arXiv:2601.07980v1 [stat.ME] 12 Jan 2026

用具有随机记忆的自激过程建模事件动态

K. Ken Peng、X. Joan Hu 和 Tim B. Swartz *

摘要

体育竞赛的事件历史数据近来在体育分析领域日益受到关注，可用于制定数据驱动的策略。这类数据往往呈现事件发生的自激现象，以及事件簇内部的依赖关系。基于间隔时间的传统事件模型可能难以捕捉这些特征。尤其是，尽管连续事件可能在短时间内发生，但先前事件引起的自激效应通常是暂时的，并会持续一段时长不确定的时间。本文提出一种扩展的霍克斯过程模型，通过随机自激持续时间刻画事件发生的动态。我们展示所提模型的示例，并介绍估计相关模型参数的程序。我们使用中国足球超级联赛2019年常规赛比赛中的角球数据，阐释该建模方法及其用途。我们还设计了用于模拟所提模型下事件过程的算法。该方法只需稍作修改，即可应用于犯罪学和传染病等许多其他研究领域。

关键词：变点分析、非泊松过程、部分隐藏的半马尔可夫过程、复发事件发生、足球分析。

*K. Ken Peng 是渥太华大学土木工程系和西门菲莎大学统计与精算科学系的博士后研究员。X. Joan Hu 和 Tim B. Swartz 是西门菲莎大学统计与精算科学系的统计学教授，地址为加拿大不列颠哥伦比亚省本拿比市 University Drive 8888 号，邮编 V5A 1S6。通讯作者：Tim B. Swartz。

1 引言

Hawkes 过程最早由 Hawkes（1971）提出，是一类点过程模型，其中事件发生的强度函数取决于过去发生的事件，且所有过去的事件都会自激励即将发生的事件。Hawkes 过程特别适合描述事件聚集现象。它一直是对多个领域中的事件发生进行建模的重要工具，包括地震（Ogata，1998；Zhuang 等人，2002；Kwon 等人，2023）、金融（Bacry 和 Muzy，2014）以及传染病（Schoenberg，2023；Garetto 等人，2021）。

人们提出了霍克斯过程的许多扩展，以适应不同的应用场景。例如，Daw 和 Pender（2022）研究了一种短暂自激霍克斯过程，而 Wu 等人（2022）研究了一种马尔可夫调制霍克斯过程，用于分析社会互动中的突发事件动态。Garetto 等人（2021）提出了一种时间调制霍克斯过程，用于模拟新冠肺炎的传播；Cai 等人（2024）则引入了一种基于霍克斯过程的框架，用于从多元事件数据中学习潜在网络结构。此外，Fierro 等人（2015）考察了激发函数变化时的霍克斯过程，并研究了其渐近性质。

在某些应用中，假设所有过去的事件都会持续影响后续事件并不现实。这促使人们发展出“单记忆”霍克斯过程，其中事件强度仅取决于最近一次事件，而非完整的历史记录。该模型与“延续效应”模型密切相关；在这类情形中，研究者通过考察事件发生后事件强度是否会暂时升高，来寻找短期影响的证据；例如，参见 C̨i˘g̨sar 和 Lawless（2012）。这种效应的延续通常无法直接观测，因此需要谨慎进行统计建模。

在体育分析领域，比赛中记录的事件数据能够揭示事件发生背后的机制，并为体育战术和博彩市场提供有价值的见解。其中，足球比赛中的角球尤其值得关注，因为它们代表着重要的得分机会，也能帮助我们了解定位球战术和比赛动态。因此，角球对于表现分析和战术规划至关重要。Peng 等人（2024）根据2019年中国足球超级联赛赛季233场比赛的事件历史数据，研究了从自然起始点到角球发生所经过的时间。Peng 等人（2024）的一项关键局限在于，他们强烈假设角球的等待时间在给定协变量的条件下相互独立——这一假设并非总是成立。应对这种潜在依赖关系的常见方法包括

连续事件之间的潜在依赖关系，常见方法包括引入脆弱性随机效应、采用联合分布模型，或将事件过程构建为增量不独立的计数过程。

分析角球时另一个特殊的复杂因素，是一次角球后不久再次出现角球的可能性。这是因为，开向球门区的角球可能会立即被防守球员挡出界外，从而造成随后再次获得角球。在计数过程框架中，对这一特征进行建模的一种自然方法是采用霍克斯过程：一次角球的发生会提高短时间内再次出现角球的可能性。然而，经典的霍克斯过程模型无法体现足球角球的一些独特特征。首先，虽然有些角球确实会引发紧随其后的快速角球，但并非所有角球都会如此：Peng 等人（2024）指出，只有约 10% 的角球会形成这种“快速”簇。其次，先前角球对后续角球的影响仅限于一个较短且未知的时间窗口。若经过了相当长的时间，合理的假设是当前角球独立于先前的角球。

受上述考虑的启发，我们提出一种扩展的霍克斯过程模型。该模型使用部分隐半马尔可夫过程来刻画先前事件的影响。我们开发了模型参数的估计方法，并通过中超联赛的角球数据加以说明。事实上，所提出的方法适用于许多具有与足球比赛中的角球相似特征的事件过程。例如，该方法可用于对呈现异质风险模式的事件时间数据建模，例如存在治疗相关变点的患者生存时间。本文其余部分安排如下：第 2 节介绍所提模型及其相关估计方法。第 3 节报告所提方法的实证研究、对 2019 年中超联赛赛季角球数据的分析，以及若干模拟实验。第 4 节总结本文的贡献，并探讨未来研究的潜在方向。

2 扩展的霍克斯过程模型

2.1 符号与建模

考虑计数过程 $\{N(t)\ t \geq 0\}$，其中 $N(t)$ 表示时间段 $(0,\ t]$ 内事件发生的次数。令 $\mathcal{H}(t)$ 表示截至

且不包括时间 t 的过程历史，它是由 $\{N(u): 0 < u < t\}$ 生成的 σ-域。用 T_1，……，T_K 表示整个随访期 $(0, E]$ 内的事件时间，其中 E 表示观测窗口的结束时间，事件总数为 $K = N(E)$，而在 $(0, E]$ 内观测到的潜在协变量记为 $\boldsymbol{Z}$。

我们提出了一种新颖的霍克斯过程模型扩展，其条件强度 $\lambda(t|\mathcal{H}(t), \tau, \boldsymbol{Z})$ 根据部分隐半马尔可夫结构在两种状态之间切换。具体而言，时刻 t 的强度为

$$\lambda(t|\mathcal{H}(t), \tau, \boldsymbol{Z}) = \begin{cases} \lambda_1(t|\mathcal{H}(t), \boldsymbol{Z}), & S(t) = 1 \\ \lambda_0(t|\mathcal{H}(t), \boldsymbol{Z}), & S(t) = 0, \end{cases} \tag{1}$$

其中，τ 是一个不可观测的随机持续时间，$S(t)$ 是定义如下的交替二元部分隐过程:

$$S(t) = \begin{cases} 1, & t - T_{N(t-)} \leq \tau \\ 0, & t - T_{N(t-)} > \tau. \end{cases}$$

这里，$N(t-)$ 表示时刻 t 之前发生的事件次数；$T_{N(t-)}$ 表示最近一次事件发生的时刻。二元过程 $S(t)$ 描述系统的潜在状态：事件发生后，系统立即切换到“热”状态（$S(t) = 1$），并在该状态中保持随机时长 τ。如果事件发生后的 τ 个时间单位内没有后续事件，系统就会返回“常规”状态（$S(t) = 0$）。请注意，尽管切换到“热”状态由事件确定性地触发，但“热”状态的持续时间由不可观测随机变量 τ 决定，因此 $S(t)$ 是一个半马尔可夫过程。模型组成部分 $\lambda_0(t|\mathcal{H}(t)\ \boldsymbol{Z})$ 是“常规”状态下的基线事件强度（即 $S(t) = 0$）；$\lambda_1(t|\mathcal{H}(t)\ \boldsymbol{Z})$ 则是“热”状态下的升高的强度（即 $S(t) = 1$）。在许多实际情形中，对于任意固定的协变量 $\boldsymbol{Z}$，$\lambda_1(t|\mathcal{H}(t)\ \boldsymbol{Z})$ 远大于 $\lambda_0(t|\mathcal{H}(t)\ \boldsymbol{Z})$，这反映了事件发生后事件发生率会暂时上升。也存在 $\lambda_1(t|\mathcal{H}(t)\ \boldsymbol{Z})$ 小于 $\lambda_0(t|\mathcal{H}(t)\ \boldsymbol{Z})$ 的例子。

