命题 4.1 的正则性条件成立，并令 $\Lambda = \Gamma^{-1} S$ 为 $\Lambda_n = \sqrt{n}(\widehat{\theta} - \theta_0)$ 的极限变量。则存在过程收敛

$$\sqrt{n}\{\widehat{A}_{(2)}(t) - A_{(2)}(t)\} \to_d \int_0^t G_{22}(s)^{-1}\, \mathrm{d}U_{(2)}(s) - \int_0^t G_{22}(s)^{-1} \phi(s)\, \mathrm{d}s\, \Lambda \tag{4.4}$$

于空间 $D[0, \tau]$ 中，该空间由 $[0, \tau]$ 上具有左极限的右连续函数构成，并配备斯科罗霍德拓扑。

证明。从 (3.3) 和 (3.4) 出发进行一些代数运算，可得

$$\begin{aligned} \mathrm{d}\widehat{A}_{(2)}(s) \;=\; & G_{n,22}(s)^{-1} n^{-1} \sum_{i=1}^{n} w_i(s) z_{i,(2)} [\mathrm{d}M_i(s) + Y_i(s) z_{i,(2)}^{\mathrm{t}} \alpha_{(2)}(s)\, \mathrm{d}s \\ & - Y_i(s) z_{i,(1)}^{\mathrm{t}} \{\alpha_{(1)}(s, \widehat{\theta}) - \alpha_{(1)}(s, \theta_0)\}\, \mathrm{d}s], \end{aligned}$$

由此可得

$$\sqrt{n}\{\mathrm{d}\widehat{A}_{(2)}(s) - \alpha_{(2)}(s)\, \mathrm{d}s\} \doteq G_{n,22}(s)^{-1} \Big\{ n^{-1/2} \sum_{i=1}^{n} w_i(s) z_{i,(2)}\, \mathrm{d}M_i(s) - \phi_n(s)\, \mathrm{d}s\, \Lambda_n \Big\}.$$

这里的 $X_n \doteq X_n'$ 表示该差值依概率趋于零。该结论可由 $D[0, \tau]$ 空间中过程收敛的一般理论推出。 □

命题 4.1 和 4.2 分别清楚地描述了我们的参数$\widehat{\quad}$和非参数估计量的大样本行为。为了对同时涉及两部分的量进行推断，例如生存曲线 $S(t \mid z)$，其中 $A(t \mid z) = z_1 A_1(t, \theta) + \cdots + z_p A_p(t, \theta) + z_{p+1} A_{p+1}(t) + \cdots + z_{p+q} A_{p+q}(t)$，我们还需要 θ 和$\sim A_{(2)}(t)$ 的联合极限分布。下面给出 $A_{(1)}(t, \theta)$ 和 $A_{(2)}(t)$ 联合极限分布的详细信息$\widehat{\quad}$ $\widehat{\quad}$。确实有

$$\sqrt{n} \begin{pmatrix} A_{(1)}(t, \widehat{\theta}) - A_{(1)}(t, \theta_0) \\ \widehat{A}_{(2)}(t) - A_{(2)}(t) \end{pmatrix} \xrightarrow{d} \mathrm{N}\big(0, \Xi(t)\big), \quad \text{with} \quad \Xi(t) = \begin{pmatrix} \Xi_{11}(t) & \Xi_{12}(t) \\ \Xi_{21}(t) & \Xi_{22}(t) \end{pmatrix}, \tag{4.5}$$

方差矩阵的公式将在后文给出。

在 (4.4) 中，第一项是方差为 $\int_0^t G_{22}^{-1} \mathrm{d}H_{22} G_{22}^{-1}$ 的高斯鞅，而第二项也是正态分布，其方差可通过命题 4.1 写出。结合命题 4.1 和 4.2，并应用 Delta方法，我们得到 (4.5)。首先，$\Xi_{11}(t) = A^*(t, \theta_0) \Gamma^{-1} \Omega \Gamma^{-1} A^*(t, \theta_0)^{\mathrm{t}}$，其中 $A^*(t, \theta)$ 是一个分量为 $\partial A_j(t, \theta)/\partial \theta$ 的 $p \times m$ 矩阵，$j = 1, \ldots, p$。其次，$\Xi_{22}(t)$ 是 (4.4) 的方差。为此，对于 (4.4) 中两项之间的协方差，我们有

$$\begin{aligned} \mathrm{E}\Big(\textstyle\int_0^t G_{22}^{-1} \mathrm{d}U_{(2)} \Big) S^{\mathrm{t}} \;&=\; \mathrm{E}\Big(\int_0^t G_{22}^{-1} \mathrm{d}U_{(2)} \Big) \int_0^\tau (\mathrm{d}U_{(1)}^{\mathrm{t}} G^{11} + \mathrm{d}U_{(2)}^{\mathrm{t}} G^{21}) V \alpha_{(1)}^* \\ &= \int_0^t G_{22}^{-1} (\mathrm{d}H_{21} G^{11} + \mathrm{d}H_{22} G^{21}) V \alpha_{(1)}^*, \end{aligned} \tag{4.6}$$

因此，(4.4) 右端的总方差为

$$\Xi_{22}(t) = \int_0^t G_{22}(s)^{-1}\mathrm{d}H_{22}(s)G_{22}(s)^{-1} + \int_0^t G_{22}(s)^{-1}\phi(s)\,\mathrm{d}s\,\Gamma^{-1}\Omega\Gamma^{-1}\Big(\int_0^t G_{22}(s)^{-1}\phi(s)\,\mathrm{d}s\Big)^{\mathrm{t}}$$
$$-2\int_0^t G_{22}(s)^{-1}\{\mathrm{d}H_{21}(s)G^{11}(s) + \mathrm{d}H_{22}(s)G^{21}(s)\}V(s)\alpha_{(1)}^*(s)\Gamma^{-1}\Big(\int_0^t G_{22}(s)^{-1}\phi(s)\,\mathrm{d}s\Big)^{\mathrm{t}}.$$

第三，(4.5) 中协方差矩阵的下方非对角块为

$$\Xi_{21}(t) = \mathrm{E}\int_0^t G_{22}^{-1}\,\mathrm{d}U_{(2)}S^{\mathrm{t}}\Gamma^{-1}\,A^*(t,\theta_0)^{\mathrm{t}} - \mathrm{E}\int_0^t \phi(s)\,\mathrm{d}s\,\Gamma^{-1}SS^{\mathrm{t}}\Gamma^{-1}\,A^*(t,\theta_0)^{\mathrm{t}}$$
$$= \int_0^t G_{22}^{-1}(\mathrm{d}H_{21}G^{11} + \mathrm{d}H_{22}G^{21})V\alpha_{(1)}^*\,\Gamma^{-1}\,A^*(t,\theta_0)^{\mathrm{t}} - \int_0^t \phi(s)\,\mathrm{d}s\,\Gamma^{-1}\Omega\Gamma^{-1}\,A^*(t,\theta_0)^{\mathrm{t}},$$

其中我们使用了 (4.6)，且 $\Xi_{12}(t) = \Xi_{21}(t)^{\mathrm{t}}$。这些协方差矩阵的估计方法显而易见，例如，在使用传统的 Aalen 估计量权重 $w_i(s) = 1$ 时。

研究 $V(s) = F^{11}(s)^{-1}$ 这一特殊情形很有意思；根据上述结果，这会带来最优的大样本性能。此时，极限过程的两个项实际上相互独立。这可由 $\mathrm{d}H(s) = F(s)\,\mathrm{d}s$ 和 $G = F$ 得出。因此，对于这种情形，(4.4) 中极限过程的协方差函数可写为

$$\int_0^{t_1\wedge t_2} F_{22}(s)^{-1}\,\mathrm{d}s + J(t_1)\Omega_0^{-1}J(t_2)^{\mathrm{t}},$$

其中 $J(t) = \int_0^t F_{22}^{-1}\phi\,\mathrm{d}s$，且 ϕ 是以下表达式的极限

$$\phi_n(s) = n^{-1}\sum_{i=1}^n Y_i(s)z_{i,(2)}\frac{z_{i,(1)}^{\mathrm{t}}\alpha_{(1)}^*(s,\theta_0)}{z_{i,(1)}^{\mathrm{t}}\alpha_{(1)}(s,\theta_0) + z_{i,(2)}^{\mathrm{t}}\alpha_{(2)}(s)}.$$

5 拟合优度评估

我们研究了参数-非参数模型 (1.4)，构造了参数分量的估计量 $\alpha_j(s,\theta)$，并推导出大样本性质，从而为所有相关量提供了推断方法，且精度优于传统的非参数方法。这些良好结果所依据的假设是参数结构确实成立。本节将构造监测过程及相关检验，以评估参数部分的适切性。

5.1 拟合优度过程

对于每个 j，我们可以考虑形如$\sqrt{n}\int_0^t K_{n,\,j}(s)\{\mathrm{d}\,\widetilde{A}_j(s) - \alpha_j(s,\widehat{\theta})\,\mathrm{d}s\}$的监测过程，其中$K_{n,\,j}$是一个合适的权重函数。更一般地，令

$$R_n(t) = \sqrt{n}\int_0^t K_n(s)\{\mathrm{d}\widetilde{A}_{(1)}(s) - \alpha_{(1)}(s,\widehat{\theta})\,\mathrm{d}s\}, \tag{5.1}$$

