上海财经大学
Shanghai University of Finance and Economics

Collaborative Translation and Continuous Updates: Advancing the Stan Chinese Documentation

Presenter: Ziyuan Zhang

Shanghai University of Finance and Economics
& Boston University

Collaborators

- Joint work with:
- Junzhu Li (16645528818@163.com)
- Shanghai University of Finance and Economics (SUFE)
- Master student of Applied Statistics

- Supervisor:
- Yixuan Qiu (qiuyixuan@sufe.edu.cn)
- Associate Professor, SUFE
- Admin of Capital of Statistics (COS, https://cosx.org/)

- Special thanks to:
- Yi Zhang (yz@yizh.org)
- Metrum Research Group
- Former member of Stan Governing Body (SGB)

Why do the Translation?

Most spoken languages, *Ethnologue*, 2024[4]

Language	First-language (L1) speakers	Second-language (L2) speakers	Total speakers (L1+L2)
English (excl. creole languages)	380 million	1.135 billion	1.515 billion
Mandarin Chinese (incl. Standard Chinese, but excl. other varieties)	941 million	199 million	1.140 billion
Hindi (excl. Urdu)	345 million	264 million	609 million
Spanish (excl. creole languages)	486 million	74 million	560 million
Modern Standard Arabic (excl. dialects)	—[a]	—	332 million

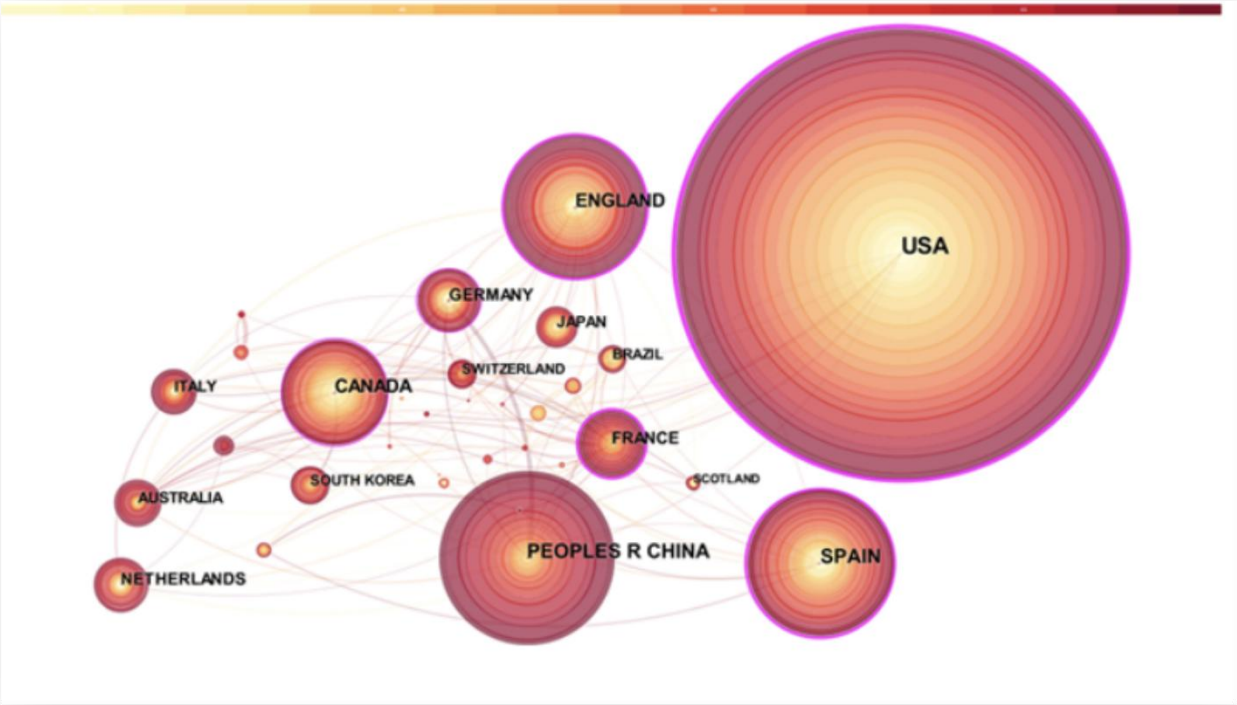
Chinese Enjoys Large Potential User base:

- The number of Chinese speakers is second only to English.
- In China, most graduate-level courses are taught in Chinese.
- Bilingual individuals prefer using their native language under similar conditions.

