arXiv:2304.04752v1 [stat.AP] 31 Mar 2023

使用 Pumas 进行药物计量学贝叶斯推断的实践者指南

MohamedTarek[1, 2*], Jose Storopoli[1,3],Casey Davis[4],Chris Elrod[1,5], JuliusKrumbiegel[1],Chris Rackauckas[1,5,6] 和 Vijay Ivaturi[1,7]1*Pumas-AI公司，美国。2悉尼大学商学院，澳大利亚。3新奥罗里藏特七月九日大学（UNINOVE）计算机科学系，巴西。4GileadSciences 公司，美国。5JuliaHub（前身为 JuliaComputing）公司，美国。6麻省理工学院计算机科学与人工智能实验室，美国。7马里兰大学巴尔的摩分校药学院，美国。

*通讯作者。电子邮件：mohamed@pumas.ai；贡献作者：jose@pumas.ai； cbdavis33@gmail.com；chrise@pumas.ai；juliusk@pumas.ai；crackauc@mit.edu； vijay@pumas.ai；

摘要

本文为使用 Pumas 工作流程的药物计量学贝叶斯实践者提供了一份全面教程。我们首先简要说明贝叶斯推断在药物计量学中的应用动机，并重点介绍现有软件的局限性以及 Pumas 如何解决这些问题。随后，我们将通过代码片段和示例，逐步介绍药物计量学标准贝叶斯工作流程中的所有步骤，包括：模型定义、先验选择、从后验分布中采样、先验与后验模拟及预测、反事实模拟与预测、收敛诊断、视觉预测检验，最后是使用交叉验证进行模型比较。最后，我们用浅显易懂的语言解释贝叶斯统计中许多高级概念的背景与直觉，其中包括用户在进行贝叶斯分析时需要牢记的许多重要观点和注意事项。本文介绍的许多算法、代码和理念广泛适用于临床研究和统计学习，但考虑到 Pumas 主要面向药物计量学家的软件定位，同时也为了缩小讨论范围，我们选择将讨论重点放在药物计量学上。

关键词： 贝叶斯推断，统计软件，药物计量学，工作流程

目录

1 动机 1

2 引言 1

2.1 药物计量学工作流程.12.2 数据.12.3 模型.12.4 算法.22.5 软件.32.6 其他软件的局限性.32.7 相关工作与贡献.32.8 论文布局与阅读计划.4

3 Pumas 中的贝叶斯工作流程 4

3.1 定义模型.53.2 先验模拟与预测.113.3 模型拟合.113.4 数值误差与调试.143.5 使用新数据更新后验分布.153.6 基本汇总统计量.163.7 需要多少样本？.163.8 诊断图.163.9 更多诊断.173.10 如果链未收敛怎么办？.193.11 高级后验查询.193.12 后验图.203.13 后验模拟与预测.213.14 视觉预测检验与模拟图.223.15 模拟查询.233.16 非房室分析（NCA）参数.253.17 交叉验证与期望对数预测密度.253.18 信息准则.29

4 背景与直觉 29

4.1 符号说明.304.2 贝叶斯统计.324.3 先验选择.344.4 马尔可夫链蒙特卡罗（MCMC）的直觉理解.374.5 无调采样器（NUTS）算法.394.6 基本汇总统计量.474.7 收敛性.474.8 交叉验证与模型选择.49

5 示例模型 51

6 结论 55

7 致谢 55

1 动机

全贝叶斯方法已成为药物计量学家工具箱中的重要工具（Lee 和 Gobburu，2011；McDade 等，2022年），因为它们能够严谨而灵活地量化模型所有参数的不确定性，并利用以往类似研究中的知识。这些方法可应用于罕见病药物审批、儿科研究、精准给药和自适应试验设计。Pumas 中实现的贝叶斯工作流程（Rackauckas 等，2020年）旨在兼具灵活性与易用性，并采用直观、简洁的语法。我们希望这篇论文能成为一份优质学习资源，帮助有兴趣了解和使用全贝叶斯方法的药物计量学家。

2 引言

本节将讨论在药物计量学中采用完全贝叶斯方法的必要性，并描述其在该领域通常使用的方法谱系中的定位。将简要介绍药物计量学中使用的标准模型、数据、算法、工作流程和软件，为后续讨论设定背景。最后，将总结 Pumas 的主要贡献以及本文的布局。

2.1 药物计量学工作流程

药物计量学是一门包含各种用于理解药物与患者之间相互作用的定量分析技术的研究领域。典型的药物计量学工作流程包括：

1. 从临床试验源数据集准备可用于分析的数据集，2. 通过汇总和图表对观测数据进行探索性数据分析和科学推理，3. 开发用于描述疾病-药物-患者相互作用的参数模型，4. 将模型参数拟合到数据并估计参数的不确定性，5. 诊断模型的拟合情况并评估拟合质量，6. 比较模型并选择最佳模型，以及

7. 使用最佳模型为现有患者或新患者预测/模拟替代场景，或回答关键的药物开发问题。

2.2 数据

药物计量学中使用的数据通常包括：

1. 患者协变量，例如年龄或性别，有时也包括时变协变量，2. 每位患者的药物给药方案和给药时间表，以及 3. 每位患者的观测响应，例如血液中测得的药物浓度和/或某些临床终点，通常在多个时间点进行观测。

2.3 模型

药物计量学中使用的模型类型通常包括：

1. 用于捕捉观测响应的结构模型，例如使用常微分方程（ODE）对药代动力学进行建模的基于动态的模型；2. 根据协变量预测观测响应的协变量模型；以及 3. 具有群体层面参数和患者特异性参数的分层模型，[1]

这些分类并非互斥。例如，基于协变量的模型也可以是包含基于动态的常微分方程组件的层次模型。

以下展示了一个经典的茶碱动力学模型，采用单房室药代动力学（PK）口服吸收模型。该模型描述药物吸收入血流及其清除的动力学。假设药物最初以单次剂量给入单个储库室中，例如，药物刚被摄入肠道时。随后，药物以吸收速率 Ka ×Depot 逐渐进入血流，其中 Depot 表示肠道中剩余的药物量。中央室

[1]在本文中，我们使用“总体层面”参数和“患者特异性”（或受试者特异性）参数这一术语，而不使用药物计量学中更为常见的“固定效应”和“随机效应”。这是因为统计学文献中对固定效应和随机效应的定义存在歧义（参见 Gelman (2005) 的第 21 页），并且在贝叶斯建模中，每个参数都使用随机变量进行建模，这进一步增加了这种歧义。

表示药物随时间在血流中吸收后的药量。还假设药物以$\mathrm{CL}\,V$ ×Central 的速率从中央室清除，其中 V 是该房室的分布容积。

该模型包含群体参数（$\boldsymbol{\theta}$,$\boldsymbol{\Omega}$, σ）和患者特异性参数 $\boldsymbol{\eta}_i$。 $\boldsymbol{\theta}$ 是一个由 4 个数字组成的向量，$\boldsymbol{\Omega}$ 是一个 3×3 正定协方差矩阵，而 σ 是误差模型的标准差参数。在该模型中，每位患者 i都有体重和性别协变量：

$$Z_i = \begin{bmatrix} \mathrm{wt}_i, \\ \mathrm{sex}_i, \end{bmatrix}, \tag{1}$$

其中，wt_i 是一个正实数，sex_i 是一个指示变量，男性取值为 0，女性取值为 1；每位患者都有个体系数：

$$\begin{bmatrix} \mathrm{Ka} \\ \mathrm{CL} \\ V \end{bmatrix} = \begin{bmatrix} \theta_1 e^{\eta_{i,1}} \\ \theta_2 (\frac{\mathrm{wt}_i}{70})^{0.75} \theta_4^{\mathrm{sex}_i} e^{\eta_{i,2}} \\ \theta_3 e^{\eta_{i,3}} \end{bmatrix}, \tag{2}$$

具有内部动力学：

$$\begin{aligned} \frac{d[\mathrm{Depot}]}{dt} &= -\mathrm{Ka}[\mathrm{Depot}] \\ \frac{d[\mathrm{Central}]}{dt} &= \mathrm{Ka}[\mathrm{Depot}] - \frac{\mathrm{CL}}{V}[\mathrm{Central}], \end{aligned}$$

其中，Depot 和 Central 分别是储库房室和中央房室；观测到的药物浓度 conc 还存在正态分布残差误差，这也称为误差模型：

$$\mathrm{conc} \sim \mathrm{Normal}\left(\frac{\mathrm{Central}}{V}, \sigma\right).$$

2.4 算法

在本节中，我们简要概述药物计量学中通常用于将模型拟合到数据的各种算法。

2.4.1 边际最大似然估计（最大似然估计）

在药理计量学中将参数模型拟合到数据时，边际最大似然估计算法最为常用。患者特异性参数被边际化，并最大化边际似然。有两类使用边际似然进行最大似然估计的算法：

1. 条件似然的近似积分（Wang，2007），包括：(a)拉普拉斯方法，以及 (b) 一阶条件估计（FOCE）。2. 随机逼近期望最大化（SAEM）（Delyon et al，1999；KuhnandLavielle，2005）。

2.4.2 边际最大后验估计（MAP）

边际最大后验估计是边际最大似然估计的一种替代方法，同样可以为群体参数提供点估计，但它不是最大化边际似然，而是最大化群体参数的先验概率与边际似然的乘积。

2.4.3 完全贝叶斯分析

边际似然最大化可以作为副产品，在给定最优群体参数的情况下，为我们提供患者特异性参数的条件后验分布 η_i 或其近似结果。然而，完全贝叶斯方法可以同时为我们提供所有群体参数和患者特异性参数的联合后验分布样本。联合后验分布的采样通常使用 MCMC 算法来完成，例如掉头采样器（NUTS）（Hoffman 和 Gelman，2014；Betancourt，2017）算法，该算法是哈密顿蒙特卡洛（HMC）（Neal，2011）MCMC算法的一个变体。我们将在本论文的后续章节中更详细地介绍贝叶斯记号和算法。

除了能够从联合后验分布中采样并轻松模拟后验不确定性外，完全贝叶斯方法还允许建模者：

1. 使用先验分布纳入领域知识和先前研究的见解。2. 量化所谓的认知不确定性，即当模型不可识别[2] 和/或可用数据点不多时，模型参数取值所存在的不确定性。

以上是贝叶斯分析的优势，而非贝叶斯工作流通常无法给出令人满意的答案。

[2]如果某个参数不可识别，就意味着无法根据现有数据唯一确定其取值。

作为一种范式，贝叶斯推断运用成熟的概率理论，以较少的模型和数据假设，严格量化参数值中的不确定性。

贝叶斯工作流程使分析人员能够在具备领域知识时利用领域知识，对模型参数的认知不确定性进行量化，从而为分析流程带来极大的灵活性。

由于贝叶斯分析在概念上是对自助法的替代，因此贝叶斯推断的性能应与自助法进行比较，而不应与使用Laplace/FOCE/SAEM进行单次模型拟合的性能比较。这一点对于设定用户的预期非常重要，因为贝叶斯推断通常会比使用Laplace/FOCE/SAEM运行单次模型拟合慢一个或多个数量级。

2.5 软件

许多软件都实现了所有经典的基于极大似然估计的分析工作流程，例如 NONMEM（Beal 等人，2009）、Monolix 和 Pumas（Rack-auckas 等人，2020年）。对于完全贝叶斯分析，有一些可选方案，包括：Pumas、Stan（Carpenter 等人，2015）、Torsten（Margossian 等人，2022年）、BUGS（Goudie 等人，2020年）、Turing.jl（Ge 等人，2018）、PyMC（Salvatier 等人，2015）、Pyro（Bingham 等人，2018）以及 NONMEM 相对较新的贝叶斯模块。

在贝叶斯统计中，当模型仅包含连续参数时（这在药代动力学模型中通常如此），Stan 已成为贝叶斯分析事实上的标准软件。多年来，许多来自不同领域、使用不同类型模型的统计学家都对 Stan 进行了测试并加以使用。在上述用于贝叶斯分析的软件中，Stan 拥有最大的社区，以及最成熟的 R 软件包生态系统，可支持完整的贝叶斯分析工作流程。

2.6 其他软件的局限性

通用概率编程语言（PPLs），例如 Stan 或 Turing.jl，在药物计量学（PMx）中的主要局限在于，它们并非针对 PMx 用户以及 PMx 中常见的模型和数据而设计。例如，截至本文撰写时，在 Stan[3] 中，受试者间并行难以实现。

Torsten 试图弥合药理计量学用户与 Stan 软件之间的差距，例如让用户更轻松地定义给药方案，并通过其群组求解器简化受试者间并行。有关 Torsten 群组求解器的更多信息，请参阅 metrumresearchgroup.github.io/Torsten/function/ode-pop/。然而，Stan 和 Torsten 的应用程序接口都无法为药理计量学用户提供最佳的使用体验，因为它们基于 C++；C 可以说是一种难以使用、非交互式的编程语言。此外，在撰写本文时，在 Torsten 中定义依赖于模型参数的动力学初始条件仍然不必要地困难，而在其上游概率编程语言 Stan 中则更容易做到这一点。

更广泛地说，一旦假设了特定的模型结构，就可以采用许多计算优化，例如自动并行化和自动常微分方程线性性检测；我们在 Pumas 中就是这样做的，但在其他更通用的软件中却很难实现。

其他软件的第二个局限是，它们不支持药理计量学中常见的其他非贝叶斯工作流，这意味着具有药理计量学背景的用户需要将模型从其他软件迁移到概率编程语言（PPL），才能比较贝叶斯和非贝叶斯结果。这通常是一个耗时且容易出错的过程。

2.7 相关工作与贡献

本文并非首篇尝试在药物计量学背景下阐述贝叶斯工作流程的论文。诸如 Wakefield 等人(1999) 和 Lunn 等人 (2002) 等早期尝试使用了 BUGS 软件（Goudie 等人，2020年）。近年来，Torsten 论文（Margossian 等人，2022年）和贝叶斯工作流程论文（Gelman 等人，2020b）也都是很好的参考资料，重点介绍如何使用 Stan（Carpenter 等人，2015）。Gelman 等人(2020b) 特别关注贝叶斯模型开发、分析和诊断的工作流程方面；尽管它并非专为药物计量学定制，Torsten 论文（Margossian

[3]为了在受试者之间进行并行化，您需要使用似然中的`reduce sum`函数，这并不是一项简单的任务。有关使用 `reduce sum` 函数的示例，您可以参考stanpmx.github.io。

等人，2022年）试图填补这一空白。此外，还有Elmokadem 等人 (2023)，其提供了使用Stan/Torsten 或 Turing.jl（Ge 等人，2018）构建生理学基础药代动力学（PBPK）模型的指南。鉴于上述优秀资源，我们将本文的两项主要贡献概括如下：

1. 提供完整的药物计量学导向的贝叶斯工作流程，包括使用 Pumas 软件的脚本和常见模型示例。
2.以直观的方式解释各种贝叶斯统计和计算概念，从而能够微调算法的选项，并在不过于技术化的情况下理解软件输出。

我们认为，Pumas 贝叶斯工作流程的实现通过以下方式解决了其他软件的许多问题：

1. 通过几行代码提供完整、易用的贝叶斯分析工作流程，包括：(a) 模型和给药方案定义，(b) MCMC 抽样，(c) 诊断，(d) 模拟，(e) 反事实预测，以及 (f) 可定制的交叉验证；2. 对贝叶斯和非贝叶斯分析中使用的基于层次结构和动力学的模型采用相同的易用、紧凑模型语法，因此无需转换模型；3. 使用 Julia 作为核心编程语言，它快速、交互性强且易于使用，能够提供出色的用户体验；4. 自动完成药理计量学模型的定义和计算优化，例如采用受试者间自动并行，以及自动进行常微分方程线性性和刚性检测，从而以简洁的语法实现高性能；5. 使用开源、模块化实现 AdvancedHMC.jl (Xu 等人，2020年)，该实现采用与 Stan 相同的 MCMC 算法，Turing.jl 中也使用了这一算法 (Ge 等人，2018)。

2.8 论文布局与阅读计划

本文其余部分的组织如下。在第3节中，我们介绍贝叶斯工作流程

Pumas。这应能帮助已经熟悉贝叶斯工作流程和理论的用户快速开始使用Pumas。在第 4 节中，我们会退后一步，对贝叶斯工作流程中涉及的所有主要概念和思想作出较为全面且注重直觉的介绍，包括所用算法背后的直觉，以及选择先验时需要注意的一般事项。本节主要使用通俗易懂的语言，并辅以一些基础数学，可以帮助用户理解高度技术性的概念，从而在使用 Pumas 时作出明智的决策。最后，第5节包含一些常见的示例模型，第 6 节则包括一些供当前和未来 Pumas 用户参考的补充阅读材料和资源。

本文旨在既作为Pumas 用户的教程论文，也作为药物计量学家学习贝叶斯推断的易读资源。如果您是经验丰富的贝叶斯研究者，并希望学习如何使用 Pumas，那么可以只阅读第3 节，跳过第4节。如果您仍在学习贝叶斯理论，我们建议您至少先阅读第4 节的部分内容。随后，您可以使用表1 切换阅读第3节中的相关部分，如果您更喜欢交织学习理论与代码。若想了解更有针对性的教程和脚本，您还可以查看 Pumas 教程网站（tutorials.pumas.ai）。

3 Pumas 中的贝叶斯工作流程

一般的贝叶斯分析工作流程在Gelman 等人 (2020年b)中进行了介绍，概述了贝叶斯统计目前的许多最佳实践。一般的贝叶斯工作流程与基于边际最大似然估计的标准药代动力学工作流程之间存在许多相似之处。不过，两者之间也存在一些值得强调的差异：

1. 贝叶斯工作流程关注的是，当模型的 MCMC 抽样失败时，建模者可以采取哪些措施。失败的原因可能是采样器完成采样耗时过长、采样器频繁触发数值误差，或样本在采样后未能通过诊断测试。简化模型、使用较少或较多的数据等做法有助于找出问题所在并改善采样。其中许多】【。

主题	第 3节	第 4节
模型定义和先验选择	3.1	4.1，4.2，4.3
先验与后验分布的预测和模拟	3.2，3.11-3.16	4.4
推断算法和选项	3.3-3.6	4.5-4.6
收敛性和诊断	3.7-3.10	4.7
交叉验证与模型选择	3.17-3.18	4.8

表1：本表展示了第3 节和第 4.

如果最大似然估计/最大后验估计优化失败，这同样适用；但与马尔可夫链蒙特卡罗以其众多失败模式之一失败相比，优化失败的情况并不常见。

2。贝叶斯工作流程包括先验选择诊断，例如先验预测检验。边际最大似然估计优化不使用群体参数的先验，而患者个体参数通常有成熟的标准先验可供使用，因此在贝叶斯框架之外，这也不太成问题。

3. 一般的贝叶斯工作流程提到使用模型和MCMC样本进行预测，但自然没有具体说明预测的类型，因为这在很大程度上取决于具体情境。在药物计量学中，我们通常关注的是确定为新患者或既有患者使用不同剂量所产生的影响，或者根据既往数据预测/模拟未来反应。

在本节中，我们将借鉴通用的贝叶斯工作流程，根据药物计量学进行定制，并介绍有助于药物计量学家在执行贝叶斯分析时遵循当前最佳实践的 Pumas 语法。此处介绍的语法使用的是 Pumas 2.4（计划于2023年6月发布），但 Pumas 2.3 已经实现了大部分工作流程。如需了解更新后的语法、选项和功能，请参阅 Pumas 文档（docs.pumas.ai）。

3.1 定义模型

3.1.1 概述

在 Pumas 中，您可以使用 @model宏 定义模型，该宏由模型块组成。本教程将介绍六个模型块：

1. @param：我们在此定义模型的人群参数及其先验。

2. @random：在此定义模型的个体特异性参数及其先验。

这是一个可选区块，单受试者模型可以省略。

3. @covariates：用于声明受试者的协变量。这是一个可选区块，对于不包含协变量的模型可以省略。4. @pre：在此执行各种预计算和其他统计变换，例如计算个体药代动力学参数。5. @dcp：在此定义模型中的剂量控制参数（DCP）。这是一个可选区块，如果模型中不存在 DCP，则可以省略。6. @dynamics：用于定义所有动力学，例如控制 PK/PD 房室之间关系的常微分方程系统。这是一个可选区块，对于不包含常微分方程的模型可以省略。7. @derived：在此定义用于计算似然的观测响应分布。8. @observed：在此定义模拟期间需要计算、但在拟合期间不计算的其他变量，例如可以在此基于浓度曲线执行非房室分析（NCA）。这是一个可选区块。

借助这些模块，您几乎可以在 Pumas 中编写所有 PK/PKPD/PBPK 模型。我们将介绍所有这些模型模块的功能。

首先，`@param` 块用于包含模型的所有群体参数。我们先指定一个 `begin ... end` 结构，并在其中每行插入一个参数。对于每个参数，我们使用 ~ 运算符，后跟一个分布来指定先验。清单 1 展示了一个包含三个参数及其先验的 `@param` 块示例。`tvcl` 的先验是 对数正态分布，对数尺度均值为 log(2.5)，对数尺度标准差为 1；`tvvc` 的先验是一个约束为正值的 正态分布

先验，均值为 70，标准差为 10；而σ 的先验是速率为 3 的指数先验。[4]

```
@param begin
    tvcl ~ LogNormal(log(2.5), 1)
    tvvc ~ Constrained(Normal(70, 10); lower =
    0)
    σ ~ Exponential(3)
end
```

列表 1：@param 块示例

接下来，@random 块包含所有受试者特异性参数及其先验。与@param 块类似，我们也从指定一个begin... end 结构开始，在其中每行插入一个参数；并为每个参数指定一个先验，使用 ~运算符，后接一个分布。在列表2中，我们给出了一个包含单个参数的@random 块示例 η ，该参数具有多元正态分布（多元正态）先验，均值为 0，协方差矩阵为单位矩阵。

```
@random begin
    η ~ MvNormal([1 0; 0 1])
end
```

列表 2：@random 块示例

使用@covariates 块指定受试者的协变量。这只是一个声明块，采用相同的方式，在begin ... end 语句中每行声明一个协变量。列表 3 中给出了一个示例，其中声明了两个受试者协变量：WT 表示受试者的体重，SEX 表示受试者的性别。

```
@covariates begin
    WT
    SEX
end
```

列表 3：@covariates 块示例

我们继续介绍@pre 代码块，其中会执行模型所需的任何预计算或统计变换。在此代码块中，我们可以使用前面代码块中声明的所有参数和协变量。到目前为止，这种处理方式与其他代码块类似：在begin ... end 语句中，每行进行一次预计算或变换。列表4展示了@pre 代码块的示例，其中我们为每位受试者计算其个体药代动力学参数。

```
@pre begin
    CL = tvcl * exp(η[1]) * (WT / 70)^0.75
    Vc = tvvc * exp(η[2]) * (WT / 70)
end
```

列表 4：@pre 块示例

第五个代码块，@dynamics，是我们指定控制模型动力学的常微分方程组的地方。在此代码块中，我们在 begin ... end 语句内每行声明一个常微分方程。方程的左端（LHS）是房室的导数，即变化率。我们可以自由地为每个房室命名。在左端，房室名称后跟一个 ’（撇号）运算符，用于表示该房室的变化率。撇号运算符是声明房室导数的一种直观方式。方程的右端（RHS）是@pre代码块中指定的房室与各个参数的某种组合。在列表 5的示例中，我们使用参数 2 中介绍的茶碱模型的动力学，其中参数为 Ka、CL 和 Vc。

```
@dynamics begin
    Depot' = -Ka * Depot
    Central' = Ka * Depot - CL/Vc * Central
end
```

清单 5：@dynamics 块示例

Pumas 支持自动检测常微分方程线性性和刚性。因此，即使是线性常微分方程，也可以采用上述易读的语法编写，而不会损失性能。此外，如果检测到常微分方程具有刚性，Pumas 将根据需要切换到刚性常微分方程求解器。

@derived 块用于定义似然项。在此块中，我们可以使用两种赋值方式：确定性赋值 =和概率性赋值 ~。对于确定性赋值，右端项是一个确定性量；而对于概率性赋值，右端项是由某个分布表示的概率性量。在清单 6中，我们定义了两个变量[5]。第一个是cp ，通过确定性赋值定义；第二个是conc，通过概率性赋值定义，同时使用 cp 作为对数正态分布的一个参数。在此示例中，conc是观测值，称为因变量。我们将 conc 的分布定义为对数正态分布，其对数尺度均值为 logcp ，且

[4]要在 Pumas中输入像 σ 和 θ 这样的 LaTeX 符号，可以先输入 LaTeX 形式，例如 $\backslash sigma$，然后按 Tab 键。

[5]我们使用宏 @. 将其右侧的所有函数调用向量化。由于 Central 和 cp 是观测时间点处的值向量。

对数尺度标准差 σ （来自我们的 @param 块）。

```
@derived begin
    cp = @. Central / Vc
    conc ~ @. LogNormal(log(cp), σ)
end
```