我们注意到，虽然强度函数 $\lambda(t|\mathcal{H}(t)\ \tau,\ \boldsymbol{Z})$ 通常取决于完整的事件历史，但模型中的各个组成部分 $\lambda_0(t|\mathcal{H}(t)\ \boldsymbol{Z})$ 和 $\lambda_1(t|\mathcal{H}(t)\ \boldsymbol{Z})$ 可以独立于历史信息 $\mathcal{H}(t)$进行设定，其影响则由隐藏的二元过程 $S(t)$ 汇总。所提模型通过这样一种简单的潜在结构，捕捉事件发生可能具有的自激行为。因此，该模型推广了传统的霍克斯过程，为事件驱动动力学建模提供了更大的灵活性，尤其适用于激发效应具有暂时性且持续时间随机的情形。

2.2 所提模型下的间隔时间分布

将连续事件时间之间的间隔时间记为 $Y_k = T_k - T_{k-1}$，其中 $k = 1$，.....，且 $T_0 = 0$。第 1 个间隔时间 Y_1 是到第 1 个事件发生的时间（即 T_1 ），其条件风险函数为 $\lambda_0(t|\boldsymbol{Z})$，其中 $t > 0$。然而，第 k 个间隔时间 Y_k （其中 $k > 1$）的分布呈现分段结构：对于 $0 < t < \tau - T_{k-1}$，即自第 $(k-1)$ 个事件发生后、处于 τ 持续时间内的时段，事件强度升高至 $\lambda_1(t|\mathcal{H}(t)\ \boldsymbol{Z})$；若经过 τ 时间仍无新事件发生，强度则恢复为 $\lambda_0(t|\mathcal{H}(t)\ \boldsymbol{Z})$。

为说明这一特征，我们考虑一个简单版本的模型：在不存在协变量的情况下， λ_0和 λ_1 为常数，

$$\lambda(t|\mathcal{H}(t), \tau) = \lambda_1 S(t) + \lambda_0 \big[1 - S(t)\big]. \tag{2}$$

直到首次事件发生的等待时间（间隔时间）服从率参数为λ_0的指数分布。由于给定激发持续时间 τ后，事件强度 λ_0, λ_1 与时间无关，因此后续等待时间 Y_k 对于 $k > 1$ 服从相同的分布，其密度函数 $f(y|\tau)$为

$$f(y|\tau) = I(y \le \tau)\lambda_1 \exp(-\lambda_1 y) + I(y > \tau)\lambda_0 \exp(-(\lambda_1 - \lambda_0)\tau) \exp(-\lambda_0 y). \tag{3}$$

(3)中的分布具有分段结构，涉及率参数 λ_0, λ_1。该分布构成一个变点模型：从前一事件发生时起直到 $y = \tau$，等待时间服从率参数为 λ_1的指数分布；在变点 $y = \tau$之后，分布转为率参数为 λ_0 的指数分布，并乘以缩放因子 $\exp(-(\lambda_1 - \lambda_0)\tau)$。

我们将式 (3) 中的分布称为在τ处具有变点的分段分布。虽然我们的建模动机来自对足球角球的分析，但这种变点模型也可应用于其他领域。例如，在某些呼吸道感染等疾病的背景下，它或许有助于对手术后患者的生存时间进行建模，并考虑治疗阈值；这类疾病可能只会给特定群体（如儿童）带来较高风险。变点模型不仅能捕捉生存时间可能存在的混合分布，还能帮助识别事件率从一个状态转变到另一个状态的时间。

2.3 模型组件的设定

2.3.1 “热”状态的随机持续时间

随机持续时间 $\tau \geq 0$ 是潜在变量，这是所提模型的独特特征之一。因此，热状态指示变量 $S(\cdot)$部分不可观测：“热”状态的开始时间可以观测到，但结束时间无法观测；而“常规”状态的结束时间可以观测到，但开始时间无法观测，除非是在时间 $t = 0$。第 3 节所示的真实数据示例假设 τ 服从伽马分布。我们建议为 τ的分布采用恰当且灵活的模型，以捕捉以下两种极端情况。

- *Case of* $\tau \equiv 0$。当事件强度 $\lambda(t|\mathcal{H}(t)\ \tau,\ \boldsymbol{Z})$ 简化为 $\lambda_0(t|\mathcal{H}(t)\ \boldsymbol{Z})$ 时，表示过去事件没有带来额外的激发。
- *Case of* $\tau \gg 0$。当变点 τ 相对于随访期长度较大时（$0,\ E]$，事件强度近似为 $\lambda_1(t|\mathcal{H}(t)\ \boldsymbol{Z})$，这表明初始事件发生后，事件过程仍处于高事件强度（即“热”）状态。

2.3.2 两组件事件强度函数

所提模型（1）在设定 $\lambda_0(t|\mathcal{H}(t)\ \boldsymbol{Z})$ 和 $\lambda_1(t|\mathcal{H}(t)\ \boldsymbol{Z})$ 方面具有相当大的灵活性，以适应当前应用中的特殊特征。可以将它们分别设定为常用的事件强度模型。下面介绍相关的特殊情形。

- *Extended Hawkes Process*。设 $\lambda_1(t|\mathcal{H}(t)\ \boldsymbol{Z})$ 为自激（即 Hawkes）过程的强度函数，例如，$\lambda_1(t|\mathcal{H}(t)\ \boldsymbol{Z}) = \lambda_0(t|\boldsymbol{Z}) + \sum^{N_{(t-)}}{}_{k=0}\ w(t - T_k)$ 或 $\lambda_1(t|\mathcal{H}(t)\ \boldsymbol{Z}) = \lambda_0(t|\boldsymbol{Z}) + \sum^{N_{(t-)}}{}_{k=0}\ w(t - T_{N(t-)})$，其中 $w(\cdot)$ 是非增且非负函数。

- *Mixture of Two Poisson Processes*。考虑 $\lambda_0(t|\mathcal{H}(t)\ \boldsymbol{Z}) = \lambda_0(t|\boldsymbol{Z})$ 和 $\lambda_1(t|\mathcal{H}(t)\ \boldsymbol{Z}) = \lambda_1(t|\boldsymbol{Z})$。可以假设这两个分量强度函数服从 Cox回归（即 Andersen-Gill）模型：$\lambda_0(t|\boldsymbol{Z}) = \lambda_{00}(t)\ \exp(\beta_0 \boldsymbol{Z})$ 和 $\lambda_1(t|\boldsymbol{Z}) = \lambda_{10}(t)\ \exp(\beta_1 \boldsymbol{Z})$。

我们通过第 3 节数值研究中双泊松过程混合的以下四个示例来展示所提模型：

模型 (a)。增加一个参数 ν 作为状态“热”的指示函数的系数，

$$\lambda(t|\mathcal{H}(t),\tau,\boldsymbol{Z}) = \lambda_{00}\exp\{\boldsymbol{\beta}\boldsymbol{Z} + \nu S(t)\}. \tag{4a}$$

模型 (b)。将 $\lambda_0(t|\boldsymbol{Z})$ 的基线函数设定为含未知参数 θ_λ 的形式，

$$\lambda(t|\mathcal{H}(t),\tau,\boldsymbol{Z}) = \lambda_0(t;\theta_\lambda)\exp\{\boldsymbol{\beta}\boldsymbol{Z} + \nu S(t)\}. \tag{4b}$$

模型 (c)。不指定 $\lambda_0(t|\boldsymbol{Z})$ 的基线函数，

$$\lambda(t|\mathcal{H}(t),\tau,\boldsymbol{Z}) = \lambda_{00}(t)\exp\{\boldsymbol{\beta}\boldsymbol{Z} + \nu S(t)\}. \tag{4c}$$

