



# Ứng dụng Bayes trong nghiên cứu KHXX

## 🌀 Buổi 1 🌀

**Đơn vị tổ chức:** Trường Đại học Hà Nội (HANU)  
Viện Nghiên cứu Cao cấp về Toán (VIASM)

**Giảng Viên:** TS Nguyễn Minh Hoàng (ISR, ĐH Phenikaa)

**Địa Điểm:** Trường Đại học Hà Nội (HANU)

**Thời Gian:** 13:30-17:00, ngày 3 tháng 11 năm 2023



– Was Kingfisher eating Fish, or... when Fish entered Kingfisher's stomach, would it also mean that Fish was eating Kingfisher?

[...]

And if there is no answer, perhaps he has to rely on empirical data: swallowing the talking partner!

In "Philosophical Chat"; *The Kingfisher Story Collection* [1]



## Nguyễn Minh Hoàng (Minh-Hoang Nguyen)

- Sơ lược bản thân: <https://sites.google.com/view/nguyen-minh-hoang/home>
- ORCID: <https://orcid.org/my-orcid?orcid=0000-0002-7520-3844>
- Google Scholar: <https://scholar.google.com/citations?user=9VbhoGYAAAAJ&hl=vi>
- Web of Science: <https://www.webofscience.com/wos/author/record/GQO-9183-2022>
- Scopus: <https://www.scopus.com/authid/detail.uri?authorId=57205713877>
- PhilPapers: <https://philpeople.org/profiles/minh-hoang-nguyen>
- ResearchGate: <https://www.researchgate.net/profile/Minh-Hoang-Nguyen-4>



# Nội dung khóa học

- **Buổi 1**


**Mục tiêu:** Người tham gia sử dụng được phần mềm

1. Sơ lược về phương pháp phân tích Bayesian Mindsponge Framework (BMF)
2. Thống kê Bayes
3. Phần mềm phân tích bayesvl
4. Ví dụ và thực hành
5. Hỏi và đáp

- **Buổi 2 và 3**

**Mục tiêu:** Tập thể hoàn hành bài nghiên cứu sử dụng BMF

1. Phương pháp phân tích Bayesian Mindsponge Framework (BMF)
2. Lý thuyết Mindsponge và các lập luận logic cơ bản
3. Tập thể triển khai bài nghiên cứu



# Buổi 1: Các nội dung chính

## Phần 1

- Sơ lược về phương pháp phân tích BMF
- Mô hình thống kê là gì?
- Thống kê Bayes là gì?
- Vì sao Thống kê Bayes lại quan trọng?
- Hỏi và đáp

## Giải lao

## Phần 2

- Ngôn ngữ R và phần mềm bayesvl
- Các bước phân tích Bayes
- Ví dụ và thực hành



# **Phương pháp phân tích Bayesian Mindsponge Framework**



- BMF là khung phân tích kết hợp sức mạnh của **Lý thuyết Mindsponge** và **phân tích Bayes** để tạo điều kiện thuận lợi cho việc điều tra các vấn đề xã hội, tâm lý và hành vi của con người - cốt lõi của khoa học xã hội và nhân văn
- Sự tương thích cao giữa Lý thuyết Mindsponge và phân tích Bayes
- Xương sống của BMF
  - Hệ Lý thuyết quản trị tri thức SM3D (Serendipity-Mindsponge-3D)
  - Bộ ba triết lý khoa học về minh bạch, chi phí, và sự chủ động
  - Khoa học mở 3O (open data, open review, open dialogue)
  - Phần mềm bayesvl
  - Văn hóa Mindsponge
- Công cụ đào tạo trên SM3D Portal



# **Hệ lý thuyết quản trị tri thức và khung khái niệm SM3D**



2013

Khái niệm “**Serendipity như lợi thế chiến lược**”

