arXiv:2302.00925v1 [stat.ME] 2 Feb 2023

评估模型对复发事件期望累积次数的预测性能

Olivier Bouaziz

巴黎西岱大学，法国国家科学研究中心，MAP5 UMR 8145，法国巴黎 F-75006

摘要

在复发事件场景中，我们提出一种新评分，用于评估给定模型对复发事件期望累积次数的预测能力。该评分能够通过患者的外部协变量考虑其个体病史，可视为单次生存时间数据布里尔评分的扩展，但适用于有或无终末事件的复发事件。我们给出了理论结果，表明在复发事件情境下，满足标准假设时，该评分渐近地可分解为模型与真实复发事件期望累积次数之间的理论均方误差，以及一个不依赖于模型的不可分离项。我们还通过模拟研究进一步展示了这种分解。此外，我们证明该评分应与参考模型进行比较，例如不包含协变量的非参数估计量。最后，我们将该评分应用于预测一组心房颤动患者数据中的住院情况，并考察了不同模型（如Cox模型或Aalen模型）的预测性能。

关键词：复发事件；预测评估；右删失；终末事件；布里尔评分。

1 引言

复发事件数据常见于随访研究中。它们可以被视为标准生存时间数据的推广，其中个体可能会随时间反复经历同一事件。典型例子包括艾滋病病毒研究，在这类研究中患者可能反复发生机会性感染；白血病患者的缓解数据，这些患者可能经历多次复发；癫痫患者的反复癫痫发作；或者住院数据，其中所关注的事件就是住院。在这些研究中，重点可能是评估协变量对复发风险的影响，或预测未来的复发。最早用于处理复发事件数据的模型是 Andersen-Gill模型 [3] ，该模型随后由 [23] 进一步扩展，以考虑复发事件过程的跳跃可能存在依赖性。此后又发展出其他模型，例如 [25]，[7]，[14]，[13] 或 [1] 中的模型，其中后四篇论文在估计过程中纳入了终末事件；也有研究使用随机效应，例如 [19]，[24], 或 [26]。特别是，在 [7]，[14]，[13]，[1]中，作者重点估计了复发事件的期望累积次数。这是一个边际量，用于计算个体在任意时间点之前所经历事件次数的期望值。该数量尤其重要，因为它概括了复发事件过程随时间的演变。在存在终末事件的情况下，它还体现了这样一个事实：终末事件发生后，患者将不再经历任何进一步的复发事件。

在某些研究中，重点更多放在模型的预测能力上，而非协变量效应的解释上。当临床医生旨在预测

未来反复发生的事件，以便提供最佳医疗护理时，情况便是如此。能够在短时间内预测任何患者未来的复发情况，也有助于预测疾病在患者一生中造成的未来负担。此外，预测模型既可作为制定医疗决策的重要工具，也可用于与患者沟通其疾病未来的发展情况。例如，在 [29]中，作者研究了患有心房颤动（一种众所周知的心脏病）的患者，试图预测患者因该疾病而未来住院的情况。由于患有这种疾病的患者通常年纪较大（研究中的年龄中位数为63岁），且心房颤动在某些情况下可能是一种严重疾病，因此这些患者也面临死亡风险。研究收集了多个协变量，并使用依赖于既往计数的Cox模型，对复发事件随时间变化的预期累积数量进行了预测。

虽然这类模型无疑受到临床医生的关注，但提出相关的诊断工具以评估所提出模型的预测性能也很重要。针对每个个体仅发生一次事件的标准生存时间数据情境，目前已有若干预测性能指标。布里尔评分由 [15] 和 [12]，提出，基本上是一种在存在删失的情况下计算生存时间均方误差的评分。随后，该评分被进一步扩展，用于处理 [31]，中的随机效应模型，或评估 [28] 中的动态预测模型性能；在这类模型中，纵向协变量所提供的信息会在每个时间点更新。还需注意，其他类型的预测准确性度量也已出现，称为区分度度量，例如 C指数（见 [17]，[11]），以及时间依赖性 ROC 曲线及其曲线下面积（例如见 [18]）。

本文旨在为复发事件模型推导一种预测准确性度量，重点关注预测能力而非区分度。我们唯一关心的量是复发事件的预期累积次数。由于在复发事件情境下尚无布里尔评分这类均方误差度量，因此本研究旨在填补这一空白，推导一种适用于复发事件的新型同类评分，同时兼顾终末事件的存在。本文表明，当每个个体最多只能发生一次事件时，该评分会退化为布里尔评分，因此可视为标准布里尔评分的直接推广。此外，由于我们的预测准则关注复发事件预期累积次数这一边际量，因此它提供了一个汇总评分，能够综合考虑对所有复发事件的预测。在存在终末事件的情境下，该评分还纳入了对终末事件的预测质量。

在第 2.1 节中，我们介绍复发事件的一般预测准则，记为 $\widehat{\text{MSE}}$。在第2.2节和第2.3节中，我们分别讨论未观测到终末事件以及存在终末事件时的建模假设。在第 2.4 节中，我们介绍一些已有的复发事件期望累积次数估计量。在第3节中，我们推导本文的主要理论结果。首先，我们引入一个理论准则，并证明它可以分解为不可分离项和不精确项，这与 [12] 中的结果类似。前者与模型无关，且无法消除；后者则恰好是复发事件过程与复发事件期望累积平均数预测模型之间的均方误差。随后，我们证明预测准则渐近收敛于理论准则。在第 4 节中，我们证明，当个体最多只能发生一次事件时，我们的预测准则等价于标准布里尔评分。在第 5.1 节中，我们开展了一项模拟研究。首先，我们展示不可分离项与不精确项的分解。由于不可分离项占主导地位，我们因此建议将所考虑模型与参考模型之间的均方误差之差作为预测 $\widehat{}$ 评分。其次，我们展示如何利用这一预测评分比较预测模型。在第 6 节中，我们研究心房颤动数据集。我们证明，依赖于既往计数的模型（即多状态模型），并按心房颤动类型进行分层，

在所考虑的所有其他模型中提供了最佳的预测性能。

2 复发事件期望累积次数的预测准则

2.1 一般框架中的预测准则

本节介绍右删失条件下两种不同复发事件场景的预测准则。我们考察仅存在右删失的情形（第 2.2 节）以及包含终末事件的情形（第 2.3 节）。在每种情形中，我们定义一个关注的计数过程 $N^*(t)$，用于统计时间 t 之前发生的复发事件次数。我们假设观测到一个多变量外部时变协变量向量 $X(t)$（外部协变量的定义见 [21] ），并定义 $\mathcal{M}$ 一类依赖于 t 和 $X(t)$ 的有界函数。对于每个 $t \geq 0$，我们定义 $\mathcal{X}_t$，即过程$\{X(u)$ 在时间点 $u = t$ 处的支撑集：$0 \leq u\}$ 。在两种情形中，都可能发生删失，因此观测到的复发事件过程为 $N(t) := N^*(t \wedge C)$，其中 C 是删失变量，$a \wedge b$ 表示 a 和 b 中的较小值。我们还定义 τ 研究终点。基于独立同分布重复样本（$N_i(t)$ $X_i(t)$：$0 \leq t \leq \tau$），我们假设存在一个估计量 $\widehat{\mu} \in \mathcal{M}$，用于估计复发事件的期望累积次数 $\mu^*(t \mid X(t)) := \mathbb{E}[N^*(t) \mid X(u)$：$0 \leq u \leq t]$。如果存在 $\mu \in \mathcal{M}$，使得对所有$t \leq \tau$，我们就称该估计量是一致的，

$$\sup_{x \in \mathcal{X}_t} |\widehat{\mu}(t \mid x) - \mu(t \mid x)| \to 0, \text{ in probability as } n \to \infty.$$

本文的主要目标是提出一种新的均方误差准则，用于评估该估计量在不同情景下的性能。

在第2.2节和第2.3节中，我们提出通过以下准则评估给定估计量$\widehat{\mu} \in \mathcal{M}$的预测能力：

$$\widehat{\mathrm{MSE}}(t, \widehat{\mu}) = \frac{1}{n} \sum_{i=1}^{n} \left(\int_0^t \frac{dN_i(u)}{1 - \hat{G}(u- \mid X_i(u))} - \widehat{\mu}(t \mid X_i(t)) \right)^2, \tag{1}$$

ˆ其中，G 是 G的估计量，即给定 $X(\cdot)$时删失ˆ ˆ变量 C 的条件累积分布函数。记号 $G(u- \mid X_i(u))$表示函数 G 在u处的左极限。接下来，我们将以下述方式假设该删失估计量的一致收敛性。

ˆ**假设1** 设 $\mathcal{G}$ 为条件删失分布的一个模型。若对所有 $t \leq \tau$，均有……，则称 G 是 $G \in \mathcal{G}$ 的一致相合估计量。

$$\sup_{x \in \mathcal{X}_t} |\hat{G}(t \mid x) - G(t \mid x)| \to 0, \ \textit{in probability as } n \to \infty.$$

针对不同的研究情境，G 的不同估计量将在第2.2节和第2.3节中讨论。

现在我们引入一个在已知删失分布时可用的理论准则。对于某个函数 $\mu \in \mathcal{M}$，令：

$$\mathrm{MSE}(t, \mu) = \mathbb{E}\left[\left(\int_0^t \frac{dN(u)}{1 - G(u- \mid X(u))} - \mu(t \mid X(t)) \right)^2 \right]. \tag{2}$$

我们的准则 (1) 背后的关键思想源于这样一个事实：

$$\mathbb{E}\left[\int_0^t \frac{dN(u)}{(1 - G(u- \mid X(u)))} \right] = \mathbb{E}[\mu^*(t \mid X(t))], \tag{3}$$

这一关系已在第2.2节和第2.3节中针对每种情形分别得到证明。在第3节中，我们针对这两种情形给出了理论结果，以证明所提出准则的合理性。特别是，第3节的命题 1 表明，理论准则可以分解为以下形式：

$$\mathrm{MSE}(t,\mu) = \mathbb{E}\left[\left(\mu^*(t \mid X(t)) - \mu(t \mid X(t))\right)^2\right] + A(t),$$

其中 $A(t)$ 不依赖于 μ。第一项是不精确项，第二项是不可分性（或残差）项，与所选模型无关。值得注意的是，这类结果类似于布里尔评分的不精确性/不可分性分解（见 [12]）。另一方面，第3节的命题2指出，如果 $\widehat{\mu}$ 是某个 $\mu \in \mathcal{M}$ 的相合估计量，那么当 n 趋于无穷时，我们的经验准则 (1) 渐近等价于在 μ处评估的理论准则 (2)。接着，当满足命题 1 中的 $\mu = \mu^*$ 时，后一个准则将达到最小值。如果 μ 不等于 μ^*，则会产生偏差，其大小等于 μ 与 μ^*之间期望差的平方。命题 1 和命题 2 成立所需的具体假设在各个情形中分别给出。

下文将介绍两种情形：仅存在右删失，以及纳入终末事件。值得注意的是，估计量 μ^* 基于率函数，该函数的定义见 [23] 或 [6]。在这种建模方法中，条件化基于协变量；但与强度函数不同，我们不以过程的完整历史为条件。因此，我们的准则 (1) 适用于一般的复发事件过程，尤其不局限于泊松过程。

2.2 存在右删失且无终末事件的情形

与前文一样，我们引入感兴趣的计数过程 $N^*(t)$，用于统计时间 t 之前发生的复发事件次数。设 $X(t)$ 为某个多变量外部协变量（见 [21]），且允许其依赖于 t。对于每个 $t \geq 0$，我们定义 $\mathcal{X}_t$，即过程 $\{X(u)$ 的支撑集：$0 \leq u\}$ 在时间点 $u = t$。此处考虑的回归建模方法定义如下：

$$\mathbb{E}[dN^*(t) \mid X(t)] = \lambda^*(t \mid X(t))dt, \tag{4}$$

其中，$\lambda^*(t \mid X(t))$ 为真实率函数。请注意，此定义非常一般，并未对复发事件过程作出任何假设，例如独立增量或泊松假设。有关该模型的更多详情，请参见 [23] 。