模型（d）。允许两种状态下的协变量效应和基线函数有所不同，

$$\lambda(t|\mathcal{H}(t),\tau,\boldsymbol{Z}) = \lambda_{00}(t)\exp\{\boldsymbol{\beta}_0\boldsymbol{Z}\}\big[1 - S(t))\big] + \lambda_{10}(t)\exp\{\boldsymbol{\beta}_1\boldsymbol{Z}\}S(t). \tag{4d}$$

上述示例中的模型彼此嵌套，因此可以进行直接的模型检验。

2.4 参数估计

设 $\{N_i(t)\ t \geq 0\}$ 是所考虑计数过程的 $i = 1,\ldots,n$ 个相互独立的观测实现。对于第 i 个研究单位，将相关协变量记为 $\boldsymbol{Z}_i$；随访区间记为 $(0, E_i]$；事件发生时间记为 T_{ik}，其中 $k = 1,\ldots,K_i = N_i(E_i)$；随机“热”状态持续时间记为 τ_i。本节介绍如何通过最大似然估计（MLE）来估计双泊松过程混合模型的模型参数，该模型的条件强度函数指定为

$$\lambda(t|\mathcal{H}(t),\tau,\boldsymbol{Z}) = \lambda_1(t;\theta_{\lambda_1})\exp(\boldsymbol{\beta}_1\boldsymbol{Z})S(t) + \lambda_0(t;\theta_{\lambda_0})\exp(\boldsymbol{\beta}_0\boldsymbol{Z})\big[1 - S(t)\big]. \tag{5}$$

上述模型包含第2.3.2节中的例1–2作为特殊情况。此外，我们假设 τ 的概率密度函数为 $f_\tau(\cdot;\theta_\tau)$。

令 $\boldsymbol{\theta} = (\theta_{\lambda_0}, \theta_{\lambda_1}, \boldsymbol{\beta}_0', \boldsymbol{\beta}_1', \theta_\tau)$。给定 τ_i，研究单位 i 对

似然函数为

$$L_i(\boldsymbol{\theta}|\tau_i) = \prod_{k=1}^{K_i} \Big\{ \lambda_1(T_{ik};\theta_{\lambda_1}) \exp(\boldsymbol{\beta}_1 \boldsymbol{Z}_i) S_i(T_{ik}) + \lambda_0(T_{ik};\theta_{\lambda_0}) \exp(\boldsymbol{\beta}_0 \boldsymbol{Z}_i) \big[1 - S_i(T_{ik})\big] \Big\}$$
$$\times \exp\Big\{ - \int_0^{E_i} \lambda_1(t;\theta_{\lambda_1}) \exp(\boldsymbol{\beta}_1 \boldsymbol{Z}_i) S_i(t) + \lambda_0(t;\theta_{\lambda_0}) \exp(\boldsymbol{\beta}_0 \boldsymbol{Z}_i) \big[1 - S_i(t)\big] dt \Big\}, \tag{6}$$

其中 $S_i(t) = I(t - T_{N_i(t-)} \leq \tau_i)$。 $\boldsymbol{\theta}$ 的似然函数与观测数据为

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} \int L_i(\boldsymbol{\theta}|\tau_i) f_\tau(\tau_i;\theta_\tau) d\tau_i. \tag{7}$$

可最大化似然函数（7），以获得 $\boldsymbol{\theta}$的最大似然估计。这一过程的计算开销可能很大。将对数变换后的完整数据似然函数记为 $l_{full}(\boldsymbol{\theta};\boldsymbol{\tau}) = \sum_{i=1}^{n} \big[\log\ L_i(\boldsymbol{\theta}|\tau_i) + \log\ f(\tau_i;\theta_\tau)\big]$ 其中 $\boldsymbol{\tau} = (\tau_1$、.....、$\tau_n)'$。一种常用的计算参数最大似然估计 $\boldsymbol{\theta}$ 的方法是按如下方式改进蒙特卡洛期望最大化算法。

算法 1： 蒙特卡洛期望最大化算法

1. L令第 d次迭代的估计值为 $\boldsymbol{\theta}^{(d)}$，其中 $\boldsymbol{\theta}^{(0)}$ 为初始值。；

2. 重复

2.1 MCE-step. Approximate $Q(\boldsymbol{\theta},\boldsymbol{\theta}^{(d)}) = E\big(l_{full}(\boldsymbol{\theta};\boldsymbol{\tau}) \mid \text{observed data}, \boldsymbol{\theta}^{(d)}\big)$ using the sample mean$\tilde{Q}(\boldsymbol{\theta},\boldsymbol{\theta}^{(d)}) = \frac{1}{M}\sum_{m=1}^{M} l_{full}(\boldsymbol{\theta};\boldsymbol{\tau}^{(m)})$, where the ith component of $\boldsymbol{\tau}^{(m)}$ is generated independently from the conditional distribution $[\tau_i \mid \text{observed data}, \boldsymbol{\theta}^{(d)}] \propto \big[L_i(\boldsymbol{\theta}^{(d)} \mid \tau_i)\big] f_\tau(\tau_i \mid \boldsymbol{\theta}_\tau^{(d)})$. ;

2.2 最大化步骤。将 $\boldsymbol{\theta}$ 的估计值更新为 $\boldsymbol{\theta}^{(d+1)} = \arg\max_{\boldsymbol{\theta}} \tilde{Q}(\boldsymbol{\theta},\ \boldsymbol{\theta}^{(d)})$。；

3. 迭代直至 $\{\boldsymbol{\theta}^{(d)}, d = 1, 2, \ldots\}$ **收敛**。

ˆ 根据 Louis (1982)，可获得最大似然估计 $\boldsymbol{\theta}$ 标准误的一致估计量：此时观测费舍尔信息矩阵为

$$E\big[- \frac{\partial^2 l^{full}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2} | \text{observed data}, \hat{\boldsymbol{\theta}}\big] - Var\big[\frac{\partial l^{full}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} | \text{observed data}, \hat{\boldsymbol{\theta}}\big].$$

当考虑式（5）中的双泊松过程混合模型，且基线函数未指定时，例如第2.3.2节模型（c）或（d），每个未指定的基线函数都可以用分段常数函数或样条（分段

多项式）函数进行近似。上述估计过程可以相应地调整，用于估计模型中的所有未知量。

3 数值研究

3.1 中国足球超级联赛2019年数据分析

3.1.1 数据与时间段构建

2019年中国足球超级联赛（中超联赛）赛季共有16支球队，每支球队都与其他每支球队交手两次（主场和客场各一次），共安排了240场比赛。由于视频或记录问题，有7场比赛的事件记录无法获取；这些比赛被排除在分析之外，并视为完全随机缺失。对于每场有记录的比赛，角球均以带时间戳的复发事件形式记录，这自然适合采用计数过程进行建模。

一个实际问题是，比赛中的角球过程会被重置比赛情境的“终止”比赛事件打断，例如进球和半场结束。此外，两支球队的角球过程在实时中相互关联。为避免直接建模球队间依赖，我们对主队和客队进行分层分析，并进一步将每场比赛划分为由终止事件隔开的时间段。

具体而言，对于比赛 i ($i = 1$、.....、233)，将时间段定义为从时间段开始事件起始、到下一个终止事件结束的区间。在我们的应用中，时间段开始事件包括半场开始和进球，终止事件包括进球和半场结束。令 $j = 1$、.....、J_i 表示比赛 i中的时间段索引。对于给定球队（主队或客队），令 $\{N_{ij}(t)\ 0 \le t \le E_{ij}\}$ 表示时间段级角球计数过程，其中 $N_{ij}(t)$表示自时间段 j开始以来，在分段时间 t 之前观测到的角球数，且 $N_{ij}(0) = 0$，随访截至 E_{ij}。令 $\boldsymbol{Z}_{ij}$ 表示时间段级协变量向量（表1）。在233场比赛中，共观测到705个进球；加上466次半场结束事件，数据集中共有1171个时间段。