---

2014

**Khái niệm Mindsponge**, 1/10 thước đo của khảo sát i2Metrix về năng lực đổi mới của doanh nghiệp

---

2015

**Quá trình sáng tạo 3D/Hệ thống lọc thông tin 3D**

**Cơ chế Mindsponge**

---

2021

**Bayesian Mindsponge Framework** được áp dụng lần đầu

Đề xuất cơ chế khởi sinh của suy nghĩ tự sát dựa trên **Cơ chế Mindsponge**

---

2022

**Hệ lý thuyết SM3D** được dùng để lý giải quá trình sách tạo vaccine COVID

Hoàn thiện phương pháp phân tích **BMF**

**Lý thuyết bản chất, cơ chế, và sự khởi sinh của Serendipity** ra đời

**Lý thuyết Mindsponge** ra đời



## **Các ứng dụng và sự phát triển của phân tích BMF/MT (tính đến tháng 7/2023)**

- Được ứng dụng cho 128 sản phẩm khoa học đã xuất bản
- Được sử dụng bởi hơn 243 nhà nghiên cứu từ 150 tổ chức của 30 quốc gia
- Trong đó, 65.4% tới từ các nước đang phát triển
- 30 nhà nghiên cứu được đào tạo thông qua chương trình SM3D Portal



## Các ứng dụng tiêu biểu

Hiện tượng cộng  
tính văn hóa

Lý thuyết về khả  
năng đàn hồi của  
tổ chức

Mô hình Chấp  
nhận công nghệ  
Mindsponge  
(MTAM)

Khung phân tích  
Ecomindsponge

Mô hình phản tư  
trong hợp tác  
chữa bệnh

Hiện tượng cận  
tự tử

Văn hóa thặng dư  
sinh thái

Nguyên lý bán  
dẫn giá trị môi  
trường-tiền tệ

Mô hình về sự  
chấp nhận trí tuệ  
nhân tạo có cảm  
xúc

Cơ chế tâm lý-tôn  
giáo đằng sau  
các vụ tấn công  
tự sát

Tính đồng nhất về  
tư tưởng phương  
Tây trong nghiên  
cứu khoa học

Quá trình tâm lý  
chính trị sinh ra  
và trao quyền cho  
nhà độc tài

Sự kỳ thị kép  
trong tiếp biến  
văn hóa

Lý thuyết về sự  
bền vững của  
doanh nghiệp

Mô hình niềm tin-  
rủi ro đối với  
thông điệp về sức  
khỏe



**Mô hình thống kê là gì?**

- Mô hình thống kê (statistical model hay gọi tắt là model) là mô hình toán học thể hiện các giả định của nhà nghiên cứu để phân tích dữ liệu.
- Ứng dụng mô hình thống kê (statistical modeling) là quá trình thử nghiệm và áp dụng mô hình thống kê lên dữ liệu thô để tìm ra mối quan hệ phù hợp nhất giữa các biến, từ đó đưa ra dự đoán thực tế.

Ví dụ như mô hình tuyến tính đơn giản:

$$y = \alpha + \beta x + \varepsilon$$

$y$  là biến phụ thuộc (biến kết quả),  $x$  là biến độc lập (biến dự báo),  $\alpha$  và  $\beta$  là tham số,  $\varepsilon$  là phần sai số (phần dư)



**Thống kê Bayes là gì?**



- Được đặt theo tên của nhà toán học người anh Thomas Bayes vào thế kỷ 18
- Suy luận Bayes đưa ra xác suất hậu nghiệm (Posterior Probability) là hệ quả của hai tiền đề: xác suất tiền nghiệm (Prior Probability) và hàm khả năng (Likelihood Function).

$$P(\theta|\mathbf{D}) = \frac{P(\mathbf{D}|\theta)P(\theta)}{P(\mathbf{D})}$$

- $\theta$  là tham số thể hiện giả thuyết cần đánh giá, hay hệ số cần xác định mà xác suất của nó chịu ảnh hưởng của dữ liệu quan sát.
- $\mathbf{D}$  là dữ liệu quan sát được, số liệu thu được từ thực tế
- $P(\theta)$  là xác suất tiền nghiệm, niềm tin hay các bằng chứng trước đây liên quan tới giả thuyết (prior probability)
- $P(\mathbf{D}|\theta)$  là xác suất quan sát được dữ liệu  $\mathbf{D}$  với điều kiện  $\theta$  (likelihood)
- $P(\mathbf{D})$  thường gọi là khả năng biên (Marginal Likelihood), một hằng số không phụ thuộc vào tham số  $\theta$
- $P(\theta|\mathbf{D})$  là xác suất hậu nghiệm, luôn tỷ lệ với xác suất tiền nghiệm và xác suất quan sát được (posterior probability)