设 $\int_0^t \lambda^*(u \mid X(u))du$ 为真实累积率函数。在不存在终末事件的情况下，该累积函数可直接解释为给定协变量过程时复发事件的期望累积次数：$\int_0^t \lambda^*(u \mid X(u))du = \mu^*(t \mid X(t))$，其中，按照第 2 节开头的定义，$\mu^*(t \mid X(t)) = \mathbb{E}[N^*(t) \mid X(u)$：$0 \leq u \leq t]$。存在删失时，会观测到一个变量 C，因此观测到的复发事件过程现在为$N(t) = N^*(t \wedge C)$。我们按以下方式假设独立删失（见 [23]）：

$$\mathbb{E}[dN^*(t) \mid X(t)] = \mathbb{E}[dN^*(t) \mid C, X(t)].$$

这一假设意味着，C 不会提供有关复发事件过程发生跳跃概率的任何额外信息。在此假设下，我们有

$$\mathbb{E}[dN(t) \mid I(C \geq t), X(t)] = I(C \geq t)\lambda^*(t \mid X(t))dt, \tag{5}$$

其中，$I(\cdot)$ 表示示性函数。最后这个等式说明了采用我们的准则 (1) 是合理的，因为 $\mathbb{E}[dN(t) \mid X(t)] = (1-G(t- \mid X(t)))\lambda^*(t \mid X(t))dt$，因此等式 (3) 成立。接下来，我们假设假设1，并作出以下附加假设。

假设 **假设 2** 我们假设存在一个常数 $\tau > 0$ 和一个常数 $c > 0$，使得 t

1. $\forall t \in [0, \tau]$, $\mathbb{P}[C \geq t \mid X(t)] \geq c$ *almost surely,*

2. $N(\tau)$ 几乎必然有界于某个常数。

请注意，条件*1.* 在 [12]中也作了假设。它比仅假设 $\mathbb{P}[C \geq \tau] \geq c$更强。条件*2.* 是复发事件数据中的标准条件，参见例如 [23]。最后请注意，通过方程（5），条件*1.* 和*2.* 意味着 $\mathbb{E}[\mu^*(\tau \mid X(\tau))] < \infty$。

基于独立同分布重复样本（$N_i(t)$ $X_i(t)$: $0 \leq t \leq \tau$），令 $\widehat{\mu} \in \mathcal{M}$ 为 μ^* 的一个估计量，其中 $\mathcal{M}$ 是一个假定有界的模型类。我们建议通过方程（1）中定义的准则均方误差$(t, \widehat{\mu})$来评估该估计量的预测能力$\widehat{}$。该准则涉及 G的一个估计量，即删失变量的条件累积分布函数。若 C 与 $X(\cdot)$相互独立，则由于这些变量均可观测，可以利用删失变量的经验累积分布函数估计 G 。若 C 依赖于 X ，则还必须对 C 的条件分布进行建模。这可以使用基于核的估计量来实现，例如针对二元响应变量 $I(C \leq t)$的Nadaraya-Watson估计量，或该模型的扩展。例如，[16], 中提出了局部逻辑回归方法和调整后的Nadaraya-Watson估计量。如果假定依赖于删失分布的协变量维数过高，则可先采用降维技术，例如使用单指标建模方法（参见 [10]）。

关于该准则有效性的理论结果将在第3节中推导。

2.3 存在右删失和终末事件的情况

本节中，我们引入一个终末事件 T^*，使得我们关注的计数过程 N^*满足（例如参见 [6]）：

$$\mathbb{E}[dN^*(t) \mid I(T^* \geq t), X(t)] = I(T^* \geq t)\lambda^*(t \mid X(t))dt, \tag{6}$$

其中 $\lambda^*(t \mid X(t))$ 是真实的率函数。模型（6）与模型（4）的区别在于，复发事件过程 N^* 会因终末事件而停止。在真实数据分析中，这种情况经常发生，其中死亡通常是终末事件。在该模型下，我们观察到

$$\mu^*(t \mid X(t)) = \mathbb{E}[N^*(t) \mid X(u) : 0 \leq u \leq t] = \int_0^t S(u \mid X(u))\lambda^*(u \mid X(u))du,$$

其中 $S(t \mid X(t)) := \mathbb{P}[T^* \geq t \mid X(t)]$ 是终末事件的条件生存函数。这意味着，为了定义 $\mu^*(t \mid X(t)) = \mathbb{E}[N^*(t) \mid X(u) : 0 \leq u \leq t]$的估计量，通常还需要对终末事件的风险率进行建模，并推导条件生存函数的估计量。因此，我们的预测准则将同时考虑生存函数的预测性能以及 N^*的率函数的预测性能，因为如果这两个估计量中的任何一个表现不佳，那么由此得到的 μ^* 估计量很可能也会表现不佳。

现在将独立删失假设表述如下：

$$\mathbb{E}[dN^*(t) \mid T^*, X(t)] = \mathbb{E}[dN^*(t) \mid T^*, C, X(t)].$$

我们用 $T = T^* \wedge C$ 表示终末事件和删失时间中的较小者，用 $Y(t) = I(T \geq t)$ 表示观测到的风险过程，用 $N(t) = N^*(T \wedge t)$ 表示观测到的计数过程。在独立删失假设下，可以证明：

$$\mathbb{E}[dN(t) \mid Y(t), X(t)] = Y(t)\lambda^*(t \mid X(t))dt. \tag{7}$$

我们假设假设1成立，并作出以下额外假设。

假设 **第3条**我们假设存在常数 $\tau > 0$ 和常数 $c > 0$，使得 t

1. $\forall t \in [0, \tau]$, $\mathbb{P}[T \geq t \mid X(t)] \geq c$ 几乎必然成立，
2. $N(\tau)$ 几乎必然被某个常数所界定。

我们还假设，在给定 $X(\cdot)$的条件下， T^* 与 C 相互独立。

这些条件在含终末事件的复发事件回归研究中是标准的，例如参见 [14] 。利用等式（7），可以很容易地观察到，在假设3中的独立删失假设下， $\mathbb{E}[dN(t) \mid X(t)] = S(t \mid X(t))(1 - G(t- \mid X(t))\lambda^*(t \mid X(t))dt$ 。由此可直接看出，方程（3）成立。

基于独立同分布重复样本（$N_i(t)$ $X_i(t)$: $0 \leq t \leq \tau$），令 $\widehat{\mu} \in \mathcal{M}$ 为μ^* 的估计量，其中 $\mathcal{M}$ 是一个假定有界的模型类。我们提出通过方程（1）中定义的准则$\text{MSE}(t, \widehat{\mu})$来评估该估计量对 $\widehat{\ }$的预测能力。该准则涉及 G的一个估计量，即删失变量的条件累积分布函数。若 C 与 $X(\cdot)$ 相互独立，则可将 C 视为因终末事件 T^*而未被完整观测的关注变量，并使用Kaplan-Meier估计量估计 G 。若 C 依赖于 X ，则还必须对 C 的条件分布进行建模。 [12] 介绍了几种可能的模型，例如Cox模型、Aalen加性模型或 [9]提出的核类型模型。或者，也可以采用针对右删失数据的单指标方法（如 [5], ），或使用 [20] 开发的随机生存森林方法。

关于该准则有效性的理论结果将在第3节推导。

2.4 复发事件的预期累积数量估计量示例

本节介绍一些预期累积平均次数的估计量。为实现这一目标，一种方法是先对率函数 λ^* 建模，然后使用代入估计量推导出 μ^*的最终估计量。例如，可以使用 Cox（见 [8]）、Aalen（见 [27]）或加速失效时间模型（见 [22]）来估计 λ^*。$\widehat{\ }$ $\widehat{\ }$我们注意到 λ 和 Λ 这类 λ 和 Λ的估计量。在右删失数据且不存在终末事件的情形下， μ^* 的估计量可表示如下（见 [6]）：

$$\widehat{\mu}(t \mid X(t)) = \int_0^t d\widehat{\Lambda}(u \mid X(u)). \tag{8}$$

存在终末事件时，一种常见方法是先对终末事件的风险率建模（例如再次使用 Cox模型），并推导出生存函数 $\widehat{S}(t \mid X(t)) = \exp(-\int_0^t \widehat{\lambda}^{T^*}(u \mid X(u))du)$ 的估计量，其中 $\widehat{\lambda}^{T^*}$ 是终末事件风险率的估计量。然后， μ^* 的最终估计量为（见 [6] 和 [1]）：

$$\widehat{\mu}(t \mid X(t)) = \int_0^t \widehat{S}(u \mid X(u)) d\widehat{\Lambda}(u \mid X(u)). \tag{9}$$

也存在直接对 μ^* 建模的其他方法。在 [14] 中，作者考虑以下 Cox 型模型： $\mu^*(t \mid X(t)) = \mu_0(t) \exp(X(t)'\beta)$。一种更一般的方法是使用单指标模型来估计 μ^*: 在 [4] 中，作者假设存在一个非参数函数 g 和一个参数 β ，使得 $\mu^*(t \mid X(t)) = g(X(t)'\beta)$。这两种方法都能直接估计感兴趣的量 μ^*。然而，当存在终末事件时，便无法再区分复发事件过程或终末事件所产生的影响。因此，解读回归参数时应谨慎。关于这一问题的讨论，另请参见 [6] 。

3 理论结果

本节针对先前讨论过的两种不同情境，对所提出的准则 (1) 给出理论结果：仅存在右删失的情境，以及同时存在右删失和终末事件的情境。我们得到两个结果。第一个结果涉及理论准则 (2)，表明将该准则应用于函数$\mu \in \mathcal{M}$ 后，会化简为 μ 与 μ^* 之间的均方误差，以及一个不依赖于 μ的项。第二个结果表明，当估量 $\hat{\mu}$ 是 μ的渐近相合估量时，MSE(t, $\hat{\mu}$) 与 MSE(t, μ) 渐近一致。

命题 1 对于仅存在右删失的情形（第 *2.2* 节），以及同时存在右删失和终末事件的情形（第 *2.3* 节），我们分别假设假设*2*和假设*3*。在这两种情形下，我们还假设独立删失。于是，对于 $\mu \in \mathcal{M}$,

$$\mathrm{MSE}(t, \mu) = \mathbb{E}\left[\Big(\mu^*(t \mid X(t)) - \mu(t \mid X(t))\Big)^2\right] + A(t), \tag{10}$$

其中，对于所有 $t \geq 0$， $A(t) \geq 0$，且 $A(t)$ 不依赖于 μ。

命题2 对于仅存在右删失的情形（第 *2.2* 节），以及同时存在右删失和终末事件的情形（第 *2.3* 节），我们分别假设假设*2*或假设*3*。在这两种情形下，我们还假设假设*1*和独立删失。于是，若估计量$\widehat{\mu} \in \mathcal{M}$ 是 $\mu \in \mathcal{M}$的一致估计量，则有

$$\sup_{t \leq \tau} \left| \widehat{\mathrm{MSE}}(t, \hat{\mu}) - \mathrm{MSE}(t, \mu) \right| \to 0, \quad \textit{in probability as } n \to \infty.$$

证明见附录第8.1节和第8.2节，其中给出了两种情形下 $A(t)$ 的显式表达式。

4 与布里尔评分的联系

布里尔评分（见 [12]）是一种常用准则，用于在右删失情形下评估回归模型对条件生存函数的预测性能，此时每个个体最多只能观察到一次事件。本节将说明，当个体最多只能经历一次事件时，如果采用我们的准则，那么记为 MSE′ 的理论准则定义如下

$$\mathrm{MSE}'(t, \pi) = \mathbb{E}\left[\left(1 - \int_0^t \frac{dN(u)}{1 - G(u- \mid X(u))} - \pi(t \mid X(t))\right)^2\right], \tag{11}$$

其中 $\pi = 1 - \mu$，该准则除去一个与模型 $\mathcal{M}$无关的项后，等同于理论布里尔评分。注意，当个体最多只能经历一次事件时，复发事件过程可简化为 $N^*(t) = I(T^* \leq t)$，而 $\mu^*(t \mid X) = \mathbb{E}[N^*(t) \mid X]$是 T^*的条件累积分布函数。由于布里尔评分是为预测条件生存函数而设计的，因此我们只需改写公式（2）中的准则，使π 表示条件生存函数的模型。

我们首先回顾一下，理论布里尔评分的定义如下（参见 [12] 中的公式（1））：

$$\mathrm{MSE}^{\mathrm{Brier}}(t, \pi) = \mathbb{E}\left[(I(T^* > s) - \pi(t \mid X))^2\right],$$

其中，期望是相对于 T^* 和 X的联合分布而言的。为简单起见，如 [12], 的论文所述，公式中的协变量 X 不随时间变化，但本节所给出的结果对于时变协变量仍然成立。在他们的研究中，

当目标是预测生存函数 $S(t \mid X) = \mathbb{P}[T^* > t \mid X]$时，作者证明了与本文命题1和命题2类似的结果。请注意，为了与全文使用的符号保持一致，我们在布里尔评分的定义中省略了对 S 的依赖。此外，在布里尔评分的定义中，π 扮演着本文中 $1-\mu$ 的角色，也就是说，它是一个拟议的条件生存估计量 $S(t \mid X)$的极限函数。下面给出相应的结果。