配有一个由权重函数构成的完整 $p\times p$ 矩阵 $K_{n,\,ij}(s)$。可以将这些过程随时间绘制出来，以判断参数建模假设是否合理。所用估计量 θ 与第 3.1 节中的相同，它取决于一个矩阵权重函数 V_n；在模型条件下，$\Lambda_n = \sqrt{n}(\theta-\theta)$ 收敛于 $\Lambda = \Gamma^{-1}S \sim \mathrm{N}_m(0,\Gamma^{-1}\Omega\Gamma^{-1})$，其定义和推导见第4.1节。

命题 5.1。假设 $K_{n,\ ij}$ 函数是可预见的，并且在 $[0,\tau]$ 上依概率一致收敛于 K_{ij} 函数；第 4 节两个命题所关联的正则性条件均成立；且参数模型对风险函数$\alpha_j(s,\theta)$（其中 $j=1,\dots,p$）成立。那么，在赋予 Skorokhod乘积拓扑的空间 $D([0,\tau]^p)$ 中，存在过程收敛，并且

$$R_n(t) \to_d R(t) = \int_0^t K[G^{-1}\,\mathrm{d}U]_{(1)} - \int_0^t K\alpha^*_{(1)}\,\mathrm{d}s\,\Lambda.$$

证明。像之前一样令 $\alpha_j^*(s,\theta) = \partial\alpha_j(s,\theta)/\partial\theta$，并使用（2.3）中的 $\mathrm{d}U_n(s)$ 所对应的表示式（2.4），此处的关键在于

$$\sqrt{n}\{\mathrm{d}\widetilde{A}_j(s) - \alpha_j(s,\widehat{\theta})\} \doteq [G_n(s)^{-1}\,\mathrm{d}U_n(s)]_j - \alpha_j^*(s,\theta)^{\mathrm{t}}\sqrt{n}(\widehat{\theta}-\theta)\,\mathrm{d}s$$

对于 $j=1,\dots,p$，通过泰勒分析可得。由第 4 节的方法和结果可知，$G_n(s)^{-1}\,\mathrm{d}U_n(s)$ 与 $\sqrt{n}(\widehat{\theta}-\theta)$ 联合分布收敛到 $G(s)^{-1}\,\mathrm{d}U(s)$ 和 $\Lambda=\Gamma^{-1}S$。令 $W(t) = \int_0^t[G(s)^{-1}\,\mathrm{d}U(s)]_{(1)}$，于是 $\mathrm{d}W(s) = G^{11}(s)\,\mathrm{d}U_{(1)}(s) + G^{12}(s)\,\mathrm{d}U_{(2)}(s)$。于是我们有

$$\sqrt{n}\{\mathrm{d}\widetilde{A}_{(1)}(s) - \alpha_{(1)}(s,\widehat{\theta})\} \to_d \mathrm{d}W(s) - \alpha^*_{(1)}(s,\theta)\Gamma^{-1}\int_0^\tau \alpha^*_{(1)}(s,\theta_0)^{\mathrm{t}}V(s)\,\mathrm{d}W(s).$$

权重函数 $K_n(s)$ 依概率一致收敛到 $K(s)$，由此我们得到 $R_n \to_d R$。其细节和方法与 Hjort（1990年，第 3–4 节）中用于类似但略有不同设定的方法相近。

□

极限过程 $R_1,\dots,R_p$ 服从联合高斯分布，均值为零。为了求出它们的协方差函数，我们利用命题证明中得到的结构。W 过程是一个具有独立增量的正态鞅，并且 $\mathrm{Var}\,\mathrm{d}W(s) = Q(s)\,\mathrm{d}s$，与之前相同，见 (4.1)。于是

$$R(t) = \int_0^t K\,\mathrm{d}W - \Psi(t)\Lambda, \quad \text{with} \quad \Lambda = \Gamma^{-1}\int_0^\tau (\alpha^*_{(1)})^{\mathrm{t}}V\,\mathrm{d}W,$$

同时将 $\Psi(t)$ 记作 $p\times m$ 矩阵 $\int_0^t K\alpha^*_{(1)}\,\mathrm{d}s$。对

$$\begin{aligned} R(t_1)R(t_2)^{\mathrm{t}} \;=\; & \int_0^{t_1} K\,\mathrm{d}W \int_0^{t_1} \mathrm{d}W^{\mathrm{t}}K^{\mathrm{t}} + \Psi(t_1)\Lambda\Lambda^{\mathrm{t}}\Psi(t_2)^{\mathrm{t}} \\ & - \int_0^{t_1} K\,\mathrm{d}W\,\Lambda^{\mathrm{t}}\Psi(t_2)^{\mathrm{t}} - \Psi(t_1)\Lambda\int_0^{t_2}\mathrm{d}W^{\mathrm{t}}K^{\mathrm{t}}, \end{aligned}$$

利用 W 的零均值独立增量性质，取其均值可得

$$\int_0^{t_1\wedge t_2} KQK^{\mathrm{t}}\,\mathrm{d}s + \Psi(t_1)\Gamma^{-1}\Omega\Gamma^{-1}\Psi(t_2)^{\mathrm{t}} - \Phi(t_1)\Gamma^{-1}\Psi(t_2)^{\mathrm{t}} - \Psi(t_1)\Gamma^{-1}\Phi(t_2)^{\mathrm{t}},$$

其中 $\Phi(t)$ 是 $p\times m$ 矩阵函数 $\int_0^t KQV\alpha^*_{(1)}\,\mathrm{d}s$。

(5.1) 框架涉及完整的权重函数矩阵，并给出用于同时监测的 p 过程。我们注意到单个 $p\times 1$ 权重函数 $K_n = (K_{n,1},\dots,K_{n,p})^{\mathrm{t}}$ 的特殊情形，此时只需考虑一个监测过程，便可从上述结果中直接读出相应结论。因此，比较增量的线性组合

$$\begin{aligned} R_n^*(t) \;&=\; \sqrt{n}\int_0^t K_n(s)^{\mathrm{t}}\{\mathrm{d}\widetilde{A}_{(1)}(s) - \alpha_{(1)}(s,\widehat{\theta})\,\mathrm{d}s\} \\ &=\; \sqrt{n}\sum_{j=1}^k \int_0^t K_{n,j}(s)\{\mathrm{d}\widetilde{A}_j(s) - \alpha_j(s,\widehat{\theta})\,\mathrm{d}s\} \end{aligned}$$

作为一个过程依分布收敛于 $R^*(t) = \int_0^t K^{\mathrm{t}}\,\mathrm{d}W - \psi(t)\Lambda$，其中此时 $\psi(t) = \int_0^t K^{\mathrm{t}}\alpha^*_{(1)}\,\mathrm{d}s$。特别地，若 $K_n = (0,\ldots,K_{n,j},\ldots,0)^{\mathrm{t}}$，我们便得到各自的监测过程

$$R_{n,j}(t) = \sqrt{n}\int_0^t K_{n,j}(s)\{\mathrm{d}\widetilde{A}_j(s) - \alpha_j(s,\widehat{\theta})\,\mathrm{d}s\}, \quad \text{for } j = 1,\ldots,p. \tag{5.2}$$

该 $R_{n,j}(t)$ 在分布上趋于

$$R_j(t) = \int_0^t K_j(s)\,\mathrm{d}W_j(s) - \psi_j(t)^{\mathrm{t}}\Gamma^{-1}S, \quad \text{with} \quad S = \int_0^\tau (\alpha^*_{(1)})^{\mathrm{t}}V\,\mathrm{d}W, \tag{5.3}$$

其中 $\psi_j(t) = \int_0^t K_j(\alpha^*_j)^{\mathrm{t}}\,\mathrm{d}s$（大小为 $m\times 1$）。通过与上述类似的计算，协方差函数 $\mathrm{cov}\{R_j(t_1)\ R_j(t_2)\}$ 可表示为

$$\int_0^{t_1\wedge t_2} K_j^2 Q_{jj}\,\mathrm{d}s + \psi_j(t_1)^{\mathrm{t}}\Gamma^{-1}\Omega\Gamma^{-1}\psi_j(t_2) - \psi_j(t_1)^{\mathrm{t}}\Gamma^{-1}\Phi_j(t_2) - \psi_j(t_2)^{\mathrm{t}}\Gamma^{-1}\Phi_j(t_1), \tag{5.4}$$

其中

$$\Phi_j(t) = \mathrm{E}\int_0^\tau (\alpha^*_{(1)})^{\mathrm{t}}V\,\mathrm{d}W\int_0^t K_j\,\mathrm{d}W_j = \int_0^t K_j(s)\alpha^*_{(1)}(s,\theta_0)^{\mathrm{t}}V(s)Q^{(j)}(s)\,\mathrm{d}s,$$

将 $Q^{(j)}(s)$ 记为矩阵 $p\times p$ (s) 的第 j 列。与 $\psi_j(t)$ 一样，$\Phi_j(t)$ 的大小为 $m\times 1$。

5.2 卡方检验

将时间观测期 $[0,\tau]$ 划分为时间窗 $I_\ell = (c_{\ell-1}, c_\ell]$（$\ell = 1,\ldots,k$），其中$c_0 = 0$ 且 $c_k = \tau$。对于每个时间窗，我们可以计算 p 维增量 $\Delta R_n(I_\ell) = R_n(c_\ell) - R_n(c_{\ell-1})$。根据命题 5.1，这些增量的集合在分布上趋于 $\Delta R(I_\ell) = R(c_\ell) - R(c_{\ell-1})$ 的分布；在模型假设下，该分布为均值为零的多元正态分布，其协方差结构可根据上述结果计算。