<此处插入表*1*>

第 2 节所提模型是在区段层面拟合的。具体而言，潜在“热状态”的持续时间允许在不同区段之间变化： τ_{ij} 被视为不可观测随机变量

，用于每个区段（$i,\ j$）。对于所有区段，我们假设 τ_{ij} 独立同分布$\sim$ Γ(形状，率)，并根据数据估计形状和率参数。在这一设定下，每个区段都构成所提事件过程的一次新实现，而终止事件（进球和半场结束）会自然地重启分段时间尺度。

3.1.2 模型设定与协变量

我们将第2.3.2节介绍的四种模型设定（模型 (a)–(d)）应用于 CSL 数据。这些模型构成一个嵌套序列，在基准强度和协变量效应的设定上灵活性逐步提高，从基准强度为常数的模型（模型 (a)），到允许基准强度随时间变化并具有状态特异性协变量效应的模型（模型 (b)–(d)）。所有模型均使用第2.4节所述的蒙特卡罗期望最大化程序进行拟合。

探索性分析显示，主队和客队的角球时间模式存在系统性差异，因此我们采用按主客场区分的协变量集合。候选协变量及其定义汇总于表1。我们根据模型 (a) 的模型设定，使用贝叶斯信息准则（BIC）进行全子集选择，以完成变量选择。该程序分别应用于主队和客队。随后，在更灵活的模型设定中也使用了相同的协变量集合。

对于主队，选定的协变量包括比分差，以及一个指示区段是否紧接在进球后开始的指示函数。对于客队，选定的协变量包括下半场指示函数、比分差、时段开始时间，以及下半场指示函数与时段开始时间之间的交互项。这些差异反映了主队和客队所面临的不同情境动态。

四种模型设定在基准强度和状态依赖的处理方式上有所不同：

- **模型 (a)：** 恒定的基准强度。
- **模型 (b)：** 通过参数化函数形式指定的时变常规状态基准强度。具体而言，对数基准强度建模为

$$\log \lambda_0(t;\theta_\lambda) = \theta_1 + \theta_2\big\{\exp(-\theta_3 t) - \exp(-\theta_4 t)\big\},$$

其中包含四个未知参数，并允许区段内呈现灵活的非单调时间模式。

- **模型 (c)：**一种半参数设定，其中常规状态的基准强度不作具体设定，而是按照第2.4节所述，用分段常数函数进行近似。根据对时段持续时间和角球时间的探索性分析，切点设为非均匀分布，在区段前段采用更高的分辨率，因为此处的时间变化更为显著。常规状态的基准强度用八个区间近似，切点设在 0, 3, 6, 9, 12, 15, 27.5, 和 40 分钟处。

- **模型（d）：**这是模型 (c) 的扩展，此外还允许热状态具有独立的基准强度。热状态的基准强度同样用分段常数函数近似，分为六个区间，切点设在 0, 5, 10, 15, 20, 和 30 分钟处。

3.1.3 参数估计与解释

表2报告了模型 (a)–(c) 的参数估计。总体而言，估计的回归效应以及潜在热状态持续时间的分布 τ 在这些模型设定下高度一致，表明结果对基准强度的选择具有稳健性。

<在此插入表*2*>

对于主队，常规状态下的估计系数表明，当球队领先以及区段在进球后立即开始时，角球发生率较低。这些效应具有统计显著性，并且在全部三种模型设定中均持续存在。这一发现与现有证据一致：领先球队往往会采取较为保守的策略，从而降低进攻压力，参见 Silva 和 Swartz（2016）；Guan 等（2023）。对于客队，估计结果表明，其在比赛下半场的角球发生率较高，而在领先时则较低。此外，比分差与区段开始时间之间的负向交互作用表明，随着比赛推进，领先带来的抑制效应会愈发明显。与主队相比，时段剩余时间对客队角球强度的影响更为重要。

潜在热状态持续时间的估计分布 τ 在不同模型和球队之间保持稳定。对于主队和客队，热状态的估计平均持续时间

接近两分钟，标准差也相近。这一发现表明，角球发生后，后续角球发生概率的任何短暂上升通常都局限于一个较短且明确的时间窗口。主队和客队之间的相似性表明，在这组数据中，此类短期激发的时间尺度基本不受球队影响。从足球角度来看，人们可能会认为此类激发效应的持续时间要短得多（例如几秒钟）；然而，观测到的事件时间包含球员站位和定位球准备，这实际上拉长了记录的角球间隔时间尺度。

表3展示了模型（d）的结果，该模型允许协变量效应因状态而异。常规状态的估计系数总体上与模型（a）–（c）的结果一致。相比之下，对于任一球队，在热状态下，候选协变量均未呈现统计显著的效应。这表明，在进入热状态的条件下，角球发生率升高主要由激发机制本身驱动，而非由额外的情境协变量所致。

<此处插入表*3*>

热状态下未发现统计显著的协变量效应，这与足球分析领域的专业认知一致。角球发生后，比赛过程具有很强的随机性，战术或情境因素难以持续发挥系统性影响。鉴于热状态持续时间较短，且快速接续的角球数量相对较少，若要检测出协变量效应，就需要更大的足球比赛样本量。事实上，一些估计效应（例如客队的比分差）幅度不可忽视，也呈现出一定的方向性趋势，但现有数据不足以支持其达到统计显著水平。例如，如果能获得更丰富的数据，就可能考察并识别球队特有的倾向，例如偏好短角球；这种倾向可能会降低紧接着再次出现角球的可能性，而以当前样本量难以检测到。

3.1.4 估计的基准强度

图1展示了四种模型设定下主队和客队的估计基准强度函数。不同模型中，估计强度的整体形态和大小大致相近。请注意，图1中的时间尺度对应的是分段时间，而非比赛时间或半场时间。实际上，只有少数

时间段可能延长至约50分钟，通常相当于一整个没有进球的半场加上加时赛。

<此处插入图*1*>

同时，模型 (b)–(d) 的结果清楚地表明，基准强度会随分段时间而变化。模型 (a) 中的固定基准假设似乎不足以刻画区段内角球强度的时间演化。模型 (b) 的时变设定提供了灵活的表示，并且与模型 (c) 和 (d) 得到的非参数估计高度吻合，这表明所选的参数形式能够充分捕捉数据中的主要时间模式。

作为参考，图1还展示了使用 Breslow 估计量从 Andersen-Gill 模型（Andersen 和 Gill，1982）估计得到的平滑基准强度（红色曲线）。Cox基准强度与所提模型估计的常规状态基准强度高度吻合，但系统性地略高一些。这种差异反映了这样一个事实：由于未区分常规状态与热状态，Cox模型会将部分瞬态激发吸收到单一基准分量中，导致估计的常规状态强度略有向上偏差，并忽略有关短期聚集的信息。

在模型（a）–（c）中，估计得到的热状态乘性效应 $\hat{\nu}$ 表明，热状态下的角球强度约为常规状态的两倍。相比之下，模型（d）允许热状态与常规状态的基准强度独立变化，且不施加比例结构。这一观察表明，热状态强度随时间变化时并非严格与常规状态强度成比例，凸显了模型（a）–（c）中比例性假设可能存在的局限性。

总体而言，估计得到的基准强度呈现出共同的时间模式：区段开始后，角球强度较低，随后在最初几分钟内上升，之后趋于稳定。在整个区段中，主队的基准角球强度通常高于客队。

3.1.5 应用：动态预测示例

除了对基准强度进行参数估计和推断外，所提出的建模框架还支持一系列标准事件时间模型难以实现的应用。例如，随着新事件的发生，可以实时更新预测。

图2通过动态预测2019年3月1日深圳与上海比赛中主队的角球强度，展示了这一能力；该预测基于模型 (b) 中的拟合模型。图中展示了六个具有代表性的比赛时间点。在每个时间点，黑色实线表示基于截至该时刻所观测到的全部信息估计的强度，而颜色较浅的延伸线则表示该区段剩余时间内的预测未来强度。每当发生新事件（例如角球或进球）时，预测都会更新。

