arXiv:2611.32026v1 [stat.ME] 24 Nov 2026

作为增广逆删失概率加权估计量的 Aalen–Johansen

Lu Mao

生物统计学与医学信息学系，威斯康星大学麦迪逊分校，610 Walnut Street，威斯康星州麦迪逊市，邮编 53726，美国。lmao@biostat.wisc.edu

摘要

我们证明，对于每个样本，广泛使用的状态占用概率 Aalen–Johansen 估计量，都是一种逆删失概率加权估计量，并以基于每位受试者当前状态的工作马尔可夫预测作为增广。这一恒等关系解释了为何 Aalen–Johansen 估计量在马尔可夫假设不成立时仍具有稳健性。它还表明，该估计量的影响函数与半参数有效影响函数之间相差两个正交分量：当前状态之外的先前历史信息，以及马尔可夫预测的误差。基于这一分解，我们提出一种实用且无模型的 Aalen–Johansen 最小二乘增广方法，即用历史评分对结果进行回归；该方法可缩小非马尔可夫过程中两者的差距，同时不会增加渐近方差。理论分析和模拟结果均验证了该方法的优势。

关键词。 影响函数；单调粗化；多状态过程；正交投影；半参数效率；状态占用概率。

1 背景

对于具有右连续样本路径的多状态过程 Y，令 $Y(t) \in \{0, \ldots, K\}$ 表示受试者在时间 t 的状态。图 1 展示了常见示例。令 $p(t) = \{p_0(t) \ldots, p_K(t)\}^\mathsf{T}$，其中$p_k(t) = \mathrm{pr}\{Y(t) = k\}$，表示状态占用概率的列向量。

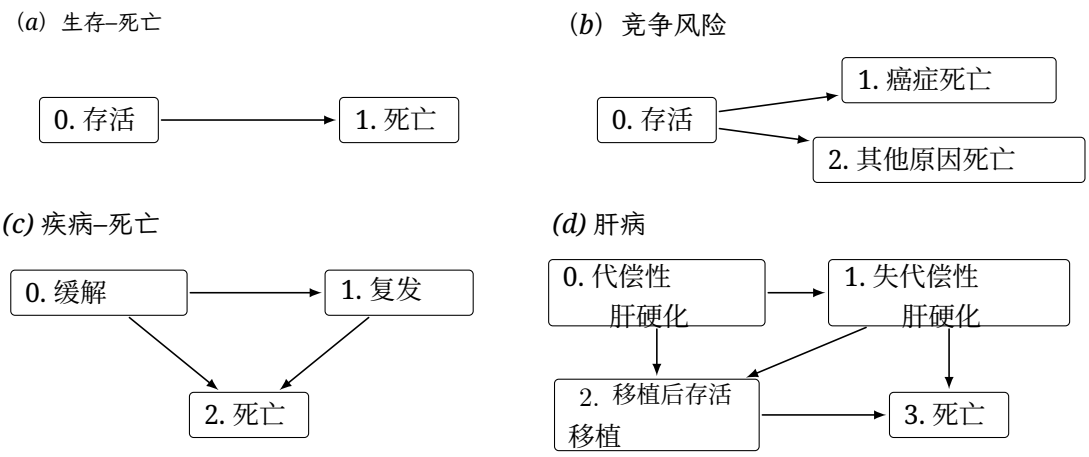

图 1：多状态模型示例。

在右删失的情况下，Aalen–Johansen（1978）在马尔可夫模型下以非参数方式估计了 $p(t)$；在该模型中，瞬时转移概率仅通过当前状态依赖于历史信息：

$$\begin{aligned}\text{Markov assumption:}\quad \operatorname{pr}\{Y(t+\mathrm{d}t)=k \mid Y(t-)=j, \mathcal{H}(t-)\} = \\ \operatorname{pr}\{Y(t+\mathrm{d}t)=k \mid Y(t-)=j\} \equiv \mathrm{d}\Lambda_{jk}(t),\end{aligned} \tag{1}$$

其中，$\mathcal{H}(t) = \{Y(s):\ 0 \le s \le t\}$，而 $\mathrm{d}\Lambda_{jk}(t)$ 是 $j \to k$ 转移强度，汇总于矩阵 $\mathrm{d}\Lambda(t)$ 中，其中 $\mathrm{d}\Lambda_{jj}(t) = -\sum_{k\neq j}\mathrm{d}\Lambda_{jk}(t)$。在（1）下，转移概率矩阵 $P(s,\ t)$ 的（$j,\ k$）元素 $P_{jk}(s,\ t) = \operatorname{pr}\{Y(t) = k \mid Y(s) = j\}$ 可表示为乘积积分

$$P(s,t) = \prod_{s<u\le t}\{I + \mathrm{d}\Lambda(u)\} \tag{2}$$

（Cook 和 Lawless，2018 年，第 2.2.1 节）。经过简单的代数运算可得 $p(t) = P(0,\ t)^{\mathsf{T}}p(0)$。

对于删失数据，我们可以使用 Nelson–Aalen 型增量 d 来估计 $\mathrm{d}\Lambda_{jk}(t)$ $\Lambda_{jk}(t)$，其定义为在 t 时观测到的 $j \to k$ 转移次数与紧接在 t 之前处于观察中的数量之比；并通过经验比例 $\widehat{p}(0)$ 估计 $p(0)$。由此得到作为代入估计量的 Aalen–Johansen 估计量：

$$\text{Aalen–Johansen:}\quad \widehat{p}(t) = \widehat{P}(0,t)^{\mathsf{T}}\widehat{p}(0), \quad \text{where } \widehat{P}(s,t) = \prod_{s<u\le t}\{I + \mathrm{d}\widehat{\Lambda}(u)\}. \tag{3}$$

尽管公式 (3) 源自马尔可夫模型，但即使马尔可夫假设 (1) 不成立，$\widehat{p}(t)$ 在独立删失下仍保持一致性（Datta 和 Satten，2001；Glidden，2002）。这种稳健性及其简便性使其成为状态占用概率的标准非参数估计量（Putter 等，2007），也成为估计占用时间等派生量的标准方法（B¨uhler 等，2023）。

事实上，在简单情形下，Aalen–Johansen 可化为人们熟悉的估计量。在图 1(a) 的生存–死亡模型中，$\widehat{p}_0(t)$ 等价于 Kaplan–Meier 生存估计量。在图 1(b) 的竞争风险模型中，$\widehat{p}_k(t)$ 是原因 k 的累积发生率的常用非参数估计量（ Andersen 等，1993 ）。二者都可以重新表述为逆概率删失加权（IPCW）估计量：通过对 $= k\}$ $I\{Y(t)$ 进行平均，横断面地估计 $p_k(t)$；平均仅针对在 t 时状态已知的受试者进行，并为每位受试者赋予其保持未删失概率的倒数作为权重，该概率通过删失时间的 Kaplan–Meier 估计量估计（Satten 和 Datta，2001；Zhang 等，2009）。

这种等价关系无法推广到更复杂的情形。毕竟，逆删失概率加权估计量只使用受试者在 t时是否处于 k 的信息，而公式 (3) 是累积的，它从 0 开始传播转移率直到 t，因此会使用沿途访问过的状态。例如，在图 1(c) 中，关于 t 时处于死亡状态的概率的逆删失概率加权估计量只考虑当前状态，而忽略受试者是在缓解状态下死亡，还是在复发后死亡；相比之下，Aalen–Johansen 通过公式 (3) 使用了先前状态的信息。

不过，Aalen–Johansen 仍然存在一种精确的加权表示：Datta 和 Satten（2001年）以及 Yan 等人（2026年）将其写成逆概率加权平均，其权重取决于每位受试者所处的状态，而不是使用单一的汇总删失权重。由于这只是一个精确的重新表达，它继承了 Aalen–Johansen 对非马尔可夫偏离的稳健性，但并未对此作出解释。

另一个独立且更具启发性的问题是：Aalen–Johansen 估计量是否属于增广逆概率删失加权估计量这一类（van der Laan 和 Robins，2003年）。这类估计量会在逆删失概率加权估计量的基础上，添加一个

均值为零的增广项；该增广项基于工作模型下对 $I\{Y(t)=k\}$ 的预测，并利用先前历史构建。无论工作模型是否正确，这一增广项都能提高效率而不会引入偏倚。最近，Maltzahn 等（2026年）表明，达到完全效率的增广项会利用 t 之前的全部历史信息。由于 Aalen–Johansen 估计量通过马尔可夫模型将当前状态向前传递，同时又能抵御非马尔可夫偏离，因此很自然会猜想，它是基于工作马尔可夫模型进行增广的逆概率删失加权估计量。

我们证明确实如此。该恒等式在每个样本中都严格成立，甚至对任意工作删失函数也成立。将工作函数设为真实删失函数后，可以得到三个推论。第一，该恒等式解释了 Aalen–Johansen 在非马尔可夫偏离下的稳健性。第二，它给出了 Aalen–Johansen 的直观影响函数，从而可以进行完整的效率分析，尤其是与半参数效率界的比较。第三，效率分析促使我们以 Aalen–Johansen 为基础，构建一种实用的无模型增广方法。

2 作为增广逆概率删失加权的 Aalen–Johansen

2.1 观测数据与逆概率删失加权

除第 1 节中定义的记号外，令 D 表示进入吸收状态（例如死亡）的时间；若不存在吸收状态，则令 $D=\infty$。通过 $D\wedge t$ 进行观测可以确定 $I\{Y(t)=k\}$，其中$a\wedge b=\min(a,b)$。令 $C\in(0,\infty)$ 表示连续分布的删失时间，并假设$C \perp\!\!\!\perp \{Y(t)\colon\ t\geq 0\}$。此外，令 $X=C\wedge D$ 表示观测随访时间，令 $\delta=I(D\leq C)$表示观测到吸收状态的指示变量。由于 $C>0$， $Y(0)$ 始终可观测。