Mô hình đầy đủ:

$$P(\theta|\mathbf{D}) = \frac{P(\mathbf{D}|\theta)P(\theta)}{P(\mathbf{D})}$$

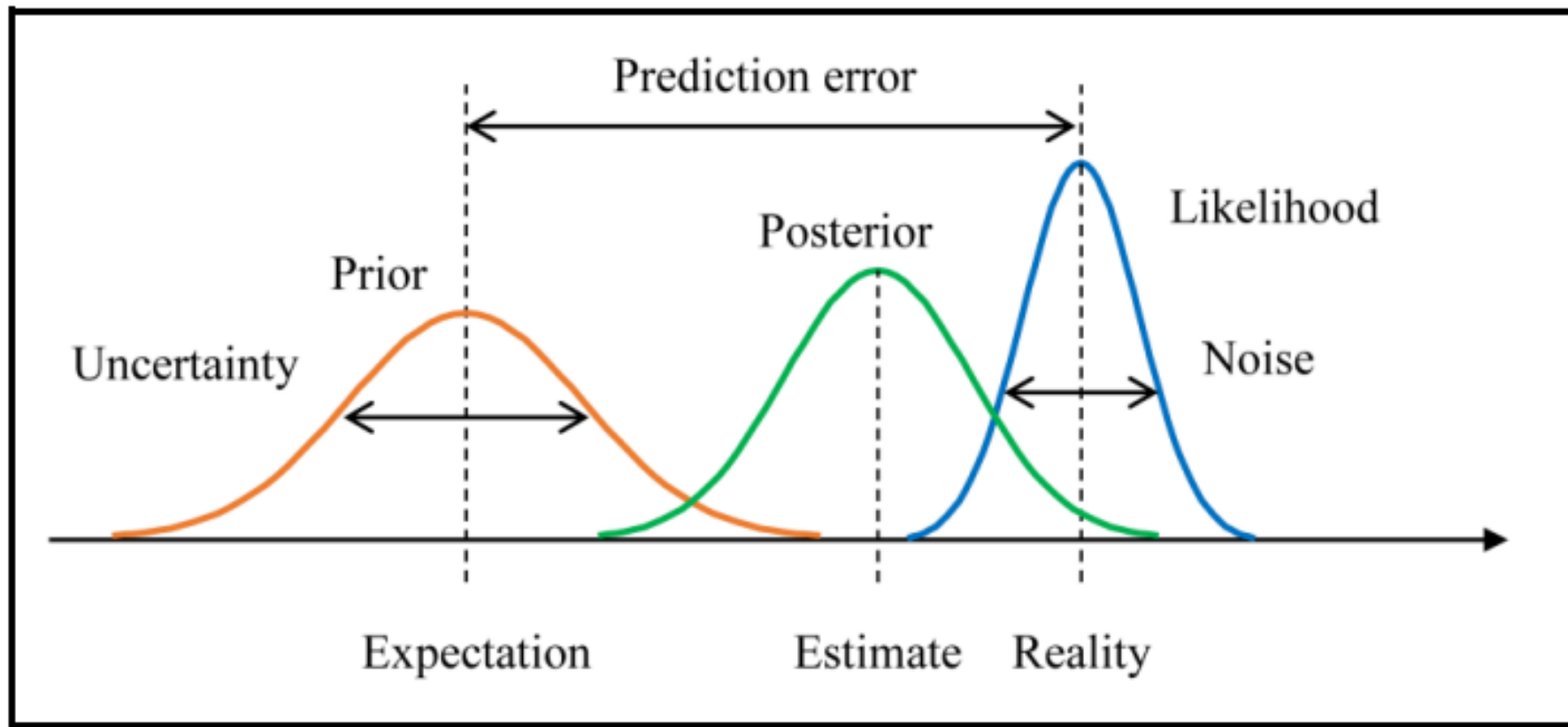
Mô hình giản hóa sau khi bỏ khả năng biên  $P(\mathbf{D})$ :

$$P(\theta|\mathbf{D}) \propto P(\mathbf{D}|\theta)P(\theta)$$

Quy tắc cơ bản nhất của suy luận Bayes:

***Posterior Probability*  $\propto$  *Prior Probability*  $\times$  *Likelihood Function***

=> Trong tất cả các mô hình Bayes, phân phối xác suất hậu nghiệm luôn tỷ lệ thuận với phân phối xác suất tiền nghiệm và hàm khả năng



**Figure 10.1:** Example of Bayesian inference with a posterior distribution, prior distribution, and likelihood function. The visualization is retrieved from Yanagisawa, Kawamata and Ueda (11) under the Creative Commons Attribution license (CC-BY).