示例 6：@derived 块示例

@observed 块可用于根据浓度曲线计算其他量，例如非房室分析参数。

```
@observed begin
    nca := @nca cp
    auc = NCA.auc(nca)
    cmax = NCA.cmax(nca)
end
```

清单 7：@observed 块示例

除了上述代码块外，还提供了更多用于以下内容的代码块：

- @init 用于手动初始化动力学，或
- @vars用于定义简写符号变量，以便在@dynamics和@derived 块中使用。

有关Pumas 模型各个组成部分的更多信息，请参阅 Pumas 文档（docs.pumas.ai）。

3.1.2 示例：药代动力学模型

为便于说明，请考虑以下具有一级吸收的一室模型：

$$\begin{aligned}\text{Depot}' &= -\text{Ka}\cdot\text{Depot}\\ \text{Central}' &= \text{Ka}\cdot\text{Depot} - \frac{\text{CL}}{V_C}\cdot\text{Central}\end{aligned}$$

其中，CL 为中央室的消除清除率； V_C 为中央室容积；Ka 为吸收速率常数。

如果我们只有一个受试者，这个模型可以在 Pumas 中使用除@random 之外的所有已见代码块进行编码，因为单受试者模型不需要该代码块。代码清单 8 展示了该模型的代码。

这是一个完整的贝叶斯 Pumas 模型。我们在 @param 块中指定参数及其先验。我们只有一名受试者，因此无需加入 @random 块，因此，在 @pre 块中定义的 PK 个体参数与群体参数相同。在 @dynamics 块中，我们声明了两个控制模型动力学的常微分方程。动力学包含两个房室，分别命名为 Depot 和 Central。最后，在 @derived 块中，我们通过确定性赋值，将 cp 计算为 Central 室除以 PK 参数 Vc 的函数，并定义观测响应 conc 服从对数正态分布，其对数尺度均值为 logcp，对数尺度标准差为 σ。

清单中的模型 8 是一个单受试者模型，但大多数情况下，数据中有多个受试者，因此需要定义一个群体模型。可以通过添加一个`@random` 块来定义个体特异性参 数 η。然后，我们可以将群体参数和个体特异性参数结合起来，在 `@pre` 块中定义个体药代动力学参数。

我们选择为 ηs 指定一个高斯先验分布，其协方差矩阵使用相关性和标准差进行参数化，具体说明见4.3.3。列表9 中的模型就是这种参数化方式的一个示例。它基于先前的单受试者药代动力学模型，增加了另外 2 个群体参数：一个相关矩阵 C 和一个标准差向量 ω，二者位于`@param` 代码块中。 C 相关矩阵采用 Cholesky 参数化的 LKJ 先验（推荐使用的先验），而 ω 标准差向量服从一个受正值约束的多元正态分布，其协方差矩阵为对角矩阵，截断前均值为 0，方差为 0.4^2。我们在 η@random 代码块中，通过使用由相关性和标准差构成的协方差矩阵，以及 Pumas 函数 `cor2cov`，为s指定多元高斯先验。最后，在 @pre 代码块中，我们将个体药代动力学参数定义为群体参数与个体特异性参数组合后的变换。

3.1.3 Selecting Prior Distributions

选择新的先验用于某个参数可能是一项令人望而生畏的任务。一般而言，不存在一种先验能够适用于所有情况；如果能找到良好的参考，遵循以往类似研究中的先验可能是一个不错的做法。然而，如果你面临选择合适先验的任务

```
@model begin
    @param begin
        tvcl ~ LogNormal(log(2), 0.25)
        tvvc ~ LogNormal(log(70), 0.25)
        tvka ~ LogNormal(log(1.5), 1)
        σ    ~ Constrained(Cauchy(), lower = 0)
    end
    @pre begin
        CL = tvcl
        Vc = tvvc
        Ka = tvka
    end
    @dynamics begin
        Depot'   = -Ka * Depot
        Central' =  Ka * Depot - (CL/Vc) * Central
    end
    @derived begin
        cp  := @. Central / Vc
        conc ~ @. LogNormal(log(cp), σ)
    end
end
```

代码清单 8：药代动力学单房室单受试者模型示例

```
@model begin
    @param begin
        tvcl ~ LogNormal(log(2), 0.25)
        tvvc ~ LogNormal(log(70), 0.25)
```

清单 9：群体药代动力学单房室一级吸收模型示例 e

对于一个全新的模型，确定先验分布通常是一个多步骤过程，包括：

1. **确定先验的支撑集**。先验分布的支撑集必须与参数域。例如，对于正参数，可以使用不同于参数介于 0 和 1 之间时所用的先验。表 2 可根据参数的域帮助缩小可用选项列表。

支撑集	分布
(0，1)	贝塔分布、KSOneSided、NoncentralBeta、Logit 正态分布
(0，∞)	贝塔分布Prime,卡分布, 卡方分布, 埃尔朗分布, 指数分布, F分布, 弗雷歇分布, 伽马分布, 逆伽马分布 逆高斯, 柯尔莫哥洛夫, 对数正态分布 NoncentralChisq, 非中心F分布, 瑞利, 威布尔分布
(−∞，∞)	柯西分布、Gumbel、Laplace、逻辑斯谛分布、正态分布、NormalCanon, 正态逆高斯分布 ，广义高斯分布 ，TDist
实向量	多元正态分布
正向量	多元对数正态分布
正定矩阵	Wishart分布，逆Wishart分布
相关矩阵	LKJ、LKJCholesky
其他	Constrained ， 截断分布 ， 位置-尺度分布 ， 均匀分布，反正弦分布, 双权重分布, 余弦分布, 埃潘尼奇尼科夫分布, 半圆分布, 对称三角分布, 三权重分布, 帕累托分布, 广义帕累托分布、广义极值分布、列维分布

表2：该表列出了 Pumas 中一些最常用的先验分布及其对应的支撑域。在 Pumas 命令行提示符中输入 ?, 后接分布名称，即可了解有关各分布的更多信息。

2. **确定先验的中心**，例如均值、中位数或众数。

3. **确定先验的强度**。这通常由相应分布函数中的标准差或尺度参数控制。较小的标准差或尺度参数意味着参数值的不确定性较低，因此先验更强。较大的标准差或尺度参数意味着不确定性较高，因此先验较弱。建议在使用任何先验分布之前，仔细评估是否采用该分布。这有助于确保先验的强度反映你对参数值的置信程度。有关先验选择的更多详情，请参阅第4.3节的讨论。

4.确定先验的形状。有些分布左偏，有些右偏，还有些是对称的。有些分布的尾部比其他分布更厚，例如，与正态分布相比，学生 t 分布以尾部更厚而闻名。概率密度函数（PDF）的形状应反映在观测数据之前对参数值的认识。

在选择新的先验时，除了第 4.3 节中的讨论外，您还可以参考 github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations，其中的提示和建议也很有用。若想深入了解各种先验选择，还可参阅 Simpson 等人 (2014)。

对于单变量分布，可以使用 PumasPlots.lines[6] 函数绘制分布的概率密度曲线，例如：

```
using PumasUtilities
dist = Normal(0.0, 1.0)
PumasPlots.lines(dist)
```

对于多变量和矩阵变量分布，可以使用 rand、mean 和 var 函数从分布中抽样，并了解这些值的分布情况。例如：

```
dist = LKJ(3, 1.0)
x = rand(dist, 100) # 100 samples
mean(x) # mean
var(x) # element-wise variance
```

以下介绍了 Pumas 中一些最常用的先验分布：

1. Normal(μ, σ)：支撑集为（$-\infty$，∞）的单变量正态分布，均值为 μ，标准差为 σ。2. LogNormal(μ, σ)：支撑集为（0，∞）的单变量对数正态分布，对数尺度均值为 μ，对数尺度标准差为 σ。3. MvNormal(μ, Σ)：均值向量为 μ、协方差矩阵为

[6]，它是 PumasUtilities 包的一部分

Σ。矩阵 Σ 也可以是对角矩阵，例如 Diagonal([1.0, 1.0])。也可以单独传入Σ 作为矩阵，例如 MvNormal(Σ)，此时均值将默认为 0。4. MvLogNormal(μ, Σ)：定义在正向量上的多元对数正态分布，其对数尺度均值向量为 μ，对数尺度协方差矩阵为 Σ，定义方式同上面的 MvNormal。5. Cauchy(μ, σ)：支撑集为（$-\infty$, ∞）的单变量柯西分布，位置参数为 μ，尺度参数为 σ。6. Constrained(dist, lower = l, upper =u)：支撑集固定为 (l, u) 的约束先验分布，基分布 dist 也固定，可以是任意单变量或多元分布。lower 和 upper 分别设置随机变量支撑集的下界和上界，默认值分别为 -Inf ($-\infty$) 和 Inf (∞)。当 dist 为单变量分布时，lower 和 upper 应为标量。约束多元分布时，lower 和 upper 可以是向量或标量。如果设为标量，则所有随机变量都使用相同的界限。此外，还有一种截断分布，它与 Constrained 不同之处在于，基分布可以是模型参数的函数，但截断分布仅支持单变量基分布。通常建议在 @param 块中使用 Constrained，在 @random 和 @derived 块中使用 truncated。示例：

- Constrained(正态分布(0.0, 1.0), lower= 0.0) 是一个半正态分布。
- Constrained(柯西分布(0.0, 1.0), lower= 0.0) 是半柯西分布。
- Constrained(多元正态分布([0.0, 0.0], [1.00.0;0.0 1.0]), 下界 = 0.0) 是约束多元正态分布。

init 关键字参数也可用于指定参数的初始值，例如 Constrained(正态分布(), lower= 0.0, init = 1.0) 7. truncated(dist, lower, upper)：类似于 Constrained，分别具有固定的下界 lower 和上界 upper，且 a

基分布 dist。设置 upper 是可选的；若未设置，默认值为 Inf (∞)。在截断分布中，基分布 dist 可以依赖于模型参数，并且每次进行对数概率计算时都会计算归一化常数。不过，lower 和 upper 必须是固定常数，而且截断分布仅支持单变量基分布。示例：truncated(Normal(0, 1), 0.0, Inf) 是半正态分布。truncated(Cauchy(), 0.0, Inf) 是半柯西分布。truncated(Normal(), -Inf, 0.0) 是负半正态分布。8. Uniform(l, u)：下界和上界分别为 l 和 u 的单变量均匀分布。9. LKJ(d, η)：定义在大小为 $d \times d$ 的相关矩阵上的矩阵变量 LKJ 先验。η 是 LKJ 先验的正形状参数。减小 η 会使样本的相关性更接近 ± 1。此外还有 LKJCholesky，它在语义上与 LKJ 相同，但具有一些优势。见下文。10. LKJCholesky(d, η)：LKJ 分布的 Cholesky 分解形式，所采样的矩阵采用因子分解形式。出于性能考虑，建议在模型中使用它，而非 LKJ。11. Wishart(ν, S)：定义在 $d \times d$ 正定矩阵上的矩阵变量 Wishart 分布，其自由度为 ν，尺度矩阵 S 为正定矩阵。12. InverseWishart(ν, Ψ)：定义在 $d \times d$ 正定矩阵上的矩阵变量逆Wishart分布，其自由度为 ν，尺度矩阵 Ψ 为正定矩阵。13. Beta(α, β)：支撑集为 0 到 1、形状参数为 α 和 β 的单变量贝塔分布。14. Gamma(α, θ)：定义在正数上的单变量伽马分布，形状参数为 α，尺度参数为 θ。15. Logistic(μ, θ)：支撑集为 $(-\infty, \infty)$、位置参数为 μ、尺度参数为 θ 的单变量逻辑斯谛分布。16. LogitNormal(μ, σ)：支撑集为 (0, 1) 的单变量 Logit 正态分布，其基正态分布的均值为 μ，标准差为 σ。

17. TDist(ν)：支撑集为 $(-\infty, \infty)$、自由度为 ν 且均值为 0 的单变量 Student t 分布。若要改变 T 分布的均值，可以使用位置-尺度分布（见下文）。

18. LocationScale(μ, σ, d)：由基分布 d 缩放和平移得到的单变量分布。基分布的随机变量先乘以 σ 进行缩放，再加上 μ 进行平移。示例：LocationScale(1.0, 2.0, TDist(2)) 是经过缩放和平移的 Student's t 分布。LocationScale 分布的均值为 $\mu + \sigma \times$ mean(d)，标准差为 $\sigma \times$std(d)。

19. Laplace(μ, σ)：支撑集为 $(-\infty, \infty)$、位置参数为 μ、尺度参数为 σ 的单变量 Laplace 分布。

20. Exponential(θ)：支撑集为 $(0, \infty)$、尺度参数为 θ 的单变量指数分布。
21. （非正规）平坦先验：除了使用分布，也可以指定一个域，例如用于向量参数的 VectorDomain、用于正定参数的 PSDDomain，或用于相关矩阵参数的 CorrDomain。在 Pumas 中，这些域会被视为平坦先验。如果域是开集，那么这就是一个非正规先验。有关域的更多信息，请参阅 Pumas 文档

（docs.pumas.ai ）.

3.2 先验模拟和预测

定义模型并指定先验后，可以运行先验预测检验，以检查先验预测或模拟与真实数据的接近程度。可以通过将模型、模型和受试者/人群、数据作为参数传入 simobs 函数来运行先验模拟。

```
sims = simobs(model, data; samples = 100)
```

清单 10：先验模拟

samples关键字参数指定要从先验分布中为数据中的每个受试者模拟的样本数量。

在 simobs函数中，还有一个关键字参数：simulateerror。如果将其设置为 true（默认值），Pumas 将从 @derived块中的响应误差模型进行采样（即模拟）；否则，它将返回误差分布的期望值（即预测）。后者等同于使用 predict 函数。

根据模拟结果，可以运行各种统计分析和查询。相关内容将在第3.15节中说明。还可以按照第3.14节所述，使用视觉预测检验（VPC）绘制模拟结果。当模拟来自先验模型时，VPC 通常称为先验预测检验。

3.3 模型拟合

现在我们已经有了模型，接下来需要使用 Pumas 对其进行拟合。这正是 fit 函数的作用，它接受四个位置参数：

1. 一个 Pumas 模型 2. 一个 Pumas 群体 3. 初始参数估计值
4. 一个拟合算法

在最大似然估计场景中，拟合算法可以是 FOCE 或 Laplace 的实例。但若要改用 MCMC 进行贝叶斯推断，则可以将其设为以下任一算法的实例：

- BayesMCMC
- MarginalMCMC

MarginalMCMC 通过先积分消去受试者特异性参数，从边际后验分布中抽取样本；而 BayesMCMC 则从联合后验分布中抽取样本。在某些情况下，MarginalMCMC 的速度可能远快于 BayesMCMC，但它目前仍处于实验阶段，未来还会进一步改进。构造算法实例时，通过关键字参数传入 BayesMCMC 和 MarginalMCMC 的选项，例如 BayesMCMC(nsamples= 2000)。BayesMCMC 和 MarginalMCMC 均可设置的主要选项包括：

- 目标接受率：NUTS 算法的目标接受率，默认为 0.8
- nsamples：要生成的马尔可夫链蒙特卡罗（MCMC）样本数，默认为 2000
- nadapts：NUTS 算法的适应步数，默认为 1000
- 链数：要采样的 MCMC 链数，默认为 4
- 每条链的有效样本量：每条链的目标有效样本量（ESS）；达到目标后将终止采样，默认为 nsamples
- 检查间隔：检查每条链的 ESS 所需的样本数

• 时间限制：以秒为单位的采样时间限制；如果达到该时间限制，采样将终止；默认为 Inf（在 Julia 中为 ∞）

• ensemblealg：可设置为 EnsembleSerial() 以进行串行采样，设置为 EnsembleThreads() 以进行多线程采样，或设置为 EnsembleDistributed()以进行多进程（即分布式并行）采样。默认情况下，如果有足够的线程/进程可用，将启用跨链和受试者的并行处理。

• 并行链：可设为 true 或 false。若设为 false，将不会并行抽取各条链。若设为 true，则会根据 ensemblealg 的值，使用多线程或多进程并行抽取各条链。默认值为 true。

• 并行受试者：可以设置为true或false。如果设置为 false，日志概率计算将不会进行并行化。当受试者数量较少时，建议采用此设置。如果设置为true，则将根据 ensemblealg 的值，使用多线程或多进程在受试者之间进行日志概率计算的并行化。当有足够的线程/进程同时对链和受试者进行并行处理时，默认值为 true。

• rng：所使用的随机数生成器
• diffeq 选项：包含所有常微分方程求解器选项的命名元组，例如，可使用 diffeq - options = -(alg = Rodas5(),) 强制 Pumas 使用刚性常微分方程求解器 Rodas5，而不是依赖 Pumas 的自动刚性检测和自动切换行为。

• constantcoef：一个 NamedTuple，其中包含采样期间要固定的参数。此参数可用于在将某些参数固定为特定值的条件下，从后验分布中采样；例如，constantcoef= (σ = 0.1,) 会将 σ 参数固定为 0.1，并从其余参数在给定 σ 的条件下的后验分布中采样。

MarginalMCMC 算法还包含一个关键字参数 marginal alg，默认值为 LaplaceI()，但也可以设为 FOCE() 或 FO()。默认情况下，BayesMCMC 和 MarginalMCMC 都会并行运行 4 条马尔可夫链，并利用剩余的计算资源对各受试者的计算进行并行处理。默认情况下会运行 2,000 次马尔可夫链蒙特卡罗迭代，并将每条链的前 1,000 个样本作为预热期。Pumas 不会自动丢弃预热期样本。因此，用户需要使用函数 丢弃 来丢弃预热期样本。如果你使用的是 Pumas ¡ 2.4，应改用 丢弃，而不是 Pumas.truncate。列表 11展示了一个贝叶斯 Pumas 模型拟合示例。我们将结果保存到 res ，并对其调用丢弃 函数，并将关键字参数 预热期设为 1,000 个样本。这样会通过丢弃每条链的前 1,000 个样本，输出截断后的拟合结果。请注意，在 Julia 中，1_000 和 1000 等效。

```
res = fit(model, pop, iparams, BayesMCMC(
ns
```

清单 11：在 Pumas 中拟合贝叶斯模型

您还可以向 ratio函数传递一个 discard 关键字参数，以丢弃 (1 - ratio) ×100% 的样本。这称为稀疏化，其工作方式是从每条链中每 1/ratio 个样本中选择 1 个样本。一般来说，在最终分析中通常不建议进行稀疏化，因为这会导致部分信息丢失。不过，在运行大量探索性模拟/预测的初始探索阶段，为了更快地进行迭代，可能需要进行稀疏化。另一个例子见清单 12 ，其中我们使用了 MarginalMCMC。

```
res = fit(model, pop, iparams, MarginalMCMC(
ns
```

列表 12：使用自定义参数在 Pumas 中拟合贝叶斯模型

使用 BayesMCMC 或MarginalMCMC拟合模型时，您可以通过如图 1所示显示的实时进度信息查看采样器的进度。在使用多线程或分布式并行时，每个字段会显示区间或比率，而不是数值。以下是图 1中显示的最重要字段的说明：

• 迭代次数指已完成的马尔可夫链蒙特卡罗迭代次数，包括适应阶段和采样阶段。

• n_步骤 是上一次提议中采取的时间步数。如果这个数值过大，NUTS 采样器将会非常缓慢且低效。有关步数的更多信息，请参见第4.5.6节。

```
[ Info: Checking the initial parameter values.
[ Info: The initial log probability and its gradient are finite. Check passed.
Sampling chain 1/4:   44%|                                  |  ETA: 0:00:08
  iterations:                   885
  time_elapsed:                 20.236122131347656 s
  n_steps:                      31
  is_accept:                    true
  acceptance_rate:              0.968519931389641
  log_density:                  -234.71908842849837
  hamiltonian_energy:           243.90493988798036
  hamiltonian_energy_error:     0.06733964736008602
  max_hamiltonian_energy_error: -0.29577139290324794
  tree_depth:                   4
  numerical_error:              false
  step_size:                    0.09892375602818866
  nom_step_size:                0.09892375602818866
  is_adapt:                     true
  mass_matrix:                  DiagEuclideanMetric([0.21619708846081584, 0.014 ...])
```

图 1：使用 Pumas 进行采样时显示的实时进度信息。

```
Iterations        = 1:1:2000
Number of chains  = 4
Samples per chain = 2000
Wall duration     = 35.26 seconds
Compute duration  = 35.26 seconds
parameters        = θ₁, θ₂, θ₃, θ₄, θ₅, Ω₁,₁, σ²

Summary Statistics
  parameters      mean       std   naive_se      mcse         ess      rhat   ess_per_sec
      Symbol   Float64   Float64    Float64   Float64     Float64   Float64       Float64

          θ₁    1.7893    0.5693     0.0064    0.0138   1863.5976    1.0025       52.8575
          θ₂    0.0855    0.0068     0.0001    0.0001   2797.6008    1.0036       79.3488
          θ₃    0.0390    0.0021     0.0000    0.0000   2874.2471    1.0036       81.5227
          θ₄    1.7363    0.5305     0.0059    0.0124   2087.4707    1.0013       59.2073
          θ₅    0.4517    0.1478     0.0017    0.0019   7151.3168    1.0005      202.8340
       Ω₁,₁    1.6356    1.0969     0.0123    0.0250   1798.0761    1.0032       50.9991
          σ²    1.1942    0.1566     0.0018    0.0021   4460.5045    0.9997      126.5140

Quantiles
  parameters      2.5%     25.0%     50.0%     75.0%     97.5%
      Symbol   Float64   Float64   Float64   Float64   Float64

          θ₁    0.7094    1.4185    1.7630    2.1317    3.0194
          θ₂    0.0731    0.0809    0.0851    0.0900    0.0994
          θ₃    0.0350    0.0376    0.0389    0.0404    0.0432
          θ₄    0.7128    1.3916    1.7213    2.0589    2.8274
          θ₅    0.1647    0.3527    0.4516    0.5491    0.7473
       Ω₁,₁    0.5233    0.9566    1.3540    1.9637    4.4267
          σ²    0.9270    1.0862    1.1840    1.2884    1.5432
```

图 2：贝叶斯拟合结果显示示例，由 Pumas 中的 fit 函数输出。

- 表示已接受：如果上一次提议已被接受，则为 true，否则为 false。有关提议和 NUTS算法的更多信息，请参阅第4.5节。
- 接受率指的是所有过去提议的平均接受率。它应收敛到与目标接受率相近的值。

适应阶段后的选项集。有关接受比的更多信息，请参阅第4.5.5节。

- 对数密度指参数和观测值的联合概率的对数。如果在采样后期该值单调增加

拟合的采样阶段，这表明采样器很可能没有收敛到该区域（或多个区域）。

适应过程中具有较高的后验概率质量，而且链可能不会收敛。有关不收敛的迹象的更多信息，请参见第4.7 节；有关对数密度单调递增（即优化行为）的更多信息，请参见第 4.5.5节。

- 树深度是生成最后一个提议时达到的最大树深度。有关详细信息，请参见第4.5.6节。
- 步长是 NUTS 算法中的时间步长，该步长在适应阶段进行调整，并在采样阶段固定。有关步长及其与 tar 的联系，

g等等 ACME ptance 比率，参见第4.5.5节。
- 在适应阶段为 true，在采样阶段为 false。

下面展示了 拟合函数结果的一个示例，见图2。系统会自动显示群体参数的多个汇总统计量。你还可以使用其他 Pumas 函数以编程方式查询特定的汇总统计量，而不只是查看显示结果。有关汇总统计量函数的更多信息，请参阅第 3.6节和第3.11节。

3.4 数值误差与调试

3.4.1 模型中的数值不稳定性

在特定参数值（η, θ）下，每次计算对数似然都需要对模型进行完整求值，包括结构模型（@pre 块）、数值求解微分方程（@dynamics 块），以及计算似然（@derived 块）。为了有效地进行贝叶斯推断，需要确保所有模型块在数值上稳定，不会产生 Inf、-Inf 或非数值。数值不稳定性可能由多种原因造成，但常见原因包括：

1.除以一个非常小的数或0，例如，如果某个参数位于分母中，并且在拟合过程中允许其为0。2. 使用较大的指数调用指数函数，例如，由于初始参数值不佳。

3. 根据您的模型，某些观测值在特定 (η, θ)下的概率可能为 0 或近似为0。例如，当使用伯努利响应分布，且伯努利分布的概率参数恰好为0（或1）时，数据中存在 1（或0）观测值。4. 常微分方程求解器无法收敛到解，因为对于特定的极端 (η, θ)，动力学并不稳定。5.响应分布的标准差为0，例如使用比例误差模型且浓度降至 0 时。 6. 对负参数取对数或平方根。

导致数值误差的初始参数值，往往比拟合过程中出现的中间不良值更值得警惕。这是因为，当中间模型导致数值误差时，系统会自动拒绝它们。发生这种情况时，可能会看到以下警告：

```
Warning: The current proposal will be rejected
                                                .se, false)
```

此警告不一定意味着出现问题或估计过程失败。它表示在某次运行中，$p(\eta, \theta|D)$ 的数值估计失败了。在许多情况下，这种警告无害且符合预期，可能会在贝叶斯估计过程中出现。例如，常微分方程模型的参数取值过于极端，可能导致动力学不稳定，从而使模拟器发散。马尔可夫链蒙特卡罗步进过程会通过拒绝该步并提出新的 (η, θ) 值进行尝试，从而恢复。因此，如果在估计过程中只看到少数此类警告，通常无需担心。

不过，如果显示了过多警告，这可能意味着许多步骤被拒绝，导致马尔可夫链蒙特卡罗过程无法有效探索后验分布，进而可能无法收敛。另一个迹象是，步进过程也显示了以下附加警告：

```
NaN dt detected. Likely a NaN value in the state
this outcome.
```

此警告意味着常微分方程求解器失败。显示此警告是因为计算

常微分方程初始时间步长 dt（即常微分方程求解过程的第一部分）时，结果为非数值。这通常是由于常微分方程参数中存在非数值或无穷大值所致。如果情况确实如此，不妨检查一下 @pre 块中的各个参数值是否合理。一个快速的方法是在 @model 定义中添加代码，打印出模拟过程中使用的当前值。例如，下面这行

```
VC = θ[3] * exp(η[2]
```

可以更改为

```
VC = @pumasdebug θ[3] * exp(η[2]
```

这将打印出每一步计算得到的值。通过这些输出，可以直接查看 @dynamics 块中使用的常微分方程参数。@pumasdebug 是 Pumas 2.4 中的一项功能，此前的版本不支持。

通过这种调试方式发现的最常见问题，有些是由于参数边界设置不正确所致。例如，如果某个参数由固定效应或随机效应相除得到，那么当分母接近零时，该参数值可能会变得（近乎）无限大。因此，解决方法是确保为导致分母（接近）为零的效应适当设置下界。

3.4.2 常微分方程求解器中的数值不稳定性

如果常微分方程参数看起来是合理的候选值，却被拒绝了，那么可能需要确保常微分方程求解过程适用于当前方程。以下任何警告通常都表明，常微分方程求解器是由于算法原因而非参数中出现非数值或 Inf 值而失败：

```
Warning: Interrupted. Larger maxiters is needed.
```

```
Warning: dt(x) <= dtmin(y) at t=z. Aborting.
```

```
Warning: Instability detected. Aborting
```

为了调试这类情况，重新使用@pre 生成的确切参数求解常微分方程，并直接调用 DifferentialEqua- tions.jl 中的常微分方程求解器函数（Rackauckas 和 Nie (2017）可能会有所帮助。不过，以下是一些常见建议：

•降低容差。通常，一组非常稳健的容差（代价是性能有所下降）为 abstol=1e-12、reltol=1e-12。可以将其作为 sampling algorithm 中的 diffeq - options 关键字参数的一部分进行设置，例如 BayesMCMC(diffeq options =(abstol=1e-12, reltol=1e-12,)).