<在此插入图2>

此示例展示了模型如何同时捕捉区段内角球强度的基线时间演化，以及近期事件发生后的短期激发。具体而言，角球发生后，预测强度会立即上升，随后随着系统恢复到常规状态而逐渐衰减；这一强度升高阶段的持续时间由潜在热状态持续时间的估计分布决定。对于给定的时间点，还可以计算下一个角球的平均时间及其对应的不确定性区间，从而为预测强度的合理性提供额外的定量检验；后续的模拟研究将进一步提供这些强度估计合理性的证据。

此类动态预测可应用于多种场景，包括实时比赛分析、战术决策支持和博彩市场应用。更广泛地说，利用拟合模型，可以推导出比赛中各种具有实际意义的指标。例如，估计的基准强度可视为衡量球队在比赛不同阶段潜在进攻压力的指标，而热状态参数则刻画短期动量以及球队维持压力的能力。加入更多协变量后，该框架还可扩展用于评估阵型变化、换人或球员特定参与等背景因素的影响，从而以一种有依据的方式将事件层面动态与战术及表现评估联系起来。

3.2 足球比赛中的角球模拟

本小节介绍一种基于第 2 节所提模型的模拟框架。此处旨在提出一种实用算法，用于在给定拟合模型参数的情况下模拟事件过程。该算法为生成

保留模型所蕴含的关键结构特征的合成事件数据，并有助于开展应用研究和情景探索。

3.2.1 模拟算法

我们考虑两种模拟算法，分别对应基准强度的不同设定。对于第 2.2 节模型（2）中的较简单情形，常规状态和热状态的强度均为时间常数，因此可以直接利用模型所隐含的变点间隔时间结构进行模拟。该过程总结于算法2，与第 2.2 节讨论的分段间隔时间分布密切相关。

算法2： 在简化模型 (2) 下模拟事件过程

输入： λ_0，λ_1，τ，E **初始化：** $t_0 = 0$，$n = 0$，$h = 0$ **当** $t_n < E$ **时执行**

- **如果** $h = 0$，**则**
  - 生成 w ∼指数分布(λ_0);
  - 设置 $t_{n+1} = t_n + w$;
- **否则** $h = 1$
  - 生成 w_1 ∼指数分布(λ_1);
  - **如果** $w_1 \le \tau$ **则**
    - 设置 $t_{n+1} = t_n + w_1$;
    - `// Event occurs within the hot window`
  - **else**
    - 生成 w_0 ∼指数分布(λ_0);
    - 设置 $t_{n+1} = t_n + \tau + w_0$;
    - `// No event during hot; switch to regular after` τ
- **如果** $t_{n+1} > E$ **则**
  - **跳出**;
  - `// Beyond the follow-up period`
- $n = n + 1$;
- `// Event count increases by 1`
- 设置 $h = 1$;
- `// Immediately after an event, enter the hot state`

返回值： $\{t_k\}_{k=1,\dots,n}$

对于基准强度随时间变化的更一般情形，已无法直接进行间隔时间模拟。此时，我们采用基于稀疏化的方法（Lewis 和 Shedler，1979），根据所提出的强度函数生成事件时间；该过程总结于算法 3。对于角球过程而言，这种额外的灵活性有充分的依据，因为真实数据分析表明，基准强度在一个区段内随时间显著变化，并非恒定。

算法 3： 模拟具有随时间变化的基线强度的事件过程（稀疏化）

输入： $\lambda_0(t)$ $\lambda_1(t)$ τ, E, $\bar{\lambda} = \sup_{0 \le t \le E} \max\{\lambda_0(t), \lambda_1(t)\}$ **初始化：** $t_0 = 0$, $n = 0$, $m = 0$, $s_0 = 0$, $h = 0$ **当** $s_m < E$ **时执行**

- 生成 $w \sim$ 指数分布$(\bar{\lambda})$;
- 设置 $s_{m+1} = s_m + w$;
- **如果** $s_{m+1} > E$ **那么**
  - **中断**;
  - `// Beyond the follow-up period`
- **如果** $s_{m+1} < h$ **那么**
  - 设置 $\lambda^\star = \lambda_1(s_{m+1})$ `// Hot state when` $s_{m+1} < h$
- **else**
  - 设置 $\lambda^\star = \lambda_0(s_{m+1})$ `// Regular state when` $s_{m+1} \ge h$
- 生成 $D \sim$ 均匀分布$(0, 1)$;
- **如果** $D \le \lambda^\star / \bar{\lambda}$ **那么**
  - $t_{n+1} = s_{m+1}$;
  - $n = n + 1$;
  - `// Event accepted`
  - $h = t_n + \tau$;
  - `// Enter hot state for` τ `time units`
- $m = m + 1$;

返回值： $\{t_k\}_{k=1,\ldots,n}$;

图3以示意图的形式展示了单个区段的模拟机制。常规状态强度 $\lambda_0(t)$ 和相应的热状态强度 $\lambda_1(t)$ 随分段时间的变化如图所示。实线表示给定时间的活跃强度，虚线表示非活跃强度。阴影区域对应于由接受事件触发的热状态

窗口。候选事件时间首先由速率为 $\lambda = \sup_{0 \le t \le E} \max\{\lambda_0(t)\ \lambda_1(t)\}$ 的 a¯齐次泊松过程生成，并以点的形式标示在时间轴上。随后，每个候选时间以概率$\lambda^{\star}(t)/\lambda$ 被接受为事件；若过程处于热状态，则 $\lambda^{\star}(t) = \lambda_1(t)$；否则为 $\lambda^{\star}(t) = \lambda_0(t)$。接受的候选事件以实心点标示，拒绝的候选事件则以叉号标示。这种构造确保接受的事件时间遵循所需的时变强度，其瞬态激发由潜在热状态持续时间决定。

<请在此处插入图*3*>

在足球角球应用中，这些事件级模拟算法嵌入于比赛级结构中。首先，每场比赛按进球和半场边界等终止事件划分为多个时间段，然后根据分段特定协变量，在每个时间段内独立模拟角球。整体比赛级模拟策略遵循 Peng 等人（2024）补充材料中所述的框架，而分段级角球过程则使用本文提出的新模拟算法生成。

3.2.2 模拟结果

为展示模拟框架，我们使用模型 (b) 下的拟合模型生成了重复模拟赛季的角球数据。具体而言，模拟过程重复进行 200 次，每次重复模拟包含 233 场比赛，与 2019 年中超联赛赛季中观测到的比赛场数相同，以便与经验数据直接比较。

进球和半场边界等终止比赛事件也按照 Peng 等人（2024）的方法进行模拟。比赛时长和进球时间根据从 CSL 数据估计的经验分布生成，并且独立于角球过程。这些终止事件将每场比赛划分为多个时间段，并在每个时间段内使用上述事件级算法模拟角球。

总体而言，在若干简单且易于解读的汇总指标上，模拟的角球过程与 CSL 数据的经验结果非常接近。对于主队和客队，模拟比赛重现了观测到的每场比赛角球平均数（主队：观测值为 5.21，模拟值为 (4.95,5.56)；客队：观测值为 4.72，模拟值为 (4.36, 4.94)），其中模拟值报告的是 200 个重复赛季中 2.5% 和 97.5% 的经验百分位数。

在每个半场前 10 分钟内，角球频率也呈现出类似的一致性。

总体而言，在若干简单且易于解读的汇总指标上，模拟得到的角球过程与 CSL 数据的经验特征十分接近。对于主队和客队，模拟比赛都重现了观测到的每场比赛平均角球数（主队：观测值为 5.21，模拟值为 (4.95,5.56)；客队：观测值为 4.72，模拟值为 (4.36, 4.94)）。这里的模拟值表示 200 个重复赛季中 2.5% 和 97.5% 的经验百分位数，各场比赛的经验离散程度也相近。另一个例子是，模拟数据在每个半场的前 10 分钟内呈现出相似的角球频率（主队：观测值为 0.98，模拟值为 (0.86, 1.15)；客队：观测值为 0.92，模拟值为 (0.76,0.99) ）。

上述汇总指标侧重于总体频率和宽泛的时间分布，而这些量大多可以直接从估计的模型参数中推导出来。模拟算法的一项关键实用优势在于，它能够考察那些仅凭参数估计难以通过闭式形式获得的感兴趣量。例如，图 4 展示了观测到的 CSL 数据和模拟比赛的簇大小分布。簇大小分布通过将预设时间阈值内连续发生的事件归为一组，汇总短时间内发生的角球数量，从而直接衡量事件过程中的短期依赖性。

