arXiv:2602.04855v1 [stat.ME] 4 Feb 2026

用于拟合疫情计数数据的动力学生存分析模型的边际似然推断

Suchismita Roy*[1]、Alexander A. Fisher†[1] 和 Jason Xu‡[2]

[1]统计科学系，杜克大学，美国北卡罗来纳州达勒姆[2]生物统计学系，加利福尼亚大学洛杉矶分校，美国加利福尼亚州洛杉矶

摘要

随机舱室模型是描述疾病传播的常用工具，但对于许多类型的监测数据而言，当边际似然函数因缺失信息而变得难以处理时，在这些模型下进行推断颇具挑战。为此，我们在动力学生存分析（DSA）范式下，为离散观测的发病计数数据构建了一个闭式似然。该方法利用大群体极限近似随机总体层面风险率，同时保留计数值随机模型，从而形成兼具计算效率和模型扩展灵活性的生存分析推断策略。通过模拟，我们表明，在部分观测情形下，参数估计的表现可与近期精确但计算成本高昂的基于似然的方法相媲美。既有研究表明，DSA 近似具有通用性；我们进一步表明，本文的推断方法也适用于具有个体异质性的模型，例如脆弱性模型。我们针对埃博拉和 COVID-19 数据开展了案例研究，考察了该模型的多种变体，包括基于网络的流行病模型以及具有易感性分布的模型，展示了其在真实、部分观测数据集上的灵活性和实用价值。

关键词：连续时间马尔可夫链；计数数据；动态生存分析；缺失数据；随机流行病模型。

*电子邮箱：suchismita.roy@duke.edu

†电子邮箱：alexander.fisher@duke.edu

‡电子邮箱：jqxu@g.ucla.edu

1 引言

传染病的数学建模在理解流行病动态和指导公共卫生干预方面发挥着重要作用。分舱模型是流行病建模的基石，它根据疾病状态将人群划分为不同的舱室，并根据机制参数描述舱室之间的转变。传统上，这些模型由常微分方程系统定义，用于描述各舱室人数如何随时间变化。易感-感染-移除模型（SIR）（Kermack 和 Mckendrick，1927）或许是最广为人知的例子。由于概念简单、易于解释且灵活多变，该模型的许多变体至今仍得到广泛应用（Britton 等人，2020年；Morozova 等人，2021；Wood 等人，2026）。

尽管确定性模型易于分析，却无法描述疾病传播固有的随机性。随机流行病模型（SEM）（Andersson 和 Britton，2000）无疑能更真实地呈现疾病动态。经典的似然方法能够在完整 流行病数据下进行严格推断；但由于实际中很少能完整观测感染与康复历程，因此面对部分观测数据时，这些方法中的一些便难以应用。例如，尽管随机 SIR 模型的轨迹遵循标准的连续时间马尔可夫链似然（Guttorp 和 Minin，2017），但实际中我们只能通过病例数、移除数据或其他部分观测的监测研究，以离散方式观测这一过程。在这种情况下，推断会困难得多，因为边际似然需要对与观测数据相符的可能轨迹进行高维积分。大量研究致力于解决这一问题，方法包括计算转移概率的数值方法（Ho 等人，2018）、利用扩散过程近似似然（Cauchemez 和 Ferguson，2008），以及采用粒子滤波器（King 等人，2016）、近似贝叶斯计算（McKinley 等人，2017）和数据增广 MCMC（Gibson 和 Renshaw，1998；O’Neill 和 Roberts，2002；Fintzi 等人，2017）。

本文从一种近期提出的思路——动力学生存分析（DSA）（KhudaBukhsh 等人，2019；Rempa la 和 KhudaBukhsh，2023）——的视角重新审视这一问题，以便在此情境下实现高效且简便的基于似然的推断。下文将对此作更详细的介绍。DSA 通过随机 SIR 模型中随机总体层面风险率的大样本极限，对其进行近似。由于它只近似风险率，近似过程仍定义在与原模型相同的离散状态上，这与会假定总体连续的扩散近似不同（Cauchemez 和 Ferguson，2008）。这种方法不仅往往能忠实地近似原始流行病模型，而且与一种便于采用简单

推断方法的生存似然相关。

本文的重点在于对缺失数据进行严格核算，以理解这些思想如何应用于现实的部分观测数据类型并评估其表现。具体而言，我们针对计数数据建立了一种闭式似然，能够精确地对感染和移除时间等未知事件进行边缘化。进而，该方法使现成的后验抽样算法得以使用，这与上文所讨论的、面向部分观测情形的现有工作形成对比，后者通常需要密集的基于模型的模拟或定制的马尔可夫链蒙特卡洛采样器。我们发现该方法与生存分析中的区间删失方法存在联系，并表明所提出的方法不仅能比现有方法实现更高效的推断，而且更容易加以扩展，以放宽易感-感染-移除模型及其变体背后的模型假设。为加以说明，我们在模拟和若干案例研究中展示了该方法的有效性，包括具有个体层面易感性变异的模型（Gomes 等人，2020年；Montalb´an 等人，2022；Yan 和 Chowell，2024年），以及以泊松随机网络替代均匀混合的情形（Wairimu et al.，2024年）。

2 背景

疾病分舱模型根据个体的疾病状态，将个体划分为互不重叠的子集。其动态通常由常微分方程（常微分方程）描述：例如，易感-感染-移除（SIR）模型假定有三种疾病状态——易感者（S）、感染者（I）和移除者（R）（Kermack 和 Mck-endrick，1927），其动态遵循 $\frac{dS(t)}{dt} = -\beta S(t)I(t)$ $\frac{dI(t)}{dt} = \beta S(t)I(t) - \gamma I(t)$。

随机 SIR 模型则用类似的表达式取代这些常微分方程，以表示马尔可夫跳跃概率或瞬时率。在这两种情况下，参数都以机制性方式描述流行病：此处，γ 是个体康复的恢复率——在随机模型中，$1/\gamma$ 因此是预期康复时间。类似地，β 可解释为易感者与感染者个体之间每次接触的传播率。模型还可加入其他分舱，以纳入潜伏期、再感染、住院等因素（Hethcote，2000；Keeling 和 Rohani，2008）。在最简单的形式中，假设易感者和感染者个体之间的感染率 β 恒定且均一，则给定流行病过程在时间 $t > 0$ 的当前状态时，瞬时感染率可表示为

$$\mathbb{P}\{(S(t+h), I(t+h)) - (S(t), I(t)) = (-1, 1)\} = \frac{\beta}{N}(S(t)I(t))h + o(h) \tag{1}$$

在一个较小的时间增量 $h > 0$ 内，恢复过程的动态则以康复率 γ 类似地定义。整体随机流行病模型演化为一个连续-

时间马尔可夫链（连续时间马尔可夫链）（Guttorp 和 Minin，2017），其中连续事件之间的逗留时间服从指数分布。随机流行病模型（SEM）——不同于其确定性对应模型——反映了疫情传播背后的随机性，并且非常适合回答与概率 疫情暴发相关的问题，对估计和预测而言具有可以说更易于解释的不确定性量化 。

我们将利用动态生存分析（DSA），从生存分析的角度解读SIR模型，从而得到一个更简单的似然，并为对缺失数据进行边缘化提供清晰的途径。这种方法基于SIR模型的个体层面视角在大群体极限下所产生的等价随机过程（Sellke，1983）。Sellke构造如下：

1. 每个 N 最初的易感者都有一个相互独立的感染阈值Q_i ，其服从均值为1的指数分布。如果累积疾病暴露超过某个易感者的感染阈值，该个体就会被感染。

2. 每个感染者以速率 β/N 向每个易感者传播疾病，直至其康复/移除；康复/移除发生在持续时间服从速率为 γ的指数分布的传染期之后。

不难证明，基于这些假设产生的流行病与使用（1）产生的流行病具有相同的分布（Britton，2019）。注意，直至时间 t，每个个体所受到的暴露总量为

$$\Lambda(t) \coloneqq \frac{\beta}{N}\int_0^t I(s)ds. \tag{2}$$

这是截至时间 t的累积风险。因此，随机个体一直保持易感状态直到时间 t 的概率为

$$\mathbb{P}(Q_i > \Lambda(t) \mid (I(s))_{s\in[0,t]}) = \exp\left(-\frac{\beta}{N}\int_0^t I(s)ds\right). \tag{3}$$