• 将常微分方程求解器更改为专为刚性微分方程设计的方法。Rodas5P() 是一种常见选择。同样，可以使用 diffeq options 关键字参数进行设置，例如 BayesMCMC(diffeq options = (alg= Rodas5P(),))

常微分方程求解器发生数值故障的一个常见原因，是常微分方程中存在一种称为刚性的性质。刚性难以严格定义，但可以粗略地理解为常微分方程中各变化率的时间尺度差异很大。例如，一个值的导数为 1，而另一个值的导数为 10^9。这可能导致常微分方程求解器出现困难，进而影响模拟和贝叶斯推断过程。默认情况下，Pumas 使用的常微分方程求解器会在一种较快但鲁棒性较低的非刚性常微分方程求解方法与一种较慢但鲁棒性较高的刚性常微分方程求解方法之间切换。然而，有时这种默认行为不如要求所有步骤都使用刚性常微分方程求解器稳定，因此第二条建议是手动切换常微分方程求解器。

3.5 使用新数据更新后验分布

有一些算法可以根据每位受试者的新观测值或新受试者，高效地更新后验样本，例如序贯蒙特卡洛。截至本文撰写时，Pumas 尚未实现这些算法。要根据现有受试者的新数据或新受试者更新后验分布，目前必须对整个数据集重新拟合模型。或者，您可以使用易于处理的分布族（例如多元高斯分布）近似后验样本，然后仅使用后验近似作为先验分布，对新数据重新拟合模型。在 Pumas 的未来版本中，我们计划实现此类高效的后验样本更新方法。有关最新功能列表，请参阅 Pumas 文档（docs.pumas.ai）。

3.6 基本汇总统计量

若要查询群体参数的若干基本汇总统计量，可以使用：

```
summarystats(tres)
```

其中，tres 是 fit 或 discard 的结果。此操作将输出样本均值、样本标准差、蒙特卡罗标准误（MCSE）、ˆ有效样本量（ESS）、 R 以及每秒 ESS。

若要获取该 i^{th} 受试者的个体特异性参数的相同汇总统计量，可以使用 subject 关键字参数：

```
summarystats(tres, subject = i)
```

3.7 需要多少个样本?

要准确估计不同的量，所需的样本数可能各不相同。例如，要估计罕见事件的概率或某些极端分位数，所需样本数会远多于估计后验分布均值所需的样本数。默认情况下，Pumas 会生成 4 条链，每条链包含 2000 个样本，其中 1000 个用于适应。根据你要估计的量，可能需要让采样器运行更长时间，并检查随着链中样本数增加，结果是否发生显著变化。

更具体地说，Vehtari 等人（2019）建议将 ESS设为 400，作为一个良好的目标值。对于 Pumas 默认运行的 4 条链，这相当于每条链的 ESS 为 100。在Gelman 等人（2013a）较早和较新的版本中，也报告了相同的ESS建议值。虽然这只是一个一般性指南，并不适用于极端分位数估计或罕见事件概率的估计，但它可以作为一个值得努力达到的良好初始目标。第 3.3 节中讨论的每条链的 ess 选项可用于设置每条链的目标 ESS。

除了目标 ESS 外，还应确保ˆ该 R 诊断值小于 1.1。如果能低于 1.01 则更好，正如ˆ在 Vehtari 等人 (2019) 中所建议的那样。估计 ESS、R 以及用于估计后验均值以外的不同量（例如参数分位数）的 MCSE，目前在 Pumas 中尚未实现，但这是一项即将推出的功能。该功能可以让用户更深入地了解其 MCMC 样本的估计准确性。

3.8 诊断图

有几种诊断图可以帮助您识别缺乏收敛性的情况，包括：轨迹图、累积均值图和自相关图。所有诊断图都需要先加载 PumasUtilities 软件包：

```
using PumasUtilities
```

假设 tres 是拟合或丢弃的输出。

3.8.1 轨迹图

参数的轨迹图显示该参数在 MCMC 算法每次迭代中的值。一个好的轨迹图应满足以下条件：

- 具有噪声，并不是一条递增或递减的线。
- 具有固定的均值。
- 具有固定的方差。
- 显示所有链彼此重叠，这也称为链混合[7]。

你可以使用 trace plot 函数绘制轨迹图，例如：

```
trace_plot(tres; parameters = [:tvcl])
```

图3 展示了参数 `tvcl`的轨迹图。你可以在 参数 关键字参数中添加更多参数名称，例如参数 `= [:tvcl, :tvvc]`，以绘制更多参数。如你所见，所示的轨迹图具备优质轨迹图的许多理想特性。

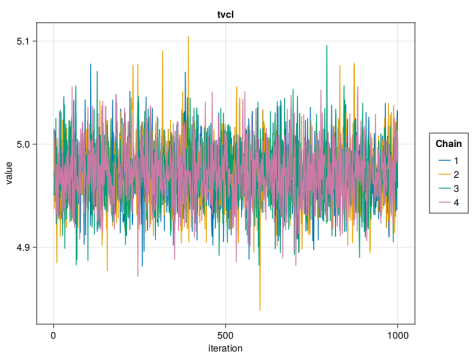

图 3：轨迹图示例。

[7]链混合指的是不同的链包含来自后验分布相同区域的样本，而不是每条链都包含来自后验分布不同区域的样本。

如果未指定参数关键字参数，则会显示所有群体参数的轨迹图。若要绘制一组受试者的个体特异性参数的轨迹图，可以改为设置 subjects 关键字参数，而不是设置参数关键字参数，例如：

```
trace_plot(tres; subjects = [1, 2])
```

有关更多详细信息和示例，请参阅 Pumas 文档（docs.pumas.ai）。

3.8.2 累积均值图

参数的累积均值图显示了每条 MCMC 链截至某次迭代时参数值的均值。收敛到平稳后验分布的 MCMC 链，其每个参数的累积均值都应收敛到一个固定值。此外，对于给定参数，所有链都应收敛到相同的均值，即后验均值。如果累积均值曲线没有收敛，或者各条链收敛到不同的均值，这就表明存在不收敛。

你可以使用以下方法绘制总体参数 tvcl 的累积均值图：

```
cummean_plot(tres; parameters = [:tvcl])
```

图 4 显示了参数 tvcl的轨迹图。与轨迹图类似，你可以在 parameters关键字参数中添加更多参数名称，也可以完全省略该参数，以绘制所有总体层级参数。同样，也可以使用 subjects关键字参数，而不是 parameters关键字参数，来绘制个体特异性参数的相同图表。

3.8.3 自相关图

马尔可夫链蒙特卡罗链中的样本容易出现自相关，因为链中的每个样本都是前一个样本的带噪函数。自相关图显示了索引为 s 的每个样本与相应的、滞后 s + 的样本之间的相关性，其中s ∈ 1:N-lag，N是样本总数。对于每个滞后值，我们都可以计算样本与其向前滞后若干步的对应样本之间的相关性度量。相关性通常是介于 0和 1 之间的值，但有时也可能介于 -1 和 0 之间。自相关图显示

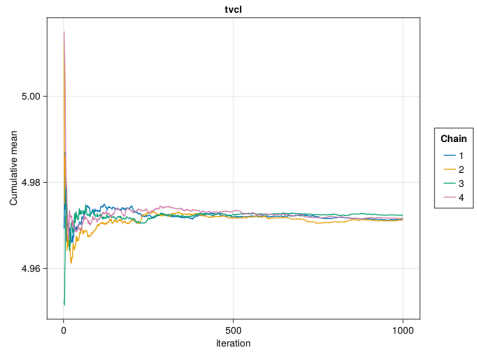

图 4：累积均值图示例。

x轴上的滞后值和 y轴上的相关值。对于表现良好的 MCMC链，随着滞后值增加，相应的相关性会越来越接近 0。这意味着，彼此距离越远的任意两个样本之间的相关性越来越小。相关性接近0时的滞后值可用于指导MCMC 样本的稀疏化，以便从自相关样本中提取出大部分相互独立的样本。可以使用discard 函数执行稀疏化，并将ratio 关键字设置为 1 / lag，其中 lag 取适当的值。

```
discard(tres; ratio = 1/lag)
```

话虽如此，一般来说，最终分析通常不建议进行稀疏化，因为这会造成一定的信息损失。不过，在初始探索阶段，运行许多探索性模拟/预测时，为了加快迭代，可能会希望进行稀疏化。

你可以绘制总体参数 tvcl 的自相关图：

```
autocor_plot(pk_1cmp_tfit; parameters = [:tvcl])
```

图 5展示了参数 tvcl的自相关图。与轨迹图类似，你可以在 parameters 关键字参数中添加更多参数名称，也可以完全省略该参数，此时将绘制所有总体层级参数。同样，可以使用 subjects 关键字参数而非parameters关键字参数来绘制个体特异性参数的相同图形。

3.9 更多诊断

还有许多其他诊断方法可以帮助你识别：

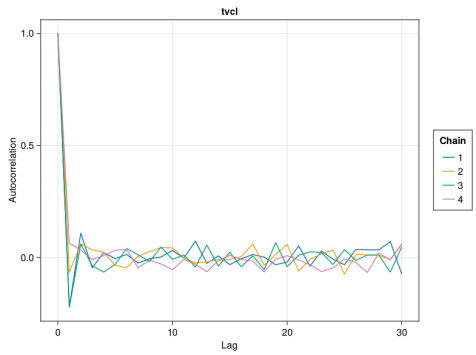

图5：自相关图示例。

- MCMC 算法尚未收敛的情况，或者
- 作为预热期丢弃多少样本。

通常，我们建议在使用 discard函数移除适应步数后运行这些诊断。其中一些诊断随后可以告诉您，在移除适应步数后还需丢弃多少样本作为预热期。discard 函数可以再次作用于其自身的输出，以移除额外的样本作为预热期。

3.9.1 Geweke 诊断

```
gewekediag(tres; subject = nothing, first = 0.1,
    last = 0.5)
```

上述函数会计算每条链的 Geweke 诊断（Geweke，1991），并为每个参数输出一个 p 值。`tres` 是来自拟合 或 丢弃 的输出，其余关键字参数采用上面所示的默认值。如果 受试者 关键字参数被设置为 `nothing`（默认值）或被省略，则诊断的是群体参数的链。如果受试者 被设置为整数索引，则诊断的是与输入索引对应的受试者的个体特异性参数的链。

Geweke 诊断使用正态分布均值差异假设检验，对整个链中两个不相交子链 X_1 和 X_2 的样本均值进行比较，其中原假设和备择假设定义如下：

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

其中，μ_1 和 μ_2 为总体均值。第一个子链 X_1 取链中样本的前（ first * 100 ） %，其中 first 是一个关键字参数，默认值为 0.1。第二个子链 X_2 取链中样本的后（last * 100）%，其中 last 是一个关键字参数，默认值为 0.5。

使用的检验统计量为：

$$z_0 = (\bar{x}_1 - \bar{x}_2) \Big/ \sqrt{s_1^2 + s_2^2}$$

其中，$\bar{x}_1$ 和 $\bar{x}_2$ 分别是 X_1 和X_2 的样本均值，s_1 和 s_2 分别是 X_1和 X_2 的马尔可夫链标准误（MCSE）估计值。假设每条子链内的样本存在自相关，但假设 X_1 中的样本与 X_2中的样本相互独立。输出的 p 值是 $P(|z| > |z_0|)$ 的估计值，其中 z 是服从标准正态分布的随机变量。

较低的 p 值表示以下情况之一：

- 链的首尾部分分别从均值不同的分布中采样，即不收敛，
- 需要丢弃一些初始样本作为预热期，或者
- 由于缺乏样本或自相关较高，需要运行采样更长时间。

较高的p值（理想情况）表明，无法得出链的前段和后段的均值存在统计显著差异的结论。然而，仅凭这一点并不能保证收敛到固定的后验分布，因为：

- X_1 和 X_2 的标准差或高阶矩可能不同，或者
- 当存在较高的自相关性时，X_1 与X_2 之间的独立性假设可能不成立。

3.9.2 Heidelberger 和 Welch 诊断

```
heideldiag(tres; subject = nothing, alpha =
    0.05, eps = 0.1, start = 1)
```

上述函数会针对每条链计算 Heidelberger 和 Welch 诊断（Heidelberger 和 Welch，1983）。如果将受试者 关键字参数设置为`nothing` （默认值）或省略该参数，

所诊断的链将是群体参数的链。如果 subject 设置为整数索引，所诊断的链将是与输入索引对应的受试者的特定参数的链。此函数的输出是一个数据框，其列如下所述。直观地说，Heidelberger 诊断尝试：

• 为每个参数确定初始瞬态阶段的截断点；此后，假设样本来自稳态分布。初始瞬态阶段可作为预热期移除。每个参数的截断点列在输出数据框的 burnin 列中。

• 在假设样本中存在稳态分布的前提下，估计每个参数稳态后验分布均值的相对置信区间。相对置信区间的计算方法是将置信区间的上下限分别除以参数的均值。置信区间较大，意味着可能尚未收敛到平稳分布，或样本数量不足。相对置信区间的一半列在输出数据框的 halfwidth 列中。如果 halfwidth 小于输入目标 eps（默认为 0.1），test 列的值为 true (1)；否则为 false (0)。请注意，均值接近 0 的参数可能因除以均值而导致相对置信区间异常偏大。因此，即使这些参数的 test 值为 false (0)，也不能据此断定其未收敛。

• 通过统计检验量化样本分布的平稳程度。返回的 p 值显示在输出数据框的 pvalue 列中，可视为均值平稳性的度量。若 p 值低于输入阈值 alpha（默认值为 0.05），则表示均值不平稳，即后验样本尚未收敛到均值固定的稳态分布。

Heidelberger 诊断仅检验分布的均值。因此，与其他诊断方法类似，它只能用于检测收敛性的缺失，而不能证明收敛。换言之，即使所有数值看起来都正常，也不能据此断定链已收敛至平稳分布，或已收敛至真实后验分布。

3.10 如果链未收敛，该怎么办?

如果链似乎未能收敛，可以尝试以下方法来帮助马尔可夫链收敛：

•将目标接受率从默认值 0.8 调低。

• 重新参数化模型，以降低参数间的依赖性。

• 将某些参数值固定为已知的合理值，例如通过最大后验估计（MAP）优化获得的值。

•使用良好的参数值初始化采样。

•在疑似良好参数值附近使用更强的先验。

• 简化模型，例如使用更简单的动力学。

•尝试使用边缘 MCMC 算法 MarginalMCMC，而不是完整联合 MCMC 算法 BayesMCMC。

3.11 高级后验分布查询

3.11.1 汇总统计量

拟合贝叶斯 Pumas模型后，您可以对 拟合 或 丢弃 函数的输出调用许多函数和图表。通常，您会希望执行后验分布查询。除了使用 `summarystats` 函数（如第3.6节所述）可以获得的基本汇总统计量外，还可以基于后验分布计算更高级的统计量。

一个常见的高级后验查询是某个参数 θ 高于 0 的概率，这可以写成一个期望问题：

$$\mathrm{E}[\theta > 0 \mid \mathrm{data}]$$

实现此操作的方法是使用均值函数和便捷的 `do` 运算符。示例13展示了一个使用`do` 运算符进行后验查询的示例，用于检验参数`tvcl` 是否高于 0。它会输出一个有效的概率估计，即 $\in [0, 1]$。

```
mean(tres) do p
    p.tvcl > 0
end
```

列表 13：使用 do 运算符进行后验查询的示例

除了 均值之外，还可以使用 var 来计算方差，或使用 cov 和 cor 分别计算协方差矩阵和相关矩阵（如果 do 代码块返回多个输出）。清单 14 展示了一个示例，其中 tvcl 和 tvvc 参数之间的相关矩阵是使用后验样本估计的。

```
cor(tres) do p
    [p.tvcl, p.tvvc]
end
```

清单 14：来自多个输出的后验分布查询

请注意，参数的任何变换都可以在do 代码块中完成。例如，我们可以使用清单C 中的代码获取相关矩阵参数的下三角 Cholesky 因子的均值，见清单15。

```
mean(tres) do p
    getchol(p.C).L
end
```

列表 15：相关矩阵的均值 Cholesky 因子

有时需要比较 Pumas 和 Stan的结- 果，因为 Stan 中与 LKJCholesky 等效的功能报告的是 MCMC 样本中的下三角因子，而不是 Pumas 所报告的实际相关矩阵。

要计算受试者 i 的个体特异性参数的汇总统计量，而非群体参数，可以使用subject关键字参数，如清单16所示。

```
mean(tres, subject = i) do p
    p.ηstd
end
```

清单 16：个体特异性参数的均值

3.11.2 分位数

您可以使用分位数函数查询群体层面或个体特异性参数的后验分位数估计值：

```
quantile(tres)
```

默认情况下，这将显示所有群体水平参数的 2.5%、25%、50%、75% 和 97.5% 分位数。要显示这些参数的分位数大小规律的

相反，您可以像这样使用 subject 关键字参数来指定受试者 i 的个体特异性参数：

```
quantile(tres, subject = i)
```

若要更改计算的分位数，也可以通过 q 关键字参数手动输入所需的分位数。例如，以下代码返回受试者 i 的个体特异性参数的 10% 和 90% 分位数：

```
quantile(tres, subject = i, q = [0.1, 0.9])
```

3.11.3 可信区间

可信区间是后验分布中包含预先指定概率质量的区间。例如，95% 可信区间的估计是一个区间，其中至少有 95% 通过马尔可夫链蒙特卡罗获得的后验样本落在该区间内。最简单的方法是将某个参数的 2.5% 分位数到 97.5% 分位数之间的区间作为 95% 可信区间。可以使用分位数函数来计算，如第3.11.2节所示。不过，更精确地说，人们可能会关注包含至少 95% 后验概率质量的最短区间。这通常称为最高后验密度区间（HPDI）。若要获得每个群体参数中包含（1 –a）% 样本的 HPDI，可以使用：

```
hpd(tres, alpha = a)
```

要获取受试者 i 的个体特异性参数的相同区间，您可以使用：

```
hpd(tres, alpha = a, subject = i)
```

3.12 后验图

有许多图可以用来可视化后验分布。在本节中，我们将介绍与参数估计相关的图：密度图、山脊图和角图。

3.12.1 密度图

密度图显示了参数值直方图的平滑版本，为每个参数的边际后验分布提供了一个近似的概率密度函数。这有助于我们直观地了解每个参数的边际后验分布形状。如果运行多个马尔可夫链，该图

将显示每个密度的重叠图，并用不同颜色加以区分。你可以使用函数 密度图 绘制密度图。清单 17和图 6 分别展示了代码和生成的密度图。如果你没有通过可选关键字参数指定想要的参数参数，绘图将自动以分面形式输出多个密度图。 参数接受参数向量。

```
density_plot(pk_1cmp_tfit; parameters = [:tvcl])
```

清单 17：密度图示例

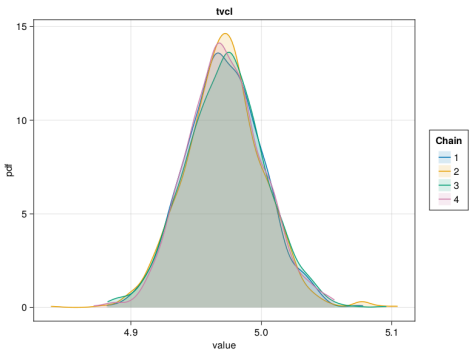

图 6：密度图示例。

3.12.2 山脊线图

另一种常见的后验图是山脊线图，它会输出一个汇总所有采样链的单一密度图，并附上与参数相关的统计信息。输出的信息包括均值、中位数、10% 和 90% 分位数，以及 95% 和 80% 的最高后验密度区间（HPDI）。你可以使用函数 `ridgeline -plot`绘制山脊线图，其语法与 密度图类似。清单 18和图 7 分别展示了代码和生成的山脊线图。

```
ridge_plot(pk_1cmp_tfit; parameters = [:tvcl])
```

清单 18：山脊线图示例

3.12.3 角图

角图是一种以组织良好的模板展示不同参数之间散点图以及边缘直方图的图表。

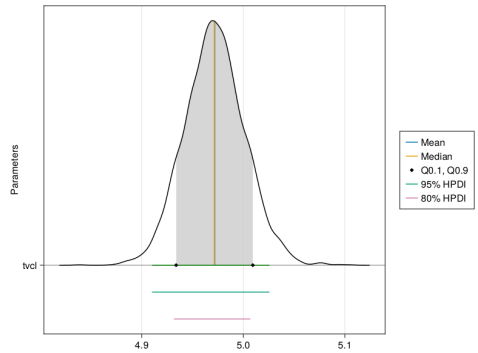

图 7：山脊线图示例。

这可用于调查参数值之间的高度相关性，而这可能是马尔可夫链蒙特卡罗采样器出现收敛问题的根源。列- 表19 展示了参数 `tvq` 和 `tvcl`的角图代码。输出见图8。

```
corner_plot(tres, parameters = [:tvq, :tvvc])
```

清单 19：角图示例

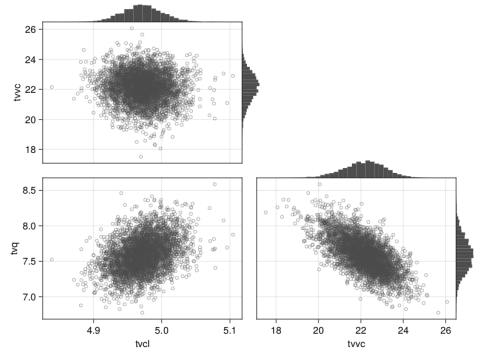

图 8：角图示例。

3.13 后验模拟与预测

3.13.1 现有受试者

你可以使用从 MCMC 结果中存储的后验分布抽取的参数值，为现有受试者模拟新的响应。这与使用从先验中抽取的参数值进行的先验预测模拟不同。可以使用 simobs 函数来实现：

```
sims = simobs(tres; samples = 100)
```

其中，tres 是 fit 或 discard 函数的输出。samples 关键字参数表示从 MCMC 样本中抽取的子样本数量。若未设置，则会使用所有 MCMC 样本。默认情况下，会模拟所有受试者。若将关键字参数 subject 设置为任意整数索引 i，则只会模拟第 i^{th} 个受试者。

在 simobs 函数中，还有一个关键字参数：simulate error。如果将其设为 true（默认值），Pumas 将从 @derived 块中的响应误差模型进行抽样（即模拟）；否则，将返回误差分布的期望值（即预测）。后者等同于使用 predict 函数。

3.13.2 新剂量或协变量

通过改变我们能够控制的某些变量来进行反事实模拟和假设分析，往往很有用。在药物计量学中，最常见的应用场景是改变剂量，但在某些情况下，也可以改变协变量。

若要更改剂量和/或协变量，并重新使用特定受试者的后验样本，您首先需要定义一个新的受试者。您可以使用`Subject` 构造函数手动定义新的受试者，也可以从数据框开始，并使用 `read pumas` 函数将其转换为 Pumas 受试者。要了解手动使用`Subject` 构造函数的方法，请参阅 Pumas 文档（docs.pumas.ai）。为展示第二种方法，假设整个群体的原始数据框为 `df`。给定数据框 `i`，如果受试者的剂量乘以 3，且剂量值位于 `df` 的第一行的`:amt` 字段中，则可以执行：

```
subj_df = copy(df[df.id .== i, :])
subj_df[1, :amt] = subj_df[1, :amt] * 3.0
new_subj = read_pumas(subj_df)[1]
```

如需了解更多有关数据框整理的信息，请参阅 Pumas 文档(docs.pumas.ai)或教程(tutorials.pumas.ai).