对于由 n 名受试者组成的随机样本，我们在表示第 i 个受试者的量时添加下标 i。令 $N_{jki}(s)$ 统计截至 s 观察到的 $j\to k$ 次转移，并令 $n_j(s)=\sum_{i=1}^n I\{X_i\geq s, Y_i(s-)=j\}$ 状态 j 中的风险集，则式 (3) 中 $\mathrm{d}\widehat{\Lambda}(s)$ 的非对角元素为 $\mathrm{d}\widehat{\Lambda}_{jk}(s)=\sum_{i=1}^n \mathrm{d}N_{jki}(s)/n_j(s)$，当 $\sim j\neq k$ 时；风险集为空时则为零。因此， $I+\mathrm{d}\Lambda(s)$ 保存了每个 $\widehat{\ }$ 起始状态内观察到的转移比例，包括停留在对角元素中的比例。根据式 $\widehat{\ }$(3) 定义的 $P(s,t)$，其第 (j,k) 个元素 $P_{jk}(s,t)$ 是拟合概率：当工作马尔可夫模型从 j 在 s 开始时，于 t 占据 k 的概率。

为刻画逆概率删失加权估计量，令 G^* 为任意删失生存函数，满足 $G^*(0)=1$、 $G^*(t)>0$，且在观测到的状态变化处或 t 处均无跳跃。这可以是真实函数 $G(t)=\Pr(C>t)$、一个估计量，甚至是其错误版本。 G^* 的具体形式在定理1所确立的恒等式中并不重要。在此 G^* 下，逆概率删失加权估计量为

$$\widehat{p}_k^{\mathrm{IPCW}}(t;G^*)=n^{-1}\sum_{i=1}^n \frac{I\{C_i\geq D_i\wedge t,\, Y_i(t)=k\}}{G^*\{(D_i\wedge t)-\}}.$$

2.2 有限样本恒等式

现在我们证明， $\widehat{p}_k(t)$ 等于 $\widehat{p}_k^{\mathrm{IPCW}}(t;\,G^*)$ 加上一个增广项。令 $M_i^C(s;\,G^*)=N_i^C(s)-\int_0^s I(X_i\geq u)\,\mathrm{d}\Lambda^*(u)$ 表示在 G^* 下的删失鞅，其中 $N_i^C(s)=I(X_i\leq s,\,\delta_i=0)$ 且 $\mathrm{d}\Lambda^*=-\mathrm{d}G^*/G^*_-$。

定理1。 对于任意这样的 G^* 以及每个 $t>0$,

$$\widehat{p}_k(t)=\underbrace{n^{-1}\sum_{i=1}^n I(C_i\geq D_i\wedge t)\frac{I\{Y_i(t)=k\}}{G^*\{(D_i\wedge t)-\}}}_{\widehat{p}_k^{\mathrm{IPCW}}(t;G^*)}+\underbrace{n^{-1}\sum_{i=1}^n\int_0^t \frac{\widehat{P}_{Y_i(s-),k}(s,t)}{G^*(s)}\,\mathrm{d}M_i^C(s;G^*)}_{Augmentation\ term}. \qquad (4)$$

定理1的证明。我们概述其思路；附录 A 给出了详细证明。(4) 式两边都等于 k 在 $t = 0$处的经验比例，完整结论可通过对观测到的转移时间进行归纳法证明。在两个这样的时间之间，左端保持不变，而受试者i对逆概率删失加权项的贡献改变了 $-I\{Y_i(t-) = k\}\,\mathrm{d}M_i^C(t;\, G^*)/G^*(t)$，这抵消了增广项的增量，因为 $P(t,t) = I$。在转移时间处，占据向量根据 (3) 通过 $\widehat{p}(t) = \{I + \mathrm{d}\widehat{\Lambda}(t)\}^{\mathsf{T}}\widehat{p}(t-)$ 更新，因此左端的第 k 个分量相应发生跳变。右端也以相同方式跳变：所有风险集中的受试者共享权重 $1/G^*(t-)$，因此$\widehat{\ }$ $\widehat{\ }$ $\widehat{\ }$转移在改变 $\widehat{p}(t)$ 的同时也改变逆概率删失加权项，而 $P(s,t) = P(s,t-)\{I + \mathrm{d}\Lambda(t)\}$ 也对增广项产生相同的作用。

□

2.3 解释与含义

定理1带来两个直接推论。首先，式 (4) 的左侧不涉及G^*，因此该恒等式是对同一数量的一族分解：随着 G^* 变化，右侧的两项会发生变化，但其和保持不变。因此，Aalen–Johansen 无需删失估计量； G^* 只是分析中的一种工具。在 $G^* = G$下，加权项是一致的，而增广项具有渐近均值为零的性质，这恢复了 Aalen–Johansen 已知的稳健性。效率分析需要付出更多努力，留待第 3 节讨论。

其次，我们可以利用式 (4) 恢复第 1 节所指出的单变量和竞争风险情形下， $\widehat{p}_k(t)$ 与逆概率删失加权$\widehat{\ }$估计量之间的数值等价性。具体而言，取 $G^* = G$，即 G的 Kaplan–Meier 估计量。随后，我们可以将式 (4) 中的增广项重新表示为一个关于$\widehat{\ }$的加权积分，其被积函数是 $P_{Y_i(s-)\,k}(s,t)$ 与其在风险集中的受试者的起始状态 0、1、.....、 K 上的平均值之差。如果所有风险集中的受试者都处于同一状态，则该差值为零；图 1(a) 和 (b) 正是这种情况，其中只有一个瞬时状态。

命题 *1*。令 $\widehat{G}(s) = \prod_{u\le s}\{1 - \sum_{i=1}^n dN_i^C(u)/r(u)\}$ 表示 G 的 ***Kaplan–Meier*** 估计量，其中 $r(u) = \sum_{j=0}^K n_j(u)$ 表示 u 时风险集中的受试者数量，并令 $P_k(s,t) = \sum_{j=0}^K\{n_j(s)/r(s)\}\widehat{P}_{jk}(s,t)$ 表示在 s 时风险集中的受试者中，$\widehat{P}_{Y_i(s-)\,k}(s,t)$ 的平均值。如果 $\nsucc G(t) > 0$，则

$$\widehat{p}_k(t) - \widehat{p}_k^{\mathrm{IPCW}}(t;\widehat{G}) = n^{-1}\sum_{i=1}^n\int_0^t \frac{\widehat{P}_{Y_i(s-),k}(s,t) - \overline{P}_k(s,t)}{\widehat{G}(s)}\,\mathrm{d}N_i^C(s). \tag{5}$$

因此，

$\widehat{\ }$*(i)* 对于给定的 k 和 t， $\widehat{p}_k(t) = \widehat{p}_k^{\mathrm{IPCW}}(t;G)$ 如果在 t 之前的每个删失时间，所有风险集中的受试者都处于相同状态；

$\widehat{\ }$ $\widehat{\ }$*(ii)* $\widehat{p}_k(t) = \widehat{p}_k^{\mathrm{IPCW}}(t;G)$ 对于所有 k 以及所有满足 $G(t) > 0$ 的 t，当且仅当在每个删失时间，所有风险集中的受试者都处于同一状态时。

证明详见附录 B。特别地，(ii) 表明，在每个样本中，只有单变量情形和具有单一瞬时状态的竞争风险情形，Aalen–Johansen 与逆概率删失加权之间才存在全局等价关系。

3 渐近分析

$\widehat{p}_k(t)$ 在 (4) 中的表示形式使其适用于渐近效率分析。事实上，鉴于 $\widehat{P}(s,\ t)$ 一致收敛于 $P(s,\ t)$，我们可以在 (4) 中用 P 替换 $\widehat{P}$，而 $\widehat{p}_k(t)$ 仅发生 $o_p(n^{-1/2})$ 阶的变化。令 $G^* = G$，即可得到 $\widehat{p}_k(t)$ 的以下线性化形式。

命题 2。 假设 $\Lambda(t)$ 有限，且 $\mathrm{d}\Lambda(t)$ 由 (1) 中的第二个等式定义。那么

$$n^{1/2}\{\widehat{p}_k(t) - p_k(t)\} = n^{-1/2}\sum_{i=1}^{n}\left[\frac{I\{C_i \geq D_i \wedge t,\, Y_i(t) = k\}}{G\{(D_i \wedge t)-\}} - p_k(t) + \int_0^t \frac{P_{Y_i(s-),k}(s,t)}{G(s)}\,\mathrm{d}M_i^C(s;G)\right] + o_p(1). \tag{6}$$

完整证明见附录 C。

(6) 中影响函数的几何结构在$\widehat{p}_k(t)$的效率分析中发挥着关键作用。为了研究这一几何结构，我们首先需要理解一般情况下 $P_{jk}(s,\ t)$ 究竟是什么。从解析角度看， $P(s,\ t)$ 是 (2) 中 $\mathrm{d}\Lambda(t)$ 的乘积积分。$\mathrm{d}\Lambda(t)$ 的各个元素定义于 (1) 的第二个等式中，可以看作跨越先前历史信息的平均强度，即在非马尔可夫情形下，$\mathrm{d}\Lambda_{jk}(t) = E\left[\mathrm{pr}\{Y(t+\mathrm{d}t) = k \mid Y(t-) = j,\ \mathcal{H}(t-)\} \mid Y(t-) = j\right]$。我们可能会以类似的方式预期 $P_{jk}(s,\ t)$ 是平均转移概率，但事实并非如此（Putter 和 Spitoni, 2018）。