Ethnologue: https://www.ethnologue.com/

Why do the Translation?

- User Guide is Essential to Stan's Success
- Lack of Chinese Bayesian Research Tools
- Effective Promotion
  - eg.Plants vs. Zombies, Python, R
- Enhance the Diversity of Stan Community



Xiao, Z.Z., Nuerbuli, Liu, H.Y. (2020). Bibliometric analysis of the research status and development trends of Bayesian networks. Computer Science and Applications, 10(3), 493-504. https://doi.org/10.12677/CSA.2020.103052

Timeline

01

- 2022.3: Initiation

SGB reached out to Chinese stat community[1], discussing how to promote Stan in non-English-speaking countries.

02

- 2023.2 Follow-up

Dr.Yixuan Qiu answered the call by exploring the initial translation

03

- 2023.6 Initial version: SUFE-Bayes

04

- 2023.7-Now Refinement

[1]Capital of Statistics (COS):an online community on statistics and data science in China

Latest Version

- Successfully completed the first round of proofreading
- Basic text can be read fluently in Chinese

Files

master

Go to file t

- gaussian-processes.qmd
- hyperspherical-models.qmd
- index.qmd
- latent-discrete.qmd
- matrices-arrays.qmd
- measurement-error.qmd
- missing-data.qmd
- multi-indexing.qmd
- odes.qmd
- one-dimensional-integrals.qmd
- parallelization.qmd
- posterior-prediction.qmd
- posterior-predictive-checks.qmd
- poststratification.qmd
- problematic-posteriors.qmd
- proportionality-constants.qmd
- references.qmd
- regression.qmd
- reparameterization.qmd
- simulation-based-calibration.q...
- sparse-ragged.qmd
- style-guide.qmd
- survival.qmd
- time-series.qmd
- truncation-censoring.qmd
- user-functions.qmd
- using-stanc.qmd
- theming

stan-docs / src / stan-users-guide / problematic-posteriors.qmd

yixuan Merge branch 'master' of https://github.com/stan-dev/docs into sync_u...

Code Blame 1047 lines (759 loc) · 58.3 KB

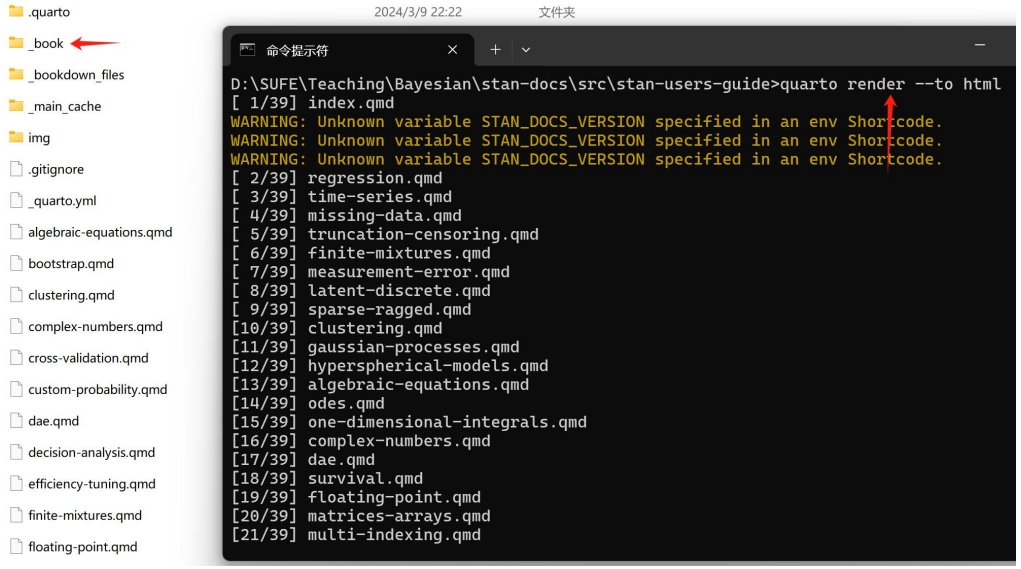
```
---
pagetitle: Problematic Posteriors
---

# Problematic Posteriors  {#problematic-posteriors.chapter}

# 有问题的后验分布 {#problematic-posteriors.chapter--cn}

本节译者：马汉腾、谈英文
本节校审：张梓源

Mathematically speaking, with a proper posterior, one can do Bayesian
inference and that's that. There is not even a need to require a
finite variance or even a finite mean---all that's needed is a finite
integral.  Nevertheless, modeling is a tricky business and even
experienced modelers sometimes code models that lead to improper
priors.  Furthermore, some posteriors are mathematically sound, but
ill-behaved in practice.  This chapter discusses issues in models that
create problematic posterior inferences, either in general for
Bayesian inference or in practice for Stan.

从数学角度来说，只要有一个适当的后验分布，就可以进行贝叶斯推断。这甚至不要求方差或均值有限，只需要一个有限的积分。然而，建模是一项棘手的任务，即使
Stan 的实践情况。

## Collinearity of predictors in regressions {#collinearity.section}

## 回归中预测变量的共线性 {#collinearity.section--cn}

This section discusses problems related to the classical notion of
identifiability, which lead to ridges in the posterior density and
wreak havoc with both sampling and inference.

本节讨论与经典可辨识性概念相关的问题，这些问题会导致后验密度中出现岭(ridge)，给采样和推断带来严重问题。

### Examples of collinearity {-}

### 共线性示例 {-}

#### Redundant intercepts {-}

#### 冗余截距项 {-}

The first example of collinearity is an artificial example involving
redundant intercept parameters.^[This example was raised by Richard McElreath on the Stan users group in a query about th
```