通过这种方式，Sellke构造将SIR模型视为个体层面或基于个体的过程，这对于模拟流行病中的异质性及其他特征非常有利。另一方面，我们可以通过该过程的可聚合性恢复基于总体的模型（Tian 和 Kannan，2006）。为此，我们为 i^{th} 个体定义一个过程

$$S_i(t) = \begin{cases} 1, \text{ if } i^{th} \text{ individual is susceptible at time } t. \\ 0, \text{ otherwise.} \end{cases}$$

类似地，定义过程 $I_i(t)$ 和 $R_i(t)$，分别表示个体 i 是否感染或康复——

——分别在时间 t 感染或康复，因此 $S_i(t)+I_i(t)+R_i(t)=1.$。求和并利用泊松过程的叠加性，我们得到群体水平过程 $\{S(t)\ I(t)\ R(t)\}$，

$$S(t)=\sum_{i=1}^{N+M} S_i(t), I(t)=\sum_{i=1}^{N+M} I_i(t), \text{ and } R(t)=\sum_{i=1}^{N+M} R_i(t). \tag{4}$$

这产生了 了一个有用的等价表示，其中包含单位速率泊松过程 $\mathcal{Y}$，$\mathcal{Z}$

$$\begin{aligned}
S(t) &= S(0)-\mathcal{Y}\left(\int_0^t \frac{\beta}{n} S(s)I(s)ds\right)\\
I(t) &= I(0)+\mathcal{Y}\left(\int_0^t \frac{\beta}{n} S(s)I(s)ds\right)-\mathcal{Z}\left(\int_0^t \gamma I(s)ds\right)\\
R(t) &= \mathcal{Z}\left(\int_0^t \gamma I(s)ds\right).
\end{aligned}$$

DSA 的关键思想是用累积随机危险率的大样本极限取代该构造中的累积随机危险率。运用泛函大数定律并取 $N\to\infty$和 $M/N\to\rho\in(0,1)$，KhudaBukhsh 等（2019）和 Izyumtseva等人（2024）证明，缩放后的过程（$\frac{S(t)}{N}$，$I_{(t)\,N}$，$R_{(t)\,N}$）收敛到常微分方程组的解，

$$\dot{s}_t=-\beta s_t\iota_t, \dot{\iota}_t=\beta s_t\iota_t-\gamma\iota_t \text{ and } \dot{r}_t=\gamma\iota_t, \tag{5}$$

其中，$s_t=\lim_{N\to\infty} S_{(t)\,N}$，$\iota_t=\lim_{N\to\infty} I_{(t)\,N}$以及$r_t=\lim_{N\to\infty} R_{(t)\,N}$的初始条件为 $s_0=1$，$\iota_0=\rho$，$r_0=0$。换言之，s_t，ι_t是易感者和感染者比例的渐近比例。常微分方程组（5）可改写为，

$$\begin{aligned}
s_t &= \exp\left(-\beta\int_0^t \iota_u du\right) = \exp(-R_0 r_t),\\
\iota_t &= \rho e^{-\gamma t}-\int_0^t \dot{s}_u e^{-\gamma(t-u)}du, \qquad r_t=\gamma\int_0^t \iota_u du,
\end{aligned} \tag{6}$$

其中 $R_0=\frac{\beta}{\gamma}$是基本再生数。DSA 的一个关键优势（下文将详细介绍）是，它能为感染时间提供闭式密度。

3 动态生存似然

3.1 已知完整感染时间和康复时间时的似然

为说明生存似然视角，我们注意到，Sellke构造 (3) 中的阈值等同于某个个体的感染时间，因此可以采用生存分析

视角。也就是说，若 $T_I^{(i)}$ 是表示 i^{th} 易感者感染时间的随机变量，则有 $\mathbb{P}(T_I^{(i)} > t \mid (I(s)_{s\in[0,t]}) = \mathbb{P}(Q_i > \Lambda(t) \mid (I(s)_{s\in[0,t]})$。利用模型的基于个体的表述与群体层面的表述之间的关系 (6)，KhudaBukhsh 等（2019）将这一观察推广为：把由常微分方程极限 (5) 得出的 s 视为随机选取的易感者感染时间的生存函数。将该时间记为 T_I，相应的生存函数可表示为

$$\mathbb{P}(T_I > t) = s_t = \exp(-\mathcal{R}_0 r_t). \tag{7}$$

这种观点也使我们能够将 $R_0 r_t = \beta \int_0^t \iota_s ds$ 解释为 T_I的累积风险函数，其风险率为 $\beta\iota_t$（参见式 (6)）。

从这里出发，我们接下来可以推导 T_I的密度。由于我们不假设所有个体在观测期结束前都会受到感染，因此生存函数不会收敛到 1，故它是一个非正规随机变量。通过考虑条件密度可以避免这一问题：对于 $T_I < \infty$， T_I 的累积分布函数和概率密度分别为

$$F_{T_I}(t_I) = \frac{1 - s_{t_I}}{\tau}, \qquad f_{T_I}(t_I) = -\frac{\dot{s}_{t_I}}{\tau}, \tag{8}$$

（KhudaBukhsh 等人，2019；Rempa la 和 KhudaBukhsh，2023）。这里， τ 表示易感个体最终受到感染的概率，并满足

$$1 - \tau = s_\infty = \exp(-R_0 r_\infty) = \exp(-R_0(\rho + \tau)). \tag{9}$$

对于观测时间截至 T的流行病，可得

$$f_{T_I}(t; T) = -\frac{\dot{s}_t}{1 - s_T}, \qquad F_{T_I}(t; T) = \frac{1 - s_t}{1 - s_T}. \tag{10}$$

最后，令 T_R 表示随机选取的一名感染者的康复时间，并令传染期 $T_R - T_I$ 服从与 T_I相互独立的指数分布，不过这一假设可以放宽。通过对 T_I 和 $T_R - T_I$进行卷积， T_R 的密度为

$$f_{T_R}(t) = \int_0^t f_{T_I}(u)\gamma e^{-\gamma(t-u)} du = \frac{\gamma}{\tau}\left(\iota_t - \rho \exp(-\gamma t)\right). \tag{11}$$

有了这些显式密度，我们将推导似然的边际形式，以适应数据的粒度或可用性。KhudaBukhsh 等（2019）指出，在感染和康复时间已知且完全观测的情形下，

$\theta :=(\beta,\ \gamma,\ \rho)$ 对于 N 个初始易感者和 M 个初始感染者的似然为

$$
\begin{aligned}
&l(\theta|\{t_i,w_i\}_{i=1}^{K},\{\epsilon_j\}_{j=1}^{\tilde{L}})\\
&=s_T^{N-K}\gamma^{L+\tilde{L}}\exp\left\{-\gamma\left(\sum_{i=1}^{K}w_i+\sum_{j=1}^{\tilde{L}}\epsilon_j+(M-\tilde{L})T\right)\right\}\prod_{i=1}^{K}(1-s_T)f_{T_I}(t_i;T).
\end{aligned}
\tag{12}
$$

此处， t_i 表示截至时间 T感染的 K 个体各自的精确感染时间， w_i表示这些个体的传染期，或表示那些在 T之后康复者的传染期尾部 $T-t_i$。L和 L分别表示初始易感者和感染者中截至时间 T康复的人数，而 ϵ_j 表示后者各自的传染期。该表达式只是指数函数的乘积，因此参数估计十分直接，但在现实场景中，我们很少能获得完整数据 t_i， w_i， ϵ_j 。下一节将聚焦于更常见的部分观测情形。

3.2 边际计数似然的推导

为推导观察性研究中常见的缺失数据的边际表达式，我们重点介绍两种情形。首先，讨论数据包含精确感染时间、但不包含康复信息的情况。随后，我们考虑更贴近现实的情形：这些感染时间也无法获得，取而代之的是按某种离散时间表（即每日或每周报告）汇总的感染计数。一个关键观察是，在DSA近似下，感染时间和康复时间在个体层面相互独立，因此在推导对缺失信息进行边际化的似然时，可以大幅简化计算。这些简化最终将带来高效而简洁的形式，使推断更便于用户使用。

举例来说，假设数据中没有任何康复或移除信息，但观察到了由 $t_1,t_2,\ldots,t_K$ 给出的 K个体的精确感染时间。在 DSA 下，样本中 K 个体的感染时间相互独立，因此，对于一场观察截至 T 的流行病， θ 的似然为

$$
l(\theta|t_{i=1}^{K})=\prod_{i=1}^{K}f_{T_I}(t_i;T)=\prod_{i=1}^{K}-\frac{\dot{s}_{t_i}}{1-s_T}=\prod_{i=1}^{K}\frac{\beta s_{t_i}\iota_{t_i}}{1-s_T}.
\tag{13}
$$

值得注意的是，我们发现该似然不依赖于 N，即初始易感者的数量，而这一数量往往未知。如果我们以已知 N 为条件，似然则变为

$$
l(\theta|t_{i=1}^{K},N)=s_T^{N-K}\times\prod_{i=1}^{K}(1-s_T)f_{T_I}(t_i;T)=s_T^{N-K}\prod_{i=1}^{K}\beta s_{t_i}\iota_{t_i}.
\tag{14}
$$

不难证明，我们也可以通过对所有未观测到的康复时间进行边缘化，从完整数据似然(12)得到(14)。

然而，在监测数据中，通常无法获得如此细粒度的精确感染时间。现在假设我们只在离散观测时间$\{\xi_0, \xi_1, \ldots, \xi_P\}$收集数据，其中 $\xi_0 = 0, \xi_P = T$。我们只能观测到每个区间（ξ_{j-1}, ξ_j）内的感染计数 Y_j，P其中 $j = 1, 2, \ldots, P$；并将时间 T之前的感染总数记为 $K = \sum_{j=1} Y_j$.