为避免混淆，我们定义所关注的量。

定义1。 对于状态 j 和 k 以及时间 $s \leq t$,

(i) 给定历史信息的转移概率为

$$Q_{jk}\{s,t;\mathcal{H}(s-)\} = \mathrm{pr}\{Y(t) = k \mid Y(s-) = j,\ \mathcal{H}(s-)\};$$

(ii) 平均转移概率为

$$q_{jk}(s,t) = \mathrm{pr}\{Y(t) = k \mid Y(s-) = j\} = E[Q_{jk}\{s,t;\mathcal{H}(s-)\} \mid Y(s-) = j];$$

(iii) 转移概率的马尔可夫替代量为 $P_{jk}(s,\ t)$，它是 $P_{jk}(s,\ t)$的渐近极限$\widehat{\ }$，也是式(2)中乘积积分的第$(j,\ k)$项。

当 $Y(s) = Y(s-)$时，根据(1)，这三个量一致；在删失时间 s处几乎必然如此。一般而言，$Q_{jk}\{s,t;\ \mathcal{H}(s-)\} \neq q_{jk}(s,\ t)$，因为 $Y(\cdot)$在 s时不一定满足马尔可夫假设。$q_{jk}(s,\ t)$与 $P_{jk}(s,\ t)$之间的差异可以从后者在式(2)中的定义看出。随着乘积积分中的 u 从 s 增加到 t ，$\mathrm{d}\Lambda(u)$不再保留条件事件$Y(s-) = j$ ，而仅以当前状态 $Y(u-)$为条件，正如马尔可夫假设所允许的那样。如果该过程在$(s,\ t]$上是非马尔可夫的，这会造成偏倚。总之，

$$Q_{jk}\{s,t;\mathcal{H}(s-)\} \underset{\text{non-Markovness at } s}{\neq} q_{jk}(s,t) \underset{\text{non-Markovness over } (s,t]}{\neq} P_{jk}(s,t). \tag{7}$$

这些关系非常重要，因为 Q_{jk} 和q_{jk} 在比较 $\widehat{p}_k\ (t)$ 的效率方面发挥着关键作用，而 $\widehat{p}_k\ (t)$ 在增广中使用了 P_{jk}。

首先，Maltzahn 等（2026）表明， $p_k\ (t)$ 的有效影响函数，就是将公式（6）中的 $P_{jk}\ (s,\ t)$ 替换为 $Q_{jk}\{s,t;\ \mathcal{H}\ (s-)\}$后的影响函数；也就是说，对于一般受试者，

$$\phi_Q^{\mathrm{eff}}(t) = \frac{I\{C \geq D \wedge t,\, Y(t) = k\}}{G\{(D \wedge t)-\}} - p_k(t) + \int_0^t \frac{Q_{Y(s-),k}\{s,t;\mathcal{H}(s-)\}}{G(s)}\,\mathrm{d}M^C(s;G). \tag{8}$$

这由 $\phi_Q^{\mathrm{eff}}(t) = \phi^{\mathrm{IPCW}}(t) - \Pi\{\phi^{\mathrm{IPCW}}(t) \mid \mathcal{S}\}$ 得到，其中 $\phi^{\mathrm{IPCW}}(t) = I\{C \geq D \wedge t, Y(t) =k\}/G\{(D \wedge t)-\} - p_k(t)$ 是中心化的逆概率删失加权贡献， $\Pi(\cdot \mid \mathcal{S})$ 表示投影到 $\mathcal{S}$ 上，

以及 $\mathcal{S} = \{ \int_0^t h\{s;\ \mathcal{H}(s-)\}\,\mathrm{d}M^C(s;\,G)\colon\ h$ 是平方可积的，$\}$ 是删失的干扰切空间。为实现（8），我们需要一个能够近似 $Q_{jk}\{s,t;\ \mathcal{H}(s-)\}$ 的增广项，而该项取决于完整历史 $\mathcal{H}(s-)$。

为了采用更简单的增广（可能会牺牲完全效率），我们可以改为将 $\phi^{\mathrm{IPCW}}(t)$ 投影到 $\mathcal{S}$ 的一个真子空间上，例如由当前状态的函数构成的鞅积分：

$$\mathcal{S}_0 = \left\{ \int_0^t a\{s; Y(s-)\}\,\mathrm{d}M^C(s;G) : a \text{ is square-integrable} \right\} \subset \mathcal{S}.$$

令 $\phi_q(t) = \phi^{\mathrm{IPCW}}(t) - \Pi\{\phi^{\mathrm{IPCW}}(t) \mid \mathcal{S}_0\}$ 表示由此得到的影响函数。

命题3. 我们有

$$\phi_q(t) = \frac{I\{C \geq D \wedge t,\, Y(t) = k\}}{G\{(D \wedge t)-\}} - p_k(t) + \int_0^t \frac{q_{Y(s-),k}(s,t)}{G(s)}\,\mathrm{d}M^C(s;G).$$

因此，当前状态增广通常使用 $q_{jk}(s,t)$，而不是 $P_{jk}(s,t)$。

证明见附录 C。

现在，令 $\phi_{\mathrm{AJ}}(t)$ 表示式 (6) 中 $\widehat{p}_k(t)$ 的影响函数。在马尔可夫模型下，$\phi_{\mathrm{AJ}}(t) = \phi_Q^{\mathrm{eff}}(t)$，因为 $P_{jk}(s,t) = Q_{jk}\{s,t;\ \mathcal{H}(s-)\}$，因此 $\widehat{p}_k(t)$ 完全有效率。一般而言，效率损失可归因于用 P_{jk} 近似 Q_{jk} 的两个步骤。首先，用 $q_{jk}(s,t)$ 替代 $Q_{jk}\{s,t;\ \mathcal{H}(s-)\}$ 会在 $Y(\cdot)$ 于 s 处为非马尔可夫时，丢失 $\mathcal{H}(s-)$ 中可能包含的信息；其次，用 $P_{jk}(s,t)$ 替代 $q_{jk}(s,t)$，若 $Y(\cdot)$ 在 $(s,t]$ 上为非马尔可夫，则会偏离最佳的当前状态增广，正如式 (7) 所概述。这两个步骤彼此正交，因为 $\phi_q(t) - \phi_Q^{\mathrm{eff}}(t)$ 与 $\mathcal{S}_0$ 正交，而 $\phi_{\mathrm{AJ}}(t) - \phi_q(t)$ 属于其中，因此它们对方差的贡献会相加。图2对此作了说明，下一条命题将对其进行量化。

命题4。令 $\Lambda_C = -\log G$。在满足命题 *2* 的条件下，

$$\begin{aligned}\mathrm{var}\{\phi_{\mathrm{AJ}}(t)\} - \mathrm{var}\{\phi_Q^{\mathrm{eff}}(t)\} = \int_0^t E\Big(& \underbrace{\mathrm{var}[Q_{Y(s-),k}\{s,t;\mathcal{H}(s-)\} \mid Y(s-)]}_{\textit{Individual variation of } Q_{jk} \textit{ from mean } q_{jk}} \\ & + \underbrace{\{P_{Y(s-),k}(s,t) - q_{Y(s-),k}(s,t)\}^2}_{\textit{Approximation error of } P_{jk} \textit{ for mean } q_{jk}} \Big) \frac{\mathrm{d}\Lambda_C(s)}{G(s)}.\end{aligned} \tag{9}$$

附录 C 给出了任意增广项的一般效率比较，其中式 (9) 中的增广项和逆概率删失加权均为特殊情形。在附录E中基于马尔可夫和非马尔可夫模型进行的模拟中，采用 Kaplan–Meier 权重的逆概率删失加权的方差比 Aalen–Johansen 高出最多 34%，这表明后者是更好的非参数基准估计量选择。

4 无模型最小二乘增广的 Aalen–Johansen

要得到（8），可以将 $Q_{jk}\{s,t;\mathcal{H}(s-)\}$ 替换为基于拟合的完整历史信息预后，并相应地对逆概率删失加权进行增广（Maltzahn 等人，2026年）。这需要为整个过程建立模型；在非马尔可夫情形下，通常必须通过模拟才能获得该预后。如果模型错误，该估计量将仅保留逆概率删失加权的一致性，无法保证优于$\widehat{p}_k(t)$。

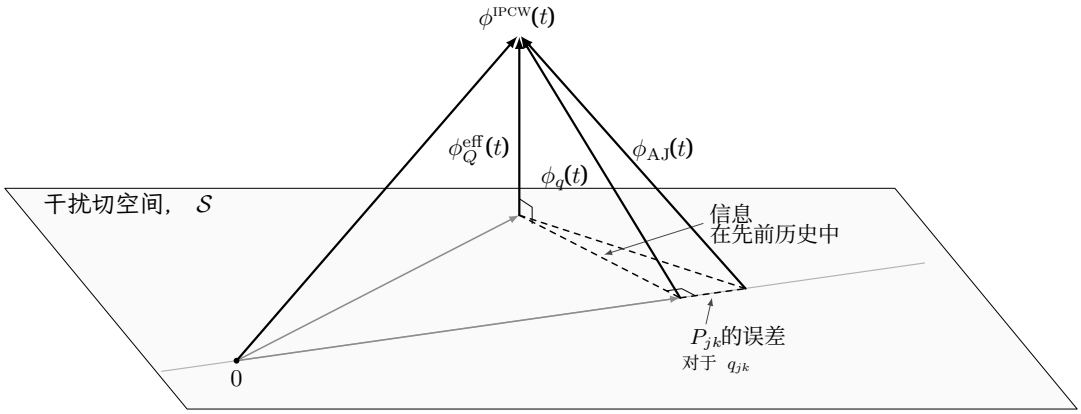

图2：影响函数的几何关系。虚线直角三角形解释了相对于 $\phi_Q^{\mathrm{eff}}(t)$， $\phi_{\mathrm{AJ}}(t)$ 的效率损失的类似勾股定理的分解，如命题4的（9）式所表述。在马尔可夫模型中，该三角形收缩为一个点，因此$\phi_{\mathrm{AJ}}(t) = \phi_q(t) = \phi_Q^{\mathrm{eff}}(t)$。