How to Access the Latest Version

Option 1: View directly on GitHub

- Search SUFE-Bayes on GitHub
- Path: stan-docs/src/stan-users-guide
- Click on any .qmd file

How to Access the Latest Version

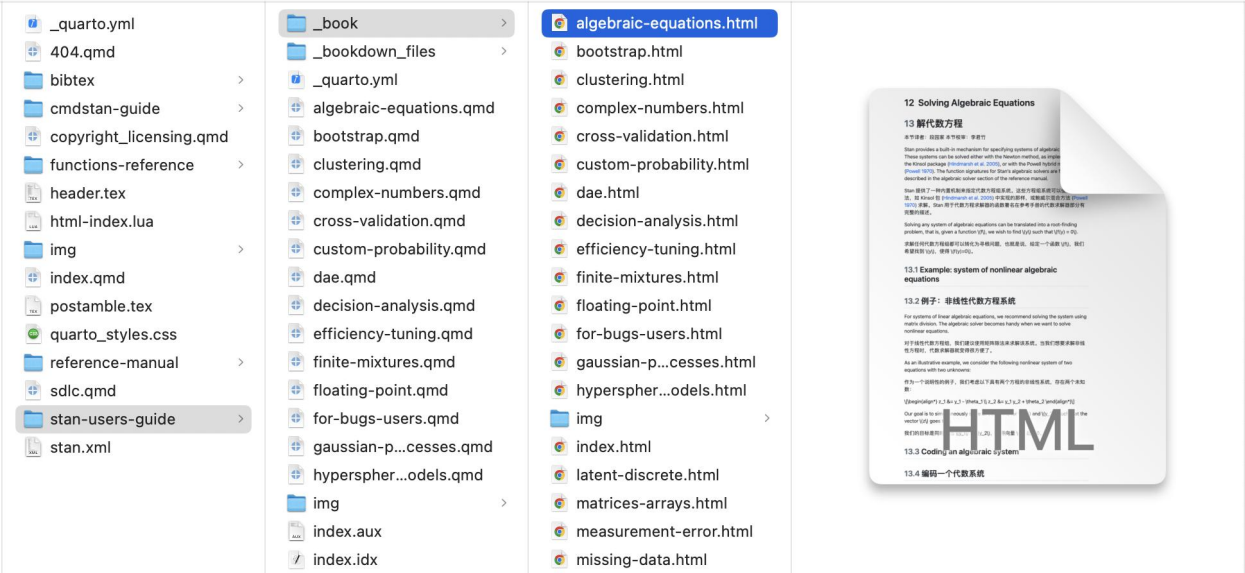


Option 2: Render the Documentation Locally

- Search SUFE-Bayes on GitHub.
- Download the source files.
- Install Quarto on your local machine.

- Enter the source directory in the command line.
- Enter the command: quarto render --to html.

- All HTML files are located in the _book folder.