<在此插入图 *4*>

图 4 展示了四个面板，分别对应不同的簇持续时间阈值。每个面板中，横轴表示每个簇中的事件数量，纵轴表示相应的经验概率质量函数。图中分别绘制了主队和客队的曲线。模拟分布与观测分布高度吻合，表明所提出的模拟框架成功再现了真实比赛中观察到的短期聚集行为。除了簇大小分布，所提出的模拟框架还可用于考察多种其他难以根据拟合模型参数进行解析推导的事件级汇总指标，例如连续事件的连续长度、给定近期历史条件下的等待时间模式，或分段累积指标。通过模拟可以考察这些量，进一步凸显了所提方法的实用价值。

4 讨论

本项目提出了一种新的复发事件建模框架，该框架采用具有随机记忆激发机制的自激计数过程。我们模型的关键特点是，将“热”状态过程 $S(t)$ 定义为部分隐半马尔可夫过程，该过程同时取决于事件历史和潜在随机变量 τ。这种设定使模型能够以灵活且由数据驱动的方式捕捉短期激发效应。与需要指定衰减激发函数的传统霍克斯过程不同，我们的模型依赖由 τ 决定的有限记忆时长，因此更易于解释和估计。我们的模型有一种特殊情形与 C˛i˘g˛sar 和 Lawless（2012）的延续效应模型相似，但我们采用随机 τ，因此具有更大的灵活性。与 Wu 等人（2022）的马尔可夫调制霍克斯过程相比，后者使用完全潜在的马尔可夫状态过程，而我们的模型提供了一种更直观的半马尔可夫结构，可将激发直接与近期事件历史联系起来。

本研究的第二项贡献在于开发了针对所提模型的模拟算法。除了参数估计之外，模拟在理解瞬态激发对可观测事件模式的影响方面也发挥着核心作用。我们提出了实用的算法，用于在基准强度固定和随时间变化这两种情况下模拟事件过程，其中包括一种能够适应潜在热点状态机制的基于稀疏化的构造。这些算法能够生成合成事件数据，同时保留拟合模型所隐含的结构特征。重要的是，这些算法还可以用于研究仅凭模型参数难以进行解析推导的量，例如短期聚集汇总指标。正如角球应用所示，模拟在模型设定与具有实证解释性的事件层面行为之间搭建了一座自然的桥梁。

在足球分析的背景下，与 Peng 等人（2024）所采用的事件时间独立假设不同，我们的模型考虑了角球事件对过去发生情况的依赖性。这种依赖性使我们能够更好地捕捉比赛进程以及事件的聚集现象。此外，我们的模型支持 Cox型设定，其中基准强度可以随时间灵活变化，因此非常适合捕捉比赛进行过程中角球发生率的变化。这一结构使我们能够纳入比赛中协变量，并更好地理解比赛条件如何影响角球的动态过程。

本研究提出了若干未来研究方向。进一步探究激发时长分布的理论性质，包括其可辨识性以及

对分布假设的敏感性，将会很有价值。在计算方面，将模拟框架扩展为联合建模多个相互作用的事件过程（例如一场比赛中的主队和客队），是自然的下一步，因为现实中的比赛动态通常包含抑制性或竞争性交互作用。

除足球外，所提出的框架还适用于各种激发具有瞬时性且由事件触发的复发事件场景。潜在应用包括流行病学、犯罪学、金融和可靠性分析。在这些领域中，短期依赖性发挥着重要作用，但长记忆激发核无法很好地捕捉这种依赖性。在此类情境下，简约的建模结构与实用的模拟工具相结合，为复杂事件动态的推断和探索性分析提供了灵活的基础。

致谢

本研究部分得到加拿大自然科学与工程研究委员会（NSERC）授予 Hu 和 Swartz 的资助，以及由 Swartz 领导的加拿大统计科学研究所（CANSSI）体育分析协作研究团队（CRT）的支持。作者感谢山东鲁能泰山足球俱乐部前技术总监 Daniel Stenz 提供本文所讨论的数据。

声明

作者声明不存在利益冲突。

参考文献

Andersen, P.K., Gill, R.D., 1982. 用于计数过程的Cox回归模型：一项大样本研究。《统计年鉴》，1100–1120。

Bacry, E., Muzy, J.F., 2014. 用于价格和交易高频动态的Hawkes模型。《Quantitative Finance》14，1147–1166。

Cai, B., Zhang, J., Guan, Y., 2024. 基于高维多元点过程的潜在网络结构学习。《美国统计学会杂志》119, 95–108。

C.i˘g.sar, C., Lawless, J.F., 2012. 评估复发事件过程中瞬态残留效应，并将其应用于慢性健康状况。Daw, A., Pender, J., 2022. 一种短暂自激点过程。《应用概率进展》54, 340–403。Fierro, R., Leiva, V., Møller, J., 2015. 具有不同激发函数的霍克斯过程及其渐近行为。《应用概率杂志》52, 37–54。Garetto, M., Leonardi, E., Torrisi, G.L., 2021. 一种用于模拟新冠肺炎传播及应对措施影响的时间调制霍克斯过程。《控制年度评论》51, 551–563。Guan, T., Cao, J., Swartz, T.B., 2023. 摆大巴。《体育定量分析杂志》19, 263–272。Hawkes, A.G., 1971. 若干自激及相互激发点过程的谱。《生物统计学》58, 83–90。Kwon, J., Zheng, Y., Jun, M., 2023. 用于模拟地震发生的灵活时空霍克斯过程模型。《空间统计学》54, 100728。Lewis, P.W., Shedler, G.S., 1979. 通过稀疏化模拟非齐次泊松过程。《海军研究后勤季刊》26, 403–413。Louis, T.A., 1982. 使用期望最大化算法时如何求得观测信息矩阵。《英国皇家统计学会会志 B辑：统计方法学》44, 226–233。Ogata, Y., 1998. 用于模拟地震发生的时空点过程模型。《统计数学研究所年刊》50, 379–402。Peng, K.K., Hu, X.J., Swartz, T.B., 2024. 足球比赛中角球发生时间：事件历史数据分析。《计算统计学》，1–17。Schoenberg, F., 2023. 使用霍克斯点过程估计新冠肺炎传播时间。《应用统计年刊》17, 3349–3362。Silva, R.M., Swartz, T.B., 2016. 足球比赛中的换人时间分析。《体育定量分析杂志》12, 113–122。

Wu, J., Ward, O.G., Curley, J., Zheng, T., 2022. 用于对社会互动中零星且突发的事件发生进行建模的马尔可夫调制霍克斯过程。《应用统计年刊》16, 1171–1190。