为推导似然，我们定义 $\{t_{j1}, t_{j2}, \ldots, t_{jY_j}\}$ 为第 Y_j 个体在区间（ξ_{j-1}, ξ_j）内的未观测感染时间，其中 j^{th}；对 $\{j = 1, 2, \ldots, P\}$ 亦如此。(14) 给出了 t 的密度，其中 $\mathbf{t} = (t_{11}, t_{12}, \ldots, t_{1Y_1}, \ldots, t_{P1}, \ldots, t_{PY_P})$。我们将进行变量变换，令 $\mathbf{t} \to \mathbf{u}$，其中 $\mathbf{u} = (u_{11}, u_{12}, \ldots, u_{1Y_1}, \ldots, u_{P1}, \ldots, u_{PY_P})$，这些变量通过以下变换相互关联:

$$u_{jl} = \frac{s_{t_{jl}} - s_{\xi_{j-1}}}{s_{\xi_j} - s_{\xi_{j-1}}}, \text{ for } j = 1, 2, \ldots, Y_p, p = 1, 2, \ldots, P, \tag{15}$$

注意，由于 $\xi_{j-1} < t_{jl} < \xi_j$ 和 $s(\cdot)$ 是时间的非递增函数，因此我们有$0 < u_{jl} < 1$。为推导$\mathbf{u}$的密度，我们首先计算雅可比矩阵。由于 u_{jl} 与 $t_{j'l'}$相互独立，对于 $j \neq j'$ 和 $l \neq l'$，雅可比矩阵为

$$\begin{aligned} J &= \det\left(\text{Diag}\left(\frac{dt_{11}}{du_{11}}, \frac{dt_{12}}{du_{12}}, \ldots, \frac{dt_{PP}}{du_{pp}}\right)\right) \\ &= \prod_{j=1}^{P}\prod_{l=1}^{Y_j} \frac{dt_{jl}}{du_{jl}} = \prod_{j=1}^{P}\prod_{l=1}^{Y_j} \frac{1}{\frac{du_{jl}}{dt_{jl}}} = \prod_{j=1}^{P}\prod_{l=1}^{Y_j} \frac{s_{\xi_{j-1}} - s_{\xi_j}}{\beta s_{t_{jl}} \iota_{t_{jl}}} \end{aligned}$$

（Rohatgi 和 Saleh，2011）。因此，记变量变换为 $t_{jl} = g(u_{jl})$，并将其代入（14），可得$\mathbf{u}$ 与发病计数数据的联合密度，

$$\begin{aligned} f(\mathbf{u}, \{Y_j\}_{j=1}^P \mid \theta) &= s_T^{N-K} \prod_{j=1}^{P}\prod_{l=1}^{Y_j} \beta s_{g(u_{jl})} \iota_{g(u_{jl})} \mathbb{I}(\xi_{j-1} < g(u_{jl}) < \xi_j) \prod_{j=1}^{P}\prod_{l=1}^{Y_j} \frac{s_{\xi_{j-1}} - s_{\xi_j}}{\beta s_{g(u_{jl})} \iota_{g(u_{jl})}} \\ &= \underbrace{s_T^{N-K} \prod_{j=1}^{P} (s_{\xi_{j-1}} - s_{\xi_j})^{Y_j}}_{f(\{Y_j\}_{j=1}^P \mid \theta)} \times \underbrace{\prod_{j=1}^{P}\prod_{l=1}^{Y_j} \mathbb{I}(0 < u_{jl} < 1)}_{f(\mathbf{u} \mid \{Y_j\}_{j=1}^P, \theta)}. \end{aligned}$$

因此，每个 u_{jl} 都服从 Unif(0，1) 分布，其密度积分为 1。值得注意的是，对 u_{jl} 积分后仅产生一个因子 1，不会留下其他项。

因此，边际似然大大简化为

$$l(\theta \mid \{Y_j\}_{j=1}^{P}) = s_T^{N-K} \prod_{j=1}^{P} (s_{\xi_{j-1}} - s_{\xi_j})^{Y_j}. \tag{16}$$

我们将这一量称为边际DSA计数似然，并强调，此处对缺失事件进行边际化，与引言中所述通常难以处理的积分步骤相同。通常，这一步必须通过大量条件模拟、滤波方法或前文概述的数据增广MCMC来完成。

与区间删失的联系推导出上述结果后，我们发现，该表达式自然引出了一种区间删失视角。为说明这一点，引入指示变量

$$\delta_{jl} = \begin{cases} 1, \text{ if } l^{th} \text{ individual became infected in the interval } (\xi_{j-1}, \xi_j) \\ 0, \text{ otherwise,} \qquad l = 1, \ldots, Y_j, \quad j = 1, \ldots, P. \end{cases}$$

那么，区间删失似然（Fleming和Harrington，1991）为：

$$\begin{aligned} l(\theta \mid \{Y_j\}_{j=1}^{P}) &= \mathbb{P}(T_I > T)^{N-K} \prod_{j=1}^{P} \prod_{l=1}^{Y_j} \mathbb{P}\left(\xi_{j-1} < T_I < \xi_j\right)^{\delta_{jl}} \\ &= s_T^{N-K} \prod_{j=1}^{P} \prod_{l=1}^{Y_j} (s_{\xi_{j-1}} - s_{\xi_j})^{\delta_{jl}} = s_T^{N-K} \prod_{j=1}^{P} (s_{\xi_{j-1}} - s_{\xi_j})^{Y_j}. \end{aligned} \tag{17}$$

也可以根据康复时间的密度（11）推导出类似的边际似然，以便在离散观测的移出或死亡数据（而非病例数据）可用时使用。

与精确推断的比较由DSA近似推导出的边际似然（16）具有显著的计算优势。用于推断时，我们只需$\{\xi_1, \xi_2, \ldots, \xi_P\}$处常微分方程（5）的解，而这些方程很容易通过数值方法求解。重要的是，该计算成本与易感人群规模 N无关，而取决于观测时间的数量，因此该方法能够很好地扩展到大型疫情暴发。相比之下，依赖似然表达式（12）或（14）的现有方法，其计算开销会随着 N 的增加而增大，因为随着易感者数量增加，感染时间会变得更加密集，因此需要在更细的网格上计算常微分方程的解。第5节通过实证结果展示了这些差异。

以下各节将展示，这种为便于说明而在简单易感-感染-移除模型设定中推导出的方法，如何立即扩展到更复杂的模型，以及推断如何进行比较

但比计算密集型黄金标准快得多。

4 Gene对其他分舱模型的可推广性

以上推导出的边际 DSA 似然为推断提供了一个优势：它所蕴含的推断策略可以立即扩展到任何变体或相关的分室模型，只要能够将其极限动力学或平均场动力学表示为常微分方程（常微分方程）系统。所幸，这方面有大量数学文献可供参考（Chowell，2017；Gumel 等，2021；Salda˜na 和 Velasco-Hern´andez，2022）；为说明这种普适性，我们重点讨论个体层面差异的脆弱性模型 和异质混合的泊松网络模型 。