定义新的受试者 new subj 后，你可以调用 simobs 的以下方法：

```
simobs(tres, new_subj, subject = i, samples =
    100)
```

将 subject 关键字参数设置为索引 i，即可使用 MCMC 样本来获取受试者 i 的参数，同时使用新受试者的剂量、协变量和时间点。请注意，索引 i 指的是传递给 fit 函数的训练人群中受试者的索引，该索引可能与受试者的 ID 字段不一致。

若要针对一个真正的新受试者 new - subj 进行模拟，其中受试者特定参数从先验分布中抽样，群体参数从后验分布中抽样，可以省略 subject 关键字参数：

```
simobs(tres, new_subj, samples = 100)
```

3.14 视觉预测检验和模拟图表

从先验分布中抽取参数值进行模拟的视觉预测检验（VPC）通常称为先验预测检验。同样地，从后验分布中抽取参数值进行模拟的视觉预测检验（VPC）通常称为后验预测检验。

调用先验或后验分布`simobs` 后，所得模拟对象`sims` 可传递给 `vpc` 函数，以计算 VPC 图所需的所有分位数。该函数还接受一个分类变量，并通过关键字参数 按.....分层来对 VPC 结果进行分层。VPC 结果可使用 VPC 图函数绘制。代码清单 20 展示了如何从模拟中运行 VPC 的代码。 -

```
vpc_res = vpc(sims)
vpc_plot(vpc_res)
```

代码清单 20：视觉预测检验

图 9 是此类代码输出的一个示例。

要绘制模拟分位数中位线，您可以将关键字参数 simquantile - medians = true 设置为 true，例如:-

```
vpc_res = vpc(sims)
vpc_plot(vpc_res, simquantile_medians = true)
```

所得图形将如图10所示。

要进一步在 VPC 图上将观测值显示为点，可以将关键字参数 observations = true 设置为，例如：

```
vpc_res = vpc(sims)
vpc_plot(vpc_res, simquantile_medians = true,
    observations = true)
```

所得图形将如图11。

有关可用的众多 VPC 选项（包括更改协变量或分层）的详细信息，

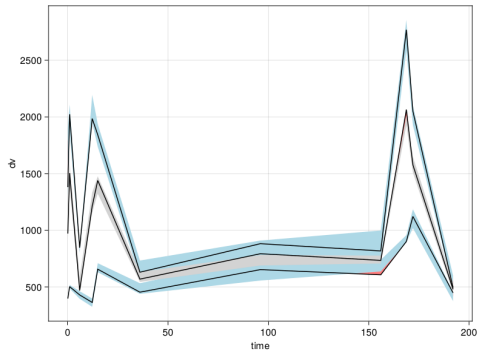

图 9：后验视觉预测检验示例。

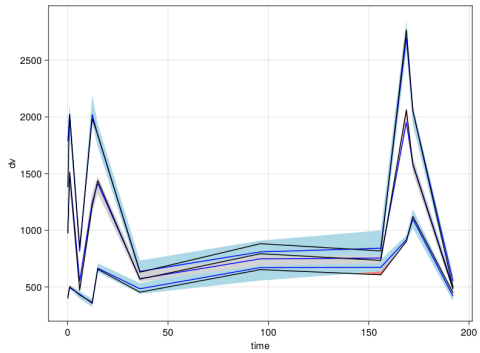

图 10：带有模拟分位数中位数的后验视觉预测检验示例。

您可以参考 Pumas 文档（docs.pumas.ai）。

除了完整的 VPC 图之外，也可以使用 sim plot 函数，仅绘制模拟响应和观测值，而不显示 VPC 色带，例如：

```
sim_plot(sims)
```

输出示例如12所示。

3.15 模拟查询

调用 simobs 的输出会存储模拟得到的观测值，以及计算出的所有中间值，例如所用的参数值、个体系数、剂量控制参数、协变量、微分方程的解等。你可以对模拟输出进行多种后处理操作，以便根据模拟计算各种查询结果和汇总统计量。

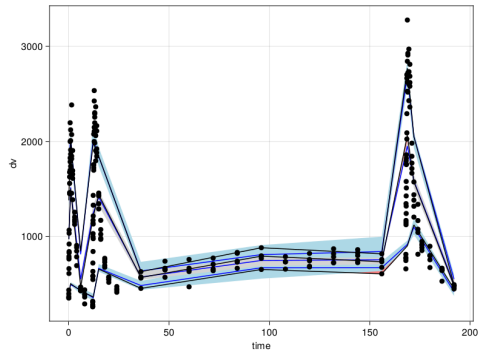

图 11：后验预测检验示例，展示模拟分位数中位数和观测值。

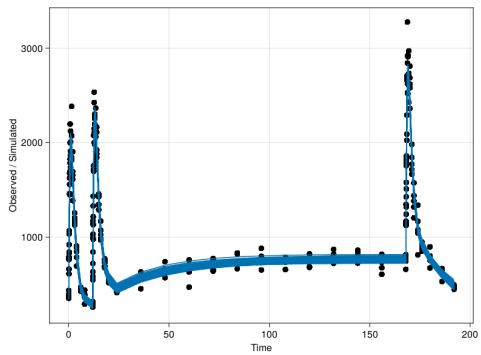

图 12：简单模拟图示例。

该后处理函数是一种强大的工具，可让你使用模拟结果进行各种查询。使用后处理函数的方式有多种。第一种方式是使用后处理函数，将模拟结果中存储的所有信息提取为由命名元组组成的向量。每个命名元组都包含模拟一次运行时计算的所有中间值。设sims 为任意一次simobs操作的输出。清单21 展示了如何提取模拟的所有中间结果。

```
generated = postprocess(sims)
```

代码清单 21：提取中间值

generated 是所有模拟运行输出组成的向量。因此，generated[i]存储第 i^{th} 次模拟运行的输出。时间依赖变量以向量形式存储，每个存在观测的时间点对应一个元素。或者，如果调用 simobs 时设置了关键字参数

obstimes，则时间依赖变量将在 obstimes 中的时间点进行评估。

使用 postprocess 的第二种方式是传入一个后处理函数。后处理函数可用于：

- 变换模拟量，或
- 将模拟结果与观测值进行比较。

我们使用 `do` 语法，它是将函数作为第一个参数传递给 `postprocess` 的简写形式。用于传递后处理函数的“do”语法如列表 22所示。

```
postprocess(sims) do gen  obs  end

```

清单 22：比较生成量与观测值

其中 `gen` 是一次模拟运行中所有生成量组成的命名元组，而 `obs` 是观测值组成的命名元组。例如，要查询模拟观测值中 `conc` 高于观测量 `conc` 且位于观测值时间点的那些值，可以使用列表 23中的代码。这有时称为贝叶斯 p 值，预计其约为 0.5。

```
postprocess(sims) do gen  obs                  ,3 end

```

代码清单 23：每次模拟的贝叶斯 p 值

gen.conc 是模拟得到的 conc向量，其长度与观测时间点的数量相同。obs.conc 是观测向量 conc。gen.conc.> obs.conc返回一个true/false向量，每个时间点对应一个元素。该向量的总和给出了模拟值高于观测值的时间点数量。除以时间点数量即可得到该比例。以这种方式使用 postprocess时，输出始终是查询结果的向量，每个模拟对应一个数值。在查询函数体中，可以选择仅使用 gen 或仅使用 obs，但标头必须始终同时包含 gen 和 obs。

使用 postprocess 函数的第三种方式，是计算模拟量或模拟量函数的汇总统计量。可以将统计函数作为 stat 关键字参数传入，以计算汇总统计量参数。例如，要估计模拟值高于观测值的概率，可以使用代码清单 24。

```
postprocess(sims, stat = mean) do gen, obs
  gen.conc .> obs.conc
end
```

列表 24：贝叶斯 p 值均值

此函数将执行两项操作：

1. 将所有模拟运行中的查询结果（例如 gen.conc .> obs.conc）串联成一个向量。2. 计算合并后向量的均值。

或者，你可以使用 均值 函数来执行相同的操作，而无需使用关键字参数。列表 25将在内部调用 后处理函数。

```
postprocess(sims) do gen  obs  end

```

清单 25：使用均值函数计算贝叶斯 p 值均值

此操作的结果将是一个标量，等于查询<一个接一个连接起来的>向量的均值。

为了获得模拟量高于观测在每个时间点的概率，可以调用均值函数，如列表26所示。

```

```

列表 26：使用均值函数计算的均值贝叶斯 p 值

这将返回一个概率向量，其长度与时间点的数量相同，而不会将所有查询连接在一起。

要计算所有生成量的汇总统计量，也可以使用清单 27中的代码。

```
postprocess(sims, stat = mean)
```

清单 27：生成量均值

无需指定后处理函数，这也等价于清单 28中的简短版本。

```
mean(sims)
```

清单 28：均值生成量——简短版

除了均值，你还可以用同样的方式使用以下汇总统计函数：

- std 表示逐元素标准差
- var 表示逐元素方差
- cor 表示多个量之间的相关性
- cov 表示多个量之间的协方差

这些函数可以作为 stat 关键字参数传递给 postprocess，也可以使用简写形式，例如：

```
generated = postprocess(
    sims, stat = std,
) do gen, obs
  ...
end
std(sims) do gen, obs
  ...
end
std(sims)

generated = postprocess(
    sims, stat = var,
) do gen, obs
  ...
end
var(sims) do gen, obs
  ...
end
var(sims)
```

cor 和 cov 统计量的独特之处在于，它们需要一个输出向量的后处理函数。例如，要估计一个房室模型中 CL 参数和 Vc 参数之间的相关性，可以使用以下任意一种方式：

```
postprocess(s, stat = cor) do gen  obs,


```

注意，gen.CL是所有时间点的模拟 CL 值向量。但由于该值在整个时间内保持不变，我们只需使用第一个元素 gen.CL[1] only。可以使用cov代替cor 来计算协方差矩阵。此操作的输出可以是相关性矩阵或协方差矩阵。

3.16 非房室分析（NCA）参数

您可以轻松地在Pumas模型中使用@observed块，将非房室分析（NCA）参数（如曲线下面积（auc）和最大浓度（cmax））集成到模拟中，例如：

```
@derived begin
    cp = @. Central / Vc
    conc ~ @. LogNormal(log(cp), σ)
end
@observed begin
    nca := @nca cp
    auc = NCA.auc(nca)
    cmax = NCA.cmax(nca)
end
```

现在，曲线下面积和最大浓度将存储在 simobs 的输- 出中，并可通过各种汇总统计量查询进行查询。例如，下面的代码将根据模拟 sims（ simobs的输出）估计曲线下面积大于 20 且最大浓度小于 30 的概率。

```


```

如需了解有关 NCA 集成和参数的更多信息，请参阅 Pumas 文档（docs.pumas.ai）。

3.17 交叉验证与期望对数预测密度

交叉验证是一种用于评估模型在未见数据（也称为样本外数据）上的预测准确性的技术。具体做法是在进行贝叶斯推断时，有系统地从训练数据中留出一部分数据[8]，然后使用留出的数据评估 MCMC 样本的平均预测准确性。交叉验证程序的每次迭代都会从训练数据中留出不同的子集，并使用其中一部分或全部数据来评估预测准确性。用于评估预测准确性的指标通常是给定未见数据时每个 MCMC 样本的（条件或边际）对数似然。随后，对交叉验证程序各次迭代的预测性能指标取平均值。

使用 Pumas，几乎可以对分层模型进行所有类型的交叉验证。在 Pumas 中执行交叉验证的主要函数是 crossvalidate 函数。cr ossvalidate 有 2 个输入：

1.来自拟合或丢弃的 MCMC 结果。我们将其称为 tres。2. 交叉验证算法。我们将其称为 cv - method。

[8]贝叶斯推断期间用于从后验分布中获取样本的数据

调用语法是 cvres = 交叉验证(res, cvmethod)。要估计给定 cv res 时的期望对数预测密度（期望对数预测密度），可以使用 elpd 函数：elpd(cv res)。

Pumas 中提供两类交叉验证算法：

1. 基于重采样的交叉验证：对每个留出的数据点或子集重新执行马尔可夫链蒙特卡罗。该算法使用 ExactCrossvalidation 结构体构建。

2. 基于 PSIS 的交叉验证（Vehtari 等人，2015）通过重要性抽样和平滑体重的方法，避免重采样的需要。该算法使用`PSISCrossvalidation` 结构体构建。

基于重采样的交叉验证方法的构造函数为：

```
cv_method = ExactCrossvalidation(; split_method, sp
```

其中，split method 和 split by 是必须设置的关键字参数，而 ensemblealg 默认使用多线程。

这定义了一个用于交叉验证的 ExactCrossvalidation 算法实例。使用该算法时，每次都会剔除一部分数据并重新进行拟合。

数据在训练集和验证集之间的拆分方式由关键字参数 split method 和 split by 决定。split method 参数可以是以下任意一种类型：

1. LeaveK：留 K 个交叉验证；2. KFold：K 折交叉验证；3. LeaveFutureK：每次留出 K 个未来观测点

关键字参数 split by 可以是以下任意一种类型：

1. BySubject：留出受试者；2. ByObservation：为每个受试者留出单个观测

下一小节将讨论每种数据划分方法。

与基于重抽样的交叉验证类似，PSIS-CV 方法的构造函数如下：

```
cv_method = PSISCrossvalidation(; split_method,
    split_by, pareto_shape_threshold = 0.7)
```

其中，拆分方法和拆分依据是必须设置的关键字参数。

这定义了`PSISCrossvalidation` 交叉验证算法的一个实例。 拆分方法 和 拆分依据 关键字参数与基于重抽样的交叉验证情形中的参数类似。帕累托分布形状阈值 = `0.7`关键字参数会导致移除所有在计算期望对数预测密度（ELPD）估计时，其帕累托分布形状参数超过 0.7 的交叉验证运行结果。这有助于避免少数糟糕的运行结果使整个 PSIS-CV 方法失效。理想情况下，对于形状参数超过阈值的交叉验证运行子集，应重新进行推断；但截至本文撰写时，Pumas 尚未实现此功能。请关注 Pumas 文档（docs.pumas.ai），了解最新功能的更新信息。

3.17.1 留 K 法

留 K 法数据划分算法的构造函数如下：

```
split_method = LeaveK(; K = 5, shuffle = false,
    rng = nothing)
```

在此算法中，数据会被多次划分为两个互不重叠的组，每次划分都从完整数据集开始。这两组通常称为训练子集和验证子集，其中验证子集包含 K 个数据点。下一轮迭代中，整个数据集会重新划分，并使用另一个包含 K 个数据点的互不重叠子集作为验证子集。重复此过程，直到几乎每个数据点都恰好出现在一个验证子集中。

数据通常是某种向量，例如观测值或受试者，而划分结果取决于顺序。执行划分之前，可以将 shuffle 关键字参数设为 true（默认为 false），以随机打乱数据向量，从而消除对数据原始顺序的敏感性。还可以传入可选的伪随机数生成器 rng，以控制伪随机性，确保结果可复现。

假设一些虚拟的原始数据 [“A”、“B”、“C”、“D”] 其类似于受试者或观测值。不进行打乱的留一法划分会产生表3中所示的数据划分。

其中每个数据点在验证子集中恰好出现一次。不进行随机打乱的留二法划分会产生表 4所示的数据划分。

训练子集	验证子集
["A", "B", "C"]	["D"]
["A", "B", "D"]	["C"]
["A", "C", "D"]	["B"]
["B", "C", "D"]	["A"]

表3：留一法划分。

训练子集	验证子集
["A", "B"]	["C", "D"]
["C", "D"]	["A", "B"]

表 4：留二法划分。

3.17.2 K 折交叉验证

K 折数据划分算法的构造函数为：

```
split_method = KFold(; K = 5, shuffle = false,
    rng = nothing)
```

在此算法中，数据被分成 2 个互不相交的组，共进行 K 次，每次都从完整数据集开始。这 2 个组通常称为训练子集和验证子集，其中验证子集包含 `floor(N / K)`[9] 个数据点，`N` 为数据点总数。在下一次迭代中，使用另一个包含 `floor(N / K)` 个不同数据点的互不相交的验证子集，对整个数据集重新进行划分，该验证子集与之前的验证子集互不相交。此过程会迭代进行，直到几乎每个数据点都出现在 1 个且仅 1 个验证子集中。如果 `N` 可被 `K`整除，则每个数据点都会出现在 1 个且仅 1 个验证子集中。否则，剩余数据点将在所有划分中都属于训练子集，并且不会出现在任何验证子集中。

数据通常是某种向量，例如观测值或受试者，并且拆分取决于顺序。在执行拆分之前，可以通过将 shuffle 关键字参数设置为 true（默认为 false）来随机打乱数据向量，从而消除对数据原始顺序的敏感性。你还可以传入一个可选的伪随机数生成器 rng，以控制伪随机性，确保结果可复现。

假设一些虚拟的原始数据 [“A”、“B”、“C”、“D”] 这类似于受试者或观测值。不进行打乱的四折划分会得到表5中所示的数据划分，

[9] `floor`是一个向下取整为整数的函数。

训练子集	验证子集
["A", "B", "C"]	["D"]
["A", "B", "D"]	["C"]
["A", "C", "D"]	["B"]
["B", "C", "D"]	["A"]

表5：4折划分。

其中每个数据点在验证子集中恰好出现一次。未进行洗牌的 2 折划分会产生表 6所示的数据划分。

训练子集	验证子集
["A", "B"]	["C", "D"]
["C", "D"]	["A", "B"]

表 6：2 折拆分。

3.17.3 留出未来 K 个数据点

leave-future-K 数据划分算法的构造函数是：

```
split_method = LeaveFutureK(; K = 1, minimum =
    2)
```

在该算法中，假设数据是一个时间序列。目标是将数据拆分为“过去”和“未来”。使用该算法时，每次都从完整数据集开始，将数据多次拆分为3 个互不相交的组，其中第三组被丢弃。前2 个组通常称为过去/训练子集和未来/验证子集，其中未来验证子集包含K个未来数据点。在下一次迭代中，使用另一个包含 K 个数据点的互不相交子集作为未来验证子集，然后对整个数据集重新进行拆分。该过程从完整数据集开始迭代执行，并沿时间向后移动，直到训练子集剩余的数据点少于预设的最小数量。使用此方法，每个数据点最多可以出现在1 个未来验证子集中。K 和最小数量的默认值分别为 1 和 2。

假设原始数据为 `["A", "B", "C","D", "E", "F"]`。采用最小 `= 2` 的留一未来点法进行拆分后，数据拆分结果如表 7 所示，剩余点将被丢弃：

留出未来 2 个受试者的划分最少 `= 2`会得到表8所示的数据划分结果。

过去子集	未来子集
["A", "B", "C", "D", "E"]	["F"]
["A", "B", "C", "D"]	["E"]
["A", "B", "C"]	["D"]
["A", "B"]	["C"]

表 7：留一未来划分。

过去子集	未来子集
["A", "B", "C", "D"]	["E", "F"]
["A", "B"]	["C", "D"]

表 8：留出未来两项划分。

3.17.4 基于受试者的划分

基于受试者的划分方法的构造函数为：

```
split_by = BySubject(; marginal = LaplaceI())
```

使用此方法时，每位受试者都被视为一个数据点。为每位受试者计算的预测对数似然可以是边际对数似然，也可以是条件对数似然。此方法有一个关键字参数 marginal。如果 marginal 设为 nothing，则使用参数的典型值计算每位受试者的预测对数似然，即条件对数似然。否则，将 marginal 用作边际化算法，计算每位受试者的预测对数似然，即边际对数似然。marginal 的默认值为 LaplaceI()，它使用拉普拉斯方法对受试者特异性参数进行积分消除。其他选项包括：FOCE() 和 FO()。

3.17.5 基于观测的划分

基于观测的划分方法的构造函数如下：

```
split_by = ByObservation(; allsubjects = true)
```

使用此方法时，每个观测值或一组观测值都被视为一个数据点。使用此方法计算预测对数似然时，所计算的是一个或多个受试者的一个或多个观测值的条件对数似然。

此方法有一个关键字参数 allsubjects。如果 allsubjects 设为 true（默认值），则每位受试者的 i^{th} 观测值都会归为一个数据点。这假设所有受试者的观测值数量相同。如果 allsubjects 设为 false，则每位受试者的每个观测值都是独立的数据点。

假设有 2 名受试者，每名受试者有 3 个观测值。使用 拆分方法 =LeaveK(K = 1)作为拆分方法，并结合 拆分依据 = 按观测值(所有受试者=false)，训练和验证拆分结果如表 9 所示。

训练子集	验证子集
Subj 1 (obs 1, 2, 3), subj 2 (obs 1, 2)	Subj 2 (obs 3)
Subj 1 (obs 1, 2, 3), subj 2 (obs 1, 3)	Subj 2 (obs 2)
Subj 1 (obs 1, 2, 3), subj 2 (obs 2, 3)	Subj 2 (obs 1)
Subj 1 (obs 1, 2), subj 2 (obs 1, 2, 3)	Subj 1 (obs 3)
Subj 1 (obs 1, 3), subj 2 (obs 1, 2, 3)	Subj 1 (obs 2)
Subj 1 (obs 2, 3), subj 2 (obs 1, 2, 3)	Subj 1 (obs 1)

表 9：使用拆分方法 = LeaveK（K = 1）且拆分依据为= 按观测值（所有受试者 = false）的训练集和验证集拆分。

另一方面，如果 所有受试者 设为true，训练集和验证集的划分如表 10所示。

训练子集	验证子集
Subj 1 (obs 1, 2), subj 2 (obs 1, 2)	Subj 1 (obs 3), subj 2 (obs 3)
Subj 1 (obs 1, 3), subj 2 (obs 1, 3)	Subj 1 (obs 2), subj 2 (obs 2)
Subj 1 (obs 2, 3), subj 2 (obs 2, 3)	Subj 1 (obs 1), subj 2 (obs 1)

表 10：使用拆分方法 = LeaveK(K = 1) 并按= 按观测值(allsubjects = true)进行训练和验证拆分。

3.17.6 示例

假设有 5 名受试者，且每名受试者有 10 个观测值，并且 res 是拟合或丢弃函数的结果。以下是上述输入可以使用的一些组合：

- 留一观测值交叉验证，每次为所有受试者留出1个观测值。

所有受试者 = true 表示，对所有受试者移除相同的观测索引，例如，在第一次运行中，所有受试者的第 10个观测用于验证；然后在第二次运行中，使用第 9 个观测进行验证，依此类推。

```
split_method = LeaveK(K = 1)

sp
 e

```

•留出两个未来观测值交叉验证：每位受试者留出 2 个未来观测值，使训练中使用的观测值不少于 4 个。allsubjects = false 表示每次只拆分一位受试者的数据。因此，在第一次运行中，使用受试者 1 的第 1 至 8 个观测值进行训练，并使用受试者 1 的第 9 和第 10 个观测值进行验证。在第二次运行中，使用受试者 1 的第 1 至 6 个观测值进行训练，并使用第 7 和第 8 个观测值进行验证。在第三次运行中，使用受试者 1 的第 1 至 4 个观测值进行训练，并使用第 5 和第 6 个观测值进行验证。在这 3 次运行中，受试者 2 至 5 的所有观测值均用于训练。接着，在第四至第六次运行中拆分受试者 2 的数据。在第七至第九次运行中拆分受试者 3 的数据，以此类推。

```
split_method = LeaveFutureK(K = 2, minimum =4)

cv_method = ExactCrossvalidation(;
sp
 e

```

•留一受试者交叉验证，在计算预测似然时使用FOCE对个体特异性参数进行边际化。使用此方法时，在第一次运行中，受试者1、2、3、4用于训练，受试者5用于验证。在第二次运行中，受试者 1 、 2 、 3 、 5 用于训练，受试者 4 用于验证，依此类推。计算得到的预测似然是被留出的受试者的边际似然，使用 FOCE 对个体特异性参数进行边际化计算。也可以使用 LaplaceI 或 FO。

```
split_method = LeaveK(K = 1)
split_by = BySubject(marginal = FOCE())
```

```
cv_method = ExactCrossvalidation(;
sp
 e