Zhuang, J., Ogata, Y., Vere-Jones, D., 2002. 时空地震发生的随机去簇。《美国统计学会杂志》97, 369–380。

A 图

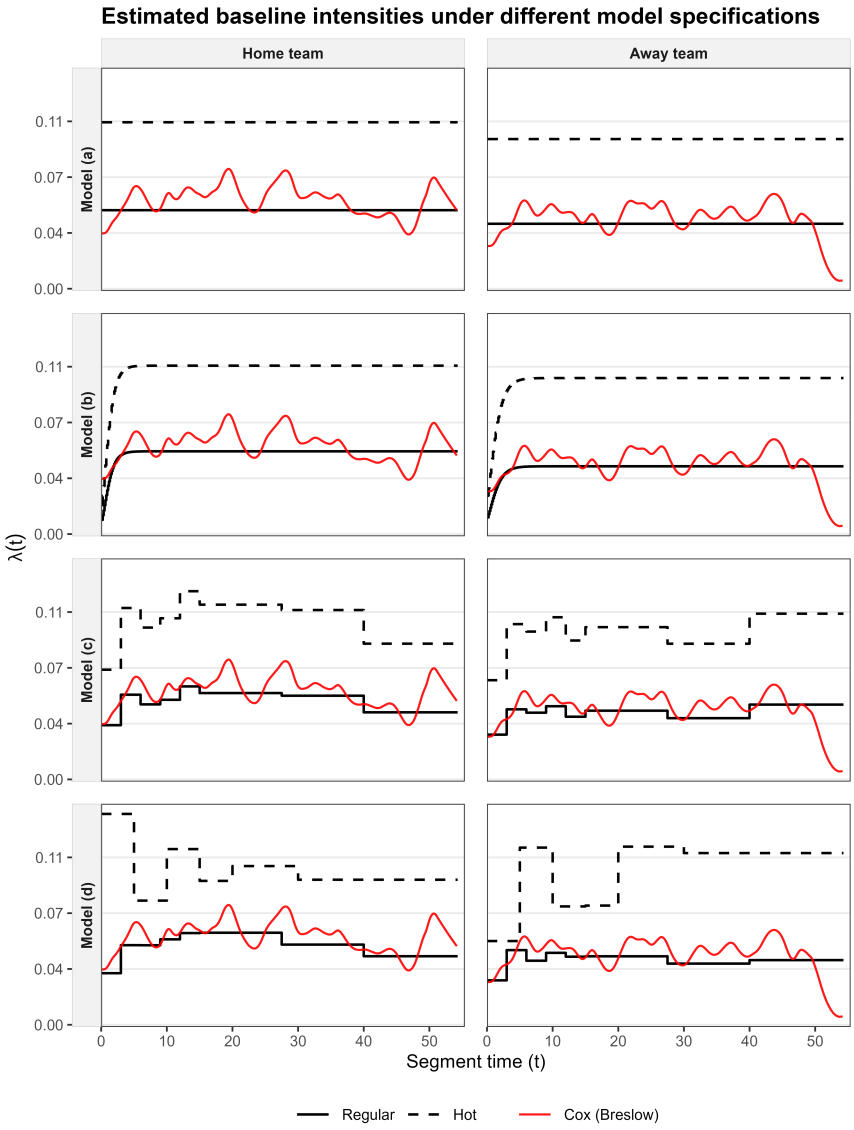

图1：四种模型设定下主队和客队的估计基准强度函数。每行对应一种不同的基线设定（模型 (a)–(d)），每列对应一支球队（主队与客队）。每个面板中，黑色实线和虚线分别表示常规状态和热状态下估计的基准强度。红色曲线表示根据 Andersen-Gill 模型并使用 Breslow 估计量得到的平滑基准强度，供参考。

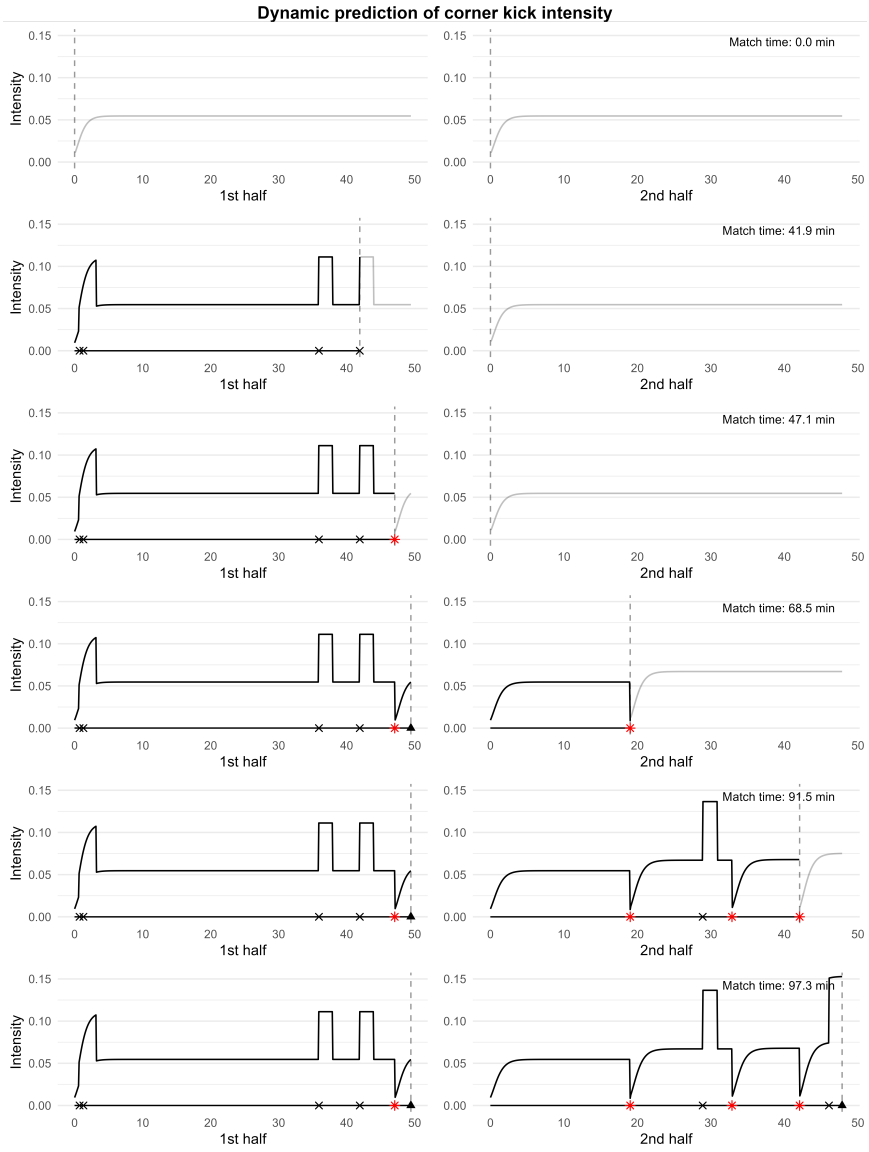

图2：足球比赛“深圳对阵上海”（2019年3月1日）中主队角球强度（纵轴）的动态预测。图中展示了六个具有代表性的比赛时间点。每一行中，黑色实线表示根据截至所示时间（竖直虚线）观察到的信息得出的估计强度，较浅的延伸线则表示在给定比赛当前状态的条件下预测的未来强度。黑色叉号表示已观察到的角球，红色星号表示进球，黑色三角形表示每个半场的结束。