相反，我们提出对 Aalen–Johansen 进行无模型最小二乘增广。根据（6）和（8）， $\widehat{p}_k(t)$ 的影响函数与有效影响函数的差异在于，通过被积函数为$Q_{Y(s-)\,k}\{s,t;\ \mathcal{H}(s-)\} - P_{Y(s-)\,k}(s,\ t)$ 的增广。令

$$\Delta_{jk}\{s,t;\mathcal{H}(s-)\} = Q_{jk}\{s,t;\mathcal{H}(s-)\} - q_{jk}(s,t) \tag{10}$$

表示先前历史信息对当前状态的增量，因此 $Q_{jk}\{s,t;\mathcal{H}(s-)\} - P_{jk}(s,t) = \Delta_{jk}\{s,t;\mathcal{H}(s-)\} + \{q_{jk}(s,t) - P_{jk}(s,t)\}$。这两项彼此正交，并分别对应（9）式中的两项。根据命题4，对 Δ_{jk} 进行近似，因而可以减小方差差距的第一项，而不影响第二项。根据定义1， $E[\Delta_{jk}\{s,t;\mathcal{H}(s-)\} \mid Y(s-) = j] = 0$。

Δ_{jk} 的工作版本也应具有这一性质。令 $h\{s;\mathcal{H}(s-)\}$ 为预先指定的评分函数，用于概括历史信息 $\mathcal{H}(s-)$，例如处于当前状态的时间长度， $h\{s;\mathcal{H}(s-)\} = s - \sup\{u < s:\ Y(u) \neq Y(s-)\}$ with $\sup \emptyset = 0$，它考虑了半马尔可夫依赖。为满足条件均值为零性质，我们在当前状态内对其进行中心化：

$$e_i(s) = h\{s;\mathcal{H}_i(s-)\} - \bar{h}_{Y_i(s-)}(s),$$
$$\text{where } \bar{h}_j(s) = n_j(s)^{-1}\sum_{i=1}^{n} I\{X_i \geq s,\, Y_i(s-) = j\}h\{s;\mathcal{H}_i(s-)\}.$$

与删失结局的协变量调整中的最小二乘增广类似（Lu and Tsiatis，2008年；Zhang 等，2008年），我们用线性工作模型 $\beta e_i(s)$ 近似 $\Delta_{Y_i(s-)\,k}\{s,t;\ \mathcal{H}_i(s-)\}$，并通过回归选择 β。由于 Δ_{jk} 是给定历史信息时 $I\{Y(t) = k\}$ 的均值，再减去其在状态内的均值，而 $e_i(s)$ 已在状态内中心化，因此这相当于在 s 时风险集中的受试者中，将 $I\{Y_i(t) = k\}$ 对 $e_i(s)$ 进行回归。在删失情况下，响应变量 $\widehat{\ }$ $\widehat{\ }$ 被 $o_i(s) = I\{C_i \geq D_i \wedge t,\, Y_i(t) = k\}G(s-)/G\{(D_i \wedge t)-\}$ 替代；该响应在 s 之后按删失 $\widehat{\ }$ 加权，并且给定历史信息时具有相同的均值。令 β 为 $o_i(s)$ 对 $e_i(s)$ 的最小二乘斜率，在增广项发挥作用的删失时间 s 上合并计算。将拟合增广项 $n^{-1}\sum_{i=1}^{n}\int_0^t \widehat{\beta} e_i(s)\,\widehat{G}(s)^{-1}\,\mathrm{d}M_i^C(s;\,\widehat{G})$ 加到 Aalen–Johansen 估计量上，我们得到

$$\widehat{p}_k^{\mathrm{aug}}(t) = \widehat{p}_k(t) + \widehat{\beta}\,\widehat{c}_h(t), \qquad \widehat{c}_h(t) = n^{-1}\sum_{i=1}^{n} I(C_i < D_i \wedge t)\,\widehat{G}(C_i)^{-1}e_i(C_i). \tag{11}$$

该过程的详细信息见附录 D。

命题5。 (11) 中的估计量具有以下性质。

(i) 如果 $h\{s\mathcal{H}(s-)\} = a\{s;\ Y(s-)\}$ 对于某个函数 a，则在每个样本中，$\widehat{p}_k^{\mathrm{aug}}(t) = \widehat{p}_k(t) \quad le.$

(ii) 对于一般评分函数 h，在附录 *D* 的正则条件下，

$$\limsup_{n\to\infty} n\big[\operatorname{var}\{\widehat{p}_k^{\mathrm{aug}}(t)\} - \operatorname{var}\{\widehat{p}_k(t)\}\big] \le 0.$$

因此，不包含超出当前状态之外任何信息的评分函数不会改变 Aalen–Johansen，而任何其他评分函数都不会增加其渐近方差。附录 D 给出了证明。在附录 E 的模拟中，h 在当前状态中停留的时间；当转移率随该时间增加时，方差下降 3–16%，否则变化至多为 1%。无需对该过程的任何部分进行建模或模拟。

致谢

本研　究得到了美国国立卫生研究院资助项目 R01HL149875 的支持　.

参考文献

O. O. Aalen 和 S. Johansen。基于删失观测的非齐次马尔可夫链经验转移矩阵。《斯堪的纳维亚统计学杂志》，5:141–150，1978年。P. K. Andersen、Ø. Borgan、R. D. Gill 和 N. Keiding。基于计数过程的统计模型。Springer-Verlag，纽约，1993年。A. B¨uhler、R. J. Cook 和 J. F. Lawless。多状态模型作为复杂过程临床试验中目标估计量设定的框架。《医学统计学》，42:1368–1397，2023年。R. J. Cook 和 J. F. Lawless。用于分析生命历程数据的多状态模型。Chapman & Hall/CRC，博卡拉顿，佛罗里达州，2018年。doi：10.1201/9781315119731。S. Datta 和 G. A. Satten。非马尔可夫模型中阶段占用概率的 Aalen–Johansen 估计量和积分转移风险的 Nelson–Aalen估计量的有效性。《统计概率通讯》，55:403–411，2001年。doi：10.1016/S0167-7152(01)00155-9。D. V. Glidden。非马尔可夫事件数据中事件概率的稳健推断。《生物统计学》，58:361–368，2002年。doi：10.1111/j.0006-341X.2002.00361.x。X. Lu 和 A. A. Tsiatis。使用辅助协变量提高对数秩检验的效率。《生物统计学》，95:679–694，2008年。N. N. Maltzahn、G. D. Luk´ats 和 K. Røysland。粗化多状态过程中占据概率的稳健估计。arXiv:2606.27952，2026年。L. Parast、L. Tian 和 T. Cai。随机临床试验中的生存和治疗效应的标志点估计。《美国统计学会杂志》，109:384–394，2014年。doi：10.1080/01621459.2013.842488。

H. Putter 和 C. Spitoni。非马尔可夫多状态模型中转移概率的非参数估计：标志点 Aalen–Johansen 估计量。《医学统计学方法》，27:2081–2092，2018年。doi：10.1177/0962280216674497。H. Putter、M. Fiocco 和 R. B. Geskus。生物统计学教程：竞争风险和多状态模型。《医学统计学》，26:2389–2430，2007年。doi：10.1002/sim.2712。D. B. Rubin 和 M. J. van der Laan。经验效率最大化：随机实验和生存分析中改进的局部有效协变量调整。《国际生物统计学杂志》，4(1)：文章 5，2008年。doi：10.2202/1557-4679.1084。O. L. Sandqvist。用于右删失结局回归和推断的双重稳健学习器。《斯堪的纳维亚统计学杂志》，53:1176–1188，2026年。doi：10.1111/sjos.70076。G. A. Satten 和 S. Datta。Kaplan–Meier 估计量作为删失逆概率加权平均。《美国统计学家》，55:207–210，2001年。doi：10.1198/000313001317098185。M. J. van der Laan 和 J. M. Robins。删失纵向数据与因果关系的统一方法。Springer，纽约，2003年。doi：10.1007/978-0-387-21700-0。T. Yan、R. Geskus 和 H. Putter。多状态模型中状态占用概率的 Aalen–Johansen 估计量的替代表述。《荷兰统计学杂志》，80:e70035，2026年。doi：10.1111/stan.70035。M. Zhang、A. A. Tsiatis 和 M. Davidian。使用辅助协变量提高随机临床试验中推断的效率。《生物统计学》，64:707–715，2008年。X. Zhang、M.-J. Zhang 和 J. P. Fine。右删失和左截断事件发生时间数据的质量再分配算法。《统计规划与推断杂志》，139:3329–3339，2009年。doi：10.1016/j.jspi.2009.03.007。

附录

定理1的证明

令 $W_i(t) = I(C_i \geq D_i \wedge t)/G^*\{(D_i \wedge t)-\}$ 表示受试者 i 的权重，使得 $\widehat{p}_k^{\mathrm{IPCW}}(t; G^*) = n^{-1}\sum_{i=1}^n W_i(t) I\{Y_i(t) = k\}$，并令

$$\widehat{U}_k(t) = n^{-1}\sum_{i=1}^n \int_0^t G^*(s)^{-1}\widehat{P}_{Y_i(s-),k}(s,t)\,\mathrm{d}M_i^C(s; G^*)$$

表示增广项，其中 $M_i^C(s;\ G^*) = N_i^C(s) - \int_0^s I(X_i \geq u)\,\mathrm{d}\Lambda^*(u)$，$N_i^C(s) = I(X_i \leq s,\ \delta_i = 0)$ 且 $\mathrm{d}\Lambda^* = -\mathrm{d}G^*/G^*_-$，如正文所述。定理1指出

$$\widehat{p}_k(t) = \widehat{p}_k^{\mathrm{IPCW}}(t; G^*) + \widehat{U}_k(t) \quad (k = 0, \ldots, K). \tag{A1}$$