我们或许可以稍显隆重地，通过 p 维 $\Delta R_n(I_j)$，检验所有 $\alpha_j(s,\theta)$ 分量均成立这一完整的联合参数假设。这里我们概述一些更简单但自然的策略，着眼于一次研究一个 $\alpha_j(s,\theta)$。为此，对于给定的权重函数 $K_{n,\,j}(s)$，我们按照 (5.2)– (5.3) 使用 $R_{n,\,j}(t) \to_d R_j(t)$。我们计算增量 $\Delta R_{n,\,j,\,\ell} = R_{n,\,j}(I_\ell)$，这些增量联合趋于增量向量 $\Delta R_j(I_\ell) = \Delta\{R_j(c_\ell) - R_j(c_{\ell-1})\}$。这是一个均值为零的多元正态分布，记作 $\mathrm{N}_k(0,\Sigma_j)$，其中 Σ_j 是由协方差函数 (5.4) 得出的相应协方差矩阵。现在我们可以通过几种方式检验 $\alpha_j(s,\theta)$ 假设。具体而言，

$$C_{n,j} = \Delta_{n,j}^{\mathrm{t}}\widehat{\Sigma}_j^{-1}\Delta_{n,j} \to_d C_j = \Delta_j^{\mathrm{t}}\Sigma_j^{-1}\Delta_j \sim \chi^2_k, \tag{5.5}$$

其中 $\Delta_{n,\,j}$ 是 $\Delta R_{n,\,j}$ 的向量，其分布趋于 Δ_j，即 $\Delta R_j(I_\ell)$ 向量；$\widehat{\ }$并且 Σ_j 是矩阵 Σ_j 的一个相合估计量。

5.3 其他检验

原则上，根据（5.1）的监测过程 R_n 构造其他检验统计量并不困难，不过其精确零分布或极限零分布可能难以制表或评估。然而，可以采用一些方法来近似这些分布，下面我们将对此加以说明。考虑（5.2）中的$R_{n,\,j}$，其中$K_{n,\,j}$取适当值，并定义

$$\|R_{n,j}\| = \max_{t\le\tau}|R_{n,j}(t)| \quad \text{for } j = 1,\ldots,p.$$

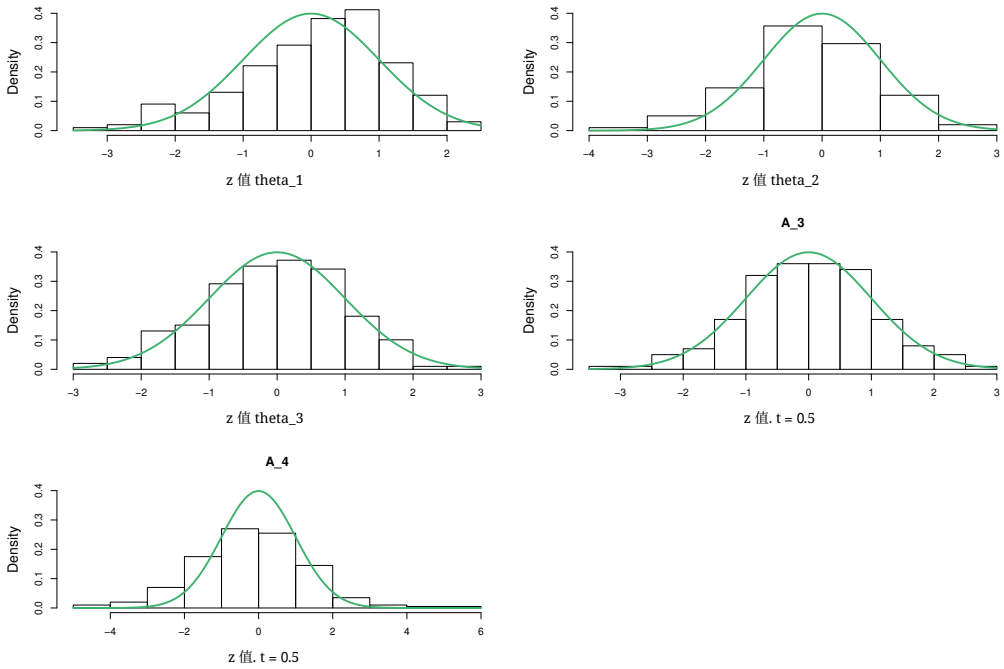


图 6.1: $\sqrt{n}(\widehat{\theta}_k - \theta_k)/\mathrm{se}(\widehat{\theta}_k)$ 和 $\sqrt{n}\{\widehat{A}_j(t) - A_j(t)\}/\mathrm{se}(\widehat{A}_j(t))$ 的直方图，其中 $k = 1$、2、3 和 $j = 3$、4。累积回归量在 $t = 0.5$ 处取值。样本量设为 $n = 2000$，直方图基于 200 次模拟。绿色曲线表示标准正态密度。

这些柯尔莫哥洛夫–斯米尔诺夫型检验具有明确的极限分布，即 $\max_{t\le\tau}|R_j(t)|$，其中 $R_j(t)$如（5.3）所示，是一个由 $W(t) = \int_0^t [G(s)^{-1}\,\mathrm{d}U(s)]_{(1)}$定义的过程。确定$\|R_{n,\ j}\|$ 零分布上临界点的方案包括以下几种：（i）可以从极限分布中模拟，并采用 K_j、 Q 和 $\alpha_j^*(s,\theta)$的估计值。通过独立正态增量模拟 W 过程，可以相对容易地实现这一点。（ii）可以从$\|R_{n,\ j}\|$ 分布中模拟，同样采用相对于 K_j、Q和 α_j^*的估计位置；具体做法是从模型中模拟完整的 N_i^*和 Y_i^*过程，其中第 i个寿命来自累积风险率为$z_{i,\ (1)}^{\mathrm{t}} A_{(1)}(t,\widehat{\theta}) + z_{i,\ (2)}^{\mathrm{t}} \widehat{A}_{(2)}(t)$的分布。这相当于在估计模型下进行半参数自助法。

请注意，上述方法也适用于同时检验统计量 $\sum_{j=1}^k \|R_{n,\ j}\|$及其相关统计量。

6 模拟与一个应用

本节将完全非参数线性风险回归模型，即 Aalen模型，与本文提出的部分参数、部分非参数线性风险回归模型进行比较。首先，在第 6.1 节中，我们基于模拟数据开展比较；第 6.2 节则分析了 $n = 312$ 名原发性胆汁性肝硬化患者的数据，这些患者于1974年1月至1984年5月期间在美国梅奥诊所参加了一项双盲随机研究。该数据集包含在 R 软件包 survival 中（Therneau & Lumley，2013）。

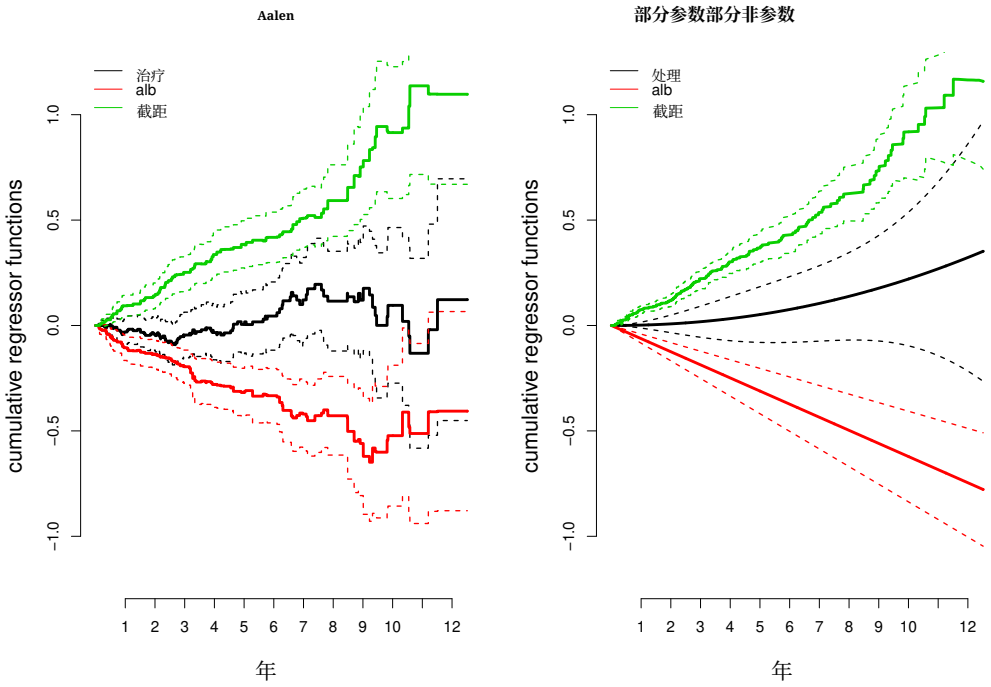


图6.2：(6.1) 式中累积回归函数的估计量，基于 PBC 数据集拟合得到。虚线表示逐点近似的 95% 置信带。

6.1 模拟

我们模拟了 200 个包含 $n = 2000$ 个可能右删失生存时间的数据集，并在这 200 次模拟中固定协变量（这反映出本文的大样本理论是在给定协变量的条件下建立的，见假设1）。第 i 个个体的真实风险率取为 $h_i(t) = \theta_1\theta_2 t^{\theta_2-1} z_{i,\ 1} + \theta_3 t z_{i,\ 2} + 0.572\, t^{2-1} + 0.123\, t z_{i,\ 4}$，其中 $\theta_1 = 0.123$、$\theta_2 = 2$ 和 $\theta_3 = 0.567$。删失时间取自 $[0, 1]$,上的均匀分布，因此约有 55% 的生存时间被观测到。对于每个数据集，我们拟合了包含四个回归变量的 Aalen 线性风险回归模型，以及一个设定正确的部分参数、部分非参数模型，即风险率为

$$h_i(t) = \theta_1\theta_2 t^{\theta_2-1} z_{i,1} + \theta_3 t z_{i,2} + \alpha_3(t) + \alpha_4(t) z_{i,4},$$