命题 3 我们假设每个个体最多只能发生一次事件，即$N(t) = I(T \le t, \Delta = 1)$，其中 $T = T^* \wedge C$是观测时间，T^*是真实事件时间，C 是删失变量，$\Delta = I(T^* \le C)$ 是删失指示变量。那么，在独立删失条件下，我们有：

$$\mathrm{MSE}'(t,\pi) = \mathrm{MSE}^{\mathrm{Brier}}(t,\pi) + B(t),$$

其中，对于所有 $t \ge 0$，$B(t) \ge 0$ 均成立，且 $B(t)$ 不依赖于 π。

证明见附录第 8.3 节，其中给出了 $B(t)$的显式表达式。由于 $B(t)$不依赖于模型 π，因此这两个准则完全等价。特别地，若将某个回归模型与参考模型进行比较，例如不包含协变量的模型，那么这两个准则会给出完全相同的数值，因为在作差时，$B(t)$项会抵消。正如我们将在下一节看到的，在实践中，我们通常建议将模型与参考模型进行比较。由于复发事件是每个个体发生单次事件情形的推广，因此我们的准则可以视为复发事件的布里尔评分的扩展。

5 模拟

5.1 存在右删失且没有终末事件的情景

对于 $i = 1, \ldots, n$，我们首先模拟一个二维协变量向量 $X_i = (X_{i,1}, X_{i,2})^\top$，其中$X_{i,1}$ 是参数为 0.5 的伯努利变量，而 $X_{i,2}$ 是期望为 2、标准差为 0.5 的高斯变量。给定 X_i，复发事件由速率为 $\lambda^*(\cdot \mid X_i)$ 的非齐次泊松过程生成，该过程遵循具有 Weibull基线和二维时间无关协变量的 Cox模型。更具体地说，

$$\lambda^*(t \mid X_i) = \lambda_0(t)\exp(\theta_0^\top X_i), \quad \lambda_0(t) = \frac{\alpha}{\beta}\left(\frac{t}{\beta}\right)^{\alpha-1},$$

其中 $\alpha = 2$ 为形状参数，$\beta = 0.39$ 为尺度参数，且 $\theta_0 = (\log(2), \log(0.5))^\top$。在此模拟设定下生成的复发事件符合公式 (4)，其真实预期复发事件数为

$$\mu^*(t \mid X_i) = \int_0^t \lambda^*(t \mid X_i)du = \left(\frac{t}{\beta}\right)^{\alpha} \exp(\theta_0^\top X_i). \tag{12}$$

我们还模拟了一个删失变量 C_i，其服从 $[0,3]$ 上的均匀分布。使用这些参数，我们观察到 30% 的个体发生 0 或 1 次复发事件，54% 的个体发生不超过 5 次事件，77% 的个体发生不超过 12 次事件。平均而言，每个个体约观察到 8 次复发事件。

基于单个模拟样本，我们首先说明命题1和命题2，其中模型类 $\mathcal{M}$ 包含 μ，并假设协变量对复发事件发生的影响为零。为此，我们分别模拟训练样本和测试样本。训练

样本用于根据公式 (8) 计算 $\hat{}\mu^{\text{train}}$，其中 $\hat{}\Lambda$ 不依赖于 X，且就是 Nelson-Aalen估计量：

$$\hat{\Lambda}(t) = \frac{1}{n}\sum_{i=1}^{n}\int_0^t \frac{dN_i(u)}{1-\hat{G}(u-)},$$

其中 $\hat{}G$ 是 C 的 Kaplan-Meier估计量。随后，我们根据大小为 n_{test} 的测试样本，利用公式 (1) 计算 $\widehat{\text{MSE}}(t,\ \hat{\mu}^{\text{train}})$；也就是说，计算基于一个独立于训练样本的样本 $(N_1(\cdot)\ X_1)\ldots,$ $(N_{n_{\text{test}}}(\cdot)\ X_{n_{\text{test}}})$ 进行。该量应能准确估计命题2中的 $\text{MSE}(t,\ \mu)$。随后，我们使用 $\mu^*(t \mid X(t))$ 的真实值，并以 $\hat{}\mu^{\text{train}}$ 替换 μ，同时以经验求和替代期望，从而利用公式 (10) 计算不精确项 $\mathbb{E}\left[\left(\mu^*(t \mid X)\ -\mu(t \mid X)\right)^2\right]$。换言之，我们计算

$$\frac{1}{n_{\text{test}}}\sum_{i=1}^{n_{\text{test}}}\left(\mu^*(t \mid X_i) - \hat{\mu}^{\text{train}}(t \mid X_i)\right)^2,$$

这应能非常准确地估计不精确项。$A(t)$ 项根据其显式表达式精确计算（见附录第 8.4 节）。图1使用$n_{\text{train}} = 200$ 和 $n_{\text{test}} = 1,\ 000$ 展示了均方误差在不精确项与不可分离项之间的分解。公式 (1) 中 G 的 Kaplan-Meier估计量由训练样本和测试样本合并计算得到。实线表示估计的均方误差，点线和虚线分别表示不可分离项和不精确项。可以看出，不可分离项与均方误差非常接近。相比之下，不精确项相对较小；由于这里估计的模型没有使用任何协变量，该项显然不为零。此图表明，比较不同模型可能较为困难，因为不可分离项在均方误差分解中占主导地位，这意味着基于两个不同模型计算出的均方误差往往会非常相似（例如，对于 $t = 2.5$，不可分离项约占均方误差的 84%）。因此，我们主张采用一个参照模型或零模型，并将给定模型的得分定义为参照模型的均方误差与该模型均方误差之差。因此，该得分表示相较于零模型，该模型所带来的预测增益。零模型通常选择不使用任何协变量的模型。得分定义为

$$\text{Score}(t, \widehat{\mu}, \widehat{\mu_0}) = \widehat{\text{MSE}}(t, \widehat{\mu_0}) - \widehat{\text{MSE}}(t, \widehat{\mu}), \tag{13}$$

其中，$\widehat{\mu}$ 是待评估模型，$\widehat{\mu_0}$ 是参考模型。这些模型通常基于训练样本构建。该评分的原理是，不可分离项会在差值中抵消，因此评分等于两个模型的不精确项之差。

图2展示了该评分的一个示例。使用与之前相同的模拟设定，我们比较基于Cox模型和Aalen模型的四种不同模型的性能，这些模型使用的协变量分别为仅第一个协变量$X_{i,\ 1}$ 或两个协变量$X_{i,\ 1}$ 和$X_{i,\ 2}$。图2展示了100个训练样本（每个样本量为50）以及一个样本量为1,000的独立测试样本所得到的预测评分。参考模型是不使用任何协变量、基于Nelson-Aalen估计量进行估计的模型。总体而言，随着时间推移，所有模型的预测性能都优于Nelson-Aalen估计量；对于分别使用一个和两个协变量的模型，这种优势尤其分别从时间等于1.5和时间等于2时开始显现。使用两个协变量的模型明显优于使用一个协变量的模型，且Cox模型的性能略优于Aalen模型。表1进一步展示了这一点，其中我们比较了这四种不同模型的平均评分

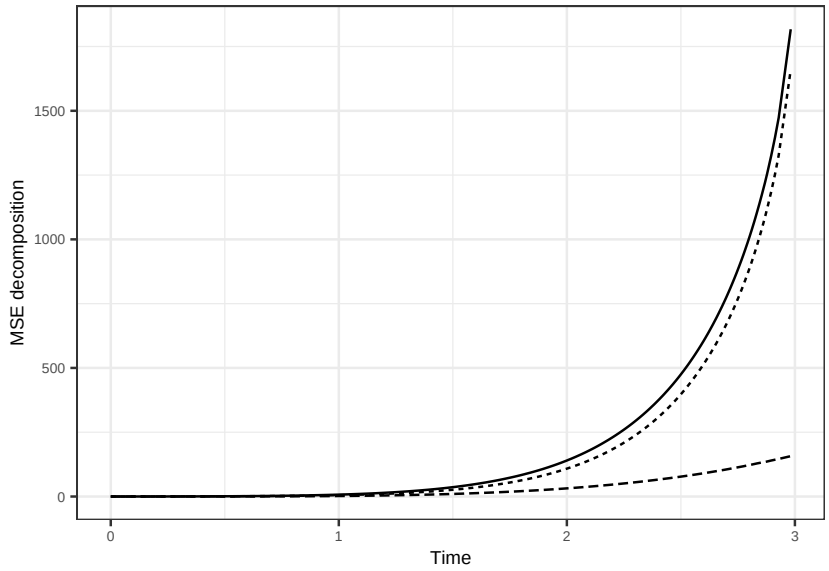


图1： 命题 1 中均方误差（实线）分解为不可分离项 $A(t)$（点线）与不精确项（虚线）之和。数据由包含两个协变量的 Cox模型模拟生成，并使用 Nelson-Aalen估计量预测复发事件的期望累积次数。训练样本（$n_{\text{train}} = 200$）用于计算 Nelson-Aalen估计量，测试样本（$n_{\text{test}} = 1$，000）用于计算均方误差。

这些模型基于500个训练样本（样本量分别为20和50）以及一个独立测试样本（样本量为1，000）。我们清楚地观察到，对于所有时间点和样本量，设定正确的双协变量Cox模型平均而言都优于其他所有模型。不过，它的标准差往往略大，尤其是对于 $n_{\text{train}} = 20$ 和 $t = 2$ 或 $t = 2.9$。

	$n_{\text{train}} = 20$			$n_{\text{train}} = 50$		
	$t = 1$	$t = 2$	$t = 2.9$	$t = 1$	$t = 2$	$t = 2.9$
单协变量 Cox	0.89 (0.33)	9.17 (6.04)	40。6 (54。93)	0.96（0.15）	10.06（2.19）	46.93 (16.06)
Cox 双重协变量	1.8 (0.56)	24.69 (8.58)	119.11 (45.56)	1.95 (0.17)	26。33 (2。54)	127.52（17.73）
Aalen 单协变量	0.14（1.33）	9.53 (4.4)	42。72 (16。27)	0.34 (0.82)	10.09 (1.99)	44。7 (9。21)
Aalen 双协变量	0。44 (1。76)	24。18 (4。8)	103.78 (25.28)	0.51 (1.22)	25.13（1.9）	109。36 (14。21)

表1：复发事件预期数量的预测评分在500次模拟中的均值和标准差（括号内）。数值越大，预测性能越好。

5.2 存在右删失和终末事件的场景

现在我们考虑一个还包含终末事件的模拟场景。复发事件过程及其协变量的模拟方式与上一节相同，参数值也相同。删失变量按照 $[0, 8]$上的均匀变量进行模拟。终末事件根据Cox模型进行模拟，其基线服从Weibull分布，形状参数为5，尺度参数为1.8。该Cox模型还包含与复发事件过程相同的两个协变量，它们对风险函数的影响也相同（即伯努利协变量和高斯协变量的影响分别为log(2)和log(0.5)）。在此设置下，每个个体平均发生8.5次事件；分别有26%、50%和77%的个体经历的事件数小于或等于3、7和12次。平均而言，28%的个体被删失。

我们是根据公式（9）估计复发事件的期望累积次数)