**Tại sao thống kê Bayes lại quan trọng?**

- Tích hợp xác suất tiên nghiệm (prior probability) vào ước lượng xác suất hậu nghiệm bên cạnh xác suất quan sát được từ dữ liệu
  - Giúp tích hợp các bằng chứng trước đây, đánh giá chuyên môn, lý thuyết, hoặc niềm tin của nhà nghiên cứu vào ước lượng
    - ❖ Cần lưu ý vấn đề thiên kiến chủ quan (subjective biases)
  - Đánh giá độ chắc chắn của kết quả ước lượng thông qua phương pháp hiệu chỉnh xác suất tiên nghiệm (prior-tweaking)
  - Giúp làm giảm vấn đề đa cộng tuyến giữa các biến phụ thuộc (multicollinearity)
  - Tăng độ chính xác của ước lượng đối với các bộ dữ liệu bé
    - ❖ Lưu ý: xác định tốt prior có thể làm tăng độ chính xác, nhưng xác định sai sẽ làm giảm độ chính xác



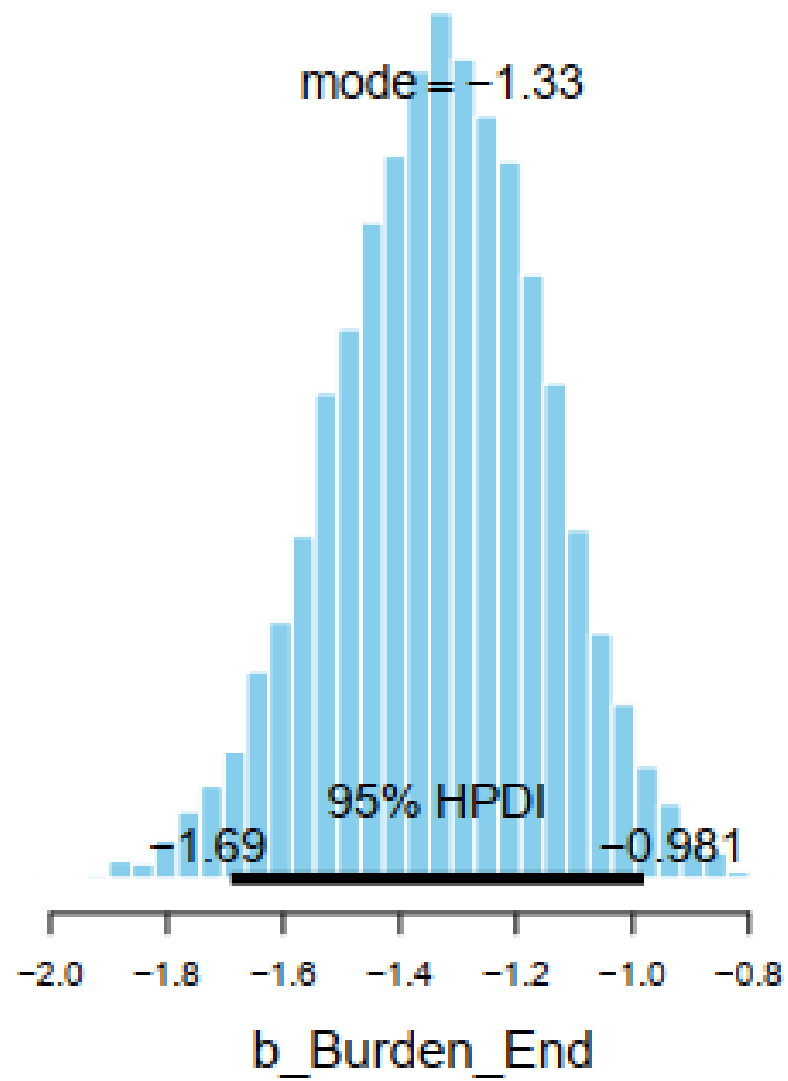
- Suy luận Bayes xem tất cả các định lượng đều là xác suất, kể cả dữ liệu không quan sát được (unobserved data) và tham số chưa biết (unknown parameters).
- Hay nói như Jeff Gill:

Với những người theo triết lý Bayes, “thế giới được chia thành: số lượng có sẵn ngay lập tức và số lượng cần được mô tả theo xác suất”