这意味着 $\alpha_3(t)$ 是“截距”函数。图 6.1 展示了 z 值（或 $\widehat{\text{Wald}}$ 统计量） $\sqrt{n}(\theta_k - \theta_k)/\text{se}(\theta_k)$（对应 $k = 1$、2、3 和 $\sqrt{n}\{A_j(t) - A(t)\}/\text{se}(A_j(t))$（对应 $j = 3$、4，其中 $\widehat{\ }$ $\widehat{\ }$后者在时间 $t = 0.5$ 处计算）的直方图。用于计算这些统计量的标准误 $\text{se}(\theta_k)$ 和 $\text{se}(A_j(t))$ 是估计量真实标准差的估计量。直方图表明，对于 $p = q = 2$ 和 $m = 3$，要使正态性真正适用于所有目标估计量，需要相当大的样本量。

6.2 实证应用

原发性胆汁性肝硬化（PBC）是一种罕见但严重的病因不明的肝病。1974年1月至1984年5月期间，美国梅奥诊所纳入了312名原发性胆汁性胆管炎患者，开展了一项比较D-青霉胺与安慰剂的双盲随机研究。在我们的分析中，我们选择将第{v1}名患者的风险率建模为

将第 i名患者的风险率建模为

$$h_i(t) = \alpha_1(t)\,\texttt{treat}_i + \alpha_2(t)\,\texttt{alb}_i + \alpha_3(t), \tag{6.1}$$

其中，treat_i 是一个指示变量，若为安慰剂则取值为零，若为D-青霉胺则取值为一；alb_i 是第 i 位患者的血清白蛋白浓度（单位：g/dl）。协变量alb_i 以其均值为中心，并按其标准差进行标准化。我们使用式（2.2）的Aalen估计量 $\widetilde{A}(t)$估计累积函数$A_j(t) = \int_0^t \alpha_j(s)\mathrm{d}s$；同时，将回归函数 $\alpha_1(t)$和 $\alpha_2(t)$参数化为 $\alpha_1(t) = \alpha_1(t,\theta) = \theta_1\theta_2 t^{\theta_2-1}$以及 $\alpha_2(t) = \alpha_2(t,\theta) = \theta_3 t$，并使用本文提出的估计方法进行估计。

图6.2绘出了累积回归函数的估计结果及其逐点近似95%置信带。对于参数累积回归量，置信带是通过应用Delta方法并利用命题 4.1 得到的。从图6.2中的两幅图很难看出，部分参数、部分非参数模型中估计量的置信带比Aalen模型的置信带更窄。因此，图 6.3绘出了累积回归函数六个估计量的逐点标准差估计，清楚地显示了效率的提升。

为了尝试评估参数函数的拟合优度，图6.5展示了在第5节中推导的式（5.2）的$R_{n,1}(t)$和$R_{n,2}(t)$函数。具体而言，蓝线表示$\sqrt{n}(\widehat{A}_1(t) - \widehat{\theta}_1 t^{\widehat{\theta}_2})$，而绿线表示$\sqrt{n}(\widetilde{A}_2(t) - \widehat{\theta}_3 t)$。我们看到，在数据的前八年中，参数回归量似乎拟合得不错；但在其余年份，Aalen 估计量与参数估计量则有些偏离。不过应当记住，对于 $n = 312$，我们在这些较晚年份拥有的数据相当~~有限，这意味着 $A_1(t)$和 $A_2(t)$的方差会增大。可以使用第5.2节的方法对参数化风险函数的适切性进行正式检验。

图 6.4分别展示了Aalen估计量和部分参数、部分非参数线性风险回归模型对应的个体生存曲线估计值，并附有逐点近似的95% 置信带（见第B.3节）。图 6.4中的两条生存曲线走势十分接近，但部分参数、部分非参数模型的置信带始终比Aalen估计量对应的置信带更窄。

7 结语

本文最后列出几点结语，其中一些指向了未来的研究方向。

A. 与广义加性模型的联系。我们研究了Aalen风险函数模型 $\sum_{j=1}^r z_j\alpha_j(s)$ 的参数-非参数模型。这与广义加性回归模型有相似之处；在这类模型中，协变量 x_1、$\ldots$、x_r的均值响应曲线建模为 $\mathrm{E}\,(Y \mid x_1,\ldots,x_r) = f_1(x_1) +$
$\cdots + f_r(x_r)$。典型的广义加性模型方法将回归函数 $f_j(x_j)$ 视为非参数函数，并有多种估计方法；参见 Hastie & Tibshirani (1990年)；Wood (2017)。本文的方法或可启发广义加性模型的参数-非参数扩展，其中部分 $f_j(x_j)$ 采用参数方式建模。

B. 拟合优度检验的局部功效。第5节中的监测函数，即 $R_n(t)$ 和 $R_{n,j}(t)$，如该节所述，可导出多类拟合优度检验，包括卡方和柯尔莫哥洛夫–斯米尔诺夫类型的检验。还可以通过将命题 5.1 扩展到真实的 $\alpha_j(s)$ 函数偏离 $O(1/\sqrt{n})$ 的情形，研究此类检验的局部功效。

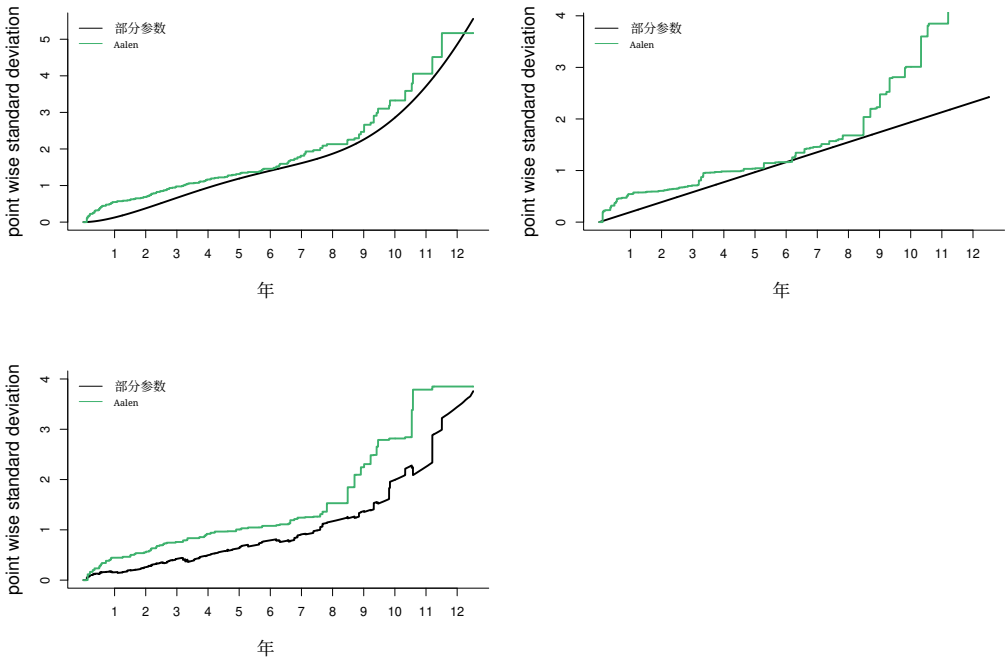


图 6.3：Aalen模型的估计量 $\widetilde{A}_j(t)$（ $j = 1$ 为 2、3）的逐点标准差估计值（绿色），以及部分参数、部分非参数模型的 $A_1(t,\theta)$、 $A_2(t,\theta)$ 和 $A_3(t)$（黑色）。

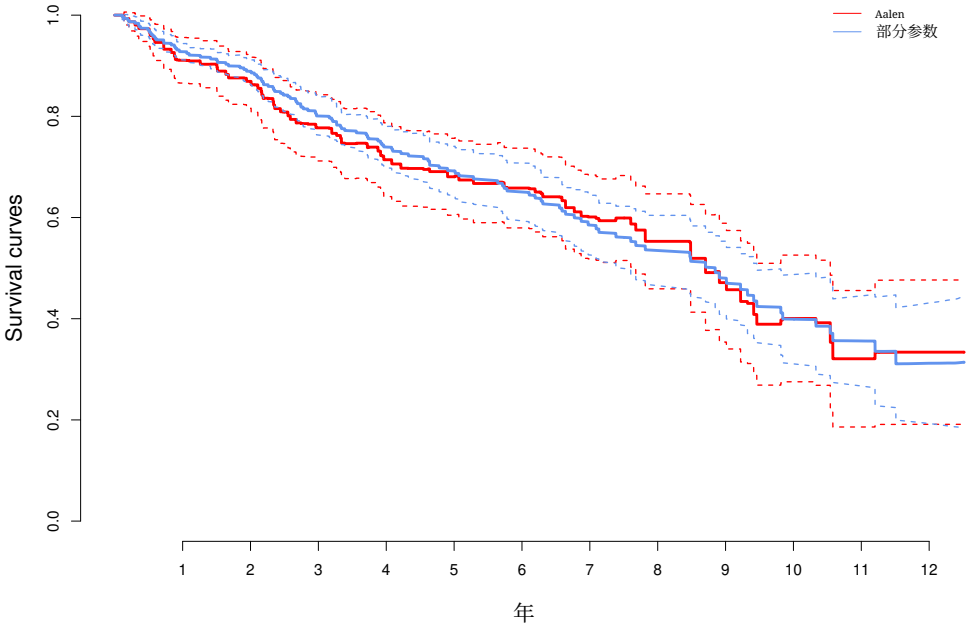


图6.4：对于 $\texttt{alb}_i$ 取其均值的未接受治疗的个体，对应于图6.2中所绘制的估计累积回归函数的估计生存曲线。虚线表示近似的95% 置信带。