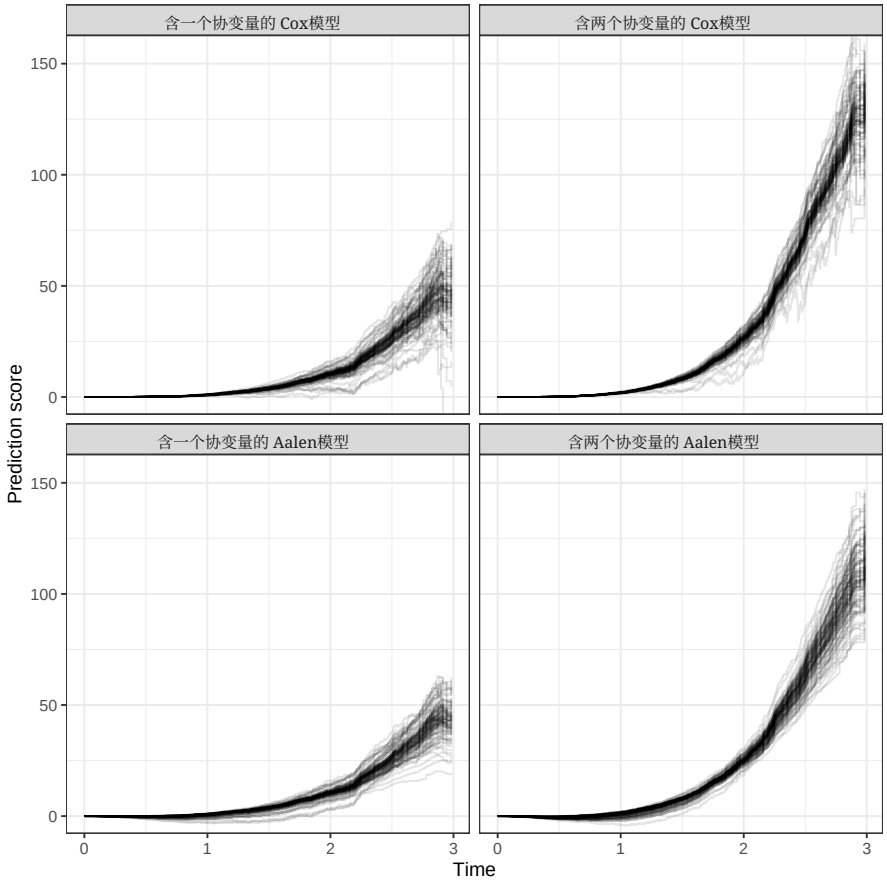


图2： 预测评分（见公式（13）），采用四种不同模型。数据由含两个协变量的 Cox模型生成，并分别使用含一个协变量的 Cox模型、含两个协变量的 Cox模型、含一个协变量的 Aalen模型和含两个协变量的 Aalen模型，预测预期复发事件累积数。参考模型不含协变量，使用 Nelson-Aalen估计量进行估计。预测评分基于 100 个训练样本计算，每个样本的大小为 $n_{\text{train}} = 50$，测试样本的大小为 $n_{\text{test}} = 1,000$。

其中，在公式中，如果终末事件的估计模型包含协变量，则使用Breslow估计量估计终末事件的条件生存函数。换言之：

$$\widehat{S}(t \mid X) = \exp\left(-\int_0^t \exp(X^\top \hat{\beta}^{T^*}) d\widehat{\Lambda}_0^{T^*}\right),$$

其中，$\hat{}\beta^{T^*}$ 是终末事件 Cox模型中的回归参数估计值，$\hat{}\Lambda_0^{T^*}$ 是其对应的基线估计量，称为 Breslow估计量。如果终末事件模型不含任何协变量，则改用 Kaplan-Meier估计量。与前文相同，我们使用公式（13）中定义的评分来评估模型的预测质量，其中参考模型 $\widehat{\mu}_0$ 定义为

$$\widehat{\mu}_0(t) = \int_0^t \hat{S}(u) d\widehat{\Lambda}(u), \tag{14}$$

$\hat{}\hat{}$使用终末事件的 Kaplan-Meier估计量以及 Λ 复发事件过程的 Nelson-Aalen估计量。相同的评分也用于预测生存函数，其中 Kaplan-Meier估计量作为参考模型。我们考虑四种不同的回归模型：一种正确设定的模型，在两个 Cox模型中均纳入这两个协变量，分别用于

复发事件过程和终末事件；一种模型在终末事件的 Cox模型中缺少高斯协变量（但复发事件的Cox模型设定正确）；以及一种在两个 Cox模型中均缺少高斯协变量的模型。

在图3中，我们模拟了100个训练样本，每个样本量为800，并在一个样本量为1,000的唯一测试样本上评估了预测评分。在下图中，我们看到，与仅包含一个协变量的模型相比，在生存模型中纳入这两个协变量可提高预测性能。此外，对于这两个模型，与Kaplan-Meier估计量相比，预测方面的增益在较小时间点更为显著，并且在大约时间2之后有所降低。大时间点下生存函数预测效率的这种损失会影响对复发事件的期望累积次数的预测。在上图中，我们看到，在生存函数的Cox 模型和复发事件模型中加入正确的协变量可提高预测性能。时间2之后，由于生存函数预测效率的损失，复发事件的期望累积次数的预测性能增益略有下降。表2给出了基于500个训练样本（样本量分别为100、200、400和800）以及一个样本量为1,000的单一测试样本得到的这三种不同模型的平均评分。对于所有样本量，我们都观察到了与图3相同的趋势。显然，增加样本量在均值方面并不会带来太多增益，尤其是在较小时间点，但确实会降低预测变量的变异性。在图4中，我们还考察了将终末事件视为删失变量的模型。其中一个模型不包含任何协变量，并使用Nelson-Aalen估计量进行估计。另一个模型包含两个正确的协变量，并基于Cox模型。参考模型与前述相同（见公式（14）），即不包含协变量，但恰当地考虑了终末事件。我们清楚地看到，这两个模型都提供了非常差的预测。这是因为当终末事件被视为删失变量时，相应的Nelson-Aalen估计量会随时间预测出过多的复发。

	$n_{\text{train}} = 100$			$n_{\text{train}} = 200$		
	$t = 1$	$t = 2$	$t = 2.9$	$t = 1$	$t = 2$	$t = 2.9$
单协变量-单协变量	0.68 (0.05)	6。47 (0。43)	6。14 (0。83)	0。69（0。03）	6.54 (0.24)	6.33 (0.45)
两个协变量减一个协变量	1.67 (0.07)	12 (0.92)	6.43 (2.71)	1.67 (0.04)	12.11 (0.6)	6.84 (1.77)
两个协变量减两个协变量	1.67 (0.07)	12.61 (0.59)	10。07 (1。18)	1。68 (0。04)	12。75 (0。3)	10.51 (0.59)
	$n_{\text{train}} = 400$			$n_{\text{train}} = 800$		
	$t = 1$	$t = 2$	$t = 2.9$	$t = 1$	$t = 2$	$t = 2.9$
单协变量-单协变量	0.69（0.01）	6.57 (0.14)	6.42 (0.28)	0.69 (0.01)	6。59 (0。08)	6.47 (0.17)
两个协变量减一个协变量	1.68 (0.02)	12.11 (0.39)	6.92 (1.22)	1。68 (0。01)	12.15 (0.27)	7.07（0.85）
两个协变量减两个协变量	1.68 (0.02)	12.82 (0.17)	10.69 (0.35)	1。68 (0。01)	12.84 (0.09)	10.78 (0.23)

表2：在存在终末事件的情况下，对复发事件预期数量的预测评分进行500次模拟所得的均值和标准差（括号内）。结果针对图3中的同三个模型给出。参考模型未使用协变量，并根据公式（14）中的非参数估计量进行估计。数值越大，表示预测性能越好。

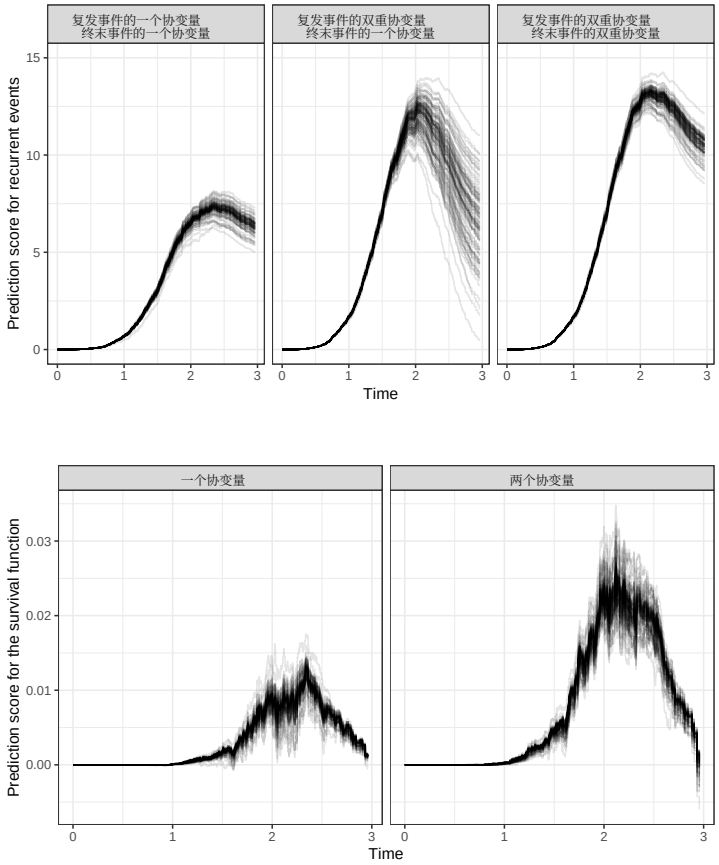


图3： 采用不同模型对复发事件（上图）和生存函数（下图）的预测评分。数据由一个包含两个协变量（$n = 200$）的Cox模型生成，用于复发事件过程；终末事件也使用相同的两个协变量。使用包含一个或两个协变量的Cox模型预测预期复发事件累积数和终末事件的生存函数。参考模型不使用协变量，其估计值分别来自公式（14）中的非参数估计量和Kaplan-Meier估计量，对应上图和下图。预测评分基于100个样本量为 $n_{\text{train}} = 800$ 的训练样本和一个样本量为 $n_{\text{test}} = 1$的测试样本计算，000。

$\times 10^4$	$n_{\text{train}} = 100$			$n_{\text{train}} = 200$		
	$t = 1$	$t = 2$	$t = 2.9$	$t = 1$	$t = 2$	$t = 2.9$
一个协变量	0.58 (0.55)	77.22 (34.77)	22.57 (8.29)	0.69（0.25）	84。71（15。22）	23.96 (3.81)
双重协变量	1.71 (1.49)	214。84 (46。73)	32。3 (18。05)	2。07（0。73）	227.83 (23.1)	38.3 (8.86)
	$n_{\text{train}} = 400$			$n_{\text{train}} = 800$		
	$t = 1$	$t = 2$	$t = 2.9$	$t = 1$	$t = 2$	$t = 2.9$
单个协变量。	0.74（0.17）	86.26 (8.65)	24.29 (2.81)	0.77（0.12）	89。74 (5。44)	27.22 (3.02)
双重协变量	2.2（0.51）	231.51（13.94）	40.49 (3.74)	2.28 (0.36)	238.41 (9.88)	46.23（4.27）

表3： 500次模拟中生存函数预测评分的均值和标准差（括号内）。结果乘以 10^4 ，并给出了包含一个或两个协变量的模型结果。参考模型不使用协变量，并由Kaplan-Meier估计量估计。数值越大表示预测性能越好。

6 真实数据分析：心房颤动数据集

本节分析一组心房颤动（AF）患者的数据集。目的是使用本研究中开发的预测评分，比较用于预测心房颤动住院预期累积数量的不同回归模型。患者于2008年1月1日至2012年12月1日期间，在丹麦希维多夫雷哥本哈根大学医院心脏病科的“Atrial Fibrillation Sur-

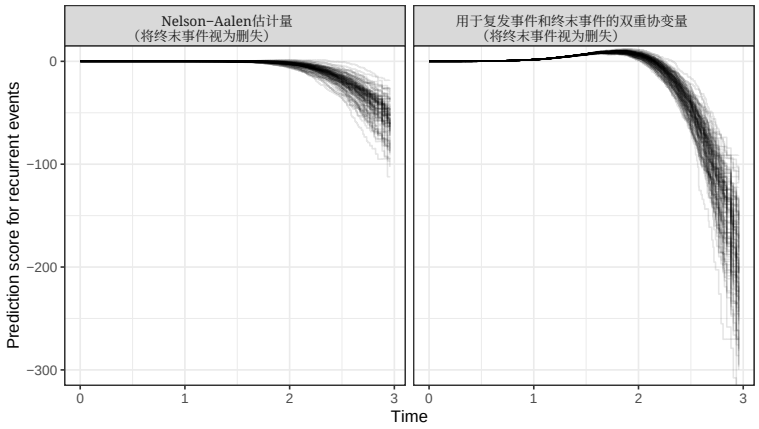


图4：忽略终末事件并将其视为删失变量时，复发事件的期望累积次数的预测评分。数据生成方式与图3相同。使用含一个或两个协变量的Cox模型预测复发事件的预期累积数。参考模型不使用协变量，其估计量基于非参数估计量的Nelson-Aalen估计量（见公式（14））。

vey–Copenhagen（ATLAS-CPH）”项目中，从住院部和门诊招募。所有患者此前均已被诊断为心房颤动，并在基线时被分为阵发性心房颤动（PAF）或持续性心房颤动（PeAF）。阵发性心房颤动定义为至少有一次记录在案的心房颤动发作，并自发转复为窦性心律；无瓣膜性心房颤动，且排除其他时间形式的心房颤动。持续性心房颤动定义为至少有一次记录在案、持续超过7天的心房颤动发作，或需要通过药物或电复律恢复窦性心律的发作（符合当时丹麦心脏病学会的心房颤动指南）。其他纳入标准包括年龄 > 18 岁、近期（< 1 个月）通过标准12导联心电图（ECG）或居家监测记录到心房颤动，以及能够提供口头和书面知情同意。共有174名患者入组，其中50名为阵发性心房颤动患者，124名为持续性心房颤动患者。时间以天为单位测量，平均随访时长为1 279天。就观察到的事件而言，患者共经历了325次心房颤动住院，其中持续性心房颤动组有305次，阵发性心房颤动组有20次。终末事件定义为进展为永久性心房颤动或发生死亡。在该数据集中，45名患者经历了终末事件，其余129名患者被删失。最后，除房颤类型外，该数据集还包括11个其他变量：性别、年龄、酒精使用量（分为两个水平：$0-5$ 和 >5）、烟草使用量（分为三个水平：“从不吸烟”“曾吸烟”“目前吸烟”）、高血压、心力衰竭、心脏瓣膜病、缺血性心脏病、糖尿病、慢性阻塞性肺疾病以及抗心律失常药物的使用情况。数据的详细介绍见 [29]。另请注意，这些数据可从《公共科学图书馆·综合》网站完整获取。