```

3.18 信息准则

从交叉验证输出计算得出的 ELPD 模型评估指标，理论上类似于所谓的广义适用信息准则（WAIC）（Vehtari 等人，2015）。更准确地说，ELPD 的 -2 倍与 WAIC 相当。ELPD 越高越好，WAIC 越低越好。使用 PSIS 留一法（LOO）交叉验证方法估计 ELPD 时，ELPD 估计值的 -2 倍有时称为 LOOIC。除了 ELPD 估计值，Pumas 还支持常见的信息准则（Burnham 和 Anderson，2002），例如赤池信息准则（AIC）、校正赤池信息准则（AICc）、贝叶斯信息准则（BIC）和 WAIC。要估计这些准则，我们首先需要根据某种“逐点”定义计算逐点对数似然。为此，你可以调用：

```
pl = loglikelihood(tres; split_method, split_by)
```

其中 tres 是 拟合 或 丢弃 的输出，拆分方法 和 拆分依据 是第 3.17 节中解释的关键字参数，它们定义了什么是一个点，以及要计算哪种对数似然。然后，您可以使用以下任一函数，利用逐点对数似然计算信息准则：

```
Pumas.aic(pl)
Pumas.aicc(pl)
Pumas.bic(pl)
Pumas.waic(pl)
```

4 背景与直觉

本节将介绍药物计量学中贝叶斯推断所用的符号和数学背景。我们将简要介绍贝叶斯统计及其在药物计量学中的用途，随后直观地解释马尔可夫链蒙特卡罗（MCMC）和无调采样器（NUTS）算法（Hoffman 和 Gelman，2014；Betancourt，2017）。然后，我们将探讨先验选择、MCMC 收敛诊断、交叉验证和模型选择背后的直觉及部分数学原理。本节旨在帮助读者做好准备

使用 Pumas 中的贝叶斯工作流程或任何标准的贝叶斯分析工具，并通过直觉和浅显的数学知识，让他们对高度专业的概念具备实用的了解。

4.1 符号说明

为方便表述，在本文其余部分中我们使用：

1. θ 表示所有群体层面参数，包括全部的（θ, Ω σ）。2. η 表示所有受试者的患者特异性参数，其中 η_i 表示受试者 i的受试者特异性参数。3. x 表示所有受试者的协变量，其中 x_i 表示受试者 i 的受试者特异性协变量。更严谨地说， x_i还包括对受试者 i进行观测的所有时间点。4. y 表示所有受试者的观测响应，其中 y_i 表示受试者 i的受试者特异性响应。5. $p(A = \alpha \mid B = \beta)$表示随机变量 A 在变量 B 取值为 β的条件下取值为 α 时的概率密度/质量。 $B = \beta$ 通常可替换为多个变量，例如 $p(A = \alpha \mid B = \beta, C = c)$。如果 A 是连续随机变量，$p(A = \alpha \mid B = \beta)$表示 $A = \alpha$ 在给定 $B = \beta$条件下的概率密度。如果 A是离散随机变量， $p(A = \alpha \mid B = \beta)$表示 $A = \alpha$在给定 $B = \beta$条件下的概率质量。省略 α 和/或 β 时，分别可用取值 A和/或 B 替代，例如 $p(A = A \mid B = B)$，具体含义可根据上下文理解。这是对符号的轻微滥用：使用相同符号 A/B 同时表示随机变量及其支撑集中的特定取值，不过这在概率论中很常见。6. $p(A, B \mid C)$表示 $p(A \mid B, C) \times p(B \mid C)$，其等于 $p(B \mid A, C) \times p(A \mid C)$；根据 A和 B的支撑集，这可能是概率密度和/或概率质量的乘积。7. D 表示所有观测数据，包括 x 和 y。8. $p(y_i \mid x_i, \eta_i \theta)$表示给定受试者 i的观测值（x_i, y_i）时（η_i, θ）的条件似然。请记住，似然是给定数据后关于参数的函数，同时它也是给定模型参数时观测到数据的概率。9. $p(y \mid x, \eta, \theta)$表示给定所有受试者的观测值（$x, y$）时（$\eta, \theta$）的条件似然。鉴于药物计量学模型的层级结构，这等于 $\prod_i p(y_i \mid x_i, \eta_i, \theta)$。10. $p(y_i \mid x_i, \theta)$表示给定受试者 i的观测值（x_i, y_i），对 η_i 进行边际化后 θ 的边际似然。它等于 $\int p(y_i \mid x_i, \eta_i \theta) \cdot p(\eta_i \mid \theta) d\eta_i$。11. $p(y \mid x, \theta)$表示给定所有受试者的观测值（x, y）时 θ 的边际似然。鉴于药物计量学模型的层级结构，这等于 $\prod_i p(y_i \mid x_i, \theta)$。

需要牢记以下一些附加假设：

1. y_i 可能不是标量，而可能（且通常确实）是个体特异性的时间序列响应，或多个此类时间序列响应。2. η_i 通常不是标量，它可以由多个个体特异性参数构成，并为每个参数指定不同的先验分布。3. x_i 通常不是标量，它可以是多个与时间无关的值，也可以是与时间无关的值和某些时间序列的组合。它还包括观测到相应 y_i 的所有时间点。4. $p(A \mid B)$将在公式中用于表示概率密度/质量函数，但在文本中，根据语境，也可能用于指代分布本身这一对象/概念。

图 13展示了在群体中有 3 名受试者时，使用上述符号表示的药物计量学中典型的模型结构。此外，图14 展示了一个示例 Pumas 模型，并标出了图 13 中各变量的定义位置。

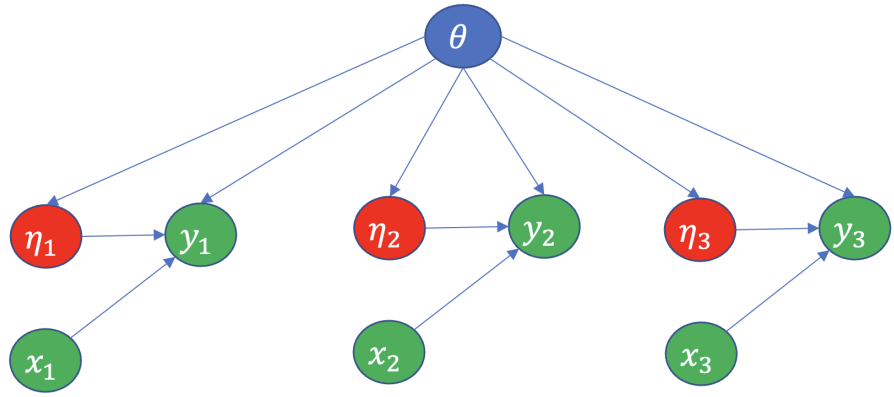

图 13：药物计量学中通常使用的模型层级结构示意图，假设群体中只有 3 名受试者。该示意图可以轻松扩展到更多受试者。有关符号的含义，请参阅符号说明部分（4.1）。

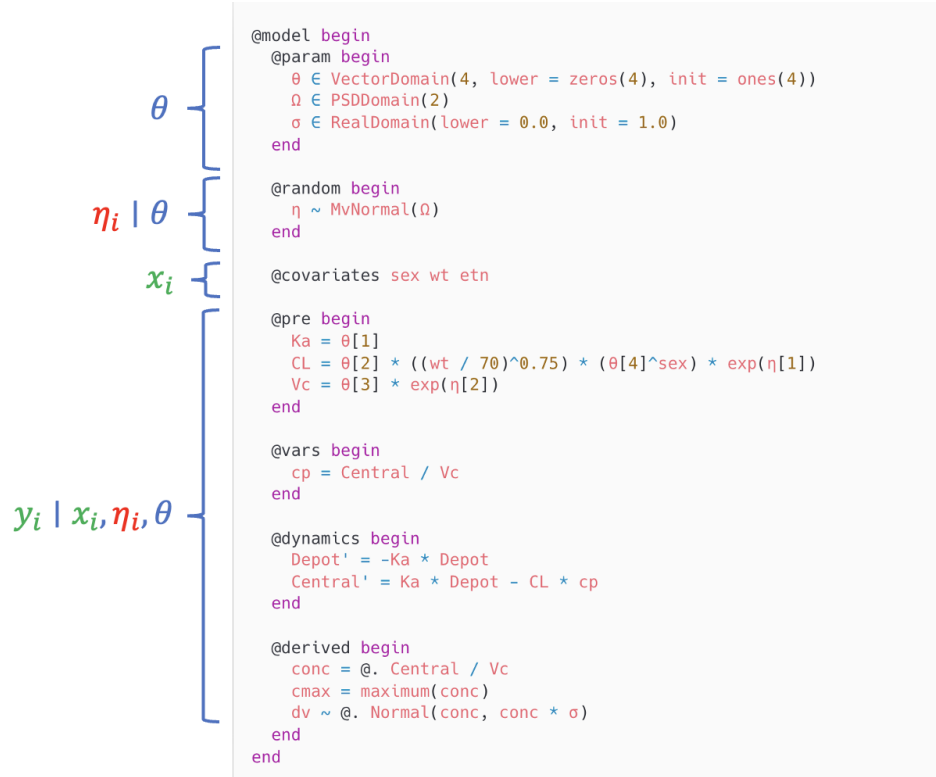

图 14：一个虚拟的 Pumas 模型，展示了图中每个变量的定义位置13 。

4.2 贝叶斯统计

贝叶斯统计是将**贝叶斯定理**作为估计感兴趣的参数或未观测数据的程序（Gelman 等人，2013a）。贝叶斯定理以托马斯·贝叶斯命名，[10]，它告诉我们如何将条件概率从“反转”p（$B \mid A,\ C$）转换为 p（$A \mid B,\ C$），其中 C 是可选的：

$$p(A \mid B, C) = \frac{p(A, C) \cdot p(B \mid A, C)}{p(B, C)} \tag{3}$$

在统计学语境中，贝叶斯法则可用于根据观测值计算每个假设为真的概率。假设我们有 10 个假设 H_1，……，H_{10} 其中每个假设都有一个先验概率 p（真实情况 $= H_i$）。我们可以使用贝叶斯法则，利用以下公式计算每个假设 H_i 的后验概率p（真实情况 $= H_i \mid$ 数据）：

$$p(\text{truth} = H_i \mid \text{data}) = \frac{p(\text{data} \mid \text{truth} = H_i) \cdot p(\text{truth} = H_i)}{p(\text{data})} \tag{4}$$

其中，分母可写为

$$p(\text{data}) = \sum_{i=1}^{10} p(\text{data} \mid \text{truth} = H_i) \cdot p(\text{truth} = H_i)$$

它是所有假设的似然之和，p（数据 $\mid$ 真值 $= H_i$），并按各自的先验概率 p（真值 $= H_i$）加权。虽然分母具有深刻的统计学意义，但也可以从实用角度将其视为一个归一化常数，其选取方式使得$\sum_{i=1}^{10} p$（真值 $= H_i \mid$ 数据）$= 1$。由于分母是所有 i的分子项之和，由此得到的后验分布概率之和必然为 1。

p（真值 $= H_i$）描述了通常所说的假设先验概率。这可以编码建模者的领域知识，在观测任何数据之前，预先为不同的假设赋予不等的概率。或者，也可以为每个假设赋予相等的概率。给定多个假设和一些数据，给定观测数据 p（真值 $= H_i \mid$ 数据）时概率最高的假设最为可信。假设通常由一个模型以及该模型参数的参数值构成。

在药物计量学语境中，将给定特定模型的每组参数值（η、θ）视为一个假设。在此情况下，我们面对的是假设连续体，而不是离散的假设集合。假设单一模型，并通过将其置于 $\mid$的右侧对其进行条件化，使用药物计量学符号，假设连续体的贝叶斯定理可写为：

$$p(\eta, \theta \mid x, \text{model}, y) = \frac{p(y \mid x, \text{model}, \eta, \theta) \cdot p(\eta, \theta \mid x, \text{model})}{p(y \mid x, \text{model})} \tag{5}$$

其中，一般形式中的 A 被替换为 (η, θ)，B 是 y，而 C 是 $(x,$ 模型 $)$[11]。请注意，在上述方程中，我们处处以 x 为条件，因为我们通常并不关心对 x 本身的概率进行建模，而是关心给定 x $(y \mid x)$ 时 y 的概率。

另请注意，$p(\eta, \theta \mid x,$模型$)$，即先验概率，可以替换为 $p(\eta, \theta \mid$ 模型$)$，因为在药物计量学中，先验通常不依赖于协变量。在 Pumas 语法中，这意味着 @covariates 块中的协变量不会用于 @param 或 @random 块中的先验设定。

当只考虑单个模型时，因此我们可以将其从方程中省略，并且先验独立于协变量，贝叶斯定理可简化为：

[10] **托马斯·贝叶斯**（1701—1761）是一位统计学家、哲学家和长老会牧师，因提出以其名字命名的定理的一个特定情形而闻名：贝叶斯定理。贝叶斯从未发表过后来成为其最著名成就的成果；他的笔记由其朋友**理查德·普莱斯**在其去世后整理并出版。该定理的正式名称是**贝叶斯–普莱斯–拉普拉斯**，因为**贝叶斯**最先发现了它，**普莱斯**获得了他的笔记，将其转写为数学记号并宣读给伦敦皇家学会，而**Laplace**则在18世纪末的法国独立地重新发现了该定理，此前彼此并无联系；当时他在拿破仑时代利用概率对人口普查数据进行统计推断。

[11] 当模型选择与之相关时，我们也可以将模型作为 A 的一部分，而不是 C，但为简洁起见，不考虑这种情况。

$$p(\eta, \theta \mid x, y) = \frac{p(y \mid x, \eta, \theta) \cdot p(\eta, \theta)}{p(y \mid x)} \tag{6}$$

为进一步简化记号，我们用 $x, y)$ D, $p(y \mid x$ 表示 $p(D)$，并用 $p(y \mid x, \eta, \theta)$ 表示 $p(D \mid \eta, \theta)$ [12]。这样就得到了更为标准形式的贝叶斯定理：

$$p(\eta, \theta \mid D) = \frac{p(D \mid \eta, \theta) \cdot p(\eta, \theta)}{p(D)} \tag{7}$$

药代动力学模型通常描述某种数据生成过程。给定：1）一组协变量 y，2）一个特定模型，以及 3）一组参数（η, θ），我们便可以据此模拟合成数据。此类模型描述响应变量的概率分布 y, $p(D \mid \eta$, $\theta)$。因此，计算 $p(D \mid \eta$, $\theta)$ 在计算上非常直接。类似地，先验概率$p(\eta$, $\theta)$ 通常用标准概率分布来定义，这些分布的概率密度函数或概率质量函数已知且易于计算。在实际应用贝叶斯定理时，主要问题在于计算分母项 $p(D) = p($ $y \mid x)$。当所有参数都是连续变量时，这可以写成一个积分：

$$p(D) = \int\int p(D \mid \eta, \theta) \cdot p(\eta, \theta)\, d\eta d\theta \tag{8}$$

这称为数据的边际概率（也称为证据或归一化常数），它是条件概率 $p(D \mid \eta$, $\theta)$ 在所有可能的 $(\eta$, $\theta)$ 取值下的加权平均，权重为其先验概率。 $p(D \mid \eta$, $\theta)$ 通常称为条件似然，而 $p(\eta$, $\theta \mid D)$ 则称为观察到 D 后 $(\eta$, $\theta)$ 的后验概率。为了更好地理解 $p(D)$，不妨重新加上对模型的条件，并将 $p(D \mid$ 模型) 视为该模型的边际似然[13]

[12]这在符号使用上略有不严谨，因为尽管 D 包含位于等号右侧的 x，我们仍将它放在等号左侧。[13]请注意，在一般统计学中，$p(D \mid \theta)$ 表示给定 θ 时 D 的概率，以及给定 D 时 $likelihood$ 的 θ。

对模型中的所有总体参数和个体特异性参数积分消去后所得的结果。

在一般情况下，计算关于 $(\eta$, $\theta)$ 的高维积分是不可行的。但我们为什么需要后验概率呢？我们往往更关注进行预测，并在预测时使用后验概率为所有可能的假设赋予权重。假设 $\hat{}y$ 为以下之一：

1.给定新受试者的已知协变量 $\hat{}x$时该受试者的未知响应，或 2. 现有受试者的未知部分响应（例如在未来时间点），该受试者具有属于 D and 的先前观测响应以及一些已知协变量 $\hat{}x$。

协变量 $\hat{}x$ 包括经典的药代动力学协变量，例如年龄和体重，但也包括响应$\hat{}y$ （如果它是时间序列）所定义的时间点。可以将 $\hat{}y$ 的平均预测（使用后验分布概率作为权重）写为：

$$E[y \mid x = \hat{x}, D] = \int \hat{y} \times p(y = \hat{y} \mid x = \hat{x}, D)\, d\hat{y} \tag{9}$$

where $p(y = \hat{y} \mid x = \hat{x}, D)$ is defined as:

$$p(y = \hat{y} \mid x = \hat{x}, D) = \int\int p(y = \hat{y} \mid x = \hat{x}, \eta, \theta) \times p(\eta, \theta \mid D)\, d\eta d\theta \tag{10}$$

其中， $p(\eta$, $\theta \mid D)$[14] 是后验概率，而 D 指此前看到的所有数据（x, y），不包括$\hat{}x$。上述积分有两个问题：

1. 我们无法使用方程 $p(\eta$, $\theta \mid D)$ 评估式 7，因为使用方程计算 $p(D)$ 式8 需要求解一个计算上难以处理的高维积分。2. 即使我们能够估计 $p(D)$，使用方程计算 $p(y = \hat{}y \mid x = \hat{}x$, $D)$ 式 10 仍需要对 $(\eta$, $\theta)$ 求解另一个高维积分。

[14]若要按标准方式使用概率的乘法法则，我们本应使用 $p(\eta$, $\theta \mid \hat{}$ x, $D)$，但由于尚未观测到$\hat{y}$，$\hat{x}$ 不会影响后验分布，因此这两项相等。

这两个问题都与高维积分难以进行可行计算有关。

当部分或全部参数是离散的时，相应的积分会变为求和。由于随着参数维度增加，η，θ 可能值组合的组合增长，参数维度较高时，求和仍然难以处理。在药物计量学中，参数通常是连续的，因此本文其余部分将聚焦于连续参数。

4.3 先验选择

本节主要关注如何理解并选择合适的先验。我们将尝试回答以下问题：

1. 先验何时有用？如何发挥作用？ 2. 先验何时有害？如何造成危害？

Pumas 中有哪些先验可用、其参数以及如何定义它们的具体信息，可在工作流部分（第 3）中找到。

4.3.1 概述

在贝叶斯统计中，参数的先验分布表示在观测到任何数据之前，对参数取值的信念状态。例如，对于一个标量参数，使用均值为 0、标准差为 2 的 $N(0, 2.0)$ 一元高斯先验分布，意味着我们认为该参数介于 -6 和 6 之间的概率质量为$\approx 99.7\%$。更一般地，概率密度函数为 $p(x)$ 的先验分布意味着，我们认为参数介于 a 和 b之间的概率质量（曲线下面积）为 $\int_a^b p(x)dx$ 。观测到数据后，使用观测值和似然值更新先验分布，从而得到后验分布。后验分布表示数据观测之后，对参数取值的信念状态。如果收集到更多数据，理论上，我们可以在仅分析新数据时，将旧的后验分布用作新的先验分布[15]。但在实践中，由于后验分布通常没有闭式解，因此将其用作任何新分析的先验可能会比较棘手

因此，可能需要使用所有新旧数据以及旧先验进行完整分析。[16]

在某些方面，先验分布与非贝叶斯工作流程中参数的定义域类似，而后验分布则类似于最大似然估计。在观测数据之前，由于我们不知道参数的值，因此只能讨论其定义域。观测数据之后，我们可以根据数据拟合模型，并确定拟合效果最佳的参数值。在非贝叶斯分析中，我们通常会为每个参数选择一个定义域，以帮助优化算法找到合理的参数值，避免浪费时间尝试合理定义域之外的值。例如，我们知道任何协方差矩阵参数都必须是正定的，因此优化算法不应探索违反此约束的协方差矩阵参数值。先验分布将这一思想推广开来：它不仅隐含了由分布决定的定义域（称为分布的支集），还允许指定对某些参数值相较于其他值的偏好程度。指定参数值偏好的能力是一种强大的工具，但如果使用不当，也可能非常危险。使用得当时，它可以让领域知识和以往类似研究的结果在当前研究中得以复用。但使用不当时，它也可能借助难以审查的复杂数学来掩盖不严谨的科学实践和不道德行为。

4.3.2 好的、坏的和无害的先验

粗略来说，先验可分为两类[17]：

1. 强先验，即信息性先验，以及

[15]我们不能在同一项分析中重复使用数据。如果旧数据已用于在新分析中定义先验，那么新分析中应仅使用新数据。

[16]有一些算法可以利用新的数据来更新先前研究中的后验样本，但本文不讨论这些算法。[17]还有一类先验通常称为 *non-informative priors*，但本文也不讨论它们，因为只有针对特定的模型参数化方式才能保证非信息性。例如，模型中针对标准差参数的非信息先验，在改用方差参数对模型重新参数化后，就会变成信息先验。因此，本文将非信息先验视为弱信息先验的一种特殊情况。最后，所谓的 *flatimproper priors* 也被有意排除在外，因为只需在固定精度浮点数所能表示的数值边界附近截断定义域，就可以轻易地将它们转换为平坦的适当先验。

2. 弱先验，即信息性较弱的先验。

这些分类比较宽泛，因为实际上，重要的只有先验相对于似然的相对强度。请回想一下，用于驱动 MCMC 抽样的联合概率为：

$$p(D, \eta, \theta) = p(D \mid \eta, \theta) \times p(\eta, \theta) \tag{11}$$

即先验概率与似然值的乘积。如果数据量很大，似然项将主导大多数先验，后验分布将主要反映能够很好地拟合数据的参数值。如果数据不足，先验及其选择会变得更加重要，因为某些先验的选择可能会使其主导上述项。当先验主导似然时，后验分布将只是先验分布向能够实际地 拟合数据的参数值方向发生的轻微偏移。在这些情况下，在分析中滥用先验的风险更高，本节将讨论可能发生这种情况的场景。

无害先验是在很大程度上模仿非贝叶斯工作流中指定参数域这一目的的先验。这些先验对参数值几乎没有偏好，甚至完全没有偏好，即使只有少量数据点也很容易压倒它们。例如，对于一个正参数而言，在 $[0, 10^{10}]$ 上的均匀分布可以被视为无害先验。在这种特定参数化下，该先验只编码参数的定义域，而不会偏好某些值。

好的先验和不好的先验有一个共同点：它们都是“信息性”的。更准确地说，好的先验具有信息性，而不好的先验则具有误导性。表 11总结了各种信息性和误导性先验。好的信息性先验具有良好的科学依据，且不会与数据相矛盾。

根据表 11，定义合理先验有两种通用方法：

1. 定义一个弱信息先验[18]，并确保随着先验越来越弱，研究结论在极限情况下不会改变。这种情况下，我们让数据自行说明，而不把我们自己的偏见强加于结果。在这种情况下，我们使用贝叶斯推断，仅仅是因为它具有概率上的严谨性，并且能够量化参数和响应的总体不确定性，即使模型不可辨识和/或我们只有少量数据点时也是如此。

2. 使用类似的先前研究来指导先验选择，并检验它是否与新数据相矛盾。在这种情况下，需要对先验分布作出更多论证，并证明先验与数据不矛盾

证明先验与数据不矛盾的一种方法是，先完全不使用观测值，并采用强先验的弱化版本，例如增大标准差。随后可以逐步增强先验（例如逐渐减小标准差），直至达到所需的强度，然后逐步将（先前移除的）观测值重新加入分析。如果每次逐步调整并重新运行分析后，所有重要统计量（例如药物有效的概率）都呈现一致的趋势，那么先验的强度就与数据所呈现的情况相符。可以通过一系列先验模拟（移除所有数据时）以及结合马尔可夫链蒙特卡罗运行和后验模拟（重新加入数据时）来完成此操作。有关检测先验与数据冲突以及模型信息不足的更多方法，请读者参阅 Kallioinen 等人 (2021)。

另一个检测数据与信息性先验之间潜在冲突的简单方法，是从以下分布中进行模拟：

$$\begin{aligned} (\eta, \theta) &\sim p(\eta, \theta) \\ y &\sim p(y \mid \eta, \theta) \end{aligned} \tag{12}$$

其中 $p(\eta, \theta)$是先验分布。然后，您可以进行模拟或检查VPC图，以评估数据与先验模拟的一致-性。如果先验提供的信息较弱或几乎不提供信息，并且我们拥有足够多的数据，使似然占据主导地位，那么在 VPC图中与数据不一-致的先验模拟可以忽略。然而，如果先验提供的信息较为丰富，它基于先前研究，并且新研究中的数据点不足以完全压倒先验，那么先验模拟的 VPC 图接下来