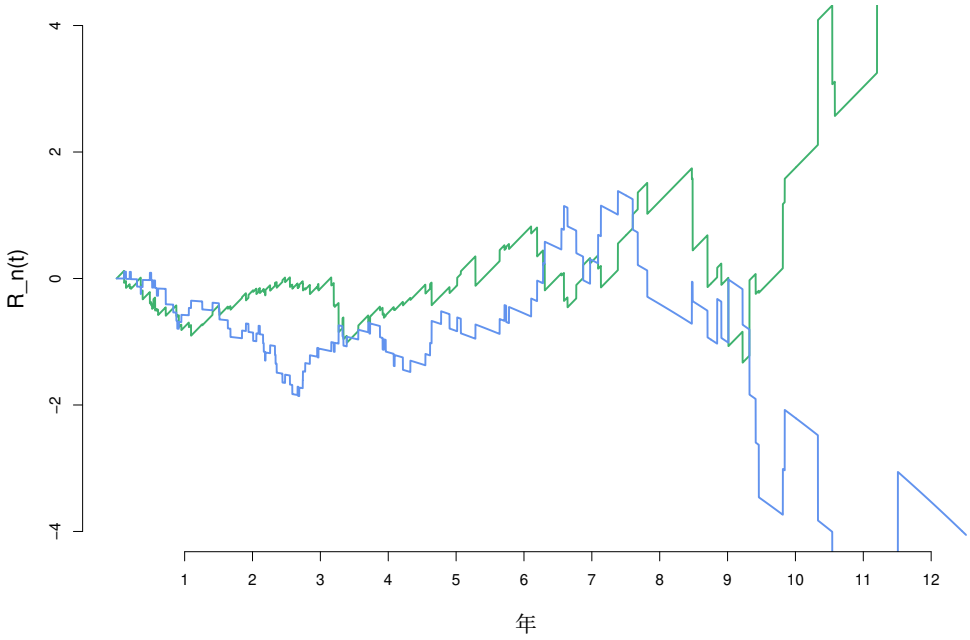


图6.5：（5.1）的 $R_{n,j}(t)$ 函数，其中权重函数为 $K_n(t) = 1$。蓝线表示 $\sqrt{n}(\widetilde{A}_1(t) - \widehat{\theta}_1 t^{\widehat{\theta}_2})$，绿线表示 $\sqrt{n}(\widetilde{A}_2(t) - \widehat{\theta}_3 t)$。

参数 $\alpha_j(s, \theta_0)$。随后可进一步利用这些结果构造权重函数$K_n(s)$，以针对某些预先设想的备择假设获得最优局部功效。

C. 模型条件之外的大样本行为。在第 4 节中，已在模型假设下推导出明确的渐近正态性结果。这些结果可扩展至真实的底层风险函数结构具有一般 Aalen 形式 $\sum_{j=1}^{r} z_j \alpha_j(s)$ 的情形，其中 $\alpha_j(s)$ 的前 p 个不一定属于参数模型，例如 $\alpha_j(s, \theta_j)$。这涉及某些最小错误参数 $\theta_{0,j}$。获得这类更一般的模型外结果，其好处部分在于可构造用于置信区间等的模型稳健方法，也可用于建立恰当的模型选择策略。

D. 用于模型选择的聚焦信息准则。我们的参数—非参数模型方法体系，是针对一组给定的参数模型分量构建的，例如，分量 $j = 1, \ldots, p$ 的 $\alpha_j(s, \theta_j)$。显然，若能进一步开发配套的模型选择方法，将会很有用；这适用于统计学家无法或不愿事先决定选取哪些分量作为参数分量，以及在这种情况下采用哪些参数结构的情形。由于不存在似然函数，因此无法使用赤池信息准则和贝叶斯信息准则这类方法。不过，可以开发聚焦信息准则（FIC）方法；关于一般性讨论，参见 Claeskens & Hjort（2008，第 6–7 章）。我们可以在当前框架下，构建沿用 Jullum & Hjort（2019）以及 Claeskens, Cunen & Hjort（2019）所提出思路的 FIC 方法。起始假设是，对于某些未知的 $\alpha_j(t)$（对应 $j = 1, \ldots, r$），非参数 Aalen 模型成立。对于某个感兴趣的量，例如 $\mu = \mu(\alpha_1(\cdot) \ldots, \alpha_r(\cdot))$，将会有一组相应的估计量，例如候选模型 M 的 $\widehat{\mu}_M$。随后，FIC 将用于估计这些 $\widehat{\mu}_M$ 的均方误差。要实现这一点，需要得到超出参数模型条件范围的大样本正态性结果，正如上文 D 点简要指出的那样。

E. 替代估计策略。我们针对参数-非参数模型 $\widehat{\ }$ 的估计量，在步骤（a）中通过最小化式（3.1）中的某个准则函数 $C_n(\theta)$；由此得到的 θ 也会在步骤（b）中用于非参数分量。步骤（a）还可以采用其他策略，包括最小化其他准则函数，使 $\alpha_{(1)}(s,\theta)$ 接近其背后的 $\alpha_{(1)}(s)$。这类思路的某些特殊版本会得到 M型估计量，其理论见 Hjort（1985，第 4 节）。然而，要将完整理论扩展到对 θ 和 $A_{(2)}(t)$ 的估计，还需要进一步研究。

F. 参数-非参数治愈模型。近年来，治愈模型受到广泛关注。综述可参见 Amico & Van Keilegom (2018)，其中也列出了相关参考文献。这类模型用于描述生存时间，其中所研究总体中有一个未知比例的个体是“治愈”的，也就是说，属于这一比例的个体永远不会经历所关注的事件。标准治愈模型的总体生存曲线形式为 $S_{\mathrm{pop.}}(t) = 1-\pi+\pi S(t)$，其中 $S(t)$ 是一个适当的生存函数（也就是说，$S(t) \to 0$ 当 $t \to \infty$），而 π 是易受所关注事件影响的概率。通常将 $S(t)$ 和 π 分别建模为协变量的函数，即 $S(t) = S(t \mid z)$ 和 $\pi = \pi(x^{\mathrm{t}}\gamma)$，其中 z 和 x 可能是不同的协变量集。Stoltenberg (2020) 中引入了一种具有 Aalen 风格线性风险回归模型 `的治愈模型；如 1.3 节所示，其（适当的）生存函数形式为 $S(t \mid z) = \exp\{-z^{\mathrm{t}}A(t)\}$，并针对 $A_j(t)$ 以及进入 $\pi(x^{\mathrm{t}}\gamma)$ 的参数提出了估计方法。受本文研究进展的启发，可以为部分参数部分非参数治愈模型开发估计方法及相应的大样本理论；该模型的总体生存函数为

$$S_{\mathrm{pop.}}(t;x,z) = 1-\pi(x^{\mathrm{t}}\gamma)+\pi(x^{\mathrm{t}}\gamma)\exp\Big[-\int_0^t \Big\{\sum_{j=1}^{p} z_{i,j}\alpha_j(s,\theta)+\sum_{j=p+1}^{p+q} z_{i,j}\alpha_j(s)\Big\}\,\mathrm{d}s\Big],$$

其中 $\pi(a)$：IR$\to[0,1]$ 某个参数函数，例如 logistic 函数。该模型中需要根据数据进行估计的未知量包括参数向量 γ 和 θ，以及非参数累积函数 $A_j(t) = \int_0^t \alpha_j(s)\,\mathrm{d}s$，其中 $j = p+1$、.....、$p+q$。这些量的估计量可以通过结合 Stoltenberg（2020）中提出的估计量与本文的两步估计程序获得。

A 渐近最优性

考虑风险率为以下形式的模型族

$$h_i(t,\theta,\eta) = z_{i,(1)}^{\mathrm{t}}\alpha_{(1)}(t,\theta)+z_{i,(2)}^{\mathrm{t}}\alpha_{(2)}(t,\eta) = \sum_{j=1}^{p} z_{i,j}\alpha_j(t,\theta)+\sum_{j=1}^{p+q} z_{i,j}\alpha_j(t,\eta), \qquad (A.1)$$

其中 $\alpha_j(t,\eta) = \sum_{l=1}^{K}\eta_{j,\,l}I_{W_l}$，对于等距划分 $W_l = [v_{l-1}, v_l)$ $0 = v_0 < v_1 < \cdots < v_{K-1} < v_K = \tau$的观察窗口 $[0,\tau]$。我们假设，对于 $j = 1,\ldots,p$，$\alpha_j(s,\theta_j)$，因此 $r \geq p$。如果 $\theta \in$IR^r，例如，那么 (A.1) 就是一个 $r+qK$ 维模型。现在这是一个完全参数模型，因此我们可以使用第 2.3 节的理论。在另行说明之前，我们假设真实模型的形式为 (A.1)，并将 K 固定。对数似然函数为 $\ell_n(\theta,\eta) = \sum_{i=1}^{n}\int_0^\tau\{\log h_i(s,\theta,\eta)\,\mathrm{d}N_i(s) - Y_i(s)h_i(s,\theta,\eta)\,\mathrm{d}s\}$。我们将得分函数分为 a$\theta$ 部分和 η 部分。我们有

$$U_n = n^{-1/2}\sum_{i=1}^{n}\int_0^\tau \frac{\alpha_{(1)}^*(s,\theta)^{\mathrm{t}}z_{i,(1)}}{h_i(s,\theta,\eta)}\,\mathrm{d}M_i(s)$$