在 [29], 中，作者采用多状态方法分析数据，设定了四种可能状态：尚未发生复发事件、发生1次复发事件、发生2次或更多次复发事件，以及表示终末事件的吸收状态。他们使用Cox模型，假设各转移强度彼此成比例，并将既往复发事件的次数纳入模型。此类包含终末事件的多状态模型已有相关描述，例如见 [6] （参见该书第6.6.4节）。分析结果表明，房颤类型、既往复发事件的次数（p值 $< 10^{-4}$）和年龄（p值$= 0.0253$）对未来心房颤动住院风险均有高度显著的影响。糖尿病的p值为0.0955。其他所有变量均被判定为不显著（p值> 0.2）。作者还采用多状态方法，以所有变量拟合了一个终末事件Cox模型（即通过比例效应纳入既往心房颤动住院的影响）

。在该模型中，只有年龄变量显著（在多变量 Cox 模型中，风险率为1.05，p值为0.0016）。在这项先前的研究中，作者随后决定，仅在复发事件过程模型中纳入房颤类型和年龄这两个协变量，并将既往心房颤动住院次数的影响设为比例效应。对于终末事件模型，他们只纳入了年龄变量。基于这些模型，可以根据患者的特征，预测其未来心房颤动住院的预期累积数量。我们将使用本文提出的预测评分，对比 [29]中所用模型与其他几种可能模型的表现。由于糖尿病是已知的心房颤动危险因素，我们还将考察纳入该变量以及房颤类型和年龄的模型。与模拟部分一样，预测评分将基于训练样本和测试样本计算，但此次采用10折交叉验证：将十分之一的观测值用作测试样本，其余观测值用于模型估计；这一过程重复十次，并对结果取平均。

我们首先评估终末事件的预测性能，分别考察仅包含年龄的 Cox 模型、包含年龄、房颤类型和糖尿病的 Cox 模型、包含年龄的 Aalen 模型，以及包含年龄的随机生存森林。参考模型采用 Kaplan-Meier 估计量，$\widehat{\ }$评分根据公式（13）计算，其中预测准则均方误差由删失变量的 Kaplan-Meier 估计量计算得出。随机生存森林通过 rfsrc 软件包实现（见 [20]）。结果见表4和图5。我们清楚地看到，随机生存森林的表现较差，尤其是在时间 1 000 之前，此时 Kaplan-Meier 的表现更好。包含年龄、房颤类型和糖尿病的 Cox 模型在所有时间点的表现都更好，而仅包含年龄的 Aalen 和 Cox 模型表现非常相似，且优于所有四种模型。我们还使用每种算法对这三个变量的不同组合进行了考察，结果相似，因此省略。接下来，我们决定使用包含年龄的 Cox 模型对终末事件建模。随后考察复发事件过程的预测性能。我们考虑以下模型：基于年龄、糖尿病和房颤类型变量的四个多变量 Cox 模型；按房颤类型分层、包含年龄和糖尿病变量的 Cox 模型；包含全部三个变量的 Aalen 模型； [29]中使用的、包含这三个变量的 Cox 多状态模型；以及按房颤类型分层的同一 Cox 多状态模型。参考模型采用非参数估计量（见公式（14）），评分仍根据公式（13）计算。结果见表5和图6（图中仅展示了五种不同的模型）。我们观察到，仅包含年龄变量的 Cox 模型预测性能较差。在此模型中加入糖尿病或房颤类型可改善模型表现，其中加入房颤类型变量带来的提升更大。此外，在同一模型中结合全部三个变量，与之前所有模型相比都能显著提升表现。另一方面，两种多状态模型对预测的改善较小，分层模型略占优势。最后，图7展示了其中一些模型对心房颤动住院预期累积数量的预测。在该图中，预测对象为两名患有糖尿病的 60 岁患者，一名患有持续性房颤，另一名患有阵发性房颤。不同模型对阵发性房颤患者的预测差异不大，但对持续性房颤患者给出的结果则各不相同。根据表5和图6的结果，按房颤类型分层的 Cox 多状态模型具有最高的预测性能，因此应当选用该模型。房颤诊断后 1 500 天，该模型预测心房颤动住院的预期次数约为 1.03。另一方面，使用多状态 Cox 模型时，预测值为 0.98；使用包含全部三个变量的多变量 Cox 模型时，预测值为 0.76。

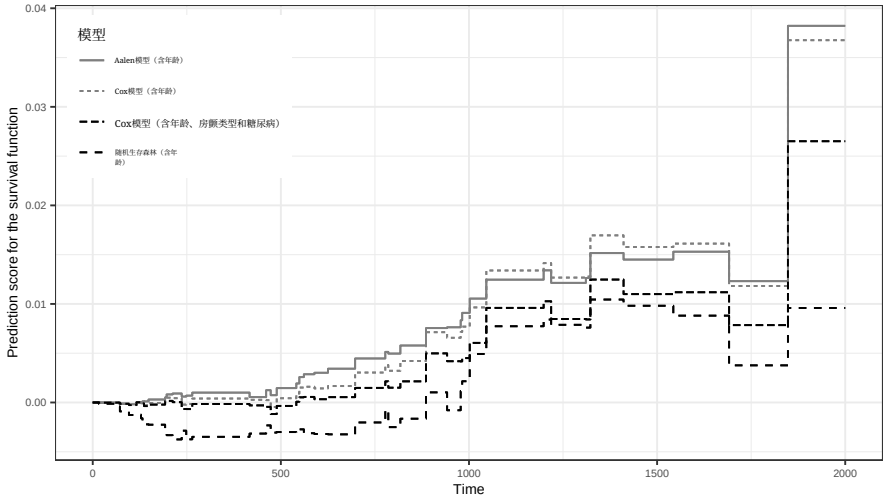


图5： 使用10折交叉验证计算的心房颤动数据集中终末事件生存函数的预测评分。以Kaplan-Meier估计量为参照，对四种不同模型进行比较。为便于可视化，我们按照这些模型在时间$t = 2000$的评分由低到高进行介绍：含年龄、房颤类型和糖尿病协变量的Cox模型（评分 $= 0.067$）、含年龄协变量的随机生存森林（RSF）模型（评分 $= 0.069$）、含年龄协变量的Aalen模型（评分 $= 0.094$），以及含年龄协变量的Cox模型（评分 $= 0.105$）。

	$t = 1000$	$t = 1500$	$t = 2000$
Aalen，包含年龄	$0.011\ [0.002, 0.017]$	$0.015\ [0.001, 0.039]$	$0.038\ [-0.001, 0.035]$
Cox，包含年龄	$0.010\ [-0.006, 0.023]$	$0.016\ [-0.001, 0.042]$	$0.037\ [-0.005, 0.053]$
Cox，包含年龄、房颤类型、糖尿病	$0.006\ [-0.009, 0.019]$	$0.011\ [-0.007, 0.04]$	$0.027\ [-0.002, 0.04]$
随机生存森林，包含年龄	$0.005\ [-0.021, 0.032]$	$0.010\ [-0.014, 0.039]$	$0.010\ [-0.031, 0.045]$

表4：心房颤动数据集中，终末事件生存函数预测评分在10折交叉验证中的均值和80%区间（以花括号表示）。以Kaplan-Meier估计量为参考，在三个不同时间点比较四种模型：协变量为年龄的Aalen和Cox模型、协变量为年龄、房颤类型和糖尿病的Cox模型，以及协变量为年龄的随机生存森林模型。

	$t = 1000$	$t = 1500$	$t = 2000$
Cox，包含年龄	$0.036\ [-0.248, 0.289]$	$0.089\ [-0.690, 0.629]$	$0.283\ [-0.719, 0.985]$
Cox，包含年龄和糖尿病	$0.166\ [-0.342, 0.538]$	$0.232\ [-0.372, 1.051]$	$0.532\ [-0.640, 1.495]$
包含年龄和房颤类型的Cox	$0.611\ [0.104, 1.138]$	$1.170\ [0.278, 2.081]$	$1.702\ [-0.027, 3.989]$
包含年龄、房颤类型和糖尿病的Cox	$0.840\ [-0.123, 1.539]$	$1.420\ [-0.033, 2.983]$	$2.121\ [-0.485, 5.683]$
包含年龄、房颤类型和糖尿病的Aalen	$0.818\ [0.143, 1.312]$	$1.482\ [0.290, 2.992]$	$2.140\ [-0.190, 5.506]$
包含年龄、糖尿病和分层（房颤类型）的Cox	$0.836\ [-0.133, 1.521]$	$1.416\ [-0.053, 3.000]$	$2.118\ [-0.468, 5.667]$
包含年龄、房颤类型和糖尿病的Cox（多状态模型）	$0.847\ [-0.041, 1.459]$	$1.439\ [0.084, 2.893]$	$2.136\ [-0.310, 5.526]$
包含年龄、房颤类型和糖尿病的Cox（多状态模型/分层）	$0.852\ [-0.014, 1.475]$	$1.454\ [0.108, 3.073]$	$2.147\ [-0.298, 5.756]$

表5： 心房颤动数据集中，复发事件预期数量预测评分在10折交叉验证中的均值和80%区间（用花括号表示）。以非参数估计量（见公式（14））为参照，在三个不同时间点比较了八种不同的（针对复发事件的）模型：五种包含年龄、糖尿病和房颤类型协变量的Cox和Aalen模型，一种按房颤类型分层的Cox模型，以及两种多状态Cox模型，即Cox（MSM）和Cox（MSM/分层）。后两种模型的区别在于，前一种假设从0次事件到1次事件，以及从“1次或以上事件”到新事件的基线转移强度成比例，而后一种则使用两个不同的基线函数。所有模型均使用以年龄为唯一协变量的Cox模型来估计生存函数。

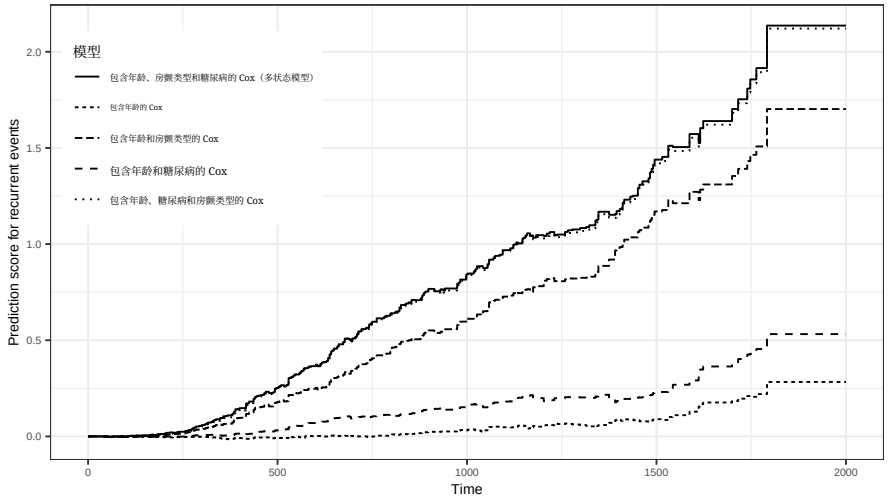


图6： 心房颤动数据集中复发事件期望累积次数的预测评分。以非参数估计量（见公式（14））为参照，对比了六种不同的模型。所有模型均使用以年龄为唯一协变量的Cox模型来估计生存函数。为便于可视化，我们按照各模型在时间 $t = 2000$ 时的评分从低到高，介绍这六种用于复发事件的模型：协变量为年龄的Cox模型（评分 $= 0.911$）、协变量为年龄和糖尿病的Cox模型（评分 $= 1.572$）、协变量为年龄和房颤类型的Cox模型（评分 $= 4.340$）、协变量为年龄、房颤类型和糖尿病的多状态（MSM）Cox模型（评分 $= 5.241$），以及协变量为年龄、房颤类型和糖尿病的Cox模型（评分 $= 5.431$）。MSM Cox模型假设从0次事件到1次事件，以及从“1次或以上事件”到新事件的转移强度成比例。