[18] 非常弱的先验有时称为弥散先验

科学依据	与数据矛盾?	评论
None	Yes	这属于作弊。使用与.....相矛盾的强先验并过度- 没有科学依据的数据可以被用来随意编造任何结果 我们想要的结果。例如，一种药物可能无效，但如果 强先验表明，该药物很可能有效。 且没有足够的数据来抵消那个强烈而错误的先验， “分析”的结论将是该药物有效。
None	No	这是糟糕的科学。使用一个与.....一致的强先验。 数据，但它会过度压制似然，且缺乏科学依据。 基础可能会导致过度自信的预测和过早的结论- 结论所用的数据点比必要的更少。这是一种 与人为多次复制数据产生的效果类似 在非贝叶斯方法中人为缩小置信区间 工作流程。
先前研究	Yes	这是一个迹象，表明使用了彼此不相似的先前研究来__日本 指导先验的选择，并且应当修订先验选择 因为它与新的数据相矛盾，并且可能会使.....产生偏差 分析的结果。
先前研究	No	谨慎使用时，来自先前研究的结果（例如，一个 其参数后验分布的近似）可 用于为一项 新的类似研究选择合适的信息性先验。这些先验不应与新的 收集到的数据相矛盾，但可能有助于我们提前结束这项新研究， 从而使用比其他情况下可能需要的更小样本量。积极地 在研究结果已经明确药物.....后提前结束研究， 根据所有可用信息判断有效，意味着更多 患者能够更早获得真正有效的药物。这一点 对于罕见病药物开发尤其重要，因为 在研究中收集更多数据往往意味着要增加数年 的临床试验时长。这种情况需要更多 监管机构与行业就界定最佳实践达成共识 在此类研究中使用信息丰富的先验分布，旨在 使患者受益。

表11：该表展示了选择**信息性先验分布**的若干种方法。只有最后一种情况是对信息性先验的良好使用。**该表仅适用于信息性先验** 可能支配似然的情况，因为被似然支配的弱信息先验通常影响没那么大。

对数据进行检查可能是一个有用的诊断方法。请参阅第 3.2 节，了解如何在 Pumas 中执行此操作。

使用弱信息先验时，你可能会想利用这类模拟图，选择一个对数据有良好（但不一定很窄）覆盖范围的先验。通常，不赞成根据数据对先验进行任何微调。这是因为这样一来，我们就会重复使用同一批数据：一次用于微调先验，另一次用于更新先验以得到后验分布。这可能导致低估后验分布的方差，也就是得到过度自信的后验分布，进而导致过度自信的后验预测。这类似于在频率学派分析流程中复制数据中的部分或全部数据点，这会低估置信区间和标准误。

如果采样因数值误差而失败，先验过弱可能是原因之一。在这种情况下，可以尝试更改先验，例如将其支撑集截断至合理范围，使其与数据更加一致。不过，最终分析中只能允许进行最小限度的此类更改。

还需要开展良好的采样后敏感性分析，以确保研究结论不会局部地受到先验影响。更一般地说，数值误差通常表明模型对某些参数值过于敏感，这可能意味着模型本身存在结构性问题。在这种情况下，还应考虑比简单截断先验更好的解决方法，例如简化模型或重新参数化，以避免数值不稳定性。Gelman 等人Gelman 等人（2020年b）的贝叶斯工作流程论文提供了极好的建议和案例研究，可供诊断马尔可夫链蒙特卡罗运行失败时参考。

4.3.3 相关性与协方差

在定义包含协方差矩阵参数的模型时（例如，药代动力学中通常用于个体特异性参数的多元正态分布先验中的协方差参数），我们总会遇到以下两种等价的参数化方式：

1. 使用协方差矩阵参数 Ω；或 2. 使用标准差向量 ω 和相关矩阵参数 C。

可以很容易地从（ω，C）求出 Ω，反之亦然。设 D_ω 为其元素是 ω 的对角矩阵，ω_i 为 ω 中的 i^{th} 元素，并设 $\Omega[i,i]$ 为 Ω 的 i^{th}对角元素。C、ω 与 Ω 之间的关系如下：

$$\begin{aligned} \Omega &= D_\omega \times C \times D_\omega \\ C &= D_\omega^{-1} \times \Omega \times D_\omega^{-1} \\ \omega_i &= \sqrt{\Omega[i,i]} \end{aligned} \tag{13}$$

在非贝叶斯背景下，这两种参数化方式是等价的。然而，在贝叶斯分析中，应为参数定义先验分布。由于先验分布旨在编码领域知识，以及对参数取值的信念状态，因此通常建议采用更直观、更易解释的参数化方式，以便更好地理解所用的先验分布。基于这个原因，有些人更倾向于使用第二种参数化方式，将标准差向量与相关矩阵分开，因为它们更易解释。我们看到了可用的先验分布示例。

用于标准差、相关性和协方差矩阵参数的第3节。

4.4 马尔可夫链蒙特卡罗（MCMC）直觉

本节将说明使用马尔可夫链蒙特卡罗的动机，展示如何利用马尔可夫链蒙特卡罗绕过高维积分的需求（第4.2节中讨论），从而满足所有实际应用的需要。

4.4.1 推断

马尔可夫链蒙特卡罗（MCMC）（Brooks 等人，2011）通过直接从后验概率分布中进行采样[19]，绕过了求解数值积分问题的需要，而只使用公式7中易于处理的分子。该分子有时被称为联合概率，因为它就是 $p(D,\ \eta,\ \theta) = p(D \mid \eta,\ \theta) \cdot p(\eta,\ \theta)$。请注意有条件项与无条件项之间的区别 |。拥有后验分布的样本后，我们可以估计如下量：

$$p(f(\eta,\theta) > 0 \mid D) \tag{14}$$

对于任意函数 f，可以通过一种简单的计数方法获得结果：检查后验分布中的所有样本，并统计其中有多少个满足所关注的特定条件。满足条件 $f(\eta,\ \theta) > 0$ 的样本比例，就是上述概率的无偏估计。更具体地说，该概率可以是 $p(\theta_i > 0 \mid D)$，其中 θ_i 对应于实验中治疗组相对于安慰剂组的效应量。

4.4.2 预测

除了使用后验样本估计事件发生的概率外，后验样本还可用于进行预测。Pumas 中用于执行后验预测的函数已在

[19]随机变量的一个样本，或其分布中的一个样本，就是该随机变量的一次取值。例如，$[0,1,0,0,1,1]$ 是来自支撑集为集合 $\{0,1\}$ 的伯努利分布的 6 个样本。分布的支撑集是随机变量的定义域。例如，集合 $\{0,1\}$ 是伯努利分布的支撑集，而所有实数构成正态分布的支撑集。后验分布中的一个样本可以理解为可能生成我们所观察到的数据的参数值。根据上下文，有时会用术语 $sample$ 指多个这样的样本。

第 3.13 节中介绍。假设我们有 N来自后验分布的样本 $p(\eta, \theta \mid D)$：

$$\{(\eta^{(j)}, \theta^{(j)}) : j \in 1 \dots N\}$$

回想一下，式9 中的难以处理项为 $p(y = \hat{} y \mid x = \hat{} x, D)$，可改写为：

$$\int\int p(y = \hat{y} \mid x = \hat{x}, \eta, \theta) \times p(\eta, \theta \mid D)\, d\eta d\theta \approx \frac{1}{N}\sum_{j=1}^{N} p(y = \hat{y} \mid x = \hat{x}, \eta = \eta^{(j)}, \theta = \theta^{(j)})$$

其中，$p(y = \hat{} y \mid x = \hat{} x, \eta = \eta^{(j)}, \theta = \theta^{(j)})$ 是在给定参数值的情况下计算得到的 $\hat{} y \mid \hat{} x$ 的条件概率。

因此，利用这些样本，可将难以处理的积分转化为可处理的 N 项均值，因为给定 $(\hat{} x, \hat{} y, \eta^{(j)}, \theta^{(j)})$ 和模型后，我们可以轻松计算 $p(\hat{} y \mid \hat{} x, \eta^{(j)}, \theta^{(j)})$。因此，式 9 中的期望项可用式 15 近似，其中 $E[y \mid x = \hat{} x, \eta^{(j)}, \theta^{(j)}]$ 只是条件分布 $p(y = \hat{} y \mid x = \hat{} x, \eta = \eta^{(j)}, \theta = \theta^{(j)})$ 的均值，可通过单次运行模型计算得到。在 Pumas 语法中，这就是 predict 函数的输出（或 simulate error = false 时 simobs 的输出），使用的参数为 $(\eta^{(j)}, \theta^{(j)})$，协变量为 $\hat{} x$。

更一般地，可以使用以下公式估计相对于后验分布的**任意函数** $g(\eta, \theta)$的期望值：

$$E[g(\eta, \theta) \mid D] = \int\int g(\eta, \theta) \times p(\eta, \theta \mid D)\, d\eta d\theta \approx \frac{1}{N}\sum_{j=1}^{N} g(\eta^{(j)}, \theta^{(j)})$$

例如，g 可以计算一些非房室分析参数（第 3.16节），这些参数基于模型的预测，或者计算任何其他作为参数的确定性函数的确定性量。

当定义 g 为 $g(\eta', \theta') = E[y \mid x = \hat{} x, \eta = \eta', \theta = \theta']$ 时，我们便得到预测这一特殊情形。而当为另一个任意函数 f 定义 g 为 $g(\eta, \theta) = \mathbb{1}_{f(\eta,\theta)>0}$ 时，其中 $\mathbb{1}_{f(\eta,\theta)>0}$ 是一个指示函数：当条件 $f(\eta, \theta) > 0$ 成立时取 1，否则取 0，我们便得到估计概率 $p(f(\eta, \theta) > 0 \mid D)$ 的特殊情形。

换句话说，后验分布中的样本几乎涵盖了你在做出决策时估计所有感兴趣量所需的一切。那么，如何才能在不计算那些难以处理的积分的情况下获得这样的样本呢？我们使用马尔可夫链蒙特卡罗。

4.4.3 模拟

我们已经看到，给定来自后验分布的 N 样本，我们可以针对函数 g的任意特定选择，计算平均预测 $E[g(\eta, \theta) \mid D] \approx \frac{1}{N}\sum_{j=1}^{N} g(\eta^{(j)}, \theta^{(j)})$。或者，您也可以获得预测的分布：

$$\{g(\eta^{(j)}, \theta^{(j)}) \text{ for } j \in 1 \dots N\} \tag{16}$$

其中 $g(\eta', \theta') = E[y \mid x = \hat{} x, \eta = \eta', \theta = \theta']$。这是对 $g(\eta, \theta)$ 的分布进行的马尔可夫链蒙特卡罗近似，其中$(\eta, \theta) \sim p(\eta, \theta \mid D)$。对于上述 g的选择，该预测的分布通常称为后验预测分布。当改为从先验中抽取(η, θ) 时，对于上述g的选择，$g(\eta, \theta)$ 的分布称为先验预测分布。

除了对预测进行采样，或者更一般地对参数的确定性函数 (η, θ) 进行采样之外，也可以从以下 $\hat{} y$ 的分布中采样：

$$(\eta, \theta) \sim p(\eta, \theta \mid D)$$
$$\hat{y} \sim p(y = \hat{y} \mid x = \hat{x}, \eta, \theta)$$

在 Pumas 语法中，这是使用后验分布参数值和协变量$\hat{} x$ 的 simobs 函数输出。或者，也可以从它们的先验分布中采样 (η, θ)，或者将其固定为特定的真实值。这些先验/后验分布/真实值模拟可用于执行以下任一操作：

1. 在使用真实数据[20]之前，先用合成数据测试 MCMC 算法。

[20]这被称为“基于模拟的校准”。Talts 等人(2020年)。

$$
\begin{aligned}
E[y \mid x = \hat{x}, D] &= \int \hat{y} \times p(y = \hat{y} \mid x = \hat{x}, D)\, d\hat{y} \\
&\approx \int \hat{y} \times \left(\frac{1}{N} \sum_{j=1}^{N} p(y = \hat{y} \mid x = \hat{x}, \eta = \eta^{(j)}, \theta = \theta^{(j)}) \right) d\hat{y} \\
&= \frac{1}{N} \sum_{j=1}^{N} \left(\int \hat{y} \times p(y = \hat{y} \mid x = \hat{x}, \eta = \eta^{(j)}, \theta = \theta^{(j)}) d\hat{y} \right) \\
&= \frac{1}{N} \sum_{j=1}^{N} E[y \mid x = \hat{x}, \eta = \eta^{(j)}, \theta = \theta^{(j)}]
\end{aligned} \tag{15}
$$

2. 通过检查先验模拟与真实数据的相似性（例如使用视觉预测检验（VPC）图，也称为先验预测检验），识别极差的先验选择，从而为先验选择提供最低限度的指导。更多关于先验选择的详细信息，请参见 4.3 。
3. 使用 VPC 图（也称为后验预测检验）将后验模拟与真实数据进行比较，并估算所谓的贝叶斯p-值，以量化模型拟合质量。

在第 3.2 节中介绍了在 Pumas 中进行先验模拟和预测的代码。同样，在第 3.13 节中介绍了进行后验模拟和预测的代码。最后，第 p 节介绍了执行 VPC、各种模拟查询（例如贝叶斯 3.14、3.11 和3.16 节分别介绍了 值和 NCA 的代码。

4.5 无调采样器（NUTS）算法

本节将直观地介绍无掉头采样器（NUTS）（Hoffman 和 Gelman，2014；Betancourt，2017）马尔可夫链蒙特卡罗算法。我们将重点建立对该算法如何运行以及如何调整其超参数的深入直觉[21]。我们不想用其他许多资料中都能找到的方程来充斥读者的头脑；要有效使用该算法，并不一定非要掌握这些方程。对于 MCMC 初学者，或者初次阅读本文的读者，可以跳过4.5.6、4.5.7 和4.5.8这几个小节，而不会对理解上下文造成明显影响。

4.5.1 高层次描述

MCMC 抽样使用随机过程（随机游走）在（η, θ）空间中迭代地从后验分布 $p(\eta, \theta \mid D)$ 中收集样本，且仅依赖“可用的局部信息”。图 15 展示了一个一维高斯随机变量的简单随机游走。在算法的第 j^{th} 次迭代中，可用的局部信息基本上就是在特定值（$\eta = \eta^{j-1}, \theta = \theta^{j-1}$）处计算的方程 7 中的分子，以及它相对于（$\eta, \theta$）的梯度[22] ；这两者都很容易计算。鉴于这个迭代算法只使用前一次迭代 $j-1$ 的信息，按定义它是一条所谓的马尔可夫链。MCMC 系列算法的目标是使各个步骤 $\{(\eta^{(j)}, \theta^{(j)}) : j \in 1 \ldots N\}$ 成为来自后验分布的有效样本。本节将重点介绍所谓的无调采样器（NUTS）算法（Hoffman 和 Gelman，2014；Betancourt，2017），它是哈密顿蒙特卡洛（HMC）（Neal，2011）这一类 MCMC 算法的变体。我们不会详细讲解这些算法，但会说明其背后的直观原理，帮助你理解它们的超参数，以便在需要时能够有根据地调整。

[21] *hyper-parameters* 通常指任何不由贝叶斯推断算法进行推断、且需要由用户在贝叶斯分析之前预先指定的参数。这些通常可以是模型超参数，例如群体参数 θ 的先验分布参数；也可以是算法超参数，例如 NUTS 算法的设置。

[22]并非所有马尔可夫链蒙特卡罗算法都会使用梯度，但能够扩展到许多参数的算法会使用。

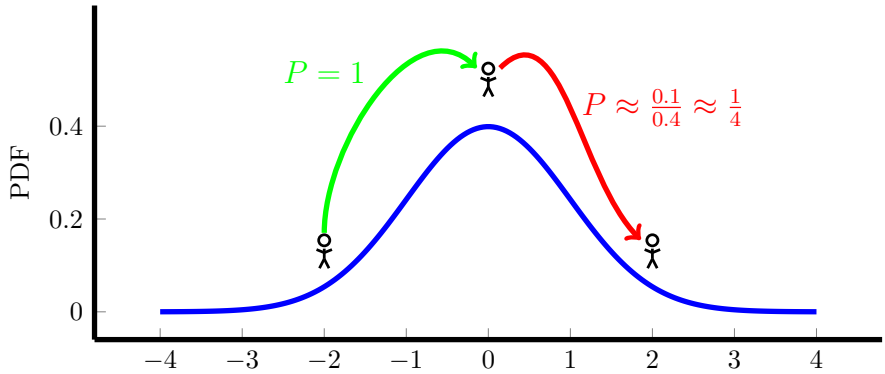

图 15：正态分布随机变量（横轴）的随机游走示意图，纵轴显示概率密度函数（PDF）。首先提出一个提议，然后根据特定的接受概率决定是否接受。当提议是让徒步者向上攀登时，该步以 1 的概率被接受；当提议是让徒步者向下攀登时，则以两个位置处概率密度函数值之比作为接受概率。

设想采样器的当前位置（$\eta^{(j-1)}$，$\theta^{(j-1)}$）是（η，θ）空间中的一个粒子。在 NUTS 算法中，随机游走过程如下所示[23]：

1. 虚拟粒子（$\eta^{(j-1)}, \theta^{(j-1)}$）被赋予一个随机方向上的随机虚拟速度，也就是说，它获得一个随机的虚拟速度。该速度从多元高斯分布中采样。2. 联合概率的对数（即式7中分子项的对数）相对于（η, θ）的梯度，在局部起到虚拟力场的作用，将虚拟粒子推向先验（对数）和／或似然（对数）较高的区域，并使其远离先验（对数）和／或似然（对数）较低的区域。3. 虚拟粒子的运动通过时间离散化和近似常微分方程求解器进行**近似**模拟[24]，总计 T 在虚拟力场作用下模拟的时间步数为[25]，每个模拟时间步的大小为 ϵ。此模拟只需初始位置和初始速度，并且能够计算任意点（η, θ）处施加的力；这等价于计算 $\log p(D, \eta, \theta)$ 相对于（η, θ）的梯度，即 $^{d}\log p(D,\eta,\theta)\, d(\eta,\theta)$。

4. 在模拟的虚构轨迹中随机选择一个位置（η, θ），该轨迹共有 T 个时间步，此位置成为**提议**。

5. 然后以经过精心选择的概率接受提议，以确保随机游走能从后验分布中为我们提供正确的样本。如果接受提议，它就成为下一个样本（$\eta^{(j)}$，$\theta^{(j)}$）；否则，再次采样前一个值，即（$\eta^{(j)}$，$\theta^{(j)}$）$=$（$\eta^{(j-1)}$，$\theta^{(j-1)}$）。[26]

上述算法仍然是随机游走，但它会偏向高概率区域

[23]实际算法包含许多此处未讨论的性能优化。[24]粒子动力学模拟归结为模拟一组常微分方程，并以 $\log p(D, \eta, \theta)$ 的梯度作为驱动力。模拟使用一种称为蛙跳法的近似常微分方程求解器。蛙跳法采用较大的步长时属于近似方法，因为其解会违反能量守恒定律，尽管它所谓的保体积性质仍然成立。不过，在此情形下，这是一种理想性质，有助于充分探索后验分布，即使高概率质量区域彼此不连通也不例外。有关时间步长对采样行为的影响，请参见第 4.5.4节。

[25]在 NUTS 算法中，T 时间步会随机分配为正向时间步和反向时间步，使得正向和反向时间步的总数为 T。反向时间步等价于将初始速度乘以 -1 后执行正向时间步。[26]在 NUTS 的最先进变体中（Betancourt，2017），提议选择与接受/拒绝步骤合并为一个步骤，从包含前一个粒子位置的虚拟轨迹中进行采样（$\eta^{(j-1)}$，$\theta^{(j-1)}$）。采样方式可确保选定的下一个位置（η^{j}，θ^{j}）是来自后验分布的正确样本。不过，为便于说明，我们在概念上将这两个步骤分开。

因为它利用对数联合概率的梯度，即使粒子以随机速度启动，也能将其推向正确方向。上述采样算法分为两个阶段：适应阶段和采样阶段。在适应阶段，采样器在执行上述采样过程的同时，会调整时间步长和初始速度分布。在采样阶段，算法的超参数不再调整，采样则继续采用上述相同的过程。通常建议在采样后丢弃适应步数，作为预热期[27]，因为这些样本可能无法代表后验分布中的样本。在 Pumas 中，可以使用丢弃 函数来实现，具体说明见第3.3节。也可以使用`nadapts`选项指定适应步数，如第3.3节所示。

4.5.2 马尔可夫链蒙特卡罗可视化

要交互式地可视化马尔可夫链蒙特卡罗采样器的工作原理，chi-feng.github.io/mcmc-demo/ 这是一个很棒的资源，你可以在其中更改目标分布和采样算法，从而直观地了解马尔可夫链蒙特卡罗采样器在不同情况下的工作方式。你可以选择“无调采样器”作为算法，然后更改目标分布，以可视化 NUTS 从不同后验分布中采样时的表现。

4.5.3 提议接受

虽然 NUTS 的最先进变体（Betan-court，2017）不使用对提议进行显式接受/拒绝检验（也称为 Metropolis-Hastings 检验），但其操作与传统的先选择提议、再进行接受/拒绝检验的过程类似。出于教学目的，我们假设这是两个独立步骤。算法第 5 步中提议的接受概率取决于：

1. 先验概率，以及 2. 似然函数（提议与数据的拟合程度）

与先前样本（$\eta^{(j-1)}$，$\theta^{(j-1)}$）相比，导致预测不佳、无法很好地拟合数据的提议，或者根据先验判断比先前样本更不可能的提议更有可能被拒绝。另一方面，与先前样本相比，能更好地拟合数据和/或根据先验判断更有可能的提议，则更有可能被接受。

4.5.4 时间步长的影响

一般来说，模拟中的时间步长越大，常微分方程求解器的近似程度就越高，提议也会越具有探索性和冒险性，从而导致接受提议的比例降低。另一方面，较小的步长通常会产生探索性较弱的提议，这些提议更有可能被接受，从而提高接受比。采样器的探索能力部分来源于近似常微分方程求解器过冲的能力：在提出提议时，从一个后验概率较高的区域跳跃到另一个[28] 区域，从而探索多个众数，即使两个众数之间存在概率为 0 的区域。概率为零的区域 $p(\eta, \theta \mid D)$意味着概率为零 $p(D, \eta, \theta)$（方程7），这将产生一股无限大的力，将粒子推离该区域。因此，精确模拟永远无法跨越这样的区域进行跳跃，这正是需要使用近似求解器和过冲的原因。

4.5.5 时间步长调整与目标接受率

在 NUTS 算法中，您无需自行设置步长。NUTS 算法会调整其步长，以使提议的平均接受比例达到某个特定值。这个目标接受比是 NUTS 的超参数。在 Pumas 中，您可以使用第`targetaccept` 节所示的3.3选项设置目标接受比。0.99 表示我们希望接受采样器生成的 99% 的提议。这通常会使提议与当前样本之间的距离较小，因为这样可以提高接受该提议的可能性。另一方面，0.1 的目标接受率表示我们希望平均只接受所生成提议的 10%。因此，NUTS 算法会尝试使用更大的步长，以确保它

[27] 也称为 $warm\text{-}up$。

[28] 后验概率较高的区域具有较高的联合概7式中的分子）。联合概率是先验概率与似然的乘积。因此，先验概率较高和/或似然较高的参数值，其联合概率和后验概率也较高。

拒绝 90% 的提议。通常建议使用 0.6–0.8 的目标接受比。Pumas 使用的默认值为 0.8。

在采样过程中，通常需要在探索和利用之间进行权衡。如果采样器过于激进，每一步都尝试远离上一个样本的提议，那么它更有可能探索整个后验分布，而不会陷入后验分布的某个局部众数附近进行采样。不过，另一方面，探索过度往往会导致许多提议被拒绝，因为数据与这些激进提议的联合概率 $p(D,\ \eta,\ \theta)$ 较低。这会降低有效样本量（ESS）[29] 与样本总数之比（也称为相对 ESS ），因为大量样本会由于提议被拒绝而彼此完全相同。

另一方面，如果减少探索，至少有 2 种可能的情况：

1. 第一种情况是，我们从后验分布的一个众数处初始化采样器。仅在上一个样本附近提出提议，可以确保我们接受大多数样本，因为后验分布众数附近的提议很可能对应良好的参数值。这种围绕已知良好参数值的局部采样行为，在这里称为利用。虽然围绕某个众数进行高强度利用所生成的样本可能无法代表整个后验分布，但仍可能足以近似后验预测分布；是否满意，可通过 VPC 图来判断。

2. 第二种情况是从不佳的参数值初始化采样器。不佳的参数值和较低的探索程度往往会导致类似优化的行为：采样器会以嘈杂的方式花费大量迭代逐渐移向某个众数。这种类似优化、寻求众数的行为会导致样本的自相关较高，因为采样器大多朝同一方向（即朝众数）移动。自相关较高意味着 ESS 较低，因为样本之间会缺乏独立性。[30] 此外，在采样器到达能够良好拟合数据的参数值之前，这些样本不太可能产生良好的后验预测分布。当适应算法因初始参数不佳且模型过于复杂、难以优化（更不用说从其后验分布中采样）而无法找到能够充分探索后验分布的合适步长时，这种情况是 MCMC 算法相当常见的一种失败模式。在这种情况下，所有样本看起来都可能存在自相关，而且样本之间的步长可能会非常小（探索程度较低）。在 Pumas 中，采样期间的实时进度信息会显示步长，如图1所示。通常，尽早发现这种失败模式并提前终止采样会有所帮助。