我们首先证明

$$W_i(t) = 1 - \int_0^t G^*(s)^{-1}\,\mathrm{d}M_i^C(s; G^*). \tag{A2}$$

由于 $\mathrm{d}(1/G^*) = \mathrm{d}\Lambda^*/G^*$ 且 $G^*(0) = 1$，

$$\int_0^t I(X_i \geq s)G^*(s)^{-1}\,\mathrm{d}\Lambda^*(s) = \int_0^{X_i \wedge t} \mathrm{d}\{1/G^*(s)\} = G^*(X_i \wedge t)^{-1} - 1,$$

因此，(A2) 的右端等于 $G^*(X_i \wedge t)^{-1} - N_i^C(t)G^*(X_i)^{-1}$。如果 $C_i \geq D_i \wedge t$，那么 $N_i^C(t) = 0$ 且 $X_i \wedge t = D_i \wedge t$，并且该表达式为 $G^*(D_i \wedge t)^{-1} = W_i(t)$，因为 G^* 在观测到的状态变化或 t 处没有跳跃。如果 $C_i < D_i \wedge t$，那么 $N_i^C(t) = 1$ 且 $X_i \wedge t = C_i$，该表达式为零，$W_i(t)$ 也是如此。事件 $C_i = t < D_i$ 的概率为零；在该事件上，$N_i^C(t) = 1$，而 $W_i(t) > 0$。

令 $0 = t_0 < t_1 < \cdots < t_m$ 表示观测到的转移时间。我们通过对 l 进行归纳法，证明 (A1) 在 $[0, t_l]$ 上成立。在 $t_0 = 0$，$W_i(0) = 1$ 且 $\widehat{U}_k(0) = 0$，因此 (A1) 两边都是 k 中的经验比例。

假设 (A1) 在 t_l 处成立，并令 $t_l \leq t < t_{l+1}$。在 $(t_l, t]$ 内未观察到转移，因此 $\widehat{p}_k(t) = \widehat{p}_k(t_l)$，且每个被观察的受试者都保持其状态。根据 (A2)，

$$\widehat{p}_k^{\mathrm{IPCW}}(t; G^*) - \widehat{p}_k^{\mathrm{IPCW}}(t_l; G^*) = -n^{-1}\sum_{i=1}^n \int_{(t_l,t]} G^*(s)^{-1} I\{Y_i(s-) = k\}\,\mathrm{d}M_i^C(s; G^*).$$

$\widehat{\ }$ $\widehat{\ }$ $\widehat{\ }$出于同样的原因，对于 $s \leq t_l$，$P(s,t) = P(s,t_l)$，而 $P(s,t)$ 是 $t_l < s \leq t$ 的单位矩阵，因此 $\widehat{\ }$ $\widehat{\ }$that $U_k(t) - U_k(t_l)$ 是上一个显示式的相反数。于是 (A1) 在 $[t_l, t_{l+1})$ 上成立。

现在令 $t = t_{l+1}$。根据正文中的乘积公式，$\widehat{p}_k(t) = \widehat{p}_k(t-) + \sum_{j=0}^K \widehat{p}_j(t-)\,\mathrm{d}\widehat{\Lambda}_{jk}(t)$. 只需证明 (A1) 右侧的两项以相同方式变化，因为 (A1) 对 $t-$ 处的每个状态都成立。由于 $\widehat{P}(s,t) = \widehat{P}(s,t-)\{I + \mathrm{d}\widehat{\Lambda}(t)\}$ 对 $s < t$ 成立，且 $M_i^C(\cdot;\ G^*)$ 在 t 处没有跳跃，因此有 $\widehat{U}_k(t) = \widehat{U}_k(t-) + \sum_{j=0}^K \widehat{U}_j(t-)\,\mathrm{d}\widehat{\Lambda}_{jk}(t)$。对于加权项，删失与观测到的状态变化不同时发生，并且 G^* 在那里没有跳跃，因此 W_i 在 t 处不发生跳跃，而状态在 t 发生变化的受试者在 $W_i(t-) = G^*(t-)^{-1}$ 时处于风险集中。由于 $\mathrm{d}I\{Y_i(t) = k\} = \sum_{j\neq k}\{\mathrm{d}N_{jki}(t) - \mathrm{d}N_{kji}(t)\}$，

$$\widehat{p}_k^{\mathrm{IPCW}}(t; G^*) - \widehat{p}_k^{\mathrm{IPCW}}(t-; G^*) = n^{-1}G^*(t-)^{-1}\sum_{i=1}^n \sum_{j\neq k}\{\mathrm{d}N_{jki}(t) - \mathrm{d}N_{kji}(t)\}. \tag{A3}$$

对于瞬时状态 j，所有满足 $X_i \geq t$ 且 $Y_i(t-) = j$ 的受试者，在 t之前都有 $D_i \geq t$，因此权重为$G^*(t-)^{-1}$，所以 $\widehat{p}_j^{\mathrm{IPCW}}(t-; G^*) = n_j(t)/\{nG^*(t-)\}$。对于 $j \neq k$，风险集因此相互抵消：

$$\widehat{p}_j^{\mathrm{IPCW}}(t-; G^*)\, \mathrm{d}\widehat{\Lambda}_{jk}(t) = \frac{n_j(t)}{nG^*(t-)} \frac{\sum_{i=1}^n \mathrm{d}N_{jki}(t)}{n_j(t)} = n^{-1} G^*(t-)^{-1} \sum_{i=1}^n \mathrm{d}N_{jki}(t).$$

对于吸收状态 j，不会观测到离开该状态，因此对所有 k，$\mathrm{d}\Lambda_{jk}(t) = 0$；在吸收时冻结的权重不会参与其中。由于 $\mathrm{d}\Lambda_{kk}(t) = -\sum_{j \neq k} \mathrm{d}\Lambda_{kj}(t)$，(A3) 的右端为 $\sum_{j=0}^K \widehat{p}_j^{\mathrm{IPCW}}(t-; G^*)\, \mathrm{d}\widehat{\Lambda}_{jk}(t)$。因此，(A1) 在 t_{l+1} 处成立，由此完成归纳法。

B 命题 1 的证明

在定理1中令 $G^* = G$。关于 G^* 的条件成立，因为 $G(0) = 1$，$G(t) > 0$，且 G 仅在删失时间发生跳变；这些时间与观测到的状态变化以及 t均不重合。因此

$$\widehat{p}_k(t) - \widehat{p}_k^{\mathrm{IPCW}}(t; \widehat{G}) = n^{-1} \sum_{i=1}^n \int_0^t \widehat{G}(s)^{-1} \widehat{P}_{Y_i(s-),k}(s,t)\, \mathrm{d}M_i^C(s; \widehat{G}), \tag{A4}$$

其中 $M_i^C(s;\, \widehat{G}) = N_i^C(s) - \int_0^s I(X_i \geq u)\, \mathrm{d}\Lambda^*(u)$，且 $\mathrm{d}\Lambda^* = -\mathrm{d}\, \widehat{G}/\widehat{G}_-$。根据 Kaplan–Meier 估计量的定义，$\mathrm{d}\Lambda^*(s) = r(s)^{-1} \sum_{l=1}^n \mathrm{d}N_l^C(s)$。因此，(A4) 的补偿过程部分为

$$\begin{aligned} & n^{-1} \sum_{l=1}^n \int_0^t \widehat{G}(s)^{-1} \left\{ r(s)^{-1} \sum_{i=1}^n I(X_i \geq s) \widehat{P}_{Y_i(s-),k}(s,t) \right\} \mathrm{d}N_l^C(s) \\ & \quad = n^{-1} \sum_{l=1}^n \int_0^t \widehat{G}(s)^{-1} \overline{P}_k(s,t)\, \mathrm{d}N_l^C(s), \end{aligned}$$

因为大括号中的平均值等于 $\sum_{j=0}^K \{n_j(s)/r(s)\} \widehat{P}_{jk}(s,t) = P_k(s,t)$。从 (A4) 中对 N_i^C 积分的部分减去该项，即得到命题 1 中的公式。如果在删失时间处风险集中的所有受试者都处于相同的状态 j，则 $P_k(s,t) = P_{jk}(s,t)$，并且对于每个 k 和 t，被积函数在s 处都为零，这就证明了 (i) 以及 (ii) 的充分性。对于 (ii) 的必要性，假设该条件不成立，并令 s 表示风险集中的受试者处于多个状态的第一个删失时间。取 $t > s$ 小于下一次观测到的转移或删失时间，并令 k 表示在 s 被删失的受试者所处的状态。在 s 之前的删失时间对命题 1 中的公式没有贡献。在 $(s,t]$ 内没有观测到转移，因此 $P(s,t)$ 是单位矩阵，而 s 处的被积函数为 $\{1 - n_k(s)/r(s)\}/G(s)$；由于某个风险集中的受试者处于另一状态，该值为正。因此 $\widehat{p}_k(t) \neq \widehat{p}_k^{\mathrm{IPCW}}(t; G)$。

使用当前状态或完整历史信息进行 C投影

令 $Y(t)$ 表示时间 t 时的状态，且 $p_k(t) = \mathrm{pr}\{Y(t) = k\}$。假设删失独立于生命历程 $\{Y(t)\colon t \geq 0\}$，其生存函数在 t 之前为正。在正文的观测方案下，随访至 $D \wedge t$（即吸收与 t 中较早者）决定 $I\{Y(t) = k\}$。令 $Q_{jk}\{s,t; \mathcal{H}(s-)\}$、$q_{jk}(s,t)$ 和 $P_{jk}(s,t)$ 如正文定义1所述。令 C 为删失时间。令 $X = C \wedge D$ 表示观测随访时间，Λ_C 表示共同删失风险率；对于一般受试者，令 $M^C(s;\, G)$ 表示正文中的删失鞅，其补偿过程为 $\int_0^s I(X \geq u)\, \mathrm{d}\Lambda_C(u)$。