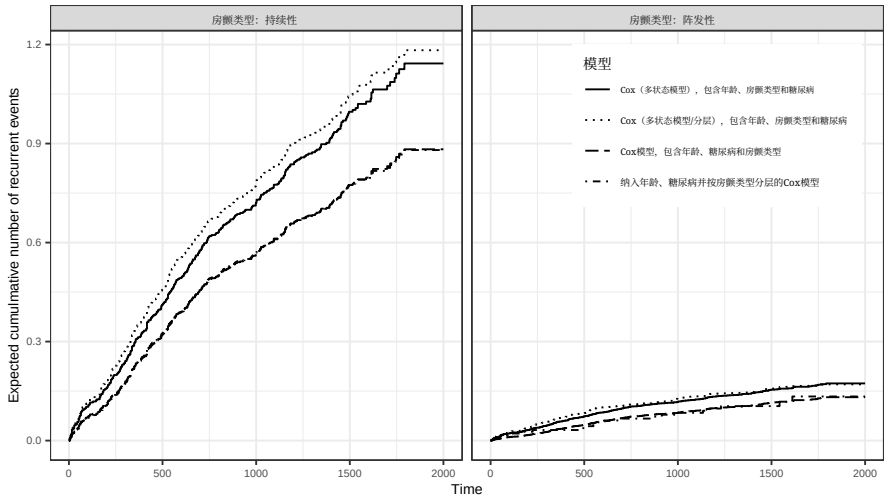


图7： 四种不同模型预测的两名60岁糖尿病患者复发事件的期望累积次数，其中一名患有持续性房颤，另一名患有阵发性房颤。这些模型的预测性能相似（见表5和图6）。

7 结论

本研究针对复发事件数据提出了一种新的预测准则。该准则用于评估复发事件的期望累积次数的预测性能，同时考虑删失和可能发生的终末事件。我们证明，该准则可以像 [15]中那样分解为不可分离项和不精确项。此外，模拟结果表明，不可分离项在分解中占主导地位。因此，我们建议使用公式（13）中定义的预测评分，即给定模型与参考模型的预测准则之差。参考模型通常是不使用协变量的模型，这样该评分便能反映所提出模型中协变量带来的绝对增益。另一种评分可以通过计算 [30]所提出的相对增益得到。这会产生一个取值范围为 0% 到 100% 的评分，并且与皮尔逊 R^2 统计量有相似之处。不过，使用这类评分时应谨慎，因为我们是相对于参考模型的预测准则进行归一化，而该准则本身可以分解为不精确项和不可分离项。因此，由于不可分离项的大小在实践中未知，这一准则可能会产生误导。

所提出的预测准则易于计算，其优点是涵盖所有复发事件。因此，它可视为一种总体性能度量，用于提供有关所提出模型在复发事件情境下整体预测能力的信息。不过，如果研究者只想评估模型对患者已经历固定次数事件后进一步发生的复发事件的预测性能，也可以对该准则进行修改。这相当于在多状态框架下，以已经历的复发事件次数为条件。此类准则类似于 [28]中提出的方法：该方法以患者存活至时间 t^* 为条件，并评估模型对时间 $s > t^*$ 的预测。另一项改进是按照 [31] 的方法纳入脆弱性模型。可以推导出一种整合脆弱性变量的边际评分。这种评分可以全面评估脆弱性模型，也是本文所提评分的自然扩展。或者，也可以针对条件（相对于脆弱性变量）下的复发事件的期望累积次数提出一种条件评分。要开发这两种评分，还需要开展更多研究。

8 附录：两种情形下复发事件的期望累积次数预测准则收敛性的证明

在命题 1 的证明中，我们将根据所研究的情形，对观测到的复发事件过程 N 采用不同的分解方式。稍微滥用一下记号，在两种情形中，我们都用同一个名称 M 表示残差过程。

- 仅存在右删失的情形（第 2.2 节）下，我们有：

$$N(t) - \int_0^t I(C \geq u)\lambda^*(u \mid X(u))du = M(t),$$

  其中 $\mathbb{E}[dM(t) \mid \{I(C \geq u),\ X(u):\ 0 \leq u\}] = 0$。

- 存在右删失和终末事件的情形（第 2.3 节）下，我们有：

$$N(t) - \int_0^t I(T \geq u)\lambda^*(u \mid X(u))du = M(t),$$

  其中 $\mathbb{E}[dM(t) \mid \{I(T \geq u),\ X(u):\ 0 \leq u\}] = 0$。

8.1 命题 1 的证明

在这两种情形下，我们可以直接得到:

$$\mathrm{MSE}(t,\mu)=\mathbb{E}\left[\left(\mu(t\mid X(t))-\mu^*(t\mid X(t))\right)^2\right]+\mathbb{E}\left[\left(\int_0^t\frac{dN(u)}{1-G(u-\mid X(u))}-\mu^*(t\mid X(t))\right)^2\right]$$
$$+2\mathbb{E}\left[\left(\int_0^t\frac{dN(u)}{1-G(u-\mid X(u))}-\mu^*(t\mid X(t))\right)\left(\mu^*(t\mid X(t))-\mu(t\mid X(t))\right)\right].$$

利用 $\mathbb{E}[\int_0^t dN(u)/(1-G(u-\mid X(u)))\mid X(t)]=\mu^*(t\mid X(t))$ 这一事实，我们得出

$$\mathrm{MSE}(t,\mu)=\mathbb{E}\left[(\mu(t\mid X(t))-\mu^*(t\mid X(t)))^2\right]+A(t),$$

其中

$$A(t)=\mathbb{E}\left[\left(\int_0^t\frac{dN(u)}{1-G(u-\mid X(u))}\right)^2\right]-\mathbb{E}\left[(\mu^*(t\mid X(t)))^2\right].\tag{15}$$

现在我们证明，在仅存在右删失的情况下， $A(t)\geq 0$。根据第 8 节开头给出的分解，我们有，

$$\mathbb{E}\left[\left(\int_0^t\frac{dN(u)}{1-G(u-\mid X(u))}\right)^2\right]=\mathbb{E}\left[\left(\int_0^t\frac{dM(u)}{1-G(u-\mid X(u))}\right)^2\right]$$
$$+\mathbb{E}\left[\left(\int_0^t\frac{\lambda^*(u\mid X(u))I(C\geq u)du}{1-G(u-\mid X(u))}\right)^2\right],$$

因为

$$\mathbb{E}\left[\int_0^t\frac{dM(u)}{1-G(u-\mid X(u))}\int_0^t\frac{\lambda^*(v\mid X(v))I(C\geq v)dv}{1-G(v-\mid X(v))}\right]$$
$$=\mathbb{E}\left[\int_0^t\frac{\mathbb{E}[dM(u)\mid\{I(C\geq v),X(v):0\leq v\}}{1-G(u-\mid X(u))}\int_0^t\frac{\lambda^*(v\mid X(v))I(C\geq v)dv}{1-G(v-\mid X(v))}\right]=0.$$

于是,

$$\mathbb{E}\left[\left(\int_0^t\frac{\lambda^*(u\mid X(u))I(C\geq u)du}{1-G(u-\mid X(u))}\right)^2\right]=2\mathbb{E}\left[\int_{0\leq u<v\leq t}\frac{\lambda^*(u\mid X(u))\lambda^*(v\mid X(v))}{1-G(u-\mid X(u))}dudv\right]$$
$$\mathbb{E}\left[(\mu^*(t\mid X(t)))^2\right]=2\mathbb{E}\left[\int_{0\leq u<v\leq t}\lambda^*(u\mid X(u))\lambda^*(v\mid X(v))dudv\right].$$

因此,

$$A(t)=\mathbb{E}\left[\left(\int_0^t\frac{dM(u)}{1-G(u-\mid X(u))}\right)^2\right]$$
$$+2\mathbb{E}\left[\left(\int_{0\leq u<v\leq t}\frac{\lambda^*(u\mid X(u))\lambda^*(v\mid X(v))G(u-\mid X(u))}{1-G(u-\mid X(u))}dudv\right)\right]\geq 0.$$

存在终末事件时的证明类似。我们发现

$$A(t)=\mathbb{E}\left[\left(\int_0^t\frac{dM(u)}{1-G(u-\mid X(u))}\right)^2\right]$$
$$+2\mathbb{E}\left[\left(\int_{0\leq u<v\leq t}\frac{1-S(u\mid X(u))(1-G(u-\mid X(u)))}{1-G(u-\mid X(u))}\right.\right.$$
$$\left.\left.\times\lambda^*(u\mid X(u))\lambda^*(v\mid X(v))S(v\mid X(v))dudv\right)\right]\geq 0.$$

8.2 命题2的证明

我们先证明存在终末事件时 $\mathbb{E}[\mu^*(\tau \mid X(\tau))] < \infty$ （没有终末事件的情形可用相同的论证得出）。对于所有 $t \in [0, \tau]$，我们有：$\mathbb{P}[C \geq t \mid X(t)] \geq \mathbb{P}[T \geq t \mid X(t)] \geq c$，这由假设3得出。根据同一假设，$N(\tau)$ 几乎必然有界于某个常数。因此，

$$\mu^*(\tau \mid X(\tau)) = \int_0^\tau \frac{\mathbb{E}[dN(t) \mid X(t)]}{1 - G(t- \mid X(t))}$$

几乎必然有界，其中等式已在第 2.3 节中得到证明。命题2在两种情形下的其余证明完全相同。

首先注意到 $F_{X(t)}(x) = \mathbb{P}[X(t) \leq x]$，我们令 $\mathcal{X}_{u,v}$ 表示联合分布 $(X(u)\ X(v))$ 的支撑集，并注意到 $F_{X(u),X(v)}(x, y) = \mathbb{P}[X(u) \leq x, X(v) \leq v]$。接着引入以下量

$$\begin{aligned}
\xi(t) = &\int_{0 \leq u,v \leq t} \int_{\mathcal{X}_{u,v}} \frac{\mathbb{E}[dN(u)dN(v) \mid X(u) = x, X(v) = y]}{(1 - \hat{G}(u- \mid x))(1 - \hat{G}(v- \mid y))} dF_{X(u),X(v)}(x, y) \\
&- 2 \int_{\mathcal{X}_t} \widehat{\mu}(t \mid x) \mu^*(t \mid x) dF_{X(t)}(x) \\
&+ \int_{\mathcal{X}_t} \left(\widehat{\mu}(t \mid x)\right)^2 dF_{X(t)}(x) =: \xi_1(t) + \xi_2(t) + \xi_3(t).
\end{aligned}$$

记作：

$$\begin{aligned}
\left|\widehat{\mathrm{MSE}^1}(t, \hat{\mu}) - \mathrm{MSE}^1(t, \mu)\right| \leq &\left|\xi(t) - \mathbb{E}\left[\left(\int_0^t \frac{dN(u)}{1 - G(u- \mid X(u))} - \mu(t \mid X(t))\right)^2\right]\right| \\
&+ \left|\frac{1}{n}\sum_{i=1}^n \left(\int_0^t \frac{dN_i(u)}{1 - \hat{G}(u- \mid X_i(u))} - \widehat{\mu}(t \mid X_i(t))\right)^2 - \xi(t)\right| \\
&=: C(t) + D(t).
\end{aligned}$$