4.5.6 可选：时间步数和掉头次数

考虑 (η, θ) 空间中的一个方向，例如某个参数的轴。对于后验分布中相对平坦的区域，沿此方向的许多取值几乎同样可能，也就是说，它们对数据的拟合程度相同，并且根据先验分布几乎同样可能；此时，即使提议值与当前样本相距较远，也仍可能在大多数情况下被接受。这种情况尤其容易出现在后验分布中模型（几乎）不可辨识、导致参数高度相关，且先验不加区分（例如先验较弱）的部分。另一方面，后验分布中高度集中在曲率较大的众数附近的区域，通常需要较小的步长才能获得合理的接受比，因为与当前样本稍有偏离的提议值，根据先验分布可能极不可能，或者可能导致非常糟糕的预测。这种情况尤其容易出现在模型对参数值高度敏感的后验分布区域，或者先验过度集中于特定参数值的情况下。

为了考虑后验分布不同区域中沿着同一方向[31]的曲率变化，

[29]ESS 是马尔可夫链在估计后验均值时所生成的“独立样本数量”的近似值。每个样本对应的 ESS 比率较低，是由 MCMC 样本中的高自相关造成的，通常是不良指标。

[30]在 Pumas中，群体参数的 ESS 值显示在 MCMC 结果的默认显示界面中，如图 2所示。[31]通过使用粒子在每次 NUTS 迭代开始时获得的初始速度的分布，可以处理不同方向上的不同曲率。

NUTS 算法采用多步提议机制，使用固定的时间步长（在适应阶段确定后保持不变）和动态的时间步数（在适应阶段和采样阶段都会动态变化）。更具体地说，采样器会模拟一条包含 T 个时间步的轨迹，然后从这条轨迹中随机选择一个提议，其中每次提出提议时，T 都不同。

NUTS 模拟的时间步数 T 由递增模拟决定：$T = 1+2+4+8+16+\ldots$ 个时间步，其中每次递增模拟的时间步数按 2 的递增幂次增长。每次递增模拟可以是：

1. 从最靠近未来的状态开始向前推进，或 2. 从最靠近过去的状态开始向后推进。

每次递增模拟的方向通过随机抽样决定，两个方向的概率各为 0.5。表12展示了在以下特定模拟方向选择下的递增模拟示例：[正向、正向、反向、反向、正向、反向]。

那么，何时停止模拟？NUTS 算法通常会在满足以下 4 个条件之一时停止模拟：

1. 出现所谓的掉头，即采样器开始从轨迹的一端向另一端折返。2. 达到预先设定的最大模拟步数。3. 某一步中的对数先验概率和/或对数似然迅速下降，降幅超过预先设定的阈值。4. 出现数值错误。

第三和第四个终止条件通常称为“发散”。

模拟终止后，从模拟轨迹上随机选取的一个点将成为下一个提议，搜索随即终止。通过发现掉头来终止通常被视为探索成功的标志。每次 NUTS 迭代中对 $\log p(D, \eta, \theta)$ 的求值次数由模拟轨迹的长度决定；若 j，该长度为 $\sum_{i=0}^{j-1} 2^i = 2^j - 1$ 需要进行增量模拟才能找到掉头[32]。

在 NUTS 算法的高效实现中，会使用深度为 j 的二叉树数据结构，以帮助从模拟过程中访问过的所有状态（η，θ）中高效选择一个提议，直到找到掉头[33]，而无需存储所有状态。这是一种效率增强技巧，但术语 树深度 保留了下来，并逐渐成为迄今为止运行的增量模拟次数的同义词，j。如果采样器即使在达到预先指定的最大 j 后仍无法找到掉头，采样器仍会终止模拟并提出一个提议。最大树深度这一术语通常用于指代在必须提出一个提议之前所允许的最大增量模拟次数 j，即使尚未找到掉头。

4.5.7 可选：初始速度的分布

回想一下，在每次 NUTS 迭代中，我们会为（η, θ）粒子采样一个随机的初始速度，然后模拟动力学以得到一个提议。假设我们已经拥有来自后验分布 $p(\eta, \theta \mid D)$ 的样本。如果你要回去使用 NUTS 重新进行采样，你会如何为这个假想的（η, θ）粒子采样初始速度，以使采样尽可能高效？一般来说，沿着后验分布中方差较大的方向移动得更快、沿着方差较小的方向移动得更慢是合理的。例如，我们可以计算每个参数轴方向上的方差，并为变化较大的参数采样更高的速度，为变化较小的参数采样更低的速度。实际上，你可以将不同参数理解为具有不同的尺度参数，其中一个参数可能处于几十的量级，而另一个参数可能处于几千的量级。在这种情况下，合理的做法是

[32]在 NUTS的高效实现中，一旦发现掉头，就会中断最后一次增量模拟，因此模型评估的次数实际上介于 2^{j-1} 和 $2^j - 1$[33]访问过的状态数量（不包括初始状态）最多为 $2^j - 1$。加上初始状态后，我们有 2^j 个可能的状态，其中任何一个都可能成为提议。这些状态理论上可以存储为深度为 j 的二叉树的叶节点，该树有 2^j 个叶节点。然而在实践中，只存储这类状态的一个子集，并利用树的思想来确保能够从全部 2^j 个可能状态中高效地随机选择提议，同时满足 MCMC 中提议的理论要求，这通常称为详细平衡条件。

增量 j	模拟方向	j 次增量后模拟的时间步间隔
0	-	$[0,0]$
1	正向	$[0,0+1]=[0,1]$
2	正向	$[0,0+1+2]=[0,3]$
3	反向	$[0-4,0+1+2]=[-4,3]$
4	反向	$[0-4-8,0+1+2]=[-12,3]$
5	正向	$[0-4-8,0+1+2+16]=[-12,19]$
6	反向	$[0-4-8-32,0+1+2+16]=[-44,19]$

表12：该表展示了针对 $j\in[1,6]$ 的 NUTS 算法增量模拟。请注意，每次递增时，正方向上增加的数值是 2 的幂，负方向上则减去 2 的幂。在递增 j 次后所经历的总时间步数（不包括初始时间点$t=0$）为 $1+2+4+8+16+\cdots=2^0+2^1+2^2+\cdots+2^{j-1}=\sum_{i=0}^{j-1}2^i=2^j-1$。**检查**：$2^1-1=2^0=1$、$2^2-1=2^0+2^1=3$、$2^3-1=2^0+2^1+2^2=7$ 等。请注意，上述区间表示时间步数。每个时间步的模拟时间步长为 ϵ。

以沿不同方向采用不同的速度，从而更高效地从后验分布中进行样本采集。更一般地，甚至可以根据可获得的（假设的）样本计算样本协方差矩阵，计算主成分，并沿着方差比其他方向更大的方向采用更高的速度。

如果我们用一个数值 s_i 来编码粒子沿各个方向 d_i 运动的缓慢程度，那么将沿该方向的速度标准差设为 $1/s_i$，即可达到所需的缓慢程度。假设每个 d_i 都是沿某个特定参数 i 的一个轴（它可能属于 η 或 θ）。沿着 d_i 的速度 v_i 分布可以是以下单变量高斯分布：

$$v_i\sim N(0,(1/s_i)^2) \tag{17}$$

其均值为 0，标准差为 $1/s_i$。从该分布采样得到的沿方向 d_i 的速度，其平均值与 s_i 成反比。

将所有参数合在一起表示，我们可以写成：

$$v\sim N(0,M^{-1}) \tag{18}$$

其中，M 是一个对角矩阵，其对角线元素为 s_i^2；M^{-1} 是速度向量 v的协方差矩阵。使用对角 M，相当于沿参数轴调整速度的标准差。而使用稠密矩阵 M，则相当于更一般地沿更优的方向 d_i 调整速度的标准差（例如，依据协方差矩阵的主成分）。

事实证明，在采样初始速度后，于 HMC/NUTS 中模拟“虚拟动力学”时，类比的动能为：

$$K(v)=v^TMv/2 \tag{19}$$

因此，在 HMC/NUTS 文献中，人们自然倾向于将上述矩阵 M 称为“质量矩阵”。回想一下，在物理学中，速度为标量 v、质量为 m 的粒子的动能为 mv^2_2 。

总而言之，“质量”较大的方向会比质量较小的方向探索得更慢。理想的质量矩阵 M 应近似于后验分布的全局 精度矩阵，也就是协方差矩阵的逆。等价地，理想的 M^{-1} 应近似于后验分布的全局协方差矩阵。

到目前为止，我们假设已经获得了来自后验分布的样本，并且能够手动调整质量矩阵。在实践中，NUTS 算法会在适应阶段为你调整质量矩阵，你只需要选择矩阵的结构，例如对角矩阵或稠密矩阵。对于大型问题，实践中通常使用对角矩阵，因为使用稠密矩阵的计算成本为 $O(D^3)$，其中 D 是 $(\eta,\ \theta)$ 中的参数总数。另一方面，使用对角矩阵的计算成本为 香港六合彩

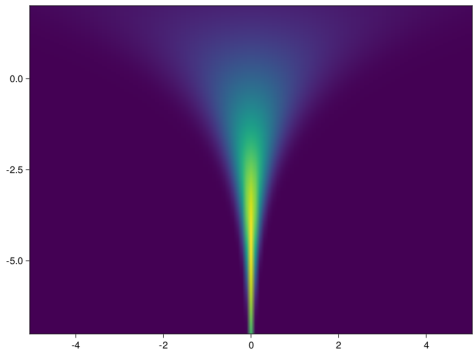

图 16：先验分布的 PDF 热图，纵轴为 logω ，横轴为 η 。

矩阵仅为 $O(D)$。当层次模型中有许多受试者时， D 可能会相当大。

在结束本节之前，需要注意的是，文献通常使用所谓的动量向量 $p = Mv$ 来解释 HMC/NUTS 算法；而在本文中，为了阐释该算法背后的直观原理，我们选择使用更直观的速度向量 v 。这两种参数化方式是等价的，但实现和 HMC/NUTS 研究中通常使用动量参数化方式。当 $v \sim N(0,\ M^{-1})$ 时，相应的动量向量 p 的分布为 $N(0,\ M)$。

4.5.8 可选：层级先验与 NUTS

考虑以下没有观测值的玩具模型：

$$\begin{aligned} \log\omega &\sim \text{Normal}(0, 1.5) \\ \eta &\sim \text{Normal}\left(0, (e^{\log\omega})\right) \end{aligned} \tag{20}$$

这是一个包含 2 个参数（logω，η）的模型，每个参数都有一个先验分布，二者之间还存在某种指数依赖关系：参数 η 的先验标准差随 logω的取值呈指数变化。图 16 展示了参数联合先验分布的概率密度函数（PDF）热图，其中 logω （y 轴）和 η （x 轴）分别对应两个参数。回想一下，NUTS 算法使用固定时间步长的多步轨迹进行虚拟动力学模拟，以处理同一方向上的曲率变化。来看图中沿 x 轴的方向。沿 x 轴的曲率会随虚拟粒子（logω，η）在 y 轴上的位置而变化。logω 的值越低，沿 η 方向的曲率就会呈指数级增大（热图中表现为颜色带更加紧密）。因此，如果尝试使用 NUTS 从该先验分布中采样，可能会出现两种糟糕的情况：

1. 采样器可能会将步长调整得非常小，以便从先验的低值区域进行采样；同时，它会使用大量时间步 T 来有效探索先验的高值区域。在这种情况下，允许的最大时间步数 $2^j - 1$ 往往不足以找到掉头点。这会降低性能，因为每次提议都需要进行大量模型评估，而且可能需要多个步骤才能完全遍历先验的高值区域。

2. 采样器可能会将步长调整到不够小的值，无法从先验的低值区域进行采样。在这种情况下，采样器可能会跳过先验中重要的低值部分，从而导致结果产生潜在偏差。

换句话说，上述先验可能导致 NUTS 采样缓慢且有偏，这显然是糟糕透顶的结果。

请注意，我们当然不会使用 NUTS 从先验中抽样，因为我们可以用更直接、高效的方法，直接从模型中的标准分布抽样。不过，研究先验的概率密度函数（PDF）及其与 NUTS 的相互作用，有助于我们理解在可用数据点较少时，NUTS 将如何与后验分布相互作用。另请注意，出于教学目的，上述模型对标准差采用了显式的对数尺度参数化。实际上，模型可以直接用标准差 ω （而非其对数）来表示，或更一般地，用多变量分布的协方差矩阵 Ω η 来表示。不过，上述示例在这些情况下仍然适用，因为 NUTS 算法的实现会在后台进行对数尺度变换，因此即使原始模型的参数被限制在不同的支撑集上，NUTS 实际上始终对无约束参数进行抽样。因此，我们针对上述模型得出的相同见解，也适用于使用 NUTS 的一般层次模型。

在药物计量学中，受试者个体特异性参数的先验（通常称为个体间变异）的标准差参数（或更一般地说，协方差矩阵）通常采用弱先验，以避免偏倚。这意味着，除非收集到足够的数据以精确识别这些参数的值，否则它们在后验分布中的取值往往具有很大的变异性。这种组合包括：

1. 对协方差矩阵参数使用较弱的先验以避免偏差，2. 数据不足，无法精确识别参数值，3. 层次模型中参数先验（标准差方面）之间的依赖关系，4. NUTS 使用的标准差和协方差参数的对数尺度域变换，以及 5. NUTS 采样器使用的固定步长，

是不幸但非常常见的一系列因素组合，可能导致错误且非常缓慢的推断。

那么解决方案是什么？一种解决方案是按如下方式重新参数化模型：

$$\begin{aligned} \log\omega &\sim \text{Normal}(0, 1.5) \\ \eta\text{std} &\sim \text{Normal}\left(0, 1\right) \\ \eta &= e^{\log\omega} \times \eta\text{std} \end{aligned} \tag{21}$$

这种重新参数化将各参数的先验解耦，并解决了 PDF 热力图中的问题。请注意，这种模型转换不会改变数据生成过程。也就是说，如果你为 η std 和 $\log\omega$ 抽样，ηsimulated 的取值将与从原始模型的先验中模拟得到的值完全相同。不过，后一种参数化对 NUTS 算法更友好。在药物计量学中，以下基于协方差的模型：

$$\begin{aligned} \theta &\sim p(\theta) \\ \Omega &\sim p(\Omega) \\ \eta_i &\sim N(\theta, \Omega) \end{aligned} \tag{22}$$

可转换为：

$$\begin{aligned} \theta &\sim p(\theta) \\ \Omega &\sim p(\Omega) \\ \eta\text{std}_i &\sim N(0, I) \\ \eta_i &= \text{chol}(\Omega) \times \eta\text{std}_i + \theta \end{aligned} \tag{23}$$

其中，chol(Ω) 是协方差矩阵 Ω 的下三角 Cholesky 因子（矩阵的一种广义平方根）。类似地，改用标准差与相关矩阵参数化后，原模型变为：

$$\begin{aligned} \theta &\sim p(\theta) \\ \omega &\sim p(\omega) \\ C &\sim p(C) \\ \eta_i &\sim N(\theta, D_\omega \times C \times D_\omega) \end{aligned} \tag{24}$$

其中，D_ω 是一个对角矩阵，其元素为标准差向量 ω。上述基于相关性的模型可以转换为以下形式，以便将先验解耦：

$$\begin{aligned} \theta &\sim p(\theta) \\ \omega &\sim p(\omega) \\ C &\sim p(C) \\ \eta\text{std}_i &\sim N(0, I) \\ \eta_i &= D_\omega \times \text{chol}(C) \times \eta\text{std}_i + \theta \end{aligned} \tag{25}$$

使用 Pumas 定义这些模型时，即使您以耦合方式编写模型，也会在后台自动执行与上述变换等效的变换。[34]

在结束本节之前，我们指出，在贝叶斯文献中，人口参数与个体特异性参数的先验相互耦合的模型有时称为中心化参数化（CP）模型。使用上述变换解除先验耦合，则通常称为非中心化参数化（NCP）。不过，这些术语可能会造成混淆，因此本文大多避免使用它们。

[34]这种等价关系适用于精确算术，但在计算机上运行计算时，使用的是浮点运算。这意味着，结果可能并不完全相同，具体取决于模型对浮点运算中舍入误差的敏感程度。

4.6 基本汇总统计量

可以通过查看一些汇总统计量来评估马尔可夫链蒙特卡罗链的收敛性。这些统计量包括：

• **有效样本量（ESS）**：在估计后验均值时，马尔可夫链生成的“独立样本数”的近似值。MCMC 样本中的自相关较高会导致每个样本的 ESS 比率较低，这通常是不良指标。

• $\widehat{R}$ **（Rhat）**：潜在尺度缩减因子，用于衡量马尔可夫链是否已混合并可能已经收敛的指标。链混合是指不同链包含来自后验分布相同区域的样本，而不是每条链各自包含来自后验分布不同区域的样本。

• **蒙特卡罗标准误（MCSE）**：标准差除以 ESS，用于衡量后验均值中的估计噪声。

估计后验均值时，有效样本量（ESS）的公式为：

$$\widehat{\eta}_{\text{eff}} = \frac{mn}{1 + \sum_{t=1}^{T} \widehat{\rho}_t}$$

其中，m 是马尔可夫链的数量，n 是每条马尔可夫链的样本总数，$\widehat{\rho}_t$ 是一个自相关估计值。

该公式是对马尔可夫链在估计参数均值时生成的“独立样本数量”的近似值。由于我们无法恢复真实的自相关 ρ，因此只能依赖估计值 $\widehat{\rho}$。对于相同数量的MCMC 样本，链中的自相关越高，有效样本量（ESS）就越低。高自相关可能源于拒绝次数过多，或采样器朝着众数移动时表现出类似优化的行为。这两者都是不收敛的迹象。不过，一般而言，仅凭高自相关本身并不能说明不收敛，因此在解读 ESS 以追溯其可能较低的根本原因时，必须谨慎。

$\widehat{R}$ 的公式为：

$$\widehat{R} = \sqrt{\frac{\widehat{\text{var}}^{+}(\psi \mid y)}{W}}$$

其中，$\widehat{\text{var}}^{+}(\psi \mid y)$是某个参数 ψ的马尔可夫链样本方差。我们通过使用链内方差 W 和链间方差 B 的加权和来计算它：

$$\widehat{\text{var}}^{+}(\psi \mid y) = \frac{n-1}{n}W + \frac{1}{n}B$$

$\widehat{R}$直观来说，如果所有链$\widehat{R}$都完全收敛，则该值为 $R = 1.0$。作为一种经验法则，如果 $R > 1.1$，你就需要担心，因为链可能尚未充分收敛。

4.7 收敛性

4.7.1 不收敛的迹象

MCMC 具有一个有趣的性质，即它会渐近收敛于目标分布。这意味着，如果时间不是有限资源，那么可以保证，无论目标分布的后验分布几何结构如何，MCMC都会给出正确答案。然而，对于所有现实场景而言，时间都是有限资源。不同的MCMC算法（如 NUTS）可以减少达到目标分布收敛性所需的采样（和适应）时间。

在理想情况下，NUTS 采样器会收敛到真实后验分布，并且不会错过任何众数。但是，我们能证明其收敛性吗？遗憾的是，通常很难证明。所有收敛诊断都只是用于检测缺乏收敛性的症状。换句话说，即使所有诊断结果都看起来正常，我们也无法证明采样器没有收敛。

存在一些不收敛的迹象：

• 任何一个矩（例如均值或标准差）都会随时间变化。可通过平稳性检验，将单条链的不同部分相互比较来诊断这一问题。

• 任一矩都对初始参数值敏感。通过使用多条链，将它们的汇总统计量相互比较，可以对此进行诊断。

虽然高自相关并不严格意味着缺乏收敛性，但具有高自相关的采样器需要更多样本，才能达到与具有低自相关的另一采样器相同的效率。因此，低自相关通常更为理想。

4.7.2 收敛性何时重要?

广义而言，我们可以使用两大类模型：

1. 因果模型，有时也称为机理模型。2. 黑箱模型，有时也称为回归模型或机器学习模型。

简而言之，因果模型／机理模型是在相关领域中具有合理性的模型。这意味着：

1. 模型中的所有变量都有其含义。2. 模型中的每种关系都经过周密考量，且尽可能精简，并基于变量子集之间所声称的因果关系。

机制模型的目标是理解所关注的系统。每个模型通常旨在回答有关模型中某些参数的问题。例如，在药代动力学模型中，我们通常有吸收和清除的个体参数。因此，在将模型拟合到数据后，可以回答某个特定个体参数大于或小于某一阈值的概率。另一个常见例子是剂量-反应模型，其中我们通常有一个代表剂量对疾病影响的系数。如果根据后验分布，该参数大于0的概率高于某一阈值，我们就可以声称这种药物是有效的。正确的因果/机制模型应当在插值和外推方面都表现良好。[35]

而在另一端，则是黑箱模型。这类模型通常具有以下特点：

1. 存在许多没有实际意义的中间变量。2. 所有变量之间关系密集，却无法预先明确说明每种关系的具体原因。

这类模型通常称为机器学习模型。可以想象一个采用最高五阶多项式基函数的线性回归模型。仅含线性项的简单线性回归模型，如果声称协变量以线性方式导致响应变量，那么可以说属于第一类因果模型。但到了三阶或四阶多项式基函数时，高阶多项式项及其系数便开始失去意义，模型也变得更加像黑箱。在贝叶斯黑箱模型中，先验分布通常是任意设定的（例如标准高斯分布），仅用于正则化。甚至可以优化先验分布的超参数，以最大化平均后验预测准确性[36]。

有许多方法可以系统地构建复杂的黑箱模型。一些示例如下：

- 使用多项式级数项作为基函数，例如泰勒多项式级数或切比雪夫多项式级数
- 使用傅里叶级数项作为基函数
- 根据需要添加神经元和层，使用深度神经网络
- 使用高斯过程进行非线性回归

这些模型在为（x、y）提供足够的数据，并且模型中包含足够多的项时，可以拟合任意函数 $y = f(x)$，而无需在模型中内置任何因果推理或实际意义。它们纯粹是预测机器，可用于插值，有时也可进行非常有限的外推。模型类拟合任意函数的能力 f，即使用足够大的模型进行拟合，有时被称为通用逼近性质，许多机器学习模型类都具备这一性质（Hornik 等人，1989）。

在实践中，有些模型可能会结合因果/机制模型和黑箱模型的组成部分。例如，可以用因果模型来确定哪些变量依赖于其他变量，但随后

[35]请注意，模型本身并不蕴含因果关系，而是基于科学家对模型及其变量的直觉和理解。换句话说，模型不过是一些工具，可用于表达科学家所设想的量之间的因果关系。

[36]使用未用于训练/推断模型参数的验证数据集

这种依赖关系的函数形式可以用黑箱模型来近似。将机制模型与黑箱模型结合有时称为科学机器学习（Rackauckas 等人，2021）。

我们之所以在这里讨论不同类型的模型，是因为所采用的诊断方法应与分析的目标相一致。如果目标是做出良好的预测，而不考虑模型的可解释性，那么我们可以将模型视为黑箱，主要依赖预测诊断。在这种情况下，只要预测良好，即使模型本身不合理，或者推断过程并不完美，也足够了。例如，在贝叶斯神经网络中，只要后验预测与数据一致，推断后验分布时通常会采用极其粗略的近似（Goan 和 Fookes，2020年；Izmailov 等人，2019）。另一方面，如果分析的目的是理解系统，并了解模型中参数的取值，因为参数本身具有重要意义，那么就应该使用因果／机制模型，并且必须格外谨慎，确保我们从后验分布中正确抽样，且先验不能过强。

4.8 交叉验证与模型选择

在贝叶斯工作流程中，通常会使用模型对样本外数据的预测能力来评估和比较模型，即未用于模型拟合或模型参数推断的数据。用于评估样本外预测准确性的一个常用模型评估指标是所谓的期望对数预测密度（ELPD）。其他常见的模型选择准则还包括各种信息准则（Burnham 和 Anderson，2002），例如广泛适用信息准则（WAIC）。有关 ELPD 以及其他模型评估准则的讨论，请参阅 Vehtari 和 Oja- nen （2012）；Piironen 和 Vehtari（2017）；Gneiting 和 Raftery （2007）。

直观地说，ELPD 是对所有后验样本的预测准确性进行平均后得到的某种度量，并且还对多个预测任务取平均。设 $\mathcal{M}$ 为参数为 (η, θ) 的药物计量学模型，用于描述以下内容的数据生成过程。观测数据 $y \mid x$。期望对数预测密度定义为：

$$\mathrm{ELPD} = \int \log p(\hat{y}|\hat{x}, D, \mathcal{M}) \cdot p_t(\hat{y}|\hat{x}) d\hat{y}$$