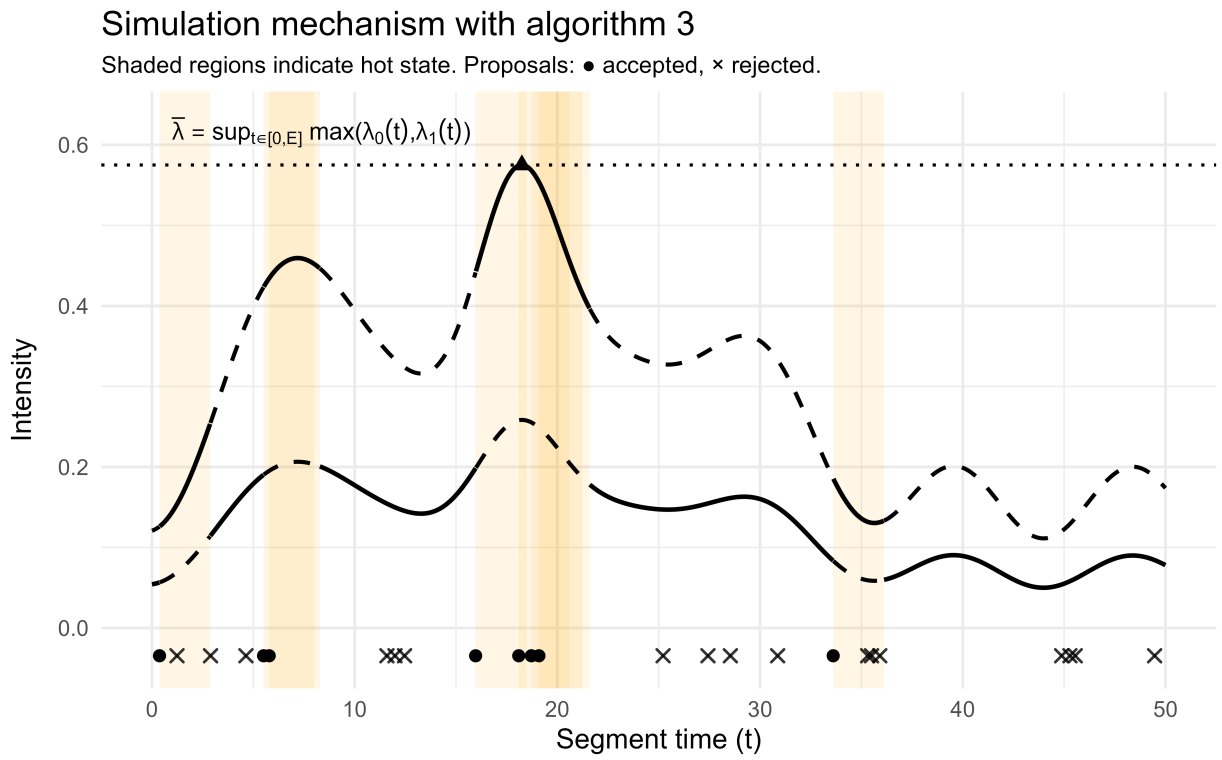

图3：具有潜在热点状态的时变强度的基于稀疏化的模拟算法示意图。实线和虚线分别表示活跃强度和非活跃强度；实线对应于在时间 t 用于稀疏化的有效强度 $\lambda^{\star}(t)$。阴影区域表示接受事件之后的热点状态窗口。在支配率下生成的候选事件以概率 $\lambda^{\star}(t)/\bar{\lambda}$ 被接受（实心点），否则被拒绝（叉号）。

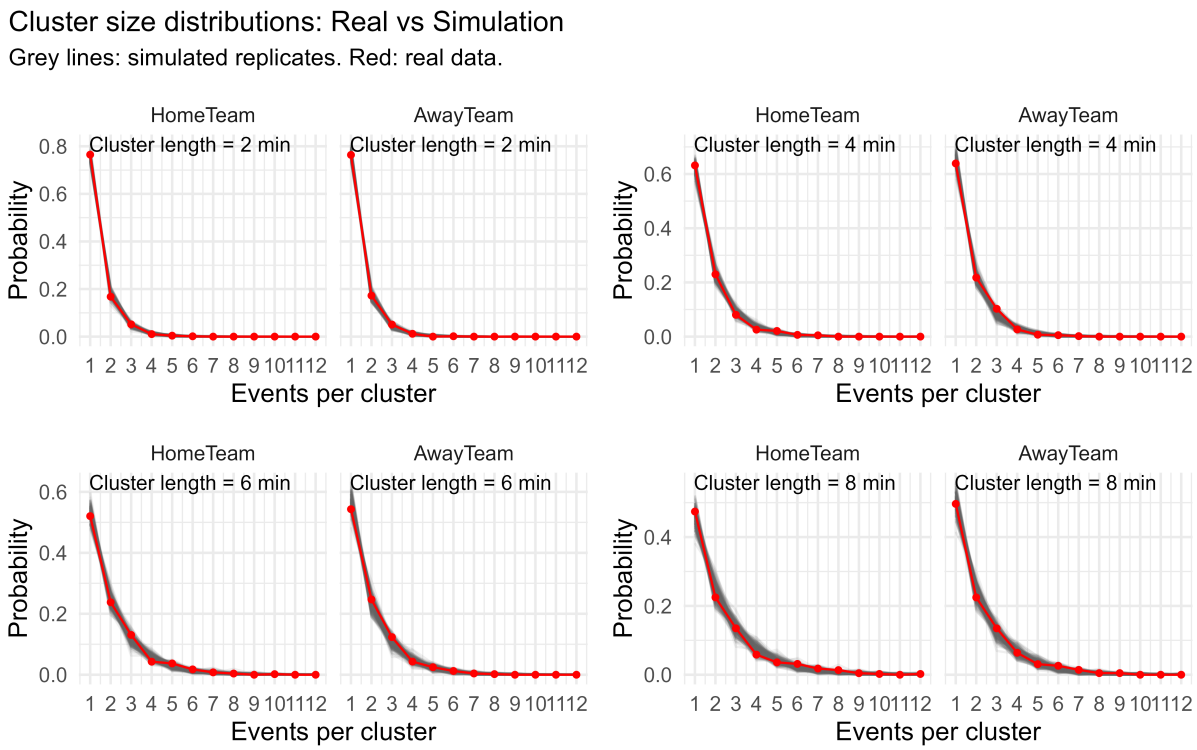

图 4：观测到的CSL比赛和模拟比赛的簇大小分布。簇大小定义为在预先指定的时间阈值内连续发生的角球次数。各面板对应不同的阈值时长，并分别以曲线表示主队和客队。红色曲线表示根据观测数据得到的经验概率质量函数，灰色曲线则对应 200 个模拟赛季。

B 表

协变量	描述
下半场指示变量 (X_1)	上半场为 0，下半场为 1
比分差 (X_2)	正数 X_2 表示该球队领先， 负数 X_2 表示该球队落后， 零 X_2 表示平局
从进球开始（X_3）	如果区段从半场的实际开始时刻起算，则为 0， 如果区段从进球开始，则为 1
片段开始时间（X_4）	一个数值变量，表示区段的开始时间 自当前半场开始以来

表1：协变量列表及描述

表2：模型 (a)、(b) 和 (c) 的估计值 β。这些模型分别对应图 (1) 中的前三行。

	模型 (a)		模型 (b)		模型 (c)	
参数	Home	Away	Home	Away	Home	Away
下半场指示变量 X_1	·	0.090 (.061)	·	0.083 (.061)	·	0.091 (.061)
比分差 X_2	-000.122 122 122 (.057057057) -. (.) -. (.)	0.032 (.034)	-000.111111111 (.052052052) -. (.) -. (.)	0.033（.034）	-000.116116116（.056056056）-. (.) -. (.)	0.036 (.034)
从球开始 X_3	-000.090 090 090 (.022022022) -. (.) -. (.)	·	-000.094 094 094 (.023023023) -. (.) -. (.)	·	-000.091 091 091 (.023023023) -. (.) -. (.)	·
片段开始时间 X_4	·	-000.005 005 005 (.002002002) -. (.) -. (.)	·	-0.004 (.003)	·	-0.004 (.003)
交互作用 X_4X_2	·	-000.006 006 006 (.002002002) -. (.) -. (.)	·	-000.006006 006 (.002002002) -. (.) -. (.)	·	-000.006006 006 (.002002002) -. (.) -. (.)
τ的分布 [1]	Γ(8.76,4.43)	Γ(8.49,4.25)	Γ(9.46, 4.81)	Γ(7.09, 4.02)	Γ(9.12, 4.65)	Γ(9.18, 4.64)
τ的均值和标准差	1.98, 0.67	2.00, 0.69	1.97, 0.64	1.76, 0.66	1.96, 0.65	1.98, 0.65

[1] Γ(形状参数，率参数)表示形状参数和率参数分别为上述值的伽马分布。

表3：模型（d）下的特定状态回归效应估计值。

	模型（d）	
参数	**Home**	**Away**
针对 $\boldsymbol{\beta}_0$ （常规状态）		
下半场指示变量 X_1	·	0.097 (.068)
比分差 X_2 -000.104 104 104 (.026026026) -. (.) -. (.)		0.019 (.038)
始于进球 X_3 -000.133 133 133 (.065065065) - . (.) - . (.)		·
片段开始时间 X_4	·	-0.005 (.003)
交互作用 X_4X_2	·	-000.006 006 006 (.002002002) -. (.) -. (.)
对于 $\boldsymbol{\beta}_1$ （热状态）		
下半场指示变量 X_1	·	0.077 (.156)
比分差 X_2 -0.012 (.066)		0.120 (.095)
从进球开始 X_3 -0.012 (.153)		·
片段开始时间 X_4	·	0.004 (.007)
交互作用 X_4X_2	·	-0.006 (.005)
τ的分布[1]	Γ(9.18, 4.60)	Γ(9.06, 4.57)
τ 的均值和标准差	1.99, 0.66	1.98, 0.66

[1] Γ(形状, 率) 表示形状参数和率参数为相应值的伽马分布。