这是一个 $r\times 1$ 列向量（其中 r 是 θ 的维数），而 $\alpha^*_{(1)}$ 是一个 $p\times r$ 矩阵，其中包含偏导数 $\partial\alpha_j(s,\theta_j)/\partial\theta_j$，适用于 $j=1,\ldots,p$。我们还有得分函数$V_n=(V_{tn,1},\ldots,V_{tn,K})^{\mathrm{t}}$，它是一个 $qK\times 1$ 列向量，其中

$$V_{n,l}=n^{-1/2}\sum_{i=1}^{n}\int_{W_l}\frac{z_{i,(2)}}{h_i(s,\theta,\eta)}\,\mathrm{d}M_i(s)\qquad\text{for } l=1,\ldots,K,$$

$\widehat{\ }$是 $q\times 1$列向量。令 J_n为（U_n， V_n）的方差过程，并令（θ， $\widehat{\eta}$）为最大$\widehat{\ }$似然估计量。在命题 2.1的条件下，我们知道，当 $n\to\infty$时，$\sqrt{n}$（$\theta-\theta_0$，$\widehat{\eta}-\eta_0$，$_K$）依分布收敛于 N_{r+qK}（0，Ω_K^{-1}），其中 Ω_K 是J_n的依概率极限，而 θ_0、η_0、$_K$ 表示参数的真实值（在第 K个模型下）。我们需要求 $\widehat{\theta}$以及累积 A_2（t， η）$=\int_0^t\alpha_2$（s， η）$\mathrm{d}s$的估计量的极限分布。引入（$p+q$）$\times$（$p+qK$）矩阵函数 H_t： $\mathrm{IR}^{p+qK}\to\mathrm{IR}^{p+q}$，使得对于每个 t， H_t（θ^{t}， η^{t}）$^{\mathrm{t}}$=（θ_1，...，θ_r， A_{p+1}（t， η）...，A_{p+q}（t， η））$^{\mathrm{t}}$。（函数 H_t也依赖于 K，但由于这对下文无关紧要，我们省略了记号中的这一点。）应用德尔塔方法可得

$$\sqrt{n}\{\widehat{\theta}-\theta,A_2(\tau,\widehat{\eta})-A_2(\tau,\eta_{0,K})\}\xrightarrow{d}\mathrm{N}_{p+q}\{0,H_\tau\Omega_K^{-1}H_\tau^{\mathrm{t}}\}.$$

（我们接下来要证明的结论并不需要完整的过程收敛版本。） $H_\tau\Omega_K^{-1}H_\tau^{\mathrm{t}}$的左上块是$\sqrt{n}(\widehat{\theta}-\theta_0)$的 $r\times r$（渐近）方差矩阵。采用以下记号

$$\Omega_K=\begin{pmatrix}\Omega_{K,11}&\Omega_{K,12}\\ \Omega_{K,21}&\Omega_{K,22}\end{pmatrix}\quad\text{and}\quad\Omega_K^{-1}=\begin{pmatrix}\Omega_K^{11}&\Omega_K^{12}\\ \Omega_K^{21}&\Omega_K^{22}\end{pmatrix},$$

$H_\tau\Omega_K^{-1}H_\tau^{\mathrm{t}}$ 的左上块为

$$\Omega_K^{11}=(\Omega_{K,11}-\Omega_{K,12}\Omega_{K,22}^{-1}\Omega_{K,21})^{-1}.$$

By d进行矩阵代数运算后，我们看到括号内的矩阵是 $r\times r$ 矩阵 ix

$$\Omega_{K,11}=\int_0^\tau(\alpha^*_{(1)})^{\mathrm{t}}F_{11}\alpha^*_{(1)}\,\mathrm{d}s,$$

$r\times qK$ 矩阵

$$\Omega_{K,12}=\int_0^\tau[(\alpha^*_{(1)})^{\mathrm{t}}F_{12}I_{W_1}\cdots(\alpha^*_{(1)})^{\mathrm{t}}F_{12}I_{W_K}]\,\mathrm{d}s,$$

以及 $\Omega_{K,21}=\Omega_{K,12}^{\mathrm{t}}$，而 $\Omega_{K,22}$ 是 $qK\times qK$ 分块对角矩阵，其各块为 $q\times q$矩阵

$$(\Omega_{K,22})_l=F_{22}I_{W_l},\quad\text{for } l=1,\ldots,K.$$

这里， F_{11}、$F_{12}=F_{21}$和 F_{22}分别是$F_{n,11}$、$F_{n,12}=F_{n,21}$和$F_{n,22}$的依概率极限，其中后者是式（2.6）中定义的矩阵 F_n 的各块。现在，我们考虑 Ω_K趋于 $K\to\infty$的情形，也就是区间长度趋于零时。在对协变量、 $n^{-1}\sum_{i=1}^n Y_i(s)$的依概率极限以及函数 $s\mapsto\alpha^*_{(1)}$（例如具有有界导数）作出适当假设后，我们有 $\Omega_{K,22}$ 的第 l个对角块为

$$(\Omega_{K,22})_l=F_{22}(v_{l-1})K^{-1}+O(K^{-2}),\quad\text{for } l=1,\ldots,K,$$

并且，类似地，

$$\Omega_{K,12}=[(\alpha^*_{(1)})^{\mathrm{t}}F_{12}(v_0)\cdots(\alpha^*_{(1)})^{\mathrm{t}}F_{12}(v_{K-1})]K^{-1}+O(K^{-2}).$$

于是我们得到

$$\Omega_{K,12}\Omega_{K,22}^{-1}\Omega_{K,21} = K^{-1}\sum_{l=1}^{K}(\alpha_{(1)}^{*}(v_{l-1}))^{\mathrm{t}}F_{12}(v_{l-1})F_{22}(v_{l-1})^{-1}F_{21}(v_{l-1})\alpha_{(1)}^{*}(v_{l-1}) + O(K^{-2}),$$

这是一个黎曼和，当 $K \to \infty$ 时收敛到 $\int_0^\tau (\alpha_{(1)}^{*})^{\mathrm{t}}F_{12}F_{22}^{-1}F_{21}\alpha_{(1)}^{*}\,\mathrm{d}s$。总之，在 K 固定时，$J_n \to_p \Omega_K$ 随 $n \to \infty$，并且

$$\Omega_{K,11}^{-1} \to \Big\{\int_0^\tau (\alpha_{(1)}^{*})^{\mathrm{t}}(F_{11} - F_{12}F_{22}^{-1}F_{21})\alpha_{(1)}^{*}\,\mathrm{d}s\Big\}^{-1},$$

当 $K \to \infty$ 时。右边的极限，记为 $\Omega_0^{-1},11$，即为 (4.3) 的表达式。

假设在 (1.4) 的部分参数、部分非参数模型下，存在 θ_0 的一个相合估计量，其方差小于 Ω_0^{-1},11。将其方差矩阵记为 V，从而 $V < \Omega_0^{-1},11$（即 $V - \Omega_0^{-1},11$ 是负定矩阵）。由于对所有 K 有 $\Omega_K^{-1},11 \le \Omega_{K+1}^{-1},11$，且当 $K \to \infty$ 时 $\Omega_K^{-1},11 \to \Omega_0^{-1},11$，这意味着存在某个 K_0，使得对所有 $K \ge K_0$ 有 $V < \Omega_K^{-1},11$。但 $\Omega_K^{-1},11$ 是在形如 (A.1) 的第 K 个参数模型下估计 θ_0 的 Cra m´er–Rao 下界，因此 $V < \Omega_K^{-1},11$ 不可能成立，从而在 (1.4) 的模型下不可能存在方差小于 $\Omega_0^{-1},11$ 的 θ_0 的相合估计量。

B 效率与相对改进计算

与非参数模型相比，建立并使用参数成分模型通常具有普遍益处，前提是能够对模型进行评估以检验其充分性；这一主题将在下一节讨论。本节我们考虑与效率有关的问题：与非参数Aalen方法相比，使用参数-非参数模型（1.4）在精度上能获得多少增益?