  - Cho phép nhà nghiên cứu xây dựng các mô hình linh hoạt theo nhu cầu nhưng không cần phải thay đổi phương pháp phân tích
    - ❖ Mô hình parsimony
    - ❖ Mô hình phức tạp, như mô hình đa lớp (hierarchical model), mô hình phi tuyến tính (non-linear model)



- Có lợi thế hơn suy luận truyền thống frequentist về mặt lý thuyết:
  - Kết hợp sức mạnh giữa toán học và common sense, giúp giải quyết các vấn đề trong nghiên cứu KHXXH, ngành có rất nhiều điều không chắc chắn.
  - Cập nhật các kết quả khoa học dựa trên các thay đổi trong thực tế.
  - Không đánh giá kết quả theo kiểu tất định (deterministic) của  $p$ -value mà sử dụng hình vẽ trực quan về phân phối xác suất để đánh giá kết quả.



# Các khó khăn khi ứng dụng thống kê Bayes

- Trở ngại vật lý:

- Yêu cầu năng lực tính toán mạnh của máy tính
- Thiếu các phần mềm và chương trình

Đều đã giải quyết được

- Trở ngại tâm lý:

- Ngại Toán học
- Ngại lập trình
- Ngại thoát khỏi vùng an toàn

Vấn đề nan giải hiện nay



# Ngôn ngữ R và phần mềm bayesvl



# R

- R & RStudio: R là nền tảng ngôn ngữ lập trình phục vụ tính toán thống kê. Ngôn ngữ R được sử dụng rộng rãi bởi các nhà thống kê học, các nhà khai thác dữ liệu nhằm phát triển các phần mềm (gọi là các package) thống kê và xử lý dữ liệu.