C.1 Aalen–Johansen 的渐近线性性

我们首先证明，在定理1的恒等式中用 P 替换 P 的一阶影响可以忽略。令 $y_j(u) = \Pr\{X \geq u, Y(u-) = j\}$。由于 C 独立于生命历程，正文公式(1)中第二个表达式定义的转移率 $\mathrm{d}\Lambda_{jl}(u)$ 等于 $\mathrm{d}E\{N_{jl}(u)\}/y_j(u)$，其中 $N_{jl}(u)$ 统计受试者截至 u 的观测 $j \to l$ 转移次数。

假设对于每个瞬时状态 j 和每个 $l \neq j$，$\Lambda_{jl}(t) < \infty$。我们首先证明，$\sup_{s \leq t} |\widehat{\Lambda}_{jl}(s) - \Lambda_{jl}(s)| \to 0$ 依概率成立。对于 $\eta > 0$，令 $B_\eta = \{u \leq t : y_j(u) < \eta\}$。给定受试者 i 在 u 时处于 j 的风险集中，其他此类受试者的数量 $n_j(u)-1$ 服从二项分布，且与受试者 i 相互独立，因此在该事件条件下，$E\{1/n_j(u)\} \leq 1/\{n y_j(u)\}$。于是 $E\{\int_{B_\eta} \mathrm{d}\widehat{\Lambda}_{jl}(u)\} \leq \int_{B_\eta} \mathrm{d}\Lambda_{jl}(u)$，当 $\eta \to 0$ 时该值趋于零，因为 $\Lambda_{jl}(t) < \infty$。在 B_η 之外，y_j 至少为 η；根据 Glivenko–Cantelli 定理，$n^{-1} n_j(u)$ 和 $n^{-1}\sum_{i=1}^n N_{jli}(u)$ 依概率一致收敛至 $y_j(u)$ 和 $E\{N_{jl}(u)\}$；由于 $E\{N_{jl}(t)\} \leq \Lambda_{jl}(t)$，转移的期望数量有限。因此，$\widehat{\ }$部分的 $\Lambda_{jl}(s)$ 在 B_η 之外累积的部分一致收敛至 $\Lambda_{jl}(s)$ 的相应部分，令 $\eta \to 0$ 即得所述结论。

对于一个由瞬态状态索引、定义在 $[0,t]$ 上的函数向量 $h = \{h_j\}$，令

$$f_h = \int_0^t G(s)^{-1} h_{Y(s-)}(s)\, \mathrm{d}M^C(s; G),$$

并令 f_h、$_i$ 表示其在受试者 i 上的取值。令 $G^* = G$，增广项的变化为 $R_n = n^{-1}\sum_{i=1}^n (f_{\widehat{h},i} - f_{h_0,i})$，其中 $\widehat{h}_j(s) = \widehat{P}_{jk}(s,t)$，且 $h_{0j}(s) = P_{jk}(s,t)$。由于 C 与生命历程独立，给定 Y 的路径后，$M^C(\cdot;\, G)$ 是鞅，因此对每个固定的 h，$E(f_h) = 0$。令 $\mathcal{F}$ 表示由 h 构成的函数类，其中每个 h_j 的绝对值不超过 1，且全变差至多为 V。每个 f_h 都是 $-\int_0^{X\wedge t} G(s)^{-1} h_Y(s-)(s)\, \mathrm{d}\Lambda_C(s)$ 与（对于在 $D \wedge t$ 之前删失的受试者）$G(X)^{-1} h_Y(X-)(X)$之和。由于一致有界的有界变差函数构成 Donsker 类，$\{f_h$ 也构成 Donsker 类：$h \in \mathcal{F}\}$。由于 $P(s-,t) - P(s,t) = \mathrm{d}\Lambda(s) P(s,t)$，且 $\widehat{P}$的各元素 $\widehat{\ }$属于 $[0,1],$，所以 h_j 在 $[0,t]$ 上的全变差至多为 $2\sum_{l\neq j}\Lambda_{jl}(t)$，因此当 $V > 2\max_j \sum_{l\neq j}\Lambda_{jl}(t)$ 时，$h \in \mathcal{F}$以趋于 1 的概率成立。根据乘积积分的连续性，$\sup_{s\leq t}|\widehat{h}_j(s) - h_{0j}(s)| \to 0$ 依概率收敛，因此在 $h = \widehat{h}$ 处取值的 $E\{(f_h - f_{h_0})^2\}$ 依概率趋于零。随后，由 Donsker 类索引的经验过程的渐近等度连续性可得

$$R_n = n^{-1}\sum_{i=1}^n \{f_{\widehat{h},i} - f_{h_0,i}\} - E(f_h - f_{h_0})\big|_{h=\widehat{h}} = o_p(n^{-1/2}).$$

根据定理1，$G^* = G$，$\widehat{p}_k(t) = \widehat{p}_k^{\mathrm{IPCW}}(t; G) + n^{-1}\sum_{i=1}^n f_{h_0,\ i} + R_n$。加权项是独立同分布项的平均值，其均值为 $E[I\{C \geq D\wedge t, Y(t) = k\}/G\{(D\wedge t)-\}] = p_k(t)$，这是由 C 的独立性所致；而 $n^{-1}\sum_{i=1}^n f_{h_0,\ i}$ 是这样一个均值为零的平均值。因此，$\widehat{p}_k(t)$ 是 $p_k(t)$ 的一致估计量，并且

$$n^{1/2}\{\widehat{p}_k(t) - p_k(t)\} = n^{-1/2}\sum_{i=1}^n \left[\frac{I\{C_i \geq D_i \wedge t, Y_i(t) = k\}}{G\{(D_i\wedge t)-\}} - p_k(t) + f_{h_0,i}\right] + o_p(1),$$

因此，Aalen–Johansen 是渐近线性的，其影响函数如正文所述。

C.2 高效与当前状态增广

给定未删失轨迹，$I\{C \geq D \wedge t, Y(t) = k\}/G\{(D \wedge t)-\}$ 的均值为$I\{Y(t) = k\}$。在具有 $D \geq s$ 的受试者中，改为给定 s 之前的历史信息，可得$Q_{Y(s-)k}\{s,t;\ \mathcal{H}(s-)\}$。给定当前状态 j，可得 $q_{jk}(s,\ t)$。

对于基于历史信息、具有可预测被积函数 $a(s)$ 的平方可积删失鞅积分，其与逆概率删失加权贡献的协方差为

$$E\left[\frac{I\{C \geq D \wedge t,\, Y(t) = k\}}{G\{(D \wedge t)-\}} \int_0^t a(s)\,\mathrm{d}M^C(s;G)\right]$$
$$= -\int_0^t E[I(D \geq s) Q_{Y(s-),k}\{s,t;\mathcal{H}(s-)\} a(s)]\,\mathrm{d}\Lambda_C(s)$$

因为鞅积分的均值为零，且删失计数过程发生跳跃时，加权示性函数为零。在补偿项中，对历史信息取条件期望会消去逆删失权重。

$M^C(\cdot;\, G)$的可预测变差过程 $\langle M^C_s\rangle$ 满足 $\mathrm{d}\langle M^C\rangle(s) = I(X \geq s)\{1 - \Delta\Lambda_C(\quad s = G\ s \quad s-\quad \mathcal{H}\ s- \quad M^C\ s\ G \quad s\ )\}\mathrm{d}$ $\Lambda_C(\,)$，其中 $\Delta\Lambda_C(\,) = \Lambda_C(\,) - \Lambda_C(\,)$。由于 $\mathrm{pr}(C \geq s) = G(\,)$ 且 $1 - \Delta\Lambda_C(\,)\,(\,)/G(\,)$，鞅等距表明，对每个 a，加入 $\int_0^t Q_{Y(s-)k}\{s,t;\,(\,)\}\mathrm{d}\,(\,;\,)/G(\,)$ 都能消去这一协方差。对于连续删失，这给出了 Maltzahn 等人（2026年，第2.3节）的全历史校正。在删失原子处，分母使用跳后生存率 $G(s)$，与 Sandqvist（2026年，补充附录B）中的做法一致。将 $a(s)$ 限定为s 的函数，并将 $Y(s-)$ 限定为状态 j 中的函数后，校正项中的 $Q_{Y(s-)k}\{s,t;\mathcal{H}(s-)\}$ 会被其在当前状态内的均值 $q_{jk}(s,t)$ 所替代。

C.3 一般增广的方差

对于一个固定的有界预测 $m_k(s,\ t)$，该预测基于 s 之前的历史信息，令 $\widehat{p}_k(t;m)$ 为以下内容的样本均值

$$\frac{I\{C \geq D \wedge t,\, Y(t) = k\}}{G\{(D \wedge t)-\}} + \int_0^t \frac{m_k(s,t)}{G(s)}\,\mathrm{d}M^C(s;G)$$

基于 n 名独立受试者。正交性和鞅等距给出 e

$$\mathrm{var}\{\widehat{p}_k(t;m)\} - \mathrm{var}\{\widehat{p}_k(t;Q)\}$$
$$= n^{-1}\int_0^t E\big(I(D \geq s)[m_k(s,t) - Q_{Y(s-),k}\{s,t;\mathcal{H}(s-)\}]^2\big)\frac{\mathrm{d}\Lambda_C(s)}{G(s)}. \tag{A5}$$