B.1 渐近相对效率

我们在前面各节已经看到，各种极限分布关键依赖于第2节中定义的 F_n、G_n、H_n的极限矩阵函数 F、G、H，以及某些相关量。现在将针对某一设定对这些函数进行研究与比较，以同时阐明相对效率的若干方面。

假设删失机制独立于生存时间，其生存函数为 $\rho(s) = P\{C_i \ge s\}$。则$\mathrm{E}\{Y_i(s) \mid z_i\} = \exp\{-z_i^{\mathrm{t}}A(s)\}\rho(s)$。以 $Y(s)$与 Z表示一般的风险指标和协变量向量，并依照所考虑的协变量分布取值，我们推得

$$\begin{aligned}
F(s) &= \mathrm{E}\,Y(s)\frac{ZZ^{\mathrm{t}}}{Z^{\mathrm{t}}\alpha(s)} = \mathrm{E}\exp\{-Z^{\mathrm{t}}A(s)\}\frac{ZZ^{\mathrm{t}}}{Z^{\mathrm{t}}\alpha(s)}\rho(s) = F_0(s)\rho(s),\\
G(s) &= \mathrm{E}\,Y(s)ZZ^{\mathrm{t}} = \mathrm{E}\exp\{-Z^{\mathrm{t}}A(s)\}ZZ^{\mathrm{t}}\rho(s) = G_0(s)\rho(s),\\
\mathrm{d}H(s) &= \mathrm{E}\,\exp\{-Z^{\mathrm{t}}A(s)\}ZZ^{\mathrm{t}}Z^{\mathrm{t}}\alpha(s)\rho(s)\,\mathrm{d}s = \mathrm{d}H_0(s)\rho(s),
\end{aligned}$$

参见假设 1。现在假设速率 α_j 为常数。于是

$$F_0(s) = \mathrm{E}\,\exp(-sZ^{\mathrm{t}}\alpha)ZZ^{\mathrm{t}}/(Z^{\mathrm{t}}\alpha),$$

其导数为 $F'_{0,j,k}(s) = -G_{0,j,k}(s)$，下一个导数为 $\mathrm{d}H_{0,j,k}(s)$。进一步假设协变量 $Z_1, \ldots, Z_r$ 相互独立，其拉普拉斯变换为 $L_j(u) = \mathrm{E}\exp(-uZ_j) = \exp\{-u\psi_j(u)\}$。于是

$$\begin{aligned} L'_j(u) &= -\mathrm{E}\, Z_j \exp(-uZ_j) = -L_j(u)\psi'_j(u), \\ L''_j(u) &= \mathrm{E}\, Z_j^2 \exp(-uZ_j) = L_j(u)\{\psi'_j(u)^2 - \psi''_j(u)\}, \end{aligned}$$

由此得到

$$G_{0,j,k}(s) = \mathrm{E}\, \exp(-sZ^{\mathrm{t}}\alpha) Z_j Z_k = \Big[\prod_{i=1}^{r} \exp\{-\psi_i(\alpha_i s)\}\Big]\{\psi'_j(\alpha_j s)\psi'_k(\alpha_k s) - \delta_{j,k}\psi''_j(\alpha_j s)\},$$

其中 $\delta_{j,\,k}$ 在 $j = k$时等于 1，否则等于 0。现在可以对这些函数进行研究和数值积分，从而得到不同情形下的 F 函数。下面仅举例说明 Z_j服从伽马分布的情形。令 Z_j 服从伽马分布（c_j，γ_j），其拉普拉斯变换为 $\gamma_j^{c_j}/(\gamma_j + u)^{c_j}$，则可得 $\psi_j(u) = c_j(\gamma_j + u)^{-1}$ 以及 $\psi''_j(u) = -c_j(\gamma_j + u)^{-2}$，因此

$$G_{0,j,k}(s) = \Big(\prod_{i=1}^{r} \frac{\gamma_i}{\gamma_i + \alpha_i s}\Big)\Big\{\frac{c_j}{\gamma_j + \alpha_j s}\frac{c_k}{\gamma_k + \alpha_k s} + \delta_{j,k}\frac{c_j}{(\gamma_j + \alpha_j s)^2}\Big\}.$$

进一步假设 α_js 均等于同一个 α，并且 Z_js 均来自同一伽马（c，γ）分布，经过一些推导可得

$$\begin{aligned} G_0(s) &= g(s)(c^{-1}I_r + e_r e_r^{\mathrm{t}}), \quad \text{where } g(s) = \frac{c^2\gamma^{cr}}{(\gamma + \alpha s)^{cr+2}}, \\ F_0(s) &= f(s)(c^{-1}I_r + e_r e_r^{\mathrm{t}}), \quad \text{where } f(s) = \frac{c^2\gamma^{cr}}{(cr+1)\alpha(\gamma + \alpha s)^{cr+1}}, \\ \mathrm{d}H_0(s) &= h(s)\,\mathrm{d}s\,(c^{-1}I_r + e_r e_r^{\mathrm{t}}), \quad \text{where } h(s) = \frac{c^2\gamma^{cr}(cr+2)\alpha}{(\gamma + \alpha s)^{cr+3}}, \end{aligned}$$

其中 $e_r = (1, \ldots, 1)^{\mathrm{t}}$ 的长度为 r，且 I_r 是大小为 $r \times r$的单位矩阵。

在这一特定设定下，假设风险率为常数且协变量相互独立，我们现在可以回答与相对效率相关的各种问题。

(i) 当 n 较大时，使用 $\breve{A}$ 估计量及估计的最$\tilde{}$优权重 $\tilde{w}_i(s)$，而非较简单的 A 估计量及普通权重 $w_i(s) = 1$，精度提高了多少（见第 2 节）？我们发现

$$F^{-1} = \frac{1}{f\rho}\Big(cI_r - \frac{c^2}{1+cr}e_r e_r^{\mathrm{t}}\Big) \quad \text{and} \quad G^{-1}\,\mathrm{d}H\,G^{-1} = \frac{h}{g^2\rho}\Big(cI_r - \frac{c^2}{1+cr}e_r e_r^{\mathrm{t}}\Big),$$

经观察，后一函数为

$$\text{a.r.e.} = \frac{cr+2}{cr+1} = \frac{\xi r + 2/\gamma}{\xi r + 1/\gamma}$$

倍于前者，记 $\xi = c/\gamma$ 为诸 Z_j 的均值。$\tilde{}$极限分布（对应于 A^* 与 A）的方差矩阵即这些函数的积分，因此渐近相对效率比等于同一常数。当 c 或 γ（$\xi = c/\gamma$ 固定）较大时，方差降低可能很小，但最大可接近 2，这发生在 c 较小或 γ 较小时。

$\hat{}$(ii) 在速率为常数 $\alpha_j(s) = \theta_j$（$j = 1, \ldots\ldots, p$）的模型条件下，参数估计量 $\theta_j t$ 对 $A_j(t)$ 的估计效果比其最佳非参数对应估计量好多少？最佳非参数估计量 $A^*_{(1)}(t)$ 的方差为

$$\mathrm{Var}_{\mathrm{nonpm}} = \int_0^t (F^{-1})_{11}\,\mathrm{d}s = \int_0^t \frac{1}{f\rho}\,\mathrm{d}s\,\Big(cI_p - \frac{c^2}{1+cr}e_p e_p^{\mathrm{t}}\Big).$$

根据第4.1节， $\sqrt{n}(\widehat{\theta}-\theta)$ 的极限方差矩阵是 $\Omega_0 = \int_0^\tau Q^{-1}\,\mathrm{d}s$ 的逆矩阵。选择最佳权重函数时，$Q=(F^{-1})_{11}$，相应地有 $Q^{-1}=(F^{11})^{-1}$，从而 Ω_0 为 $\int_0^\tau f\rho\,\mathrm{d}s$ 的 $\{cI_p - c^2(1+cr)^{-1}e_pe_p\}^{-1}$ 倍。因此，$\widehat{\theta}_j t$ 估计量的极限方差矩阵变为

$$\mathrm{Var}_{\mathrm{pm}} = t^2\Omega_0^{-1} = \frac{t^2}{\int_0^\tau f\rho\,\mathrm{d}s}\Big(cI_p - \frac{c^2}{1+cr}e_pe_p^{\mathrm{t}}\Big).$$

为了得到更具体的比较，我们令删失分布为平移帕累托型 $\rho(s)=(1+\alpha s/\gamma)^{-k}$，其密度为$(k\alpha/\gamma)(1+\alpha s/\gamma)^{-(k+1)}$。我们还令 $\tau=\infty$。该分布随 k减小而随机递增，中位数等于$(\gamma/\alpha)(2^{1/k}-1)$。无删失情形对应于 $k=0$，而较大的 k 对应于更重的删失。可得

$$\int_0^\infty f\rho\,\mathrm{d}s = \frac{c^2}{\alpha^2}\frac{1}{(cr+1)(cr+k)} \text{ and } \int_0^t \frac{1}{f\rho}\,\mathrm{d}s = \frac{cr+1}{cr+k+2}\frac{\gamma^2}{c^2}\Big\{\Big(1+\frac{\alpha}{\gamma}t\Big)^{cr+k+2}-1\Big\}.$$

渐近无效率比变为

$$\frac{\mathrm{Var}_{\mathrm{nonpm}}}{\mathrm{Var}_{\mathrm{pm}}} = \frac{1}{(cr+k)(cr+k+2)}\frac{(1+u)^{cr+k+2}-1}{u^2}, \quad \text{where } u=(\alpha/\gamma)t.$$

注意，这两个矩阵彼此仅成比例，且该比值与p无关。

(iii) 与非参数 $A^*_{(2)}$相比，第4.2节中构造的 A_{p+1}，……， A_{p+q}的半参数估计量 $A_{(2)}$ 有多大改进？后者的极限分布方差矩阵为

$$\int_0^t (F^{-1})_{22}\,\mathrm{d}s = \int_0^t \frac{1}{f\rho}\,\mathrm{d}s\,\Big(cI_q - \frac{c^2}{1+cr}e_qe_q^{\mathrm{t}}\Big).$$