4.1 通过脆弱性实现变化的易感性

在生存分析文献中，可以通过引入一个捕捉个体间未观测异质性的脆弱性变量来对偏离比例风险的情况进行建模（Balan 和 Putter，2020；Hougaard，1995），这类似于加入随机效应项。这一思想在分 compartment 建模中同样有用——许多模型，包括标准 SIR 模型，都假设所有个体具有相同的易感性，而这通常并不现实。易感性可能因多种因素而异，包括年龄和健康状况等可测量属性，或健康的社会决定因素等难以观测的驱动因素（Marmot 等人，2008；Zelner 等人，2022）。为了在易感人群中引入异质性，可以引入一个非负脆弱性随机变量（Gomes 等人，2020；Rempala，2023a）。Montalb´an 等人（2022）推导出了当脆弱性变量服从伽马分布时 SEIR 模型的闭式 ODE。遵循他们的方法，假设脆弱性变量 X 服从均值为 1、标准差为 ν 的伽马分布，以便说明。我们的模型可以自然地扩展以适应这一设定，并且只需进行少量修改即可联合推断其参数。在模型中，我们只需将 Sellke 构造修改为现在从 Exp(X_i) 分布中抽取感染阈值Q_i ，而不是统一从 Exp(1) 中抽取。我们可以通过类比随机效应模型，理解每个个体的易感性是如何通过 $X = x$ 建模的。在给定 $X = x$的条件下，我们可以将 $S(t, x)$ 定义为脆弱性为 x 的个体在时间 t 仍保持易感的概率；类似地，可以定义 $I(t, x)$ $R(t, x)$。

就控制方程而言，我们通过对脆弱性分布进行积分来加以修改；对于伽马分布等指数族，这一积分可以解析地求出。

回顾式（2），易感者在时间 t 之前一直保持易感状态的概率为

$$P(Q_i > \Lambda(t) \mid (I(s))_{s\in[0,t]}, x) = S(t,x) = \exp\left(-\frac{x\beta}{N}\int_0^t I(s)ds\right).$$

这里感染者数量的定义类似 $I(t) = \int I(t,x)dx$；因此，

$$\dot{S}(t,x) = -\exp\left(-\frac{x\beta}{N}\int_0^t I(s)ds\right) \times \frac{x\beta}{N}I(t) = -\frac{x\beta}{N}S(t,x)I(t).$$

对于给定的个 体，其脆弱性变量为 x，我们现在有如下常微分方程组 b y

$$\begin{aligned}
\dot{S}(t,x) &= -\frac{x\beta}{N}S(t,x)I(t)\\
\dot{I}(t,x) &= \frac{x\beta}{N}S(t,x)I(t) - \gamma I(t)\\
\dot{R}(t,x) &= \gamma I(t),
\end{aligned} \tag{18}$$

这与 SEIR 模型在 (Gomes 等人, 2020年; Montalb´an 等人, 2022) 中的方程相对应。对于伽马分布$(1,\nu)$，对脆弱性分布进行积分可得 $\int xS(t,x)dx = S(t)^{1+\nu^2}$. 因此，当 $N\to\infty$ 和 $M/N\to\rho$ 时，大规模人群行为可由以下内容刻画

$$\dot{\tilde{s}}_t = -\beta\tilde{s}_t^{1+\nu^2}\tilde{\iota}_t, \dot{\tilde{\iota}}_t = \beta\tilde{s}_t^{1+\nu^2}\tilde{\iota}_t - \gamma\tilde{\iota}_t \text{ and } \dot{\tilde{r}}_t = \gamma\tilde{\iota}_t. \tag{19}$$

我们看到，通过这种方式，只需将其代入 (8) 中给出的密度 f_{T_I}，即可立即使用基于 DSA 的似然进行推断，从而得到

$$\tilde{f}_{T_I}(t;T) = \frac{\beta\tilde{s}_t^{1+\nu^2}\tilde{\iota}_t}{1-\tilde{s}_T}. \tag{20}$$

推断 ˜e 的推导过程类似： $\theta = (\beta,\gamma,\rho,\nu)$ 的完整似然 (12) 为

$$\begin{aligned}
&l(\tilde{\theta}|\{t_i,w_i\}_{i=1}^K,\{\epsilon_j\}_{j=1}^{\tilde{L}})\\
&= \tilde{s}_T^{N-K}\gamma^{L+\tilde{L}}exp\left\{-\gamma\left(\sum_{i=1}^K w_i + \sum_{j=1}^{\tilde{L}}\epsilon_j + (M-\tilde{L})T\right)\right\}\prod_{i=1}^K(1-\tilde{s}_T)\tilde{f}_{T_I}(t_i;T),
\end{aligned} \tag{21}$$

及其边际 发病数据下的 al 形式会修改计数似然 (16)，使其变为

$$l(\tilde{\theta} \mid \{Y_j\}_{j=1}^P) = \tilde{s}_T^{N-K}\prod_{j=1}^P(\tilde{s}_{\xi_{j-1}} - \tilde{s}_{\xi_j})^{Y_j}. \tag{22}$$

在下一节的模拟研究中，我们将使用 (22) 来估计参数。

4.2 使用泊松随机网络放宽均匀混合

为说明该模型的可扩展性，我们进一步考虑 Wairimu 等人（2024）研究的泊松随机网络上的 SIR模型。通过随机接触网络放宽了 SIR 背后的均匀混合假设——该假设基本上认为接触发生在完全图上——其中随机接触网络是随机图 $\mathcal{G}(N+M,\ p)$ 上的连续时间马尔可夫链（CTMC），$N+M$ 为图的规模，其度分布记为 p。在假设 p服从平均度为 μ的泊松分布的情况下，Wairimu 等人（2024）；Rempala（2023b）表明，网络SIR模型由以下方程组刻画

$$\dot{S} = -\tilde{\beta} S(1 + \rho - S + \frac{\tilde{\gamma}}{\tilde{\beta}} logS), \tag{23}$$

$$\dot{I} = -\dot{S} - \gamma I, \qquad R = 1 + \rho - S - I. \tag{24}$$

˜在初始条件 $S(0) = 1,\ I(0) = \rho$下，其中 $\beta = \mu\beta,\ \tilde{\gamma} = \beta + \gamma,$ 以及 $\beta,\ \gamma,\ \rho,$ μ 如上所定义。与之前一样，满足（23）的解曲线 S 作为感染时间的生存函数代入，因此，只需将 s 替换为常微分方程（23）、（24）的解，即可立即应用边际计数似然（16）的相同形式。

5 模拟研究

鉴于以往关于动态生存分析的研究已考察过完全观测情形下的近似质量，我们的实证研究将重点评估：在部分观测数据上使用 DSA 边际计数似然的表现，与完整数据可用时相比如何。此外，我们还将本方法与精确方法进行比较。本节还包括脆弱性设定下的模拟研究，展示该方法对不同模型变体的适应性和有效性。

5.1 DSA 下的推断与前向动态

在进行推断或估计之前，我们首先考察随机SIR过程与支配 DSA 的近似动力学下疫情暴发分布的差异。为此，我们使用 Sellke构造，从精确的随机SIR过程模拟了 500 次实现，观测区间长度为 $T = 10$ 。我们按照一项研究设定真实参数 $\theta = (\beta, \gamma, \rho) = (2, 0.5,$ $0.05), (2, 1, 0.05), (1.5, 1, 0.05)$，该研究

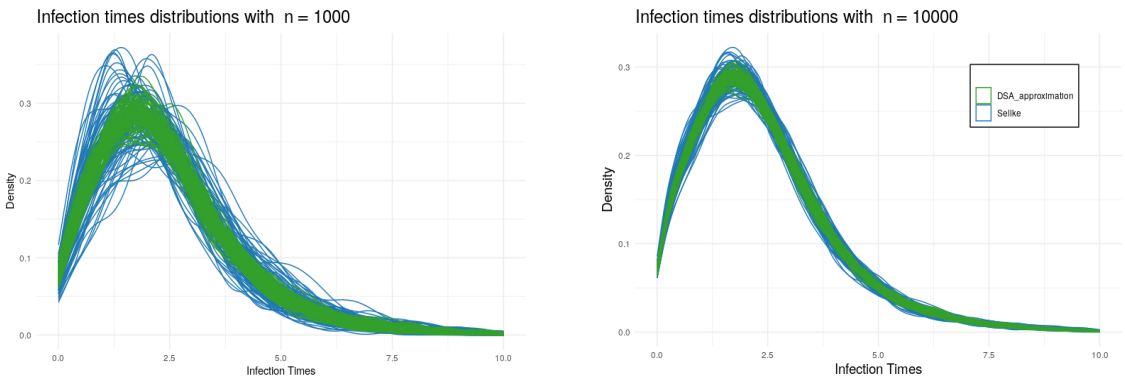

图1：比较使用 Sellke 构造（蓝色）模拟得到的 100 个感染时间经验密度与使用 DSA近似（绿色）模拟得到的 100 个感染时间经验密度。

作者：KhudaBu khsh 等人（2019），并针对规模为 $N = 250, 1000, 10000$ 的人群重复实验 .