将平方项分解为另外三项后，我们以下述方式对 $C(t)$进行界定：$C(t) \leq |C_1(t)| + |C_2(t)| + |C_3(t)|$ 其中

$$\begin{aligned}
C_1(t) = &\int_{0 \leq u,v \leq t} \int_{\mathcal{X}_{u,v}} \frac{(1 - G(u- \mid x))(1 - G(v- \mid y)) - (1 - \hat{G}(u- \mid x))(1 - \hat{G}(v- \mid y))}{(1 - \hat{G}(u- \mid x))(1 - \hat{G}(v- \mid y))(1 - G(u- \mid x))(1 - G(v- \mid y))} \\
&\mathbb{E}[dN(u)dN(v) \mid X(u) = x, X(v) = y] dF_{X(u),X(v)}(x, y), \\
C_2(t) = &-2 \int_{\mathcal{X}_t} (\widehat{\mu}(t \mid x) - \mu(t \mid x)) \mu^*(t \mid x) dF_{X(t)}(x), \\
C_3(t) = &\int_{\mathcal{X}_t} \left(\left(\widehat{\mu}(t \mid x)\right)^2 - \left(\mu(t \mid x)\right)^2\right) dF_{X(t)}(x).
\end{aligned}$$

对于 $C_1(t)$，我们有

$$\begin{aligned}
&(1 - G(u- \mid x))(1 - G(v- \mid y)) - (1 - \hat{G}(u- \mid x))(1 - \hat{G}(v- \mid y)) \\
&= \left(\hat{G}(u- \mid x) - G(u- \mid x)\right) + \left(\hat{G}(v- \mid y) - G(v- \mid y)\right) \\
&\quad + G(u- \mid x)\left(G(v- \mid y) - \hat{G}(v- \mid y)\right) + \hat{G}(v- \mid y)\left(G(u- \mid x) - \hat{G}(u- \mid x)\right),
\end{aligned}$$

and 我们可以用相同的方法处理所有四项。例如，对于第一项 ，

$$\int_{0\le u,v\le t}\int_{\mathcal{X}_{u,v}} \frac{\left(\hat{G}(u- \mid x)-G(u- \mid x)\right)\mathbb{E}[dN(u)dN(v)\mid X(u)=x,X(v)=y]}{(1-\hat{G}(u- \mid x))(1-\hat{G}(v- \mid y))(1-G(u- \mid x))(1-G(v- \mid y))}dF_{X(u),X(v)}(x,y)$$

$$\le \int_0^t\int_{\mathcal{X}_u} \frac{\left|\hat{G}(u- \mid x)-G(u- \mid x)\right|\mathbb{E}[dN(u)\mid X(u)=x]}{(1-\hat{G}(u- \mid x))(1-G(u- \mid x))}dF_{X(u)}(x),$$

利用 $\int_0^t dN(v)/((1-\hat{}G(v- \mid y))(1-G(v- \mid y)))$ 有界这一事实。然后，由于 $\int_0^t \mathbb{E}[dN(u)/(1-G(u- \mid X(u)))\mid X(u)=x]=\mu^*(t\mid X(t))$ 且 $(1-\hat{}G(u- \mid x))^{-1}$ 渐近上一致有界，我们利用删失估计量的一致收敛，得出 $|C_1(t)|$ 依概率趋于 0。

对于 $C_2(t)$，我们利用 $\widehat{\mu}$ 的一致性以及 $\mathbb{E}[\mu^*(t\mid X(t))]$ 有限这一事实，证明 $|C_2(t)|$ 依概率趋于 0。

对于 $C_3(t)$，我们直接写出 $(\widehat{\mu}(t\mid x))^2-(\mu(t\mid x))^2=(\widehat{\mu}(t\mid x)-\mu(t\mid x))(\widehat{\mu}(t\mid x)+\mu(t\mid x))$，并利用 $\mu(t\mid x)$ 有界这一事实以及 $\widehat{\mu}$ 的一致性，证明 $|C_3(t)|$ 依概率趋于 0。

与 $C(t)$ 类似，我们得到以下界： $D(t)\le |D_1(t)|+|D_2(t)|+|D_3(t)|$ 其中

$$D_1(t)=\frac{1}{n}\sum_{i=1}^n\int_{0\le u,v\le t}\frac{dN_i(u)dN_i(v)}{(1-\hat{G}(u- \mid X_i(u)))(1-\hat{G}(v- \mid X_i(v)))}-\xi_1(t),$$

$$D_2(t)=-\frac{2}{n}\sum_{i=1}^n\int_0^t\frac{dN_i(u)}{1-\hat{G}(u- \mid X_i(u))}\widehat{\mu}(t\mid X_i(t))-\xi_2(t),$$

$$D_3(t)=\frac{1}{n}\sum_{i=1}^n\left(\widehat{\mu}(t\mid X_i(t))\right)^2-\xi_3(t).$$

现在我们使用界 $|D_1(t)|\le |D_{1,\ 1}(t)|+|D_{1,\ 2}(t)|+|D_{1,\ 3}(t)|$ 其中

$$D_{1,1}(t)=\frac{1}{n}\sum_{i=1}^n\int_{0\le u,v\le t}\frac{dN_i(u)dN_i(v)}{(1-G(u- \mid X_i(u)))(1-G(v- \mid X_i(v)))}$$
$$-\int_{0\le u,v\le t}\int_{\mathcal{X}_{u,v}}\frac{\mathbb{E}[dN(u)dN(v)\mid X(u)=x,X(v)=y]}{(1-G(u- \mid x))(1-G(v- \mid y))}dF_{X(u),X(v)}(x,y),$$

$$D_{1,2}(t)=\frac{1}{n}\sum_{i=1}^n\int_{0\le u,v\le t}\chi(u,v,X_i(u),X_i(v))dN_i(u)dN_i(v)$$

$$D_{1,3}(t)=-\int_{0\le u,v\le t}\int_{\mathcal{X}_{u,v}}\chi(u,v,x,y)\mathbb{E}[dN(u)dN(v)\mid X(u)=x,X(v)=y]dF_{X(u),X(v)}(x,y)$$

and

$$\chi(u,v,x,y)=\Big\{\left(\hat{G}(u- \mid x)-G(u- \mid x)\right)+\left(\hat{G}(v- \mid y)-G(v- \mid y)\right)$$
$$+\left(\hat{G}(v- \mid y)-G(v- \mid y)\right)+G(u- \mid x)\left(G(v- \mid y)-\hat{G}(v- \mid y)\right)$$
$$+\hat{G}(v- \mid x)\left(G(u- \mid y)-\hat{G}(u- \mid y)\right)\Big\}$$
$$\times\frac{1}{(1-\hat{G}(u- \mid x))(1-\hat{G}(v- \mid y))(1-G(u- \mid x))(1-G(v- \mid y))}.$$

根据强大数定律，项 $|D_{1,1}(t)|$ 依概率收敛于 0。项 $|D_{1,2}(t)|$ 有界于

$$\sup_{u,v,x,y}|\chi(u,v,x,y)|\frac{1}{n}\sum_{i=1}^n\int_{0\le u<v\le t}dN_i(u)dN_i(v),$$

ˆ $\sup_{u,v,x,y} |\chi(u,v,x,y)|$ 根据 G 的一致收敛性收敛于 0，而另一项根据大数定律收敛于一个有界值。同样的论证适用于 $|D_{1,3}(t)|$，它也依概率收敛于 0。

对于 $D_2(t)$，我们将 $|D_2(t)| \le |D_{2,1}(t)| + |D_{2,2}(t)| + |D_{2,3}(t)| + |D_{2,4}(t)|$ 记为

$$
\begin{aligned}
D_{2,1}(t) &= -\frac{2}{n}\sum_{i=1}^{n}\int_0^t \frac{dN_i(u)}{1-G(u- \mid X_i(u))}\mu(t \mid X_i(t)) + 2\int_{\mathcal{X}_t}\mu(t \mid x)\mu^*(t \mid x)dF_{X(t)}(x),\\
D_{2,2}(t) &= \frac{2}{n}\sum_{i=1}^{n}\int_0^t \frac{dN_i(u)}{1-G(u- \mid X_i(u))}(\mu(t \mid X_i(t)) - \widehat{\mu}(t \mid X_i(t)))\\
D_{2,3}(t) &= 2\int_{\mathcal{X}_t}(\widehat{\mu}(t \mid x) - \mu(t \mid x))\mu^*(t \mid x)dF_{X(t)}(x),\\
D_{2,4}(t) &= \frac{2}{n}\sum_{i=1}^{n}\int_0^t \frac{(G(u- \mid X_i(u)) - \hat{G}(u- \mid X_i(u)))dN_i(u)}{(1-G(u- \mid X_i(u)))(1-\hat{G}(u- \mid X_i(u)))}\widehat{\mu}(t \mid X_i(t)).
\end{aligned}
$$

根据大数定律，$D_{2,1}(t)$ 项依概率收敛至 0。对于$D_{2,2}(t)$、$D_{2,3}(t)$ 和$D_{2,4}(t)$，我们利用 $\widehat{\mu}$ 的一致性、$\sum_i \int_0^t dN_i(u))(1-G(u- \mid X_i(u)))/n$ 的依概率收敛性、$\mathbb{E}[\mu^*(t \mid X(t))]$, 的有界性、ˆ$G$ 的一致一致性以及 $\widehat{\mu}$ 和 $(1-\hat{G}(u- \mid x))^{-1}$ 的渐近有界性，证明这三项均依概率收敛至 0。

最后，对于 $D_3(t)$，我们写作

$$
\begin{aligned}
D_3(t) = &\frac{1}{n}\sum_{i=1}^{n}\Big(\mu(t \mid X_i(t))\Big)^2 - \int_{\mathcal{X}_t}\Big(\mu(t \mid x)\Big)^2 dF_{X(t)}(x)\\
&+ \frac{1}{n}\sum_{i=1}^{n}\left(\Big(\widehat{\mu}(t \mid X_i(t))\Big)^2 - \Big(\mu(t \mid X_i(t))\Big)^2\right)\\
&+ \int_{\mathcal{X}_t}\left(\Big(\widehat{\mu}(t \mid x)\Big)^2 - \Big(\mu(t \mid x)\Big)^2\right) dF_{X(t)}(x).
\end{aligned}
$$

根据大数定律，第一项依概率收敛于 0；根据 $\widehat{\mu}$ 的一致收敛性，另外两项也依概率收敛于 0。

8.3 命题 3 的证明

首先，注意到布里尔评分可以写成如下形式：

$$
\mathrm{MSE}^{\mathrm{Brier}}(t,\pi) = \mathbb{E}[S(t \mid X)] - 2\mathbb{E}[S(t \mid X)\pi(t \mid X)] + \mathbb{E}[(\pi(t \mid X))^2].
$$

现在我们来研究预测评分 MSE′(t, π)。根据标准的鞅性质（例如参见 [2]），可直接得到 $\mathbb{E}[dN(t) \mid X] = H(t \mid X)\lambda^*(t \mid X)dt$，其中，在独立删失条件下，$H(t \mid X) = \mathbb{P}[T > t \mid X] = S(t \mid X)(1-G(t \mid X))$，且 λ^* 是 T^*的风险率。因此，

$$
\mathbb{E}\left[\int_0^t \frac{dN(u)}{1-G(u-)} \mid X\right] = \int_0^t S(u \mid X)\lambda^*(u \mid X)du = 1 - S(t \mid X), \tag{16}
$$

因为 $S(u \mid X)\lambda^*(u \mid X)$ 等于 T^* 的条件密度函数。此外，值得注意的是

$$
\mathbb{E}\left[\left(\int_0^t \frac{dN(u)}{1-G(u-)}\right)^2\right] = \mathbb{E}\left[\int_0^t \frac{dN(u)}{(1-G(u-))^2}\right] = \mathbb{E}\left[\int_0^t \frac{S(u \mid X)}{1-G(u-)}\lambda^*(u \mid X)du\right],
$$

其中，第一个等式是因为 $\int_0^t dN\ u\ /\ 1-$ 只能 $G\ u-\ \ ^2$ N $\Delta I\ T \le t\ /\ 1-G\ T-\ \ ^2$ 跳变一次，因此 (( ) (( )) 就等于 ( ) (( ))。现在，

$$
\begin{aligned}
\mathrm{MSE}'(t,\pi) &= \mathbb{E}\left[\left(1-\int_0^t \frac{dN(u)}{1-G(u-)}\right)^2\right] - 2\mathbb{E}\left[\left(1-\int_0^t \frac{dN(u)}{1-G(u-)}\right)\pi(t\mid X)\right] \\
&\quad + \mathbb{E}[(\pi(t\mid X))^2] \\
&= 1 - 2\mathbb{E}[(1-S(t\mid X))] + \mathbb{E}\left[\left(\int_0^t \frac{dN(u)}{1-G(u-)}\right)^2\right] - 2\mathbb{E}[S(t\mid X)\pi(t\mid X)] \\
&\quad + \mathbb{E}[(\pi(t\mid X))^2] \\
&= \mathrm{MSE}^{\mathrm{Brier}}(t,\pi) + B(t),
\end{aligned}
$$

with

$$
B(t) = -\mathbb{E}[1-S(t\mid X)] + \mathbb{E}\left[\int_0^t \frac{S(u\mid X)}{1-G(u-)}\lambda^*(u\mid X)du\right].
$$