# bayesvl

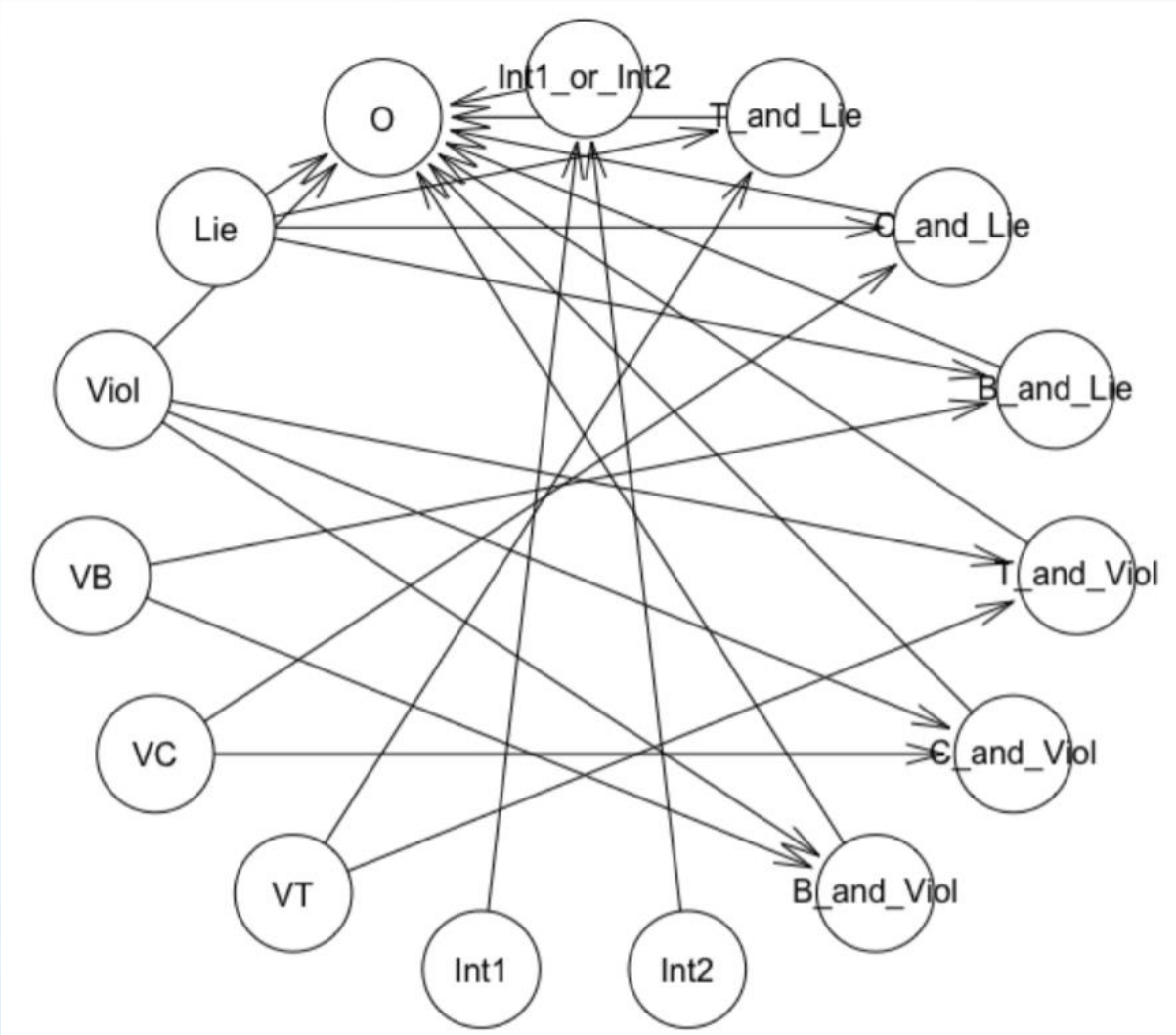
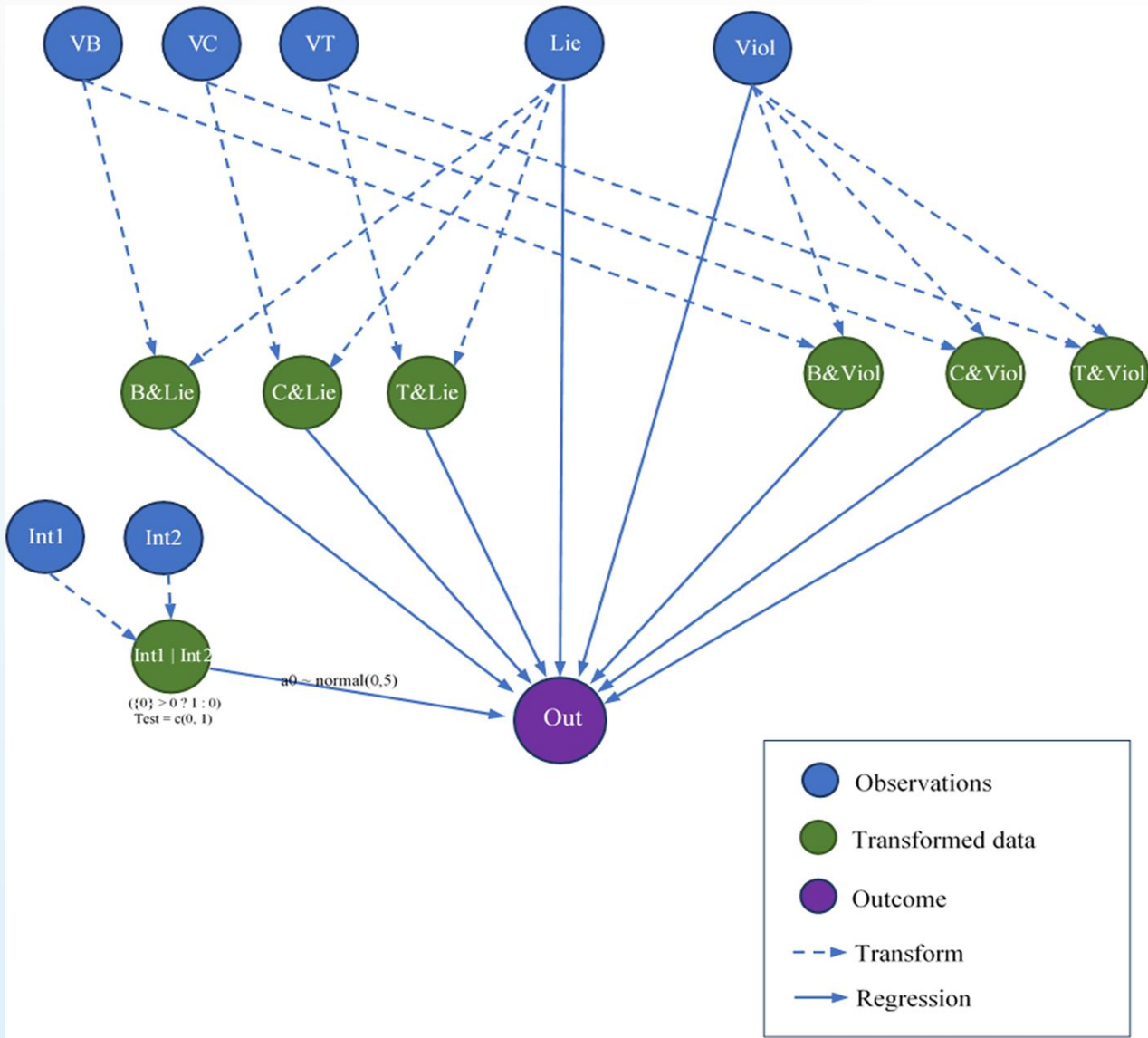
- Phần mềm phát triển bởi kỹ sư Lã Việt Phương và TS Vương Quân Hoàng
- Chính thức xuất bản trên Comprehensive R Archive Network (CRAN) ngày 24-5-2019 (v0.8.3)
  - <https://cran.r-project.org/web/packages/bayesvl/index.html>
- Hiện tại chương trình đang được đánh số phiên bản v.1.0 (github)
  - <https://github.com/sshpa/bayesvl>