算法 1 生成感染时间 t^* 的伪代码， f_{T_I}输入：参数（β, γ, ρ）以及最终观测时间 T。**步骤：**

1：给定参数（β, γ, ρ），在时间区间（$0, T$] 上以固定的小步长求解常微分方程（5），并计算 s_T。2：生成一个随机样本 $u \sim$ Unif(0,1)。3：令 $F_{T_I}(t^*; T)$ 等于 u，计算 $s_{t^*} = 1 - u(1 - s_T)$。4：通过对给定的常微分方程解 s_t 进行插值，为映射 $s_t \mapsto t$ 构造一个插值函数 t。5：使用插值函数，通过逆概率抽样从 s_{t^*} 中抽取 t^* 。

输出：从密度 f_{T_I} 中随机模拟得到的感染时间 t^* 。

接下来，我们以类似的方式重复这项研究，但用 DSA 下对应的近似动力学（详见算法 1）替代精确的易感-感染-移除模型动力学。图1展示了从精确的易感-感染-移除模型和DSA近似中各抽取 100 条轨迹进行比较的结果。图中显示，DSA近似所反映的变异性存在经验上的差异；我们观察到，随着 N 增大，这些轨迹越来越接近精确的易感-感染-移除模型动力学下的轨迹，但两者之间的差距并未消失。另一方面，均值趋势非常相似。

为了理解这些差异可能如何转化为估计上的差异，我们将注意力转向推断。我们比较了基于所提出的边际似然拟合的模型与几种同行方法：首先，使用式（12）中的感染和康复时间完整信息考察参数估计。接下来，我们将其与 KhudaBukhsh 等（2023）、Lauro 等（2022）此前应用于病例计数数据的一种策略进行比较；该策略将缺失的感染时间填补为均匀随机变量，以便基于似然（14）进行推断

。作为金标准，我们将其与一种最新的数据增广MCMC方法（Morsomme和Xu，2025）进行比较；该方法通过专门设计的数据增广采样器，在精确SIR模型下对缺失时间的适当分布进行边际化。

表1：真实参数为（β，γ，ρ）=（2，0。5，0。05）时的参数估计。M1：使用精确的感染和康复时间；M2：使用均匀近似的感染时间；M3：计数似然；M4：PDSIR。

		N = 250				N = 1000				N = 10000			
		avg	sd	95%覆盖率	每秒有效样本量	avg	sd	95%覆盖率	每秒有效样本量	avg	sd	95%覆盖率	每秒有效样本量
β	M1	2.06	0.29	0.82	74.51	1.99	0.07	0.85	21.84	2.00	0.02	0.83	3.38
	M2	1.62	0.22	0.58	23.56	1.71	0.11	0.30	10.83	1.73	0.04	0.00	1.77
	M3	1.95	0.25	0.83	45.96	1.98	0.12	0.84	53.15	2.00	0.04	0.84	104.37
	M4	1.98	0.24	0.94	2.78	1.99	0.12	0.94	3.91	2.00	0.04	0.92	0.71
γ	M1	0.80	0.26	0.93	96.28	0.50	0.02	0.91	28.44	0.50	0.01	0.93	4.43
	M2	0.39	0.09	0.75	26.55	0.42	0.05	0.59	12.10	0.43	0.02	0.01	1.83
	M3	0.48	0.11	0.89	51.86	0.49	0.05	0.88	59.37	0.50	0.02	0.91	116.38
	M4	0.48	0.11	0.93	1.02	0.50	0.05	0.93	1.30	0.50	0.02	0.92	0.22
ρ	M1	0.11	0.05	0.70	77.18	0.05	0.01	0.74	22.85	0.05	0.00	0.69	3.55
	M2	0.10	0.03	0.54	24.38	0.08	0.01	0.37	11.62	0.07	0.00	0.00	1.77
	M3	0.06	0.02	0.79	47.29	0.05	0.01	0.78	57.12	0.05	0.00	0.77	113.03

我们对所有参数 $\beta.\gamma,\rho$均采用伽马分布 Gamma(0.1,0.1) 作为先验，其中比例 ρ的上限为 1。对于基于 DSA 的方法，我们通过 Stan 使用哈密顿蒙特卡洛抽取 10^5 后验样本（Neal，2012；Carpenter 等，2017），并将前 50% 的样本作为预热期丢弃。对于精确数据增广采样器，我们调整每个 Metropolis 步骤中更新的缺失数据比例，以达到 $20-25$% 的接受率。对于第一组参数组合的三个不同 N，我们考虑将增广数据的更新比例（论文中记作 ρ）设为 (0.4,0.08, 0.01)；对于另外两组参数，该比例分别设为 (0.4, 0.15,0.02) 和 (0.4,0.35, 0.07)。结果汇总于表1、表2、表3，其中列出了后验均值和标准差、95%可信区间的频率学派覆盖率，以及每秒的平均有效样本量。总体来看，在各种设置下，后验均值都接近真实参数；例外是使用 DSA模型进行推断且将缺失事件时间按均匀分布插补的情况。随着 N 增加，这种观察到的偏差并未消失，而其他所有方法的表现均如预期般改善。这凸显了正确考虑缺失信息分布的必要性；显然，在 SIR模型中，缺失信息具有高度结构性，并非随机缺失。我们发现，采用 DSA近似的方法，其可信区间覆盖率低于名义 95% 水平；鉴于我们在图1中观察到的变异性降低，这或许并不意外。值得注意的是，按均匀分布插补事件时间时，随着 N 增加，覆盖率会进一步下降

表 2：真实参数为 (β，γ，ρ) =(2，1，0.05) 时的参数估计。M1：使用精确的感染和康复时间；M2：使用均匀近似的感染时间；M3：计数似然；M4：PDSIR。

		N = 250				N = 1000				N = 10000			
		avg	sd	95%覆盖率	每秒有效样本量	avg	sd	95%覆盖率	每秒有效样本量	avg	sd	95%覆盖率	每秒有效样本量
β	M1	2.02	0.21	0.74	85.28	1.99	0.07	0.75	53.40	2.00	0.02	0.77	3.14
	M2	1.77	0.25	0.63	28.16	1.87	0.12	0.61	9.59	1.86	0.04	0.22	2.09
	M3	1.99	0.27	0.70	44.40	2.02	0.13	0.65	83.33	2.00	0.04	0.63	77.99
	M4	2.03	0.37	0.95	2.07	2.02	0.18	0.94	1.08	2.00	0.06	0.93	0.80
γ	M1	1.12	0.16	0.82	100.02	1.00	0.03	0.82	62.81	1.00	0.01	0.83	3.69
	M2	0.90	0.14	0.65	30.10	0.94	0.07	0.65	10.22	0.94	0.02	0.27	2.23
	M3	1.01	0.15	0.69	47.45	1.02	0.07	0.70	83.22	1.00	0.02	0.63	83.10
	M4	1.02	0.22	0.96	1.66	1.01	0.11	0.94	0.84	1.00	0.03	0.94	0.59
ρ	M1	0.08	0.03	0.59	94.61	0.05	0.01	0.61	59.71	0.05	0.00	0.64	3.51
	M2	0.09	0.03	0.64	27.86	0.06	0.01	0.65	10.13	0.06	0.00	0.22	2.23
	M3	0.07	0.03	0.66	44.21	0.05	0.01	0.64	83.64	0.05	0.00	0.68	83.34

而基于 DSA 的方法仍保持合理的覆盖率；对于 $R_0 = \beta/\gamma$较小的疫情暴发，这种偏差有所减轻。为确认这种覆盖不足是模型近似所致，而非推断方法所致，我们还从模型的后验分布中进行抽样，其中数据按照第 1 节所述由 DSA近似模拟得到。表 4 显示，在这种情况下，达到了名义覆盖率。

表 3：真实参数为 (β, γ, ρ) =(1.5, 1,0.05) 时的参数估计。M1：使用精确的感染和康复时间；M2：使用均匀近似的感染时间；M3：计数似然；M4：PDSIR。