Quaro installation:https://quarto.org/docs/download/

1 Regression Models

2 回归模型

本节译者：于曙光、王才兴、王超 本章校审：张梓源

Stan supports regression models from simple linear regressions to multilevel generalized linear models.

Stan 支持从简单线性回归到多层次广义线性模型的回归模型。

2.1 Linear regression

2.2 线性回归

The simplest linear regression model is the following, with a single predictor and a slope and intercept coefficient, and normally distributed noise. This model can be written using standard regression notation as

最简单的线性回归模型包含自变量、斜率、截距以及服从正态分布的噪声。使用标准回归符号，可以将该模型写成如下形式：

$$y_n = \alpha + \beta x_n + \epsilon_n \quad \text{where} \quad \epsilon_n \sim \text{normal}(0, \sigma).$$

This is equivalent to the following sampling involving the residual,

这相当于涉及残差的以下采样，

$$y_n - (\alpha + \beta X_n) \sim \text{normal}(0, \sigma),$$

and reducing still further, to

并可以进一步简化为如下形式：

$$y_n \sim \text{normal}(\alpha + \beta X_n, \sigma).$$

This latter form of the model is coded in Stan as follows.

Table of contents

Problems and Solutions

Before Translation Main Challenges:

- Needs to have a solid background of Bayesian analysis,etc.

Translator:
Students taking Bayesian Statistics course

Problems of Initial Version:

- Machine translation
- Not unified

Refinement

Challenges in Proofreading: Chapter Allocation

How?

Advantages:

- Efficient
- Convenient

Evenly distribute all chapters? → The chapter lengths are inconsistent

Translate separately, chapter by chapter → Difficult to assess the workload

Two people may working on the same chapter

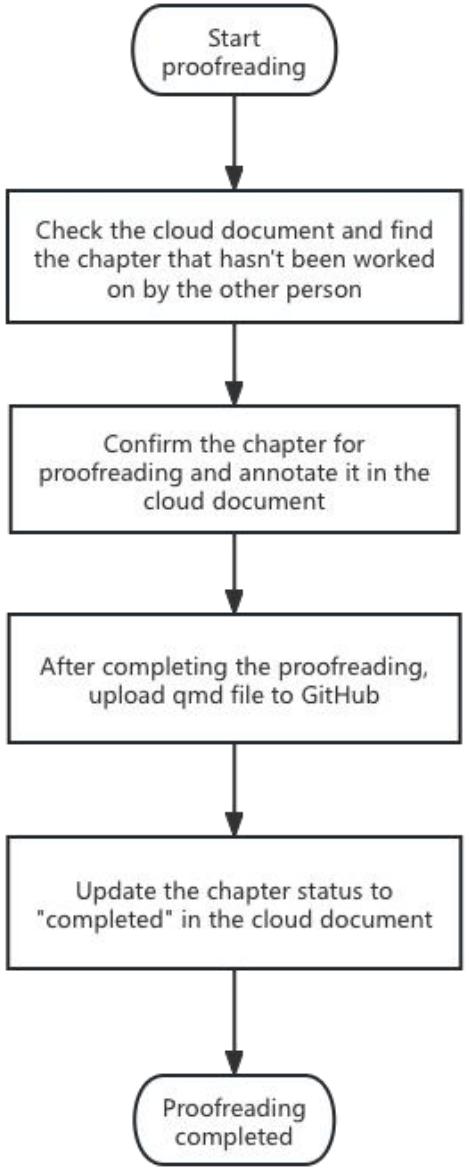
Challenges in Proofreading: Version Synchronization

Version Conflict happens when:

- Same text edited at once
- File uploads overwrite others
- Local copy out of sync with GitHub
- System differences cause merge issues
- ...

Solution: Workfolw

- Create a Lark progress table
- Check status before proofreading
- Proofreading
- Sync to GitHub
- Update status



Translation：Simple?

- Machine translation
- Time-consuming, but no skill needed.