这需要与公式 (4.5) 进行比较。它涉及 $\phi(s)$，在此情形下可见其即为 $F_{21}(s)$，因此

$$J(t) = \int_0^t F_{22}^{-1}F_{21}\,\mathrm{d}s = \int_0^t \frac{1}{f\rho}f\rho\,\mathrm{d}s\Big(cI_q - \frac{c^2}{1+cq}e_qe_q^{\mathrm{t}}\Big)e_qe_p^{\mathrm{t}} = t\frac{c}{1+cq}e_qe_p^{\mathrm{t}}.$$

发现方差矩阵公式 (4.5) 等于

$$\int_0^t \frac{1}{f\rho}\,\mathrm{d}s\Big(cI_q - \frac{c^2}{1+cq}e_qe_q^{\mathrm{t}}\Big) + \frac{t^2}{\int_0^\tau f\rho\,\mathrm{d}s}\frac{c^3p}{(1+cq)(1+cr)}e_qe_q^{\mathrm{t}}.$$

在上述删失机制下，可得渐近无效率比（对应于 A_j 估计量的非参数极限方差除以参数极限方差）为

$$\frac{1+c(r-1)}{1-cr}v(u)\Big/\Big\{\frac{1+c(q-1)}{1-cq}v(u)+c^2\kappa(c,k)u^2\Big\},$$

其中同样有 $u=(\alpha/\gamma)t$,

$$v(u)=(1+u)^{cr+k+2}-1, \quad \text{and} \quad \kappa(c,k)=\frac{(cr+k+2)p(cr+k)}{(cq+1)(cr+1)}.$$

现在可以针对不同的参数组，将这个比率作为 t 中的一条曲线来研究。在某些设定下，参数估计量的精度可能显著优于非参数估计量。

B.2 朴素权重估计量的效率改进

在上一小节中，我们关注了理论上最优估计量这一特殊情形，$\widehat{\quad}$ $\widehat{\quad}$ $\widehat{\quad}$对于 θ和 A_{j+1}，.. ., A_{p+q}。这需要使用一些繁复的最优权重 $w_i(s) = 1/z_i^{\mathrm{t}}\widetilde{\alpha}(s)$，在最小化（ 3.1 ）中的准则函数 $C_n(\theta)$ 时，还需要配套使用最优 $V_n(s)$ 矩阵，等等。这使得相关相对效率比的公式相对清晰、易于理解。

不过，在实践中，我们更关心与我们偏好的默认选择 $V_n(s) = n^{-1}\sum_{i=1}^n Y_i(s)z_i z_i^{\mathrm{t}}$ 相似的效率比。命题 4.1 和 4.2 也可用于求出这种情况下的极限方差表达式，其中涉及

$$n^{-1}\sum_{i=1}^n Y_i(s)z_i z_i^{\mathrm{t}} \to_p \mathrm{E}\,\exp\{-Z^{\mathrm{t}}A(s)\}ZZ^{\mathrm{t}}\,\rho(s),$$

$$n^{-1}\sum_{i=1}^n Y_i(s)z_i z_i^{\mathrm{t}} z_i^{\mathrm{t}}\alpha(s) \to_p \mathrm{E}\,\exp\{-Z^{\mathrm{t}}A(s)\}ZZ^{\mathrm{t}}\,Z^{\mathrm{t}}\alpha(s)\,\rho(s),$$

以及其他一些量；同样，期望是相对于协变量的遍历分布而言，见假设1。对于给定的协方差分布设置和 $A_j(t)$ 函数，可以数值计算这些表达式以及其他所需量。不过，效率比没有明确的公式，因此与上一小节中的比较相比，比较更加困难，也不够直观。我们在一些简单情形下进行了数值计算，发现所得比值大体上与前文相近。

B.3 给定问题的改进潜力

对于给定的数据集和给定的关注参数 $\mu = \mu(A_1, \ldots, A_r)$，统计学家可以计算两个$\widetilde{\quad}$ $\widetilde{\quad}$估计量：使用（2.2）估计累积回归函数的 $\widehat{\mu}_{\mathrm{Aalen}} = \mu(A_1, \ldots, A_r)$，以及

$$\widehat{\mu}_{\mathrm{Partly}} = \mu(A_1(\cdot,\widehat{\theta}), \ldots, A_p(\cdot,\widehat{\theta}), \widehat{A}_{p+1}, \ldots, \widehat{A}_{p+q}).$$

至关重要的是，也可以使用第 4 节中推导的方差公式来计算它们的标准误，即估计标准差。例如，在生存分析中，一个特别受关注的关注参数是在某个固定时间点 t 处取值的生存函数。对于协变量为 $z_0 = (z_{0,\,(1)}^{\mathrm{t}}, z_{0,\,(2)}^{\mathrm{t}})^{\mathrm{t}}$ 的个体，关注参数为 $\mu = \exp\{-z_0^{\mathrm{t}}A(t)\}$，我们有$\sqrt{n}(\widehat{\mu}_{\mathrm{Partly}} - \mu) \to_d \mathrm{N}(0, \mu^2 z_0^{\mathrm{t}}\Xi(t)z_0)$，其中 $\Xi(t)$ 是式 (4.5) 中出现的矩阵。

针对给定关注参数直接比较 $\widehat{\mu}_{\mathrm{Aalen}}$ 和 $\widehat{\mu}_{\mathrm{Partly}}$ 的标准误，可以看出，对于所讨论的 $\alpha_j(s)$ 分量，从非参数方法转向参数方法是否能带来明显收益。图 6.3 展示了这一点，其中给出了两个累积函数估计量 $A_1(t)$ $A_2(t)$ $A_3(t)$ 的标准误对比图，直观呈现了实际效果；图 6.4 的生存曲线图也提供了一个例子：部分参数生存曲线的置信带明显比 Aalen估计量对应的置信带更窄。

致谢

本文的研究得益于与 Ian McKeague 和 Ingrid Van Keilegom 的讨论；我们也感谢两位审稿人提出的建设性意见。

参考文献

Aalen, O. O. (1980). 计数过程非参数回归分析模型。《统计学讲义》2, 1–25。Aalen, O. O. (1989). 寿命分析的线性回归模型。《医学统计学》8, 907–925。Aalen, O. O. (1993). 生存分析中非参数线性回归模型的进一步结果。《医学统计学》12, 1569–1588。Aalen, O. O., Borgan, Ø.& Gjessing, H. (2008). 《计数过程统计模型》。柏林：施普林格出版社。Amico, M. & Van Keilegom, I. (2018). 生存分析中的治愈模型。《统计学及其应用年度评论》5, 311–342。Andersen, P. K., Borgan, Ø., Gill, R. D. & Keiding, N.(1993). 《生存与事件史分析：过程视角》。柏林：施普林格出版社。Borgan, Ø. (1984). 参数计数过程模型中的最大似然估计及其在删失失效时间数据中的应用。《斯堪的纳维亚统计学杂志》11, 1–16。Borgan, Ø., Fiaccone, R. L., Henderson, R. & Barreto, M. L.(2007). 利用缺失观测数据动态分析复发事件数据，并应用于巴西婴儿腹泻。《斯堪的纳维亚统计学杂志》34, 53–69。Claeskens, G., Cunen, C. & Hjort, N. L. (2019). 利用复杂生态学与进化数据的聚焦信息准则进行模型选择。《生态学与进化前沿》7, 415–428。Claeskens, G. & Hjort, N. L. (2008). 《模型选择与模型平均》。剑桥：剑桥大学出版社。Hastie, T. & Tibshirani, R. J. (1990). 《广义加性模型》。伦敦：Chapman and Hall/CRC。Hjort, N. L. (1985). 对 Andersen 和 Borgan《生命史数据的计数过程模型：综述》的讨论。《斯堪的纳维亚统计学杂志》12, 141–150。Hjort, N. L. (1986). 参数计数过程模型中的贝叶斯估计量与渐近效率。《斯堪的纳维亚统计学杂志》13, 63–85。Hjort, N. L. (1990). 基于累积风险率的生命史数据模型拟合优度检验。《统计年刊》18, 1221–1258。Hjort, N. L. (1992). 参数生存数据模型中的推断。《国际统计评论》60, 355–387。Huffer, F. W. & McKeague, I. (1991). Aalen加性风险模型的加权最小二乘估计。《美国统计学会杂志》86, 114–129。

Jullum, M. & Hjort, N. L. (2017). 参数模型或非参数模型：聚焦信息准则方法。《统计学报》27, 951–981.Jullum, M. & Hjort, N. L. (2019). 半参数 Cox 回归的代价是什么？《寿命数据分析》25, 406–438.Martinussen, T. & Scheike, T. H. (2002a). 基于当前状态数据的加性风险回归高效估计。《生物计量学》89, 649–658.Martinussen, T. & Scheike, T. H. (2002b). 灵活的加性乘法风险模型。《生物计量学》89, 283–298.Martinussen, T. & Scheike, T. H. (2007). 生存数据的动态回归模型。柏林：施普林格。Martinussen, T. & Scheike, T. H. (2009a). 高维回归变量下的加性风险模型。《寿命数据分析》15, 330–342.Martinussen, T. & Scheike, T. H. (2009b). 半参数加性风险模型的协变量选择。《斯堪的纳维亚统计学杂志》36, 602–619.McKeague, I. W. & Sasieni, P. D. (1994). 部分参数加性风险模型。《生物计量学》81, 501–514.Stoltenberg, E. A. (2020). 具有线性风险的标准治愈模型。arXiv 预印本 arXiv:2011.12858 .Therneau, T. & Lumley, T. (2013). R 生存分析软件包。Wood, S. W. (2017).《广义加性模型：R 语言导论》 *[2版]*。伦敦：Chapman and Hall/CRC。