其中$\hat{y}$ 是未观测数据，例如未来的数据点，$p_t(\hat{y} \mid \hat{x})$ 是$\hat{y}$的真实数据生成分布（实践中未知），而 $p(\hat{y}|\hat{x}, D, \mathcal{M})$ 是后验预测密度，定义为：

$$p(\hat{y}|\hat{x}, D, \mathcal{M}) = \int p(\hat{y}|\hat{x}, \eta, \theta, \mathcal{M}) \cdot p(\eta, \theta|D, \mathcal{M}) d\theta$$

其中，$p(\eta, \theta|D, \mathcal{M})$ 描述了基于先前观测到的数据D 和模型 $\mathcal{M}$，(η, θ) 的后验分布。由于真实的数据生成分布未知，通常会用观测数据上的经验分布来近似期望对数预测密度。一种此类估计量是逐点对数预测密度（lppd）。

令 (x_i, y_i) 为通过对数据进行任意拆分得到的 i^{th} 个观测（不一定按受试者拆分），分成 D 块，并令 S ,$\eta^{(j)}$ 为 $\theta^{(j)}$从后验分布中抽取的j^{th} 样本（$p, \eta, \theta|D, \mathcal{M}$），其中 $j \in 1, ..., N$。可使用公式 26 计算 lppd。

lppd 的一个缺点是，它不能代表模型对未见数据的预测准确性，因为 (x_i, y_i) 既用于对后验分布进行推断，也用于评估模型的样本外表现。

4.8.1 留 K 交叉验证

交叉验证通过确保在评估 $y_i \mid x_i$ 的样本外表现时，不使用 (x_i, y_i) 对后验分布进行推断，从而解决了这个问题。将数据划分为样本内和样本外子集的最简单方法是留一交叉验证（loo）：在每次外层迭代中，将一个数据点视为样本外，其余数据点视为样本内。留一法对数预测密度（loo-lpd）在公式 27 中定义，其中 D_{-i} 是排除 (x_i, y_i) 后的所有数据，而 $(\eta_{-i}^{(j)}, \theta_{-i}^{(j)})$ 是从后验分布 $p(\eta, \theta|D = D_{-i}, \mathcal{M})$ 中抽取的第 j^{th} 个样本。这可以推广为留 K 交叉验证，其中 (x_i, y_i) 被视为 K 个观测值，例如 K 个受试者或 K 个药物浓度观测值。

$$
\begin{aligned}
\text{lppd} &= \frac{1}{S}\sum_{i=1}^{S}\log p(y=y_i|x=x_i,D,\mathcal{M}) \\
&= \frac{1}{S}\sum_{i=1}^{S}\log\int p(y=y_i|x=x_i,\eta,\theta,\mathcal{M})p(\eta,\theta|D,\mathcal{M})d\theta \\
&\approx \frac{1}{S}\sum_{i=1}^{S}\log\Big(\frac{1}{N}\sum_{j=1}^{N}p(y=y_i|x=x_i,\eta=\eta^{(j)},\theta=\theta^{(j)},\mathcal{M})\Big)
\end{aligned}
\tag{26}
$$

$$
\begin{aligned}
\text{loo-lpd} &= \frac{1}{S}\sum_{i=1}^{S}\log p(y=y_i|x=x_i,D=D_{-i},\mathcal{M}) \\
&= \frac{1}{S}\sum_{i=1}^{S}\log\int p(y=y_i|x=x_i,\eta,\theta,\mathcal{M})\cdot p(\eta,\theta|D=D_{-i},\mathcal{M})d\theta \\
&\approx \frac{1}{S}\sum_{i=1}^{S}\log\Big(\frac{1}{N}\sum_{j=1}^{N}p(y=y_i|x=x_i,\eta=\eta_{-i}^{(j)},\theta=\theta_{-i}^{(j)},\mathcal{M})\Big)
\end{aligned}
\tag{27}
$$

4.8.2 留未来K折交叉验证

处理时间序列数据时，根据模型仅利用既往值进行训练并预测未来值的能力来评估模型，往往更有用。由此产生了另一种交叉验证变体，称为留未来点交叉验证（lfoo），其对应的 lfoo-lpd 定义于公式 28，其中 t 是用于训练/推断的数据点最小数量，$D_{1:i-1}$ 是既往数据，而 $(\eta_{-}^{(j)}{}_{(i:S)}\ \theta_{-}^{(j)}{}_{(i:S)})$ 是从后验分布 $p(\eta,\theta|D=D_{1:i-1},\mathcal{M})$ 中抽取的第 j^{th} 个样本，该后验分布是通过在推断中排除未来数据 $D_{i:S}$ 得到的。

4.8.3 层次模型的交叉验证

在层次模型中进行交叉验证时，有多种方法可以衡量模型的预测能力。例如，在层次药物计量建模中，目标是学习一个群体模型，以便对新患者进行预测，同时学习个体特异性模型，以便在观察到特定受试者过去的响应后，对其未来的响应进行预测。

针对药物。这些模型对于新患者或现有患者的剂量选择和剂量调整非常有用，其目标是在避免毒性的同时最大化治疗效果。

根据所关注的预测任务，可以选择将每个时间观测视为一个数据点，或将每位患者/受试者的全部数据视为一个数据点。如果目标是评估模型预测新受试者响应的能力，则应使用留一受试者交叉验证。或者，如果目标是评估模型预测未来药物浓度或模型所预测的任何其他可观测的时间依赖量的能力，那么对每位受试者留出未来观测值会更合理。这将称为留一观测值或留一未来观测值交叉验证。

在进行交叉验证时，选择留出哪些点会影响预测似然的计算方式：

$$
p(y=y_i|x=x_i,\eta=\theta^{(j)},\theta=\theta^{(j)},\mathcal{M})
$$

在排除受试者时，我们通常会对给定总体参数 θ 的后验分布样本下该受试者的边际似然感兴趣，并将受试者特异性的

$$\begin{aligned}\text{lfoo-lpd} &= \frac{1}{S-t}\sum_{i=t+1}^{S}\log p(y=y_i|x=x_i, D=D_{1:i-1}, \mathcal{M}) \\ &= \frac{1}{S-t}\sum_{i=t+1}^{S}\log\int p(y=y_i|x=x_i,\eta,\theta,\mathcal{M})\cdot p(\eta,\theta|D=D_{1:i-1},\mathcal{M})d\theta \\ &\approx \frac{1}{S-t}\sum_{i=t+1}^{S}\log\frac{1}{N}\sum_{j=1}^{N}p(y=y_i|x=x_i,\eta=\eta^{(j)}_{-(i:S)},\theta=\theta^{(j)}_{-(i:S)},\mathcal{M})\end{aligned} \tag{28}$$

参数 η。或者，对于某些默认值或典型值的个体特异性参数，例如先验分布的众数，也可以使用条件似然。为了边际化个体特异性参数，可以使用 LaplaceI和FOCE等近似积分方法来获得边际似然。另一方面，当在单个受试者中留出观测值时，感兴趣的量通常是：在给定该受试者先前数据的情况下，给定来自个体受试者总体参数和个体特异性参数联合后验分布的每个样本的条件似然。

4.8.4 帕累托平滑重要性抽样交叉验证

评估留一法对数预测密度或 lfoo-lpd 的开销很大，因为需要从 N 或 $N-t$ 个不同的后验分布中抽取样本，例如，为计算留一法对数预测密度，需要从 $p(\eta, \theta|D=D_{-i}, \mathcal{M})$ 中抽样。通常使用马尔可夫链蒙特卡罗来完成这项工作，例如 NUTS 算法。尽管近来取得了进展，但当参数数量较多，且后验分布沿一个或多个维度的曲率不均匀时，计算开销仍然很大。

克服这一困难的一种方法是用于留一交叉验证（PSIS-LOO-CV）的帕累托平滑重要性抽样方法（Vehtari 等人，2015）。在 PSIS-LOO-CV 中，只需在完整数据上运行一次马尔可夫链蒙特卡罗。随后在交叉验证的每次外层迭代中重复使用相同的样本，但采用不同的体重。通过重要性抽样（IS）比较留出一个数据点时的似然与完整数据集的似然，从而确定这些体重。然后将原始重要性体重拟合到广义帕累托分布，以进行平滑处理。平滑后的体重随后可用于估计每个数据点对期望对数预测密度的贡献。

除了能够近似期望对数预测密度外，PSIS-LOO-CV 还提供了一项有用的诊断指标：对每个数据点进行留一处理时，拟合原始权重所得帕累托分布的形状参数。移除后导致形状参数较大的数据点，比导致形状参数较小的数据点影响更大。对于影响很大的数据点，若帕累托分布的形状参数高于 0.7，则该点对期望对数预测密度的贡献可视为不可靠。在这种情况下，建议移除该影响点后，从后验分布中重新抽样。

5 示例模型

[{"id":0,"output":"清单 29、30、31 和 32 是药物计量学中一些常见模型的示例。"}]

```
@model begin
    @param begin
        tvcl ~ LogNormal(log(10), 0.25) # CL
        tvq  ~ LogNormal(log(15), 0.5)   # Q
        tvvc ~ LogNormal(log(35), 0.25) # V1
        tvvp ~ LogNormal(log(105), 0.5) # V2
        tvka ~ LogNormal(log(2.5), 1)   # ka
        σ    ~ truncated(Cauchy(), 0, Inf) # sigma
    end
    @pre begin
        CL = tvcl
        Vc = tvvc
        Q  = tvq
        Vp = tvvp
        Ka = tvka
    end
    @dynamics begin
        Depot'      = -Ka * Depot
        Central'    =  Ka * Depot  - (CL+Q)/Vc * Central + Q/Vp * Peripheral
        Peripheral' =                    Q/Vc * Central - Q/Vp * Peripheral
    end
    @derived begin
        cp  := @. Central/Vc
        conc ~ @. LogNormal(log(cp), σ)
    end
end
```

清单 29：单个受试者药代动力学模型

```
@model begin
    @param begin
      tvcl ~ LogNormal(log(10), 0.25) # CL
      tvq  ~ LogNormal(log(15), 0.5)   # Q
      tvvc ~ LogNormal(log(35), 0.25) # V1
      tvvp ~ LogNormal(log(105), 0.5) # V2
      tvka ~ LogNormal(log(2.5), 1)   # ka
```

代码清单 30：群体药代动力学模型

```
@model begin
    @param begin
        θKa   ~ LogNormal(log(0.8), 1)
        θKe   ~ LogNormal(log(0.15), 1)
        θVd   ~ LogNormal(log(100), 1)
        θn    ~ LogNormal(log(2.0), 1)
        θδ    ~ LogNormal(log(0.20), 1)
        θc    ~ LogNormal(log(7.0), 1)
        θEC50 ~ LogNormal(log(0.12), 1)

        ω²Ka   ~ truncated(Normal(0.25, 1), 0.0, Inf)
        ω²Ke   ~ truncated(Normal(0.25, 1), 0.0, Inf)
        ω²Vd   ~ truncated(Normal(0.25, 1), 0.0, Inf)
        ω²n    ~ truncated(Normal(0.25, 1), 0.0, Inf)
        ω²δ    ~ truncated(Normal(0.25, 1), 0.0, Inf)
        ω²c    ~ truncated(Normal(0.25, 1), 0.0, Inf)
        ω²EC50 ~ truncated(Normal(0.25, 1), 0.0, Inf)

        σ²PK ~ truncated(Normal(0.04, 1), 0.0, Inf)
        σ²PD ~ truncated(Normal(0.04, 1), 0.0, Inf)
    end
    @random begin
        ηKa   ~ LogNormal(0.0, sqrt(ω²Ka))
        ηKe   ~ LogNormal(0.0, sqrt(ω²Ke))
        ηVd   ~ LogNormal(0.0, sqrt(ω²Vd))
        ηn    ~ LogNormal(0.0, sqrt(ω²n))
        ηδ    ~ LogNormal(0.0, sqrt(ω²δ))
        ηc    ~ LogNormal(0.0, sqrt(ω²c))
        ηEC50 ~ LogNormal(0.0, sqrt(ω²EC50))
    end
    @pre begin
        # constants
        p = 100.0
        d = 0.001
        e = 1e-7
        s = 20000.0

        Ka   = θKa * ηKa
        Ke   = θKe * ηKe
        Vd   = θVd * ηVd
        n    = θn * ηn
        δ    = θδ * ηδ
        c    = θc * ηc
        EC50 = θEC50 * ηEC50
    end
    @init begin
        T = (c * δ) / (p * e)
        I = max(0, s * e * p - d * c * δ) / (p * δ * e)
        W = max(0, s * e * p - d * c * δ) / (c * δ * e)
    end
    @vars begin
        # short-hand notation for A / Vd
        conc = A / Vd
    end
    @dynamics begin
        X' = -Ka * X
        A' = Ka * X - Ke * A
        T' = s - T * (e * W + d)
        I' = e * W * T - δ * I
        W' = p / ((conc / EC50 + 1e-100)^n + 1) * I - c * W
    end
    @derived begin
        log10W := @. log10(W)
        yPK    ~ @. truncated(Normal(conc, sqrt(σ²PK)), 0, Inf)
        yPD    ~ @. truncated(Normal(log10W, sqrt(σ²PD)), 0, Inf)
    end
end
```

列表 31：丙型肝炎病毒模型

```
@model begin
    @param begin
        λ₁ ~ LogNormal(0.0, 2.0) # basal hazard
        β  ~ LogNormal(0.0, 2.0) # fixed effect DOSE
        ω  ~ LogNormal(0.0, 2.0) # inter-subject variability
    end
    @random begin
        η ~ LogNormal(0.0, ω)
    end
    @covariates DOSE
    @pre begin
        _λ₁ =  λ₁ * η       # basal hazard with inter-subject variability
        _λ₀ = _λ₁ * β^DOSE # total hazard
    end
    @vars begin
        λ = _λ₀
    end
    @dynamics begin
        Λ' = λ
    end
    @derived begin
        dv ~ @. TimeToEvent(λ, Λ)
    end
end
```

清单 32：生存时间模型

6 结论

在本教程中，我们介绍了使用 Pumas 进行贝叶斯分析的完整工作流程。我们尽可能采用直观优先的方法，讲解了所有语法和相关理论，并提供了大量交叉链接。如果您是现有的 Pumas 用户并有其他问题，可以通过 Pumas Discourse 平台（discourse.pumas.ai）联系我们。您还可以在 Pumas 教程网站（tutorials.pumas.ai）找到更有针对性的教程和完整脚本，以便继续学习。

读完本文后，如果您想进一步阅读和学习贝叶斯统计，以下是一些可用于继续学习的优秀资源：

1. 贝叶斯数据分析（Gelman 等人，2013b）2. 《统计学再思考：附 R 和 Stan 示例的贝叶斯课程》（McElreath，2020年），3. 回归及其他故事（Gelman 等人，2020a），4. 使用回归和多层/层次模型进行数据分析（Gelman 和 Hill，2006），5. 概率机器学习：导论（Murphy，2022年）6. 《概率论：科学的逻辑》（Jaynes，2003）

7 致谢

我们谨向所有审阅本作品早期草稿并提出宝贵反馈的审阅者致谢。特别感谢**HadenBunn** （Pumas-AI Inc.）、**Joga Gobburu** （Pumas-AI Inc.及马里兰大学巴尔的摩分校）、**Yoni Nazarathy** （Pumas-AI Inc.及昆士兰大学）、**Vaibhav Dixit** （Pumas-AI Inc.）、**Russell Tsuchida**（CSIRO Data61）、**Anas-tasios Panagiotelis** （悉尼大学）以及**Mutaz Jaber** （Gilead Sciences Inc.）。

参考文献

Beal S、Boeckmann L、Bauer R 等（2009）NONMEM 用户指南。（1989–2009）

Betancourt M (2017) 哈密顿蒙特卡罗的概念性介绍。arXiv:1701.02434

Bingham E, ChenJP, Jankowiak M 等（2018）Pyro：深度通用概率编程。https://doi.org/10.48550/ARXIV.1810.09538，网址 https://arxiv.org/abs/1810.09538

BrooksS,GelmanA, JonesG, 等(2011)马尔可夫链蒙特卡罗手- 册。CRC出版社

BurnhamKP,Anderson DR (2002) 模型选择与多模型推断，第2版。施普林格，纽约，NY Carpenter B, HoffmanMD, BrubakerM, 等（2015）Stan 数学库：C语言中的反向模式自动微分++。arXiv：150907164 [cs]网址 http://arxiv.org/abs/1509.07164，arXiv：1509.07164 Delyon B,LavielleM, MoulinesE（1999）随机逼近版本EM算法的收敛性。《统计年鉴》27(1):94–128。网址 http://www.jstor.org/stable/120120，出版商：数理统计学会Elmokadem A, Zhang Y,Knab T, 等（2023）使用r/Stan/Torsten和Julia/SciML/Turing.jl进行贝叶斯生理药代动力学建模。CPT：药理计量学与系统药理学 https://doi.org/10.1002/psp4.12926，网址 https://doi.org/10.1002/psp4.12926Ge H, Xu K, Ghahramani Z（2018）Turing：一种用于灵活概率推断的语言。载于：Storkey A, Perez-Cruz F（编）第二十一届人工智能与统计国际会议论文集，《机器学习研究论文集》，第84卷。PMLR，第1682–1690页，网址 https://proceedings.mlr.press/v84/ge18b.html。

Gelman A(2005) 方差分析——为何它比以往任何时候都更加重要。《统计年刊》33(1)。https://doi.org/10.1214/009053604000001048

Gelman A,Hill J (2006)《回归与多层次／层级模型数据分析》. 剑桥大学出版社, https://doi.org/10.1017/cbo9780511790942,网址 https://doi.org/10.1017/cbo9780511790942

GelmanA、Carlin JB、SternHS 等（2013a）《贝叶斯数据分析》。Chapman and Hal-l/CRC

Gelman A、Carlin JB、Stern HS 等（2013b）《贝叶斯数据分析》。Chapman and Hall/CRC, https://doi.org/10.1201/b16018，网址 https://doi.org/10.1201/b16018Gelman A、Hill J、Vehtari A（2020a）《回归及其他故事》。剑桥大学出版社, https://doi.org/10.1017/9781139161879，网址 https://doi.org/10.1017/9781139161879Gelman A、Vehtari A、Simpson D 等（2020b）贝叶斯工作流程。arXiv:2011.01808Geweke J（1991）评估基于抽样的方法计算后验矩的准确性。https://doi.org/10.21034/sr.148，网址 https://doi.org/10.21034/sr.148Gneiting T、Raftery AE（2007）严格恰当的评分规则、预测与估计。《美国统计学会杂志》102(477):359–378。https://doi.org/10.1198/016214506000001437，网址 https://doi.org/10.1198/016214506000001437, https://arxiv.org/abs/https://doi.org/10.1198/0162145060000001437Goan E、Fookes C（2020）贝叶斯神经网络：导论与综述。《应用贝叶斯数据科学案例研究：CIRM Jean-Morlet讲席，2018年秋季》，1，(2020)45-87, https://doi.org/10.1007/978-3-030-42553-1 3, arXiv:2006.12024Goudie RJ、Turner RM、De Angelis D 等（2020）Multibugs：BUGS建模框架的并行实现，可加快贝叶斯推断。统计软件期刊 95

Heidelberger P、Welch PD（1983）存在初始瞬态时的模拟运行长度控制。《运筹学》31(6):1109–1144。https://doi.org/10.1287/opre.31.6.1109，网址https://doi.org/10.1287/opre.31.6.1109Hoffman MD、Gelman A（2014）无调采样器：在哈密顿蒙特卡洛中自适应设定路径长度。《机器学习研究杂志》15(1):1593–1623。网址http://arxiv.org/abs/1111.4246Hornik K、Stinchcombe M、White H（1989）多层前馈网络是通用逼近器。《神经网络》2(5):359–366。https://doi.org/10.1016/0893-6080(89)90020-8Izmailov P、Maddox WJ、Kirichenko P 等（2019）贝叶斯深度学习的子空间推断。arXiv:1907.07504Jaynes ET（2003）《概率论：科学的逻辑》。剑桥大学出版社，网址https://bayes.wustl.edu/etj/prob/book.pdfKallioinen N、Paananen T、B¨urknerPC 等（2021）通过幂缩放检测和诊断先验与似然敏感性。arXiv:2107.14054Kuhn E、Lavielle M（2005）非线性混合效应模型中的最大似然估计。《计算统计与数据分析》49(4):1020–1038 Lee JY、Gobburu JV（2011）用于指导帕金森病早期药物开发的贝叶斯定量疾病–药物–试验模型。《AAPS期刊》13(4):508–518。https://doi.org/10.1208/s12248-011-9293-6，网址https://doi.org/10.1208/s12248-011-9293-6Lunn DJ、Best N、Thomas A 等（2002）《药代动力学与药效动力学杂志》29(3):271–307。https://doi.org/10.1023/a:1020206907668，网址https://doi.org/10.1023/a:1020206907668Margossian CC、Zhang Y、Gillespie WR（2022）使用 Stan 和 Torsten 进行灵活高效的贝叶斯药物计量学建模（第一部分）。《CPT：药物计量学与系统药理学》p psp4.12812。https://doi.org/10.1002/psp4.12812McDade E, Cummings JL, Dhadda S, 等（2022年）仑卡奈单抗用于早期阿尔茨海默病患者：二期概念验证研究随机阶段及开放标签延长期中生物标志物、认知和临床效应的详细结果。《阿尔茨海默病研究与治疗》14(1)。https://doi.org/10.1186/s13195-022-01124-2，网址https://doi.org/10.1186/s13195-022-01124-2McElreath R（2020年）《统计学再思考：附 R 和 Stan 示例的贝叶斯课程》。CRC出版社，网址https://xcelab.net/rm/statistical-rethinking/Murphy KP（2022年）《概率机器学习：导论》。麻省理工学院出版社，网址probml.aiNeal RM（2011）使用哈密顿动力学进行马尔可夫链蒙特卡罗。载于：Brooks S、Gelman A、Jones GL 等（编）《马尔可夫链蒙特卡罗手册》。Chapman and Hall/CRC。Piironen J、Vehtari A（2017）用于模型选择的贝叶斯预测方法比较。《统计与计算》27：711–735。Rackauckas C、Nie Q（2017）DifferentialEquations.jl——用于求解 Julia 微分方程的高性能、功能丰富的生态系统。《开放研究软件杂志》5(1)。https://doi.org/10.5334/jors.151，网址https://app.dimensions.ai/details/publication/pub.1085583163andhttp://openresearchsoftware.metajnl.com/articles/10.5334/jors.151/galley/245/download/，于 2019/05/05 从 https://app.dimensions.ai 导出。Rackauckas C、Ma Y、Noack A 等（2020年）借助 Pumas（一种高性能药物建模与模拟平台）加速预测性医疗分析。bioRxivhttps://doi.org/10.1101/2020.11.28.402297Rackauckas C、Ma Y、Martensen J 等（2021）面向科学机器学习的通用微分方程。https://doi.org/10.48550/arXiv.2001.04385Salvatier J、Wiecki T、Fonnesbeck C（2015）在 Python 中使用概率编程

PyMC。 https://doi.org/10.48550/ARXIV.1507.08050, URL https://arxiv.org/abs/1507.08050

Simpson DP、Rue H、Martins TG 等（2014）惩罚模型组件复杂度：构建先验的一种有原则且实用的方法。arXiv:1403.4630

Talts S、Betancourt M、Simpson D 等（2020年）使用基于模拟的校准验证贝叶斯推断算法。arXiv https://doi.org/10.48550/arXiv.1804.06788

Vehtari A, Ojanen J (2012) 贝叶斯预测方法在模型评估、选择和比较中的综述。StatisticsSurveys 6(none).https://doi.org/10.1214/12-ss102，网址 https://doi.org/10.1214/12-ss102

Vehtari A、Gelman A、Gabry J（2015）使用留一交叉验证和 WAIC 进行实用的贝叶斯模型评估。统计学与计算，2017年，第27卷，第5期，第1413-1432页，https://doi.org/10.1007/s11222-016-9696-4

’ arXiv:1507.04544

Vehtari A、Gelman A、Simpson D 等（2019）秩正态化、折叠与局部化：^用于评估马尔可夫链蒙特卡罗 R 收敛性的改进方法。https://doi.org/10.1214/20-BA1221, arXiv:1903.08008

Wakefield J、Aarons L、Racine-Poon A（1999）群体药代动力学/药效动力学模型的贝叶斯方法。载于：Gatsonis C、Kass RE、Carlin B 等（编）《贝叶斯统计案例研究》。施普林格，纽约，纽约州纽约市，第 205–265 页

Wang Y（2007）各种 NONMEM 估计方法的推导。《药代动力学与药效动力学杂志》34(5)：575–593。https://doi.org/10.1007/s10928-007-9060-6，网址 http://link.springer.com/10.1007/s10928-007-9060-6

Xu K、Ge H、Tebbutt W 等（2020年）：Advancedhmc.jl：高级 HMC 算法的稳健、模块化且高效实现。载于《近似贝叶斯推断进展研讨会》，PMLR，第 1–10 页。