对于工作时刻 作中的 Aalen–Johansen 状态 j 预后，其 (A5) 中的条件平方误差为

$$\mathrm{var}[Q_{jk}\{s,t;\mathcal{H}(s-)\} \mid Y(s-) = j] + \{P_{jk}(s,t) - q_{jk}(s,t)\}^2.$$

第一项衡量历史信息中超出当前状态的预后信息。第二项衡量工作状态概率与真实状态概率之间的差异。使用真实 G 的逆删失概率加权估计量是情形 $m_k(s,\ t) \equiv 0$，此时 (A5) 中的被积函数为 $E[I(D \geq s)Q_{Y(s-)k}\{s,t;\mathcal{H}(s-)\}^2]$。

为得到正文图2中的几何关系，我们在精确的 Aalen–Johansen恒等式中设定 $G^* = G$，即真实的删失生存函数。正如附录 C.1 所示，用其极限 P 替换 P 会使平均删失积分改变 $o_p(n^{-1/2})$。因此，Aalen–Johansen的影响函数等于上方的中心化被加项，其中 $m_k(s,\ t) = P_{Y(s-)k}(s,\ t)$。当

$P_{jk}(s,\ t) = q_{jk}(s,\ t)$ 在状态 j 的删失支持集上时，加权影响函数与 Aalen–Johansen影响函数之差与后者正交。为实现完全效率而进行的进一步校正使用乘子 $[Q_{Y(s-)k}\{s,t;\ \mathcal{H}(s-)\} - q_{Y(s-)k}(s,\ t)]/G(s)$，根据上述协方差计算，该校正与有效影响函数正交。根据附录 C.1，Aalen–Johansen的渐近方差比有效估计量的渐近方差大 n 倍于 (A5) 右端，其中 $m_k(s,\ t) = P_{Y(s-)k}(s,\ t)$；而下方给出的 (A5) 条件平方误差分解则得到正文中的方差公式。

D 无模型最小二乘增广

在本节及附录E中，我们将正文第 4 节的评分函数简记为 $h_i(s) = h\{s;$ 对于受试者 i，记为 $\mathcal{H}_i(s-)\}$；对于一般受试者，记为 $h(s) = h\{s;\ \mathcal{H}(s-)\}$。为估计 $p_k(t)$的$^-$，令评分函数有界且可预测，并令 $h_j(s)$表示在 s之前处于 j 状态的风险集中的受试者$\widehat{\ }$的样本均值。令 G 表示合并的 Kaplan–Meier 删失生存函数估计量。按照正文第 4 节的做法，以 $e_i(s) = h_i(s) - h_{Y_i(s-)}(s)$表示在当前$\widehat{\ }$状态内中心化后的评分函数，我们将 $\beta\widehat{c}_h(t)$加到 Aalen–Johansen估计值中，其中

$$\widehat{c}_h(t) = n^{-1}\sum_{i=1}^{n}\int_0^t \frac{h_i(s) - \bar{h}_{Y_i(s-)}(s)}{\widehat{G}(s)}\,\mathrm{d}N_i^C(s).$$

在每个删失时间，各状态内中心化后的评分函数之和为零。因此，当用拟合或真实的删失鞅增量替换 $\mathrm{d}N_i^C$时，补偿过程会相互抵消。在每个状态内均为常数的评分函数不会改变 Aalen–Johansen估计值。

正文命题 5(ii) 的正则条件包括命题 2 的条件、一个有界的可预测得分函数，以及下文所述的一阶替代条件。在这些条件下，$n^{1/2}\{\widehat{p}_k(t) - p_k(t)\}$ 和 $n^{1/2}\{\widehat{p}_k^{\mathrm{aug}}(t) - p_k(t)\}$ 渐近正态，其方差分别为 $\mathrm{var}(a)$ 和 $\mathrm{var}(a) - \mathrm{cov}(a,\ c)^2/\mathrm{var}(c) \le \mathrm{var}(a)$，符号含义见下文。命题 5(ii) 中的表达式就是极限分布方差之间的不等式。若进一步假设两个标准化估计量的平方一致可积，那么该不等式也适用于估计量本身的方差（乘以 n），因为此时它们的二阶矩收敛于相应极限的二阶矩。假设 Aalen–Johansen估计值和评分函数校正项关于 $p_k(t)$ 和零联合渐近线性，其影响函数分别为 a 和 c，且方差有限。这些函数考虑了删失曲线和状态内均值的拟合。关于非马尔可夫律下 Aalen–Johansen估计值的推断，参见 Glidden (2002)。令 $\mu_j(s) = E\{h(s) \mid X \ge s,\ Y(s-) = j\}$。若在平均鞅积分中用极限值替换拟合均值和权重所产生的误差为 $o_p(n^{-1/2})$，则

$$c = \int_0^t \frac{h(s) - \mu_{Y(s-)}(s)}{G(s)}\,\mathrm{d}M^C(s;G).$$

对于删失时间构成的固定网格，若风险集均为正，则可通过展开经验比率得出这一结果。固定 β 时，增广估计的影响函数为 $a + \beta c$。若 $\mathrm{var}(c) > 0$，则最小化可得

$$\beta^* = -\frac{\mathrm{cov}(a,c)}{\mathrm{var}(c)}, \qquad \mathrm{var}\{a + \beta^* c\} = \mathrm{var}(a) - \frac{\mathrm{cov}(a,c)^2}{\mathrm{var}(c)}.$$

如果 $\mathrm{var}(c) = 0$，则我们设定 $\beta = 0$。Rubin 和 van der Laan（2008年）以 Kaplan–Meier 作为基准，采用方差最小化方法；Parast 等人（2014年）则将生存估计量投影到删失鞅上。

为了估计 β^* 而无需重新拟合，我们计算这两个矩。其中， a 是经中心化的逆概率删失加权贡献，再加上将 $P_{Y(s-)\,k}(s,t)/G(s)$ 关于 $M^C(s;\,G)$ 积分所得的结果。后者与c不相关，因为 $P_{Y(s-)\,k}(s,t)$ 仅取决于当前状态，而 $h(s)-\mu_Y(s-)(s)$ 在每个状态内的均值均为零。根据附录 C 中乘数为 $\{h(s)-\mu_{Y(s-)}(s)\}/G(s)$ 时的协方差计算，并且由于 $Q_{Y(s-)\,k}\{s,t;\ \mathcal{H}(s-)\}$ 是 $I\{Y(t)=k\}$ 在给定历史信息下的条件均值，

$$\operatorname{cov}(a,c)=-\int_0^t E[I(D\ge s)\{h(s)-\mu_{Y(s-)}(s)\}I\{Y(t)=k\}]\frac{\mathrm{d}\Lambda_C(s)}{G(s)},$$
$$\operatorname{var}(c)=\int_0^t E[I(D\ge s)\{h(s)-\mu_{Y(s-)}(s)\}^2]\frac{\mathrm{d}\Lambda_C(s)}{G(s)}.$$

因此， β^* 是将 $I\{Y(t)=k\}$ 对中心化评分函数进行回归所得的斜率，并在各个删失时间上汇总。在分子中，期望等于 $I(X\ge s)\{h(s)-\mu_{Y(s-)}(s)\}^2/G(s-)$ 的期望；在分母中，则等于$I\{C\ge D\wedge t,\,Y(t)=k\}G(s-)/G\{D\wedge t-\}$ 的期望。将 $\mathrm{d}\Lambda_C(s)$ 替换为 Nelson–Aalen 增量$r(s)^{-1}\sum_{l=1}^n \mathrm{d}N_l^C(s)$，可得

$$\widehat{\beta}=\frac{\sum_{l=1}^n I(C_l<D_l\wedge t)\,\omega(C_l)\sum_{i=1}^n I(X_i\ge C_l)\,e_i(C_l)\,o_i(C_l)}{\sum_{l=1}^n I(C_l<D_l\wedge t)\,\omega(C_l)\sum_{i=1}^n I(X_i\ge C_l)\,e_i(C_l)^2},$$

其中， $e_i(s)=h_i(s)-h_{Y_i(s-)}(s)$是回归变量， $o_i(s)=I\{C_i\ge D_i\wedge t,\,Y_i(t)=k\}G(s-)/G\{(D_i\wedge t)-\}$是响应变量， $\omega(s)=\{r(s)\,\widehat{G}(s)\widehat{G}(s-)\}^{-1}$，而 $r(s)=\sum_{i=1}^n I(X_i\ge s)$。这就是正文第 4 节所述的加权最小二乘斜率。如果这些均值在$s\le t$上统一收敛，那么 $\widehat{\beta}$ 是 β^*的一致估计。由于 $\widehat{c}_h(t)=O_p(n^{-1/2})$，估计该斜率会使估计值改变 $o_p(n^{-1/2})$。对于基于同一组结局训练得到的评分函数，其拟合过程也必须纳入联合线性化。

假设在附录 C 的独立删失和一阶替换条件下，评分函数 $h_i(s)$ 是给定保留当前状态的历史分组时，在 t 时处于状态 k 的真实概率。其在状态 j 内的均值为 $q_{jk}(s,\ t)$，因此 c 使用对比项 $h_i(s)-q_{jk}(s,\ t)$。在附录 C 的删失内积下，该对比项与任何当前状态预测正交；将完整历史概率 $Q_{Y(s-)\,k}\{s,t;\mathcal{H}(s-)\}$ 对该分组取条件期望后，得到 $h_i(s)$。因此，即使工作用的当前状态预测有误，$\operatorname{cov}(a,\ c)=-\operatorname{var}(c)$；而对于 $\operatorname{var}(c)>0$，最优系数为 1，并且在 $n^{1/2}$ 尺度上，方差降低量为 $\operatorname{var}(c)$。