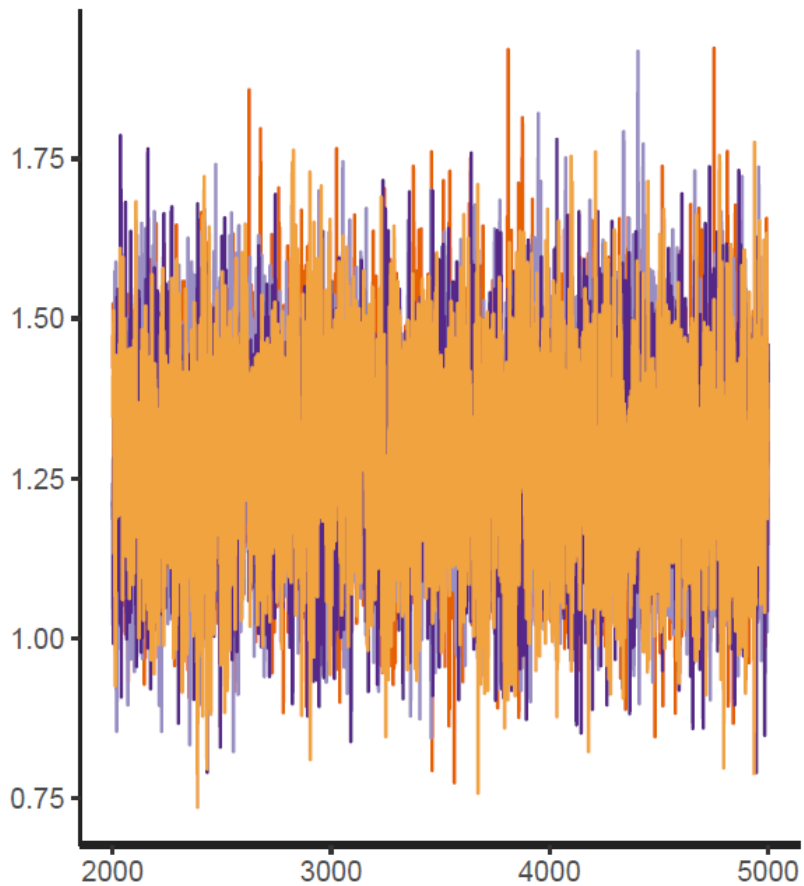
## Lợi thế của bayesvl

- Chức năng xây dựng cây quan hệ (“relationship tree”) hay sơ đồ logic
  - Hỗ trợ tư duy xây dựng mô hình nghiên cứu
  - Dễ sử dụng và tái lập mô hình phân tích
- Kết quả được trực quan hóa có độ thẩm mỹ cao
  - Hỗ trợ đọc hiểu kết quả thông qua hình vẽ
- Hỗ trợ nhiều chức năng phân tích
  - Tương thích cao với các phần mềm liên quan khác, như **ggplot2**, **Loo**, etc.
  - Cho phép thực hiện các nghiên cứu có độ phức tạp cao, như mô hình đa tầng, mô hình phi tuyến tính, v.v.