		N = 250				N = 1000				N = 10000			
		avg	sd	95%覆盖率	每秒有效样本量	avg	sd	95%覆盖率	每秒有效样本量	avg	sd	95%覆盖率	每秒有效样本量
β	M1	1.54	0.18	0.81	86.93	1.50	0.06	0.83	59.63	1.50	0.02	0.83	1.99
	M2	1.25	0.31	0.56	16.17	1.41	0.16	0.57	23.72	1.44	0.05	0.52	1.88
	M3	1.50	0.35	0.55	83.90	1.47	0.16	0.57	37.01	1.51	0.05	0.61	42.92
	M4	1.52	0.37	0.94	1.65	1.53	0.20	0.94	0.88	1.50	0.06	0.94	0.23
γ	M1	1.09	0.13	0.84	95.05	1.01	0.04	0.86	64.87	1.00	0.01	0.85	2.17
	M2	0.86	0.21	0.58	16.83	0.95	0.10	0.58	24.27	0.96	0.03	0.50	1.92
	M3	1.02	0.24	0.54	86.52	0.98	0.11	0.60	37.85	1.01	0.03	0.60	43.79
	M4	1.04	0.28	0.93	1.43	1.02	0.15	0.93	0.77	1.00	0.05	0.94	0.20
ρ	M1	0.07	0.03	0.47	97.49	0.05	0.01	0.59	67.67	0.05	0.00	0.55	2.26
	M2	0.16	0.10	0.65	15.63	0.07	0.02	0.68	23.51	0.06	0.00	0.58	1.94
	M3	0.16	0.09	0.58	75.62	0.07	0.02	0.67	36.52	0.05	0.00	0.66	44.39

最后，基于 DSA 的方法在计算效率上实现了数量级的提升，这得益于 HMC采样下较低的自相关。与

使用完整数据进行的推断相比——而完整数据在真实数据环境中不可获得——我们基于边际似然的推断显示出估计方差的膨胀，这正是移除完整事件时间后我们所预期看到的。重要的是，其计算负载不会随 N 增长，这与在完整数据似然下对常微分方程系统进行数值积分形成对比，后者需要随 N 缩放的更小步长。

表 4：仅使用感染计数对由 DSA 近似模拟的数据集进行参数估计

	N = 250				N = 1000				N = 10000			
	Avg	SD	95%覆盖率	每秒有效样本量	Avg	SD	95%覆盖率	每秒有效样本量	Avg	SD	95%覆盖率	每秒有效样本量
$\beta = 2$	1.94	0.25	0.93	36.47	1.98	0.12	0.95	16.37	2.00	0.04	0.95	55.17
$\gamma = 0.5$	0.47	0.10	0.92	41.22	0.49	0.05	0.95	18.30	0.50	0.02	0.95	61.49
$\rho = 0.05$	0.06	0.02	0.93	37.33	0.05	0.01	0.96	17.59	0.05	0.00	0.95	59.80
$\beta = 2$	1.98	0.26	0.93	43.06	1.99	0.13	0.96	47.04	2.00	0.04	0.94	48.79
$\gamma = 1$	0.99	0.14	0.94	46.07	0.99	0.07	0.95	50.14	1.00	0.02	0.95	51.93
$\rho = 0.05$	0.06	0.02	0.94	43.15	0.05	0.01	0.96	49.73	0.05	0.00	0.94	52.19
$\beta = 1.5$	1.38	0.34	0.91	112.28	1.49	0.16	0.92	41.01	1.50	0.05	0.95	49.56
$\gamma = 1$	0.93	0.22	0.91	115.51	0.99	0.10	0.92	41.86	1.00	0.03	0.94	50.54
$\rho = 0.05$	0.09	0.07	0.91	89.17	0.05	0.01	0.93	40.86	0.05	0.00	0.94	51.26

5.2 脆弱性模型

我们现在转而在脆弱性模型框架下开展模拟研究。在本实验中，数据集通过修正的Sellke 构造进行模拟，其中 Q_i 服从Exponential(X)的次序统计量，且 $X \sim Gamma(1/\sqrt{\nu}, 1/\sqrt{\nu})$。这通过 X 引入了变异性的层次结构，并增加了一个描述脆弱性分布的额外参数 ν ，其余模拟设计保持不变。我们在此考虑三种参数组合：$(\beta, \gamma, \rho, \nu)$ =(2, 0.5,0.05, 0.1), (2, 1, 0.05, 1),(1.5, 1,0.05, 0.5)。对于每种设定，我们再次考虑三种不同的初始易感者人数， $N = 250$,1000, 10000，并对每种情形使用修正的Sellke 构造模拟500个数据集，仅保留每日感染数，而忽略连续的感染时间和康复时间。随后，我们使用(22)估计参数，其中 $\tilde{s}_t$是常微分方程(19)的解。表5汇总了全部四个参数的结果。

我们注意到可能会出现轻微的实际不可辨识性。由于脆弱性分布引入了额外参数，因此成功估计所有参数可能需要更多数据。尽管如此，我们发现，改用稍微不那么弥散的伽马分布(1，1)先验分布基本上能够解决这一问题。即使在估计脆弱性参数时，对关键流行病参数 β 和 γ 的估计仍然成功，且覆盖率与均质模型相近；不过，

表5：仅使用感染计数时脆弱性模型的参数估计，针对三组参数：(β，γ，ρ，ν) =(2，0。5，0。05，0。1)；(2，1，0。05，1)；(1。5，1，0。05，0。5)。

	N = 250				N = 1000				N = 10000			
	Avg	SD	95%覆盖率	每秒有效样本量	Avg	SD	95%覆盖率	每秒有效样本量	Avg	SD	95%覆盖率	每秒有效样本量
$\beta = 2$	1.93	0.26	0.82	22.32	1.99	0.13	0.83	49.48	2.00	0.04	0.86	30.10
$\gamma = 0.5$	0.35	0.14	0.78	17.74	0.42	0.09	0.82	19.23	0.48	0.03	0.94	11.27
$\rho = 0.05$	0.06	0.02	0.82	22.75	0.05	0.01	0.83	43.78	0.05	0.00	0.80	24.03
$\nu = 0.1$	0.29	0.17	0.98	17.73	0.23	0.13	0.97	19.87	0.13	0.07	0.97	32.71
$\beta = 2$	1.91	0.52	0.73	52.15	2.08	0.28	0.71	14.61	2.05	0.08	0.68	38.57
$\gamma = 1$	1.09	0.37	0.87	53.72	1.18	0.27	0.77	10.52	1.10	0.13	0.69	34.01
$\rho = 0.05$	0.13	0.09	0.76	44.46	0.06	0.02	0.76	14.01	0.05	0.01	0.82	40.00
$\nu = 1$	0.78	0.44	0.98	60.31	0.76	0.34	0.87	9.91	0.90	0.15	0.73	31.28
$\beta = 1.5$	1.58	0.35	0.72	77.23	1.47	0.21	0.68	15.30	1.52	0.09	0.68	3.66
$\gamma = 1$	0.81	0.32	0.74	57.50	0.86	0.23	0.79	11.04	1.00	0.12	0.80	3.08
$\rho = 0.05$	0.05	0.02	0.88	81.13	0.07	0.03	0.72	16.27	0.05	0.00	0.64	6.16
$\nu = 0.5$	0.90	0.45	0.89	66.53	0.67	0.34	0.94	11.79	0.47	0.20	0.86	3.11

例如在较小的总体中，ν等参数的估计会变得更加困难。总体而言，这项模拟证实了计数似然用于异质性脆弱性模型的有效性，不过，对于参数化更丰富的模型，使用信息量更高的先验分布有助于避免实际不可辨识性。

6 案例研究

我们现在将该方法应用于研究两个数据集——刚果民主共和国的2018–2020年埃博拉疫情，以及中国2020年的早期COVID-19疫情。

埃博拉疫情暴发的泊松网络SIR模型本节通过分析刚果民主共和国（DRC）2018–2020年埃博拉疫情的第三波疫情，展示泊松网络SIR模型计数似然的表现。此前有两项研究使用DSA分析了这些数据：两种方法都假设或插补了所有未观测到的流行病时间信息，其中Sharma等人（2022）拟合了SIR模型，而Wairimu 等人（2024）则使用泊松随机网络重新分析了这些数据。借助边际似然，我们可以重新分析这些数据，并直接考察每日病例数。这样做不必强行假设能够根据数据确定感染和康复的准确时间；这些数据基于住院记录、记录在案的症状出现时间，或患者自述的居家康复时间。关于该数据集的更详细说明，请参阅Sharma等人（2022）和Wairimu 等人（2024）。