E 数值实验

E.1 疾病–死亡模型

我们将采用 Kaplan–Meier 权重的逆概率删失加权估计量 $\widehat{p}_k^{\mathrm{IPCW}}(t;G)$ 与 Aalen–Johansen 估计量 $\widehat{p}_k(t)$ 进行比较，每种配置重复试验 5000 次。第一个实验采用正文图 1(c) 中的疾病 – 死亡模型，其中状态 0（缓解）、状态 1（复发）和状态 2（死亡）。受试者从状态 0 开始，从该状态出发的 $0\to1$ 和 $0\to2$ 转移具有恒定风险率 1 和 0.3。状态 1 在死亡前的驻留时间服从尺度参数为 1、形状参数为 κ的威布尔分布，因此当 $\kappa=1$ 时该过程为马尔可夫过程，否则为半马尔可夫过程；对于 $\kappa<1$，其 $1\to2$ 风险率随复发后的时间增加而降低，而对于 $\kappa>1$则随之增加。删失时间服从风险率为 0.3 或 0.8 的指数分布。目标是在 $t=1.5$时的 $p_1(t)$，在 κ 的四个取值下，该目标介于 0.26 和 0.27 之间。

^表 A1 显示， $\widehat{p}_1^{\text{IPCW}}(t;G)$ 和 $\widehat{p}_1(t)$ 均基本无偏，且前者的方差从未小于后者。方差增幅最高达到 17%，并且在重度删失条件下以及对于 $\kappa \leq 1$ 时最大。

表 A1：疾病–死亡模型： $\widehat{p}_1^{\text{IPCW}}(t;\widehat{G})$、 $\widehat{p}_1(t)$ 和 $\widehat{p}_1^{\text{aug}}(t)$ 在$t = 1.5$时的偏倚和标准差，乘以 10^3

			Bias			SD			比值	
κ	删失	n	$\widehat{p}_1^{\text{IPCW}}$	$\widehat{p}_1$	$\widehat{p}_1^{\text{aug}}$	$\widehat{p}_1^{\text{IPCW}}$	$\widehat{p}_1$	$\widehat{p}_1^{\text{aug}}$	$\widehat{p}_1^{\text{IPCW}}$	$\widehat{p}_1^{\text{aug}}$
0.5	25	200	0.5	0.3	−0.3	36.9	35.9	36.0	1.06	1.00
0.5	25	400	0.0	0.0	−0.2	26.3	25.5	25.5	1.06	1.00
0.5	50	200	0.2	−0.2	−3.2	50.8	47.2	47.3	1.16	1.01
0.5	50	400	0.0	0.0	−1.4	36.2	33.5	33.5	1.17	1.00
1	26	200	0.5	0.4	−0.1	37.3	36.6	36.6	1.04	1.00
1	26	400	−0.1	0.0	−0.3	26.5	26.0	26.0	1.04	1.00
1	53	200	0.5	0.5	−2.1	51.9	49.4	49.4	1.10	1.00
1	53	400	0.0	0.2	−0.9	36.7	34.8	34.8	1.11	1.00
5	29	200	0.3	0.4	0.3	37.5	37.3	36.6	1.01	0.97
5	29	400	0.6	0.7	0.6	26.8	26.7	26.1	1.00	0.95
5	58	200	0.0	0.4	−0.5	52.4	51.4	49.6	1.04	0.93
5	58	400	1.0	1.3	0.7	37.9	37.2	35.5	1.04	0.91
10	29	200	0.3	0.4	0.3	38.0	37.9	37.0	1.01	0.95
10	29	400	0.2	0.3	0.3	27.0	27.0	26.1	1.00	0.93
10	59	200	−0.3	0.2	−0.6	53.0	52.3	49.5	1.03	0.90
10	59	400	0.6	0.8	0.3	38.3	37.7	35.4	1.03	0.88

标准差，标准差 n；删失，在死亡和 t 中较早者之前被删失的受试者百分比；比值，方差 e
除以 $\widehat{p}_1($ $t)$.

表 A1 还报告了正文第 4 节中的增广估计量 $\widehat{p}_1^{\text{aug}}(t)$，其中以在状态 1 中度过的^时间作为评分函数 $h_i(s)$，并用附录 D 的最小二乘公式计算 β。对于 $\kappa \geq 5$，其方差为 $\widehat{p}_1(t)$ 的 0.88 至 0.97 倍，此时 $1 \to 2$ 风险率随该时间急剧上升；而对于 $\kappa \leq 1$，其方差为 $\widehat{p}_1(t)$ 的 1.00 至 1.01 倍，此时对持续时间的依赖较弱或不存在。其偏倚相对于标准差很小，并且随 n 增加而减小。为增强对持续时间的依赖，我们将 $0 \to 1$ 风险率提高至 3，使受试者更早进入状态 1 并在其中停留更久，并取 $\kappa = 5$ 或 10，此时 $p_1(t) = 0.16$ 或 0.15。表 A2 显示， $\widehat{p}_1^{\text{aug}}(t)$ 的方差此时为 $\widehat{p}_1(t)$ 的 0.84 至 0.96 倍；在重度删失情况下， $\kappa = 10$ 处的降幅最大，为 16%。综合表 A1 和表 A2， $\kappa \geq 5$ 的降幅为 3% 至 16%。

E.2 渐进模型

第二个实验使用图 A1 的渐进模型，其中状态 0（缓解）、1（复发）、2（转移）和 3（死亡）。受试者从状态 0 开始。 $0 \to 1$ 和 $0 \to 3$ 转移的风险率分别为恒定的 1 和 0.1， $1 \to 2$ 和 $1 \to 3$ 转移的风险率分别为 0.8 和 0.15，而 $2 \to 3$ 转移的风险率为 0.8。如果复发发生在时间 0.5 之前，则所有后续风险率均乘以 e^{θ}，因此该过程仅对 $\theta = 0$而言是马尔可夫过程。删失时间服从风险率为 0.3 或 0.6 的指数分布，目标是在 $t = 2$时的 $p_k(t)$ $(k = 0, 1, 2, 3)$。

^表 A2：风险率为 $0 \to 1$ 3 的疾病–死亡模型： $t = 1.5$时 $\widehat{p}_1^{\text{IPCW}}(t; G)$、$\widehat{p}_1(t)$ 和 $\widehat{p}_1^{\text{aug}}(t)$ 的偏倚和标准差，乘以 10^3

				Bias			SD			比值	
κ	删失		n	$\widehat{p}_1^{\text{IPCW}}$	$\widehat{p}_1$	$\widehat{p}_1^{\text{aug}}$	$\widehat{p}_1^{\text{IPCW}}$	$\widehat{p}_1$	$\widehat{p}_1^{\text{aug}}$	$\widehat{p}_1^{\text{IPCW}}$	$\widehat{p}_1^{\text{aug}}$
5		27	200	−0.4	−0.4	−0.6	31.7	31.6	30.8	1.00	0.95
5		27	400	0.2	0.2	0.0	21.9	21.8	21.4	1.01	0.96
5		56	200	−0.2	−0.6	−2.4	44.3	44.0	41.2	1.02	0.88
5		56	400	0.8	0.6	−0.3	30.5	30.3	28.5	1.02	0.88
10		28	200	−0.2	−0.2	−0.4	31.1	31.0	30.0	1.00	0.93
10		28	400	0.2	0.2	0.0	21.5	21.4	20.8	1.01	0.94
10		57	200	0.5	−0.1	−1.9	43.9	43.5	39.9	1.02	0.84
10		57	400	0.2	0.0	−0.8	30.1	29.8	27.3	1.02	0.84

标准差，标准差；删失，在死亡和 t中较早者之前被删失的受试者百分比；比值，方差与 $\widehat{p}_1(t)$ 方差之比。

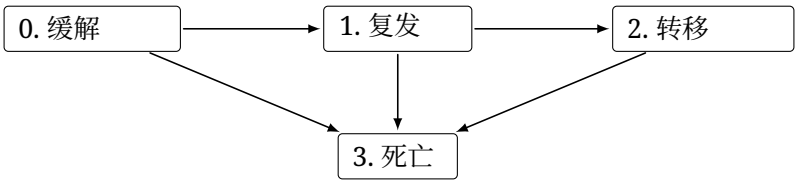

图 A1：渐进模型。

两种估计量的绝对偏倚均不超过 0.002。表 A3 显示，对于瞬态状态 $k = 0,1,2$，$\sim\widehat{p}_k^{\text{IPCW}}(t;\ G)$ 的方差比 $\widehat{p}_k(t)$ 高 8% 至 34%；对于 $k = 3$，则高 2% 至 12%。这些结果与附录 C 一致：两种估计量都通过忽略先前历史信息的预测进行增广，但 $\widehat{p}_k^{\text{IPCW}}(t; G)$ 还忽略了当前状态。

表格 A3：渐进模型：在 $t = 2$ 时，$\widehat{p}_k^{\mathrm{IPCW}}(t; G)$ 的方差与 $\widehat{p}_k(t)$ 的方差之比

θ	删失	n	$k = 0$	$k = 1$	$k = 2$	$k = 3$
0	38	200	1.20	1.11	1.10	1.02
0	38	400	1.18	1.10	1.10	1.02
0	61	200	1.34	1.21	1.19	1.04
0	61	400	1.34	1.20	1.18	1.04
1.5	31	200	1.18	1.09	1.09	1.06
1.5	31	400	1.16	1.09	1.09	1.06
1.5	51	200	1.29	1.18	1.17	1.10
1.5	51	400	1.30	1.18	1.15	1.12
2.5	29	200	1.17	1.08	1.09	1.02
2.5	29	400	1.15	1.08	1.10	1.03
2.5	46	200	1.28	1.16	1.19	1.04
2.5	46	400	1.29	1.16	1.17	1.06

删失，即在死亡和 t 两者中较早发生者之前被删失的受试者百分比。