Poor-quality translations appear everywhere

Translation：Challenging

- Choose appropriate words and technical terms
- Difficult to translate between different language systems.
  - eg.German vs English, Japanese vs Chinese, and German vs Chinese
- Consider the context
- Low error tolerance for technical documentation
- Formula and equation

Solution：Glossary

1	Computing One Dimensional Integrals					
2		单词	采用翻译	其他	含义	备注
3		the normalizing constant	标准化常量	归一化常数	先验概率	
4		argument	实参/实际参数		调用的参数	
5		parameters	形参/形式参数		函数中的参数	
6		function signature	函数签名			
7		call	调用		调用子程序	
8		Integration	积分		泛指	
9		integral	积分		特指	
10		integrator	积分器			
11		evaluate	求解		求（方程式，公式，函数）的数值	
12		norm	范数			
13		machine epsilon	机械极小值		舍入误差	
14		quadrature	求积	正交；求积；弦		
15		numerator	式子（意译）	分子	上下文中是定积分式子	
16						
17						
18	2 回归模型					
19		multilevel generalized linear models	多层次广义线性模型			
20	2.2 线性回归					
21		predictor	自变量			
22		outcome	因变量			
23		sampling	采样			（有疑问，感觉改成“抽样”会不会更好一点）
24		overloaded	重载			
25		improper priors	不合适的先验			
26	2.10 逻辑回归和概率回归					
27		link function	链接函数			（这里参考周志华老师的西瓜书里翻译成“联系函数”是不是更好呢）
28		Logit Parameterization				（这里是要翻译成“逻辑参数化”还是“对数几率参数化”呢）
29	2.18 分层logistic回归					
30		pooling	汇集			
31	2.26 分层模型的多元先验					

Glossary of Key Terms

- Original and alt translations
- Definitions
- Debates on contentious words

Application for Grant

Previously Funded Programs

Column visibility | Copy | CSV | Excel | PDF | Print

Show 10 entries Search: stan

Project	Proposal Title	Amount	Year
Stan	Non-English versions of Stan user guide and language reference	$10,000.00	2023
Stan	Commumity building through StanCon 2023	$10,000.00	2023
Stan	Update mc-stan.org	$4,800.00	2021
Taskflow	Standard GPU Algorithms with Task Graph Parallelism	$5,000.00	2021

Application:

- Dr. Zhang reported this work to SGB.
- Applying for the Small Development Grants from NumFocus (https://numfocus.org/programs/small-development-grants)

Approval:

- The end of 2023
- The proposal "Non-English versions of Stan user guide and language reference" was approved for full funding in the amount of $10,000.

Uses:

- Reimburse previous contributors of the project
- Recruit translators to refine the translation

More than Translation: Building Community

- Provid a platform for Stan communication in Chinese.
- Train new Stan users.
- Enhance interaction between the Chinese and English Stan communities, based on the same programming language.
- Contribute to the development of Stan.

Development Prospects

Continuous Updates and Future Plans

- Synchronous updates with the official Stan documentation
- Future launch of the official documentation
- Expansion of Chinese community resources (such as forums, tutorials, etc.)

Prospects for the Stan Chinese Community Development

- Call for more volunteer participation
- Plans to build an active Chinese user community
- Guidance from bilingual experts in statistics and Bayesian fields

Acknowledgements

Capital of Statistics(COS)
https://cosx.org/

Stan
https://mc-stan.org/

School of Statistics and Management, SUFE
https://ssm.sufe.edu.cn/

Acknowledgements

sufe-bayes

Overview | Repositories 3 | Projects | Packages | Teams | People 31

sufe-bayes

Tanzs
TanTzs

Zhi Yang
tobi0520

Caixing Wang
WangCaixing-96

wangchao-afk

Xin Guo
XinGuo-Bill

Jingyuan Yang
YangJingyuanSufe

Yixuan Qiu
yixuan

yuanjiafirst

Yuanying Chen
Yuanying1215

Shuguang Yu
yusg-sufe

zhouyw1217

Xin Guo
XinGuo-Bill

nidimajiabie

ZiliangShen
pinzek

QiaolingLi
QiaolingLi

xiangyu hu
rapostate

书宁
ShaneneCheng

suiyangsoo

sutianyuan

Songhua Tan
Tansonghua-sufe

Tanzs
TanTzs

Zhi Yang
tobi0520

Caixing Wang
WangCaixing-96

wangchao-afk

bobothebest

dc-wangjn

Liying
duducaida

Lost_drunk_bird
ForestKnowsWhy

hanteng_ma
hanteng-ma

HiWangQinYi

Shizhe Hong
hsztsuna

JiaoRY

JosieYuuu

lalalahhh888

LinyuchangSufe

Zhen Ma
Mazing2023

Q&A