继瓦　　irimu 等人（2024年），我们通过$\hat{N} = \frac{K_T}{1-S_T}$，估计未知规模~~$N$~~，其中e

K_T 是截至时间 T 的感染者总数， S 是（23）的解。疫情暴发在 $T = 108$ 天的窗口内被观测，共包含1068次感染事件。我们还˜遵循其先验设定，分布为 β， $\tilde{\gamma} \sim$ 伽马分布 (0.02，0.02)，且 ρ 服从˜ ˜均匀分布，约束为 $\beta \in$ (0.1， ∞) $\tilde{\gamma} \in$ (0， β) $\rho \in$ (0，0.015)。表6比较了Wairimu 等人（2024）论文中使用完整感染和康复时间得到的估计与我们使用计数似然得到的估计。由于推断现在仅基于

表6：比 较埃博拉数据的泊松网络模型参数估计值

参数	完整信息	计数似然
$\tilde{\beta}$	0.229（0.209，0.259）	0.227(0.155, 0.447)
$\tilde{\gamma}$	0.215（0.197，0.242）	0.205（0.134，0.426）
ρ	0.0055 (0.0046, 0.0073)	0.009（0.002，0.015）
μ	39.48（7.93，93.00）	-
$\tilde{R}_0$	1.071 (1.034， 1.109)	1.121 (1.044， 1.200)

在不对事件时间作任何额外假设的情况下，边际分析所得的后验可信区间反映了更大的不确定性。这些结果与以往分析一致：以往分析得出的区间更窄，且包含在我们的可信区间内，后验均值也相近。这些更宽的区间或许更能反映数据，因为在连续时间模型下，将症状出现时间视为精确感染时间可能并不现实。我们的分析得出的疫情暴发最终规模估计值较小，为 3502；此前分析的估计值为 3773，两者的可信区间有所重叠。

中国的 COVID-19接下来，我们将所提出的方法应用于分析中国的 COVID- 数据（Lab, 2020）。我们考察了从 53 个城市获得的发病数据，其中包括 2020 年 1 月 25 日至 3 月 5 日期间的每日感染数。为确保公平比较，我们参照 Yerlanov 等（2023）的研究，固定 $\gamma = 1/6$ 并采用类似的预处理方法，足以消除实际可辨识性问题。城市人口规模至少为 50 万，相较于不足 3,000 的样本量足够大，因此 DSA 背后的渐近假设是合理的。

我们分别使用边际似然（16）对随机 SIR 模型拟合 DSA近似，并使用似然（22）拟合一种脆弱性模型变体，以允许易感人群存在异质性。我们在 $\beta.\nu$ 服从伽马分布 (0.2,0.2)先验、 ρ服从均匀分布$(0, 1)$先验的条件下，通过 Stan 使用 HMC 从后验分布中抽样。为评估模型表现，我们将感染时间的密度与数据中的每日感染数原始直方图叠加比较。密度图是基于分别对这两个模型运行哈密顿蒙特卡洛后得到的模型参数后验均值估计绘制的；

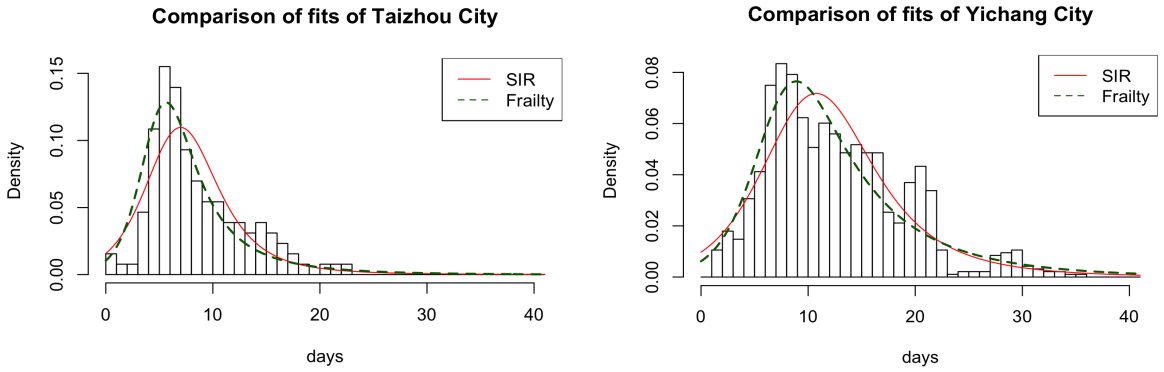

图2：每日感染数的直方图（归一化至面积为1），并叠加标准 SIR 模型和脆弱性模型的感染时间估计密度。

这些结果与 Yerlanov 等人（2023）的发现一致。

我们的分析表明，允许易感性变化所得到的拟合效果不逊于标准 SIR 模型，甚至更优。两种模型的参数估计大体一致，但在若干城市中，估计得到的脆弱性参数 ν 与零明显不同。在这些情况下，采用脆弱性模型可以改善拟合效果；图2展示了其中两个城市的结果。可以看出，与均质模型相比，变化的易感性模型所得的密度图与原始计数拟合得更好。值得注意的是，标准 SIR 模型与脆弱性模型的参数估计相近，但这些城市的基本再生数 R_0 的后验均值估计并未落在标准 SIR 模型拟合的95%可信区间内。这表明，尽管两种模型在总体层面上的密度图看起来相似，但关键流行病参数的估计差异在统计上可能具有显著性。更多详情、完整的后验估计和模型检查可在补充材料中链接的 GitHub代码仓库中查看。

由于脆弱性模型似乎能产生更优的拟合，这表明当忽略个体间易感性差异时，所推断的流行病参数可能存在偏倚。两种模型在一定程度上都捕捉到了流行病的总体趋势和形态。然而，在几乎所有城市中，仍有显著变异未能得到解释。这表明可以探索更精细的模型，以考虑接触网络、干预措施和病毒演化等因素（Bu et al., 2022; Hethcote, 2000; Rasmussen et al., 2011），从而更全面地理解流行病动力学。

7 讨论

本文提出一种闭式似然，旨在填补在缺失数据情形下拟合随机流行病模型的方法学空白。我们重点关注发病计数数据，但该方法同样适用于离散移除数据，或两者均可用且仅部分观测的情形。尽管本文侧重于贝叶斯框架下的后验抽样，但我们关于DSA近似下推断以及采用不同方法处理缺失数据之影响的见解，同样适用于频率学派似然框架。据我们所知，这是DSA近似下针对计数数据的首个闭式似然，并表明它提供了一种有吸引力的建模范式，使得即使在考虑模型变体时，边缘化及后续推断仍然便于用户使用。

许多现有的计数数据拟合方法都是针对特定流行病模型开发的，但我们已经看到，只要大样本行为能够用常微分方程系统描述，所提出的直接边际似然推断方法就很容易扩展到许多其他分舱模型。尽管该方法不如基于模拟的“即插即用”方法那样通用，但计算负担明显更轻，在标准笔记本电脑上只需数秒即可运行。虽然我们的实证结果令人鼓舞，但未来的工作可以尝试推导针对“缺失”方差的校正因子，以进一步改善可信区间的覆盖性质。研究采用线性噪声近似等易于处理的随机过程的相关方法之间的联系，将很有价值（Fintzi 等人，2022）。

同样，案例研究表明，尽管DSA框架下的标准 SIR 模型以及脆弱性修正都能较好地刻画总体流行病趋势，但仍存在相当大的变异性。这表明，纳入接触网络、干预措施、详细免疫特征或病毒演化的更现实模型，可能进一步提升模型的现实性与表达能力。由于更复杂的模型可能带来一定程度的实际可辨识性，审慎考虑先验信息或收缩的程度或可指导实践，而整合多来源数据也可能有助于实现这一目标。我们的论述聚焦于马尔可夫模型，但该方法可方便地扩展至非马尔可夫情形Lauro 等（2022），同时大体保留其简洁性与易用性。由于该框架具有相当的可扩展性与模块化特征，我们希望这些贡献能够引发疾病动力学数学建模、生存分析与随机过程建模交叉领域的更多视角。

补充材料

COVID-19数据可在 https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/MR5IJN获取。所有其他数据以及包含开源代码的完整实现可在 https://github.com/Suchismita-Statistics/DSA_CountData获得。

致谢

本研究部分由多尺度免疫系统建模卓越中心（美国国立卫生研究院 1U54-AI191253）和美国国家科学基金会 RAISE项目资助 DMS-2230074 资助。

参考文献

[1] Andersson, H.和Britton, T.（2000）。《随机流行病模型及其统计分析》，统计学讲义第*151*卷，第151卷。

[2] Balan, T. A. 和 Putter, H.（2020年）。脆弱性模型教程。《医学研究统计方法》，29(11):3424–3454。PubMed文献编号：32466712。

[3] Britton, T.（2019）。用于推断的随机流行病模型。

[4] Britton, T., Ball, F., andTrapman, P.（2020）。一个数学模型揭示了人口异质性对SARS-CoV-2 群体免疫的影响。《科学》，369(6505):846–849。

[5] Bu, F., Aiello, A. E., Xu, J., and Volfovsky, A.（2022）。动态网络中部分观测流行病的基于似然的推断。《美国统计学会杂志》，117(537):510–526。

[6] Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., and Riddell, A.（2017）。Stan：一种概率编程语言。《统计软件杂志》，76(1):1–32。

[7] Cauchemez, S. 和 Ferguson, N.（2008）。基于似然的连续时间流行病模型估计：应用于伦敦的麻疹传播时间序列数据。《皇家学会会刊·界面》/ 皇家学会，5：885–97。

[8] Chowell, G. (2017). 拟合流行病暴发的动态模型并量化不确定性：参数不确定性、可辨识性与预测入门。传染病建模，2(3):379–398.