现在，利用公式（16），我们可以将 $B(t)$ 改写如下：

$$
\begin{aligned}
B(t) &= -\mathbb{E}\left[\int_0^t S(u\mid X)\lambda^*(u\mid X)du\right] + \mathbb{E}\left[\int_0^t \frac{S(u\mid X)}{1-G(u-)}\lambda^*(u\mid X)du\right] \\
&= \mathbb{E}\left[\int_0^t \frac{G(u-)}{1-G(u-)}S(u\mid X)\lambda^*(u\mid X)du\right].
\end{aligned}
$$

这表明 $B(t) \ge 0$，且该量不依赖于 π。

8.4 模拟部分中 $A(t)$ 的计算

虽然无法根据真实数据计算不可分离项，但在已知所有变量的分布时，可以得到其表达式。本节针对第 5.1 节的模拟设定，给出不可分离项 $A(t)$ 的显式表达式。为此，我们使用公式（15）。首先注意，对于 $u < v$, $dN(u)dN(v) = I(C \ge v)dN^*(u)dN^*(v)$，并且

$$
\begin{aligned}
\mathbb{E}[dN(u)dN(v)] &= \mathbb{E}[I(C\ge v)\mathbb{E}[dN^*(u)dN^*(v)\mid X, C] \\
&= \mathbb{E}[I(C\ge v)\lambda_0(u)\lambda_0(v)\exp(2\theta_0^\top X)]dudv \\
&= (1-G(v-))\lambda_0(u)\lambda_0(v)\mathbb{E}[\exp(2\theta_0^\top X)]dudv,
\end{aligned}
$$

其中我们利用了以下事实：$\mathbb{E}[dN^*(u)dN^*(v) \mid X,\ C] = \mathbb{E}[dN^*(u) \mid X]\mathbb{E}[dN^*(v) \mid X] = \lambda_0(u)\lambda_0(v)\exp(2\theta_0^\top X)dudv$；因为在我们的模拟方案下，$N^*$ 独立于 C，且在给定 X的条件下，$dN^*(u)$ 独立于 $dN^*(v)$，对于 $u \ne v$。令 $\gamma = 3$，使得 C服从 $[0,\gamma]$上的均匀分布。对于 $t < \gamma$，我们有：

$$
\mathbb{E}\left[\left(\int_0^t \frac{dN(u)}{1-G(u-\mid X(u))}\right)^2\right] = A_1(t) + A_2(t),
$$

其中

$$
A_1(t) = 2\iint_{0<u<v<t} \frac{\lambda_0(u)\lambda_0(v)}{1-G(u-)}dudv\mathbb{E}[\exp(2\theta_0^\top X)], \tag{17}
$$

$$
A_2(t) = \int_0^t \frac{\mathbb{E}[dN(u)]}{(1-G(u-))^2}. \tag{18}
$$

现在我们计算

$$\int_0^v \frac{\lambda_0(u)}{1-G(u-)}du = 2\times\frac{\gamma}{\beta^2}\int_0^v \frac{u}{\gamma-u}du = 2\times\frac{\gamma}{\beta^2}\times\left(\gamma\log\left(\frac{\gamma}{\gamma-v}\right)-v\right),$$

其中，我们将 λ_0 替换为形状参数为 $\alpha=2$、尺度参数为 β 的 Weibull分布的风险函数，最后一个等式则通过变量替换 $w=\gamma-u$ 得出。接下来需要计算 $A_1(t)$ 中的以下积分：

$$\begin{aligned}\int_0^t \left(\gamma\log\left(\frac{\gamma}{\gamma-v}\right)-v\right)\lambda_0(v)dv &= \frac{2}{\beta^2}\int_0^t\left(\gamma\log\left(\frac{\gamma}{\gamma-v}\right)-v\right)vdv\\ &= \frac{2}{\beta^2}\left(\frac{\gamma t^2}{2}\log(\gamma)-\gamma\int_0^t v\log(\gamma-v)dv-\frac{t^3}{3}\right).\end{aligned}$$

最后一个积分通过分部积分计算，然后再使用变量替换 $w=\gamma-v$：

$$\begin{aligned}\int_0^t v\log(\gamma-v)dv &= \int_0^t \frac{v^2}{2}\frac{dv}{\gamma-v}+\frac{t^2}{2}\log(\gamma-t)\\ &= \int_{\gamma-t}^{\gamma}\frac{(\gamma-w)^2}{2w}dw+\frac{t^2}{2}\log(\gamma-t)\\ &= \frac{\gamma^2}{2}\log\left(\frac{\gamma}{\gamma-t}\right)-\gamma t+\frac{\gamma^2}{4}-\frac{(\gamma-t)^2}{4}+\frac{t^2}{2}\log(\gamma-t)\\ &= \frac{t^2-\gamma^2}{2}\log(\gamma-t)-\frac{\gamma t}{2}-\frac{t^2}{4}+\frac{\gamma^2}{2}\log(\gamma).\end{aligned}$$

汇总 $A_1(t)$ 中的所有部分，得到：

$$\begin{aligned}&A_1(t)\\ &= \frac{8\gamma}{\beta^4}\left(\frac{\gamma t^2}{2}\log(\gamma)-\frac{\gamma}{2}(t^2-\gamma^2)\log(\gamma-t)+\frac{\gamma^2 t}{2}+\frac{\gamma t^2}{4}-\frac{\gamma^3}{2}\log(\gamma)-\frac{t^3}{3}\right)\mathbb{E}[\exp(2\theta_0^\top X)].\end{aligned}$$

另一方面，利用关系式（参见第 2.2 节）$\mathbb{E}[dN(t)\mid X(t)] = (1-G(t-))\lambda^*(t\mid X(t))dt$，计算 $A_2(t)$ 很简单。对于 $t<\gamma$，我们有

$$\begin{aligned}A_2(t) &= \int_0^t \frac{\lambda_0(t)}{1-G(t-)}dt\mathbb{E}[\exp(\theta_0^\top X)]\\ &= 2\times\frac{\gamma}{\beta^2}\times\left(\gamma\log\left(\frac{\gamma}{\gamma-t}\right)-t\right)\mathbb{E}[\exp(\theta_0^\top X)].\end{aligned}$$

最后，根据公式（15），我们需要计算 $A_3(t)=\mathbb{E}\left[(\mu^*(t\mid X(t)))^2\right]$。由公式（12）可直接得到

$$A_3(t)=\left(\frac{t}{\beta}\right)^{2\alpha}\mathbb{E}[\exp(2\theta_0^\top X_i)].$$

总之，$A(t)=A_1(t)+A_2(t)-A_3(t)$，而该公式中涉及的项（包括 $\mathbb{E}[\exp(\theta_0^\top X_i)]$ 或 $\mathbb{E}[\exp(2\theta_0^\top X_i)]$）都可以通过蒙特卡罗模拟轻松计算。

参考文献

[1] Andersen, P. K., Angst, J.& Ravn, H. (2019). 终末事件存在时复发事件研究中边际特征的建模。《终生数据分析》 **25**, 681–695.

[2] Andersen, P. K., Borgan, Ø., Gill, R. D. & Keiding, N. (1993). 基于计数过程的统计模型。Springer统计学丛书。纽约：施普林格出版社。[3] Andersen, P. K. & Gill, R. D. (1982). 计数过程的Cox回归模型：一项大样本研究。《统计年刊》，1100–1120。[4] Bouaziz, O., Geffray, S. & Lopez, O. (2015). 通过单指标模型对复发事件过程进行半参数推断。《统计学》49, 361–385。[5] Bouaziz, O. & Lopez, O.(2010). 删失单指标回归模型中的条件密度估计。《伯努利》16, 514–542。[6] Cook, R. J. & Lawless, J.(2007). 复发事件的统计分析。Springer科学与商业媒体。[7] Cook, R. J. & Lawless, J. F. (1997). 复发事件和终止事件的边际分析。《医学统计学》16, 911–924。[8] Cox, D. R. (1972). 回归模型与生命表。《皇家统计学会杂志：B辑（方法论）》34, 187–202。[9] Dabrowska, D. M. (1989). 核条件Kaplan-Meier估计的一致收敛。《统计年刊》，1157–1167。[10] Delecroix, M., Härdle, W. & Hristache, M. (2003). 条件单指标回归中的有效估计。《多元分析杂志》86, 213–226。[11] Gerds, T. A., Kattan, M. W., Schumacher, M. & Yu, C. (2013). 协变量依赖删失下生存预测模型时间依赖一致性指数的估计。《医学统计学》32, 2173–2184。[12] Gerds, T. A. & Schumacher, M.(2006). 右删失事件时间的一般生存模型中期望Brier评分的一致估计。《生物统计学杂志》48, 1029–1040。[13] Ghosh, D. & Lin, D. (2003). 依赖删失存在时复发事件数据的半参数分析。《生物统计学》59, 877–885。[14] Ghosh, D.& Lin, D. Y. (2002). 复发事件和终末事件的边际回归模型。《中华统计学志》，663–688。[15] Graf, E., Schmoor, C., Sauerbrei, W. & Schumacher, M. (1999). 生存数据预后分类方案的评估与比较。《医学统计学》18, 2529–2545。[16] Hall, P., Wolff, R. C. & Yao, Q. (1999). 条件分布函数的估计方法。《美国统计学会杂志》94, 154–163。[17] Harrell Jr, F. E., Lee, K. L. & Mark, D. B. (1996). 多变量预后模型：模型开发、假设与充分性评估以及误差测量与减少中的问题。《医学统计学》15, 361–387。[18] Heagerty, P. J. & Zheng, Y.(2005). 生存模型预测准确性与ROC曲线。《生物统计学》61, 92–105。[19] Hougaard, P. & Hougaard, P. (2000). 多变量生存数据分析，第564卷。施普林格。

[20] Ishwaran, H., Kogalur, U. B., Blackstone, E. H.& Lauer, M. S. (2008). 随机生存森林。《应用统计年鉴》2, 841–860。[21] Kalbfleisch, J. D. & Prentice, R. L. (2002). 失效时间数据的统计分析。威利概率与统计丛书。Wiley-Interscience (John Wiley & Sons), 新泽西州霍博肯，第2版。[22] Lin, D., Wei, L. & Ying, Z. (1998). 计数过程的加速失效时间模型。《生物统计学》85, 605–618。[23] Lin, D. Y., Wei, L.-J., Yang, I. & Ying, Z.(2000). 复发事件均值函数和率函数的半参数回归。《皇家统计学会会志：B辑（统计方法学）》62, 711–730。[24] Liu, L., Wolfe, R. A. & Huang, X. (2004). 复发事件和终末事件的共享脆弱性模型。《生物统计学》60, 747–756。[25] Prentice, R. L., Williams, B. J. & Peterson, A. V.(1981). 多变量失效时间数据的回归分析。《生物统计学》68, 373–379。[26] Rondeau, V., Mathoulin-Pelissier, S., Jacqmin-Gadda, H., Brouste, V. &Soubeyran, P. (2007). 使用最大惩罚似然估计的复发事件与死亡联合脆弱性模型：在癌症事件中的应用。《生物统计学》8, 708–721。[27] Scheike, T. H. (2002). 作为计数过程率函数的加性非参数与半参数Aalen模型。《终生数据分析》8, 247–262。[28] Schoop, R., Schumacher, M. & Graf, E. (2011). 含纵向协变量的生存数据预测误差度量。《生物统计学杂志》53, 275–293。[29] Schroder, J., Bouaziz, O., Agner, B. R., Martinussen, T., Madsen, P. L., Li,D. & Dixen, U. (2019). 复发事件生存分析可预测阵发性和持续性心房颤动患者未来的住院风险。《公共科学图书馆·综合》14, e0217983。[30] Steyerberg, E. W., Vickers, A. J., Cook, N. R., Gerds, T., Gonen, M., Obuchowski, N., Pencina, M. J. & Kattan, M. W. (2010). 评估预测模型的性能：针对若干传统和新颖度量的框架。《流行病学》（马萨诸塞州剑桥）21, 128。[31] Van Oirbeek, R. & Lesaffre, E. (2016). 通过布里尔评分探讨脆弱性生存模型的聚类效应。《统计学通讯—模拟与计算》45, 3294–3306。