[9] Fintzi, J., Cui, X., Wakefield, J., and Minin, V. N. (2017). 高效数据增强：将随机流行病模型拟合至患病率数据。计算与图形统计学杂志，26(4):918–929.[10]Fintzi, J., Wakefield, J., and Minin, V. N. (2022). 拟合至部分观测发病数的随机流行病模型的线性噪声近似。生物统计学，78(4):1530–1541.[11] Fleming, T. and Harrington, D. (1991). 计数过程与生存分析。概率与数理统计 Wiley 丛书。Wiley，纽约 [等]。[12] Gibson, G. J. and Renshaw, E. (1998). 使用马尔可夫链方法估计随机分舱模型中的参数。数学医学与生物学：*IMA*期刊，15(1):19–40.[13]Gomes, M. G. M., Corder, R. M., King, J. G., Langwig, K. E., Souto-Maior, C., Carneiro, J., Gonˌcalves, G., Penha-Gonˌcalves, C., Ferreira, M. U., and Aguas, R. (2020). 对 SARS-CoV-2 易感性或暴露的个体差异会降低群体免疫阈值。$medRxiv$。[14] Gumel, A. B., Iboi, E. A., Ngonghala, C. N., and Elbasha, E. H. (2021). 运用数学理解 COVID-19 动态的入门指南：建模、分析与模拟。传染病建模，6:148–168.[15] Guttorp, P. and Minin, V. (2017). 科学数据的随机建模，第二版。Chapman & Hall/CRC 统计科学丛书。CRC出版社。[16]Hethcote, H. W. (2000). 传染病数学。*SIAM*评论，42(4):599–653.[17] Ho, L. S. T., Crawford, F. W., and Suchard, M. A. (2018). 对离散观测的传染病随机分舱模型进行直接的基于似然的推断。[18] Hougaard, P. (1995). 生存数据的脆弱性模型。终生数据分析，1(3):255–273. [19] Izyumtseva, O., KhudaBukhsh, W. R., and Rempa la, G. A. (2024). 含随机效应的流行病模型的泛函大数定律。[20] Keeling, M. J. and Rohani, P. (2008). 人类与动物传染病建模。普林斯顿大学出版社，普林斯顿，新泽西州。

[21] Kermack, W. O. 和 Mckendrick, A. G.（1927）。对流行病数学理论的贡献。《英国皇家学会学报*A*辑：数学、物理与工程科学》，115:700–721。[22] KhudaBukhsh, W. R.、Bastian, C. D.、Wascher, M.、Klaus, C.、Sahai, S. Y.、Weir, M. H.、Kenah, E.、Root, E.、Tien, J. H. 和 Rempa la, G. A.（2023）。预测俄亥俄州的新冠肺炎病例与医院负担。《理论生物学杂志》，561:111404。[23] KhudaBukhsh, W. R.、Choi, B.、Kenah, E. 和 Rempala, G.（2019）。生存动态系统：基于群体层面流行病模型的个体层面生存分析。《界面焦点》，10:20190048。[24] King, A. A.、Nguyen, D. 和 Ionides, E. L.（2016）。通过 therpackagepomp 对部分可观测马尔可夫过程进行统计推断。《统计软件杂志》，69(12)。[25] Lab, S. D.（2020）。带底图的中国新冠肺炎每日病例。[26] Lauro, F.、KhudaBukhsh, W. R.、Kiss, I.、Kenah, E.、Jensen, M. 和 Rempala, G.（2022）。非马尔可夫流行病模型的动态生存分析。《英国皇家学会界面杂志》，19。[27] Marmot, M.、Friel, S.、Bell, R.、Houweling, T. A. J. 和 Taylor, S.（2008）。弥合一代人的差距：通过行动改善健康公平，关注健康的社会决定因素。《柳叶刀》，372(9650):1661–1669。[28]McKinley, T.、Vernon, I.、Andrianakis, I.、McCreesh, N.、Oakley, J.、Nsubuga, R.、Goldstein, M. 和 White, R.（2017）。复杂随机流行病模型的近似贝叶斯计算与基于模拟的推断。《统计科学：数理统计学会评论期刊》，33。[29] Montalb´an, A.、Corder, R. 和 Gomes, G.（2022）。个体差异与再感染条件下的群体免疫。《数学生物学杂志》，85。[30] Morozova, O.、Li, Z. R. 和 Crawford, F. W.（2021）。建模与预测新冠肺炎传播的一年：为康涅狄格州政策制定者提供支持。《科学报告》，11(1):20271。[31] Morsomme, R. 和 Xu, J.（2025）。针对部分观测发病数据拟合随机流行病模型的精确贝叶斯推断。《应用统计年刊》，19。[32] Neal, R.（2012）。使用哈密顿动力学的马尔可夫链蒙特卡罗。《马尔可夫链蒙特卡罗手册》。

[33] O’Neill, P. D. 和 Roberts, G. O.（2002）。针对部分观测随机流行病的贝叶斯推断。英国皇家统计学会会刊 *A*辑：社会统计学，162(1):121–129。[34] Rasmussen, D.、Ratmann, O. 和 Koelle, K.（2011）。利用谱系和时间序列对非线性流行病学模型进行推断。*PLOS*计算生物学，7:e1002136。[35] Rempala, G.（2023a）。动态生存分析：流行病的生存模型。讲座，西蒙斯计算理论中心，伯克利，加利福尼亚州。[36] Rempala, G. A.（2023b）。质量作用与泊松网络 SIR 流行病模型的等价性。[37] Rempala, G. A. 和 KhudaBukhsh, W. R.（2023）。用于流行病建模的动态生存分析，第1–17页。施普林格国际出版，查姆。[38] Rohatgi, V. 和 Saleh, A.（2011）。概率与统计学导论。威利概率与统计丛书。Wiley。[39] Salda˜na, F. 和 Velasco-Hern´andez, J. X.（2022）。新冠肺炎疫情建模：数理流行病学入门与综述。*SeMA*期刊，79(2):225–251。[40] Sellke, T.（1983）。随机流行病规模的渐近分布。应用概率期刊，20(2):390–394。[41] Sharma, P.、Suleman, S.、Farooqui, A.、Ali, W.、Narang, J.、Malode, S. J. 和 Shetti, N. P.（2022）。埃博拉病毒检测的分析方法。微化学期刊，178:107333。[42] Tian, J. P. 和 Kannan, D.（2006）。马尔可夫过程的可聚合性与交换性。随机分析与应用，24(3):685–702。[43] Wairimu, J. K.、Gothard, A. 和 Rempala, G. A.（2024）。泊松网络 SIR 流行病模型。[44] Wood, S. N.、Wit, E. C.、McKeigue, P. M.、Hu, D.、Flood, B.、Corcoran, L. 和 Abou Jawad, T.（2026）。COVID-19应对的若干统计学方面。英国皇家统计学会会刊 *A*辑：社会统计学，189(1):1–31。[45] Yan, P. 和 Chowell, G.（2024）。通过未观测个体异质性对次指数流行病增长动力学建模：一种脆弱性模型方法。数学生物科学与工程，21:7278–7296。

[46] Yerlanov、M.、Agarwal、P.、Colijn、C. 和 Stockdale、J.（2023）。简单传染病模型中的有效种群大小。《数学生物学杂志》，87。

[47] Zelner、J.、Masters、N. B.、Naraharisetti、R.、Mojola、S. A.、Chowkwanyun、M. 和 Malosh、R.（2022）。不存在平等机会感染者：流行病学建模者必须重新思考我们应对感染风险不平等的方法。*PLOS*计算生物学，18(2)：e1009795。