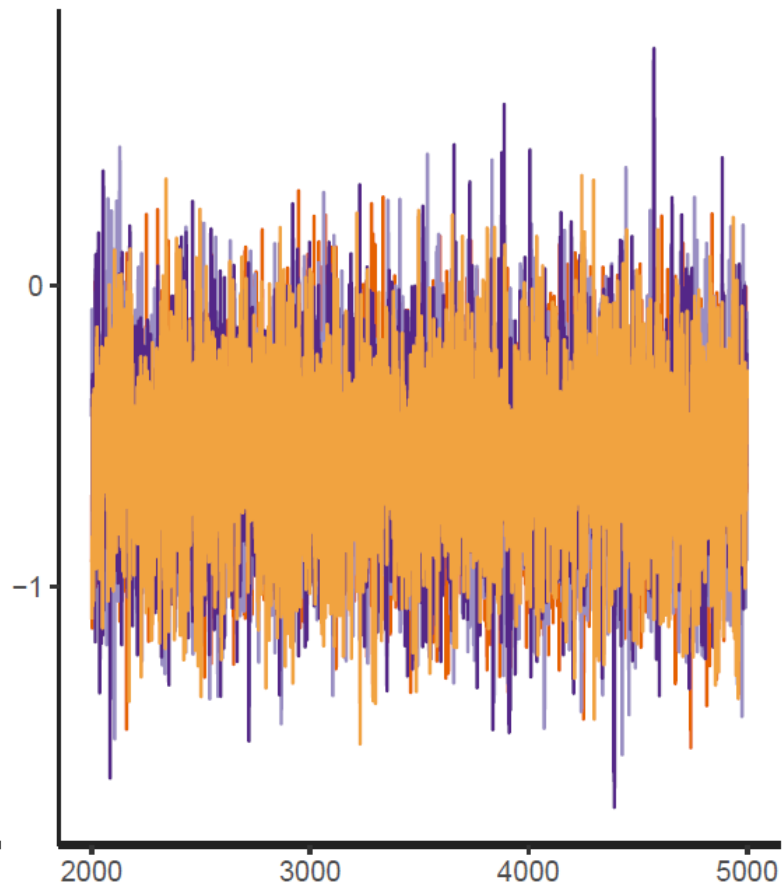


<https://www.nature.com/articles/s41599-020-0442-3>

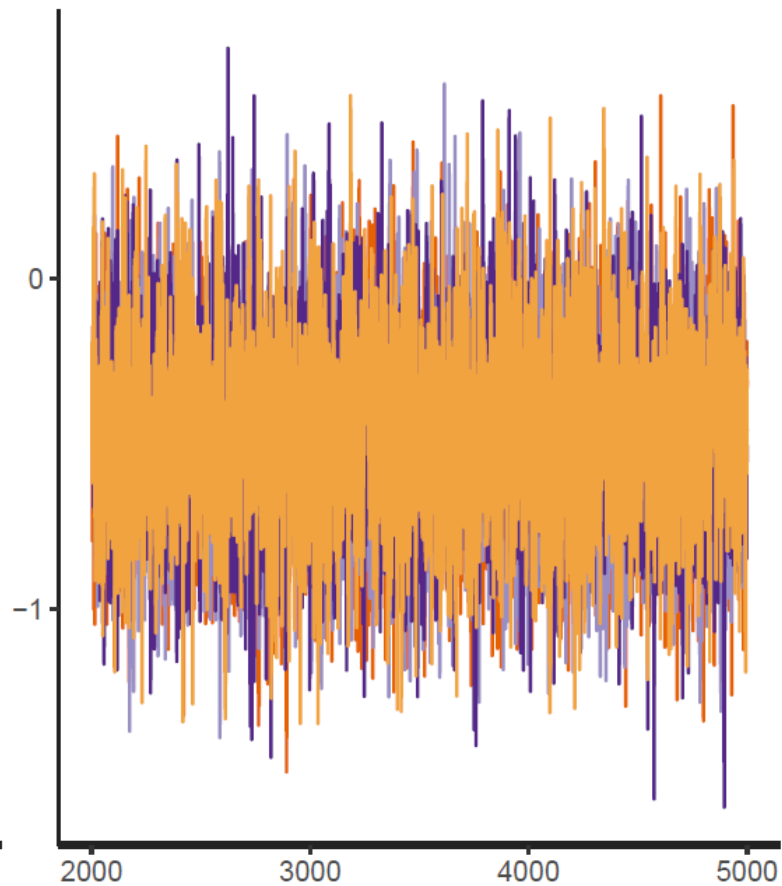
a\_O



b\_Viol\_O



b\_Lie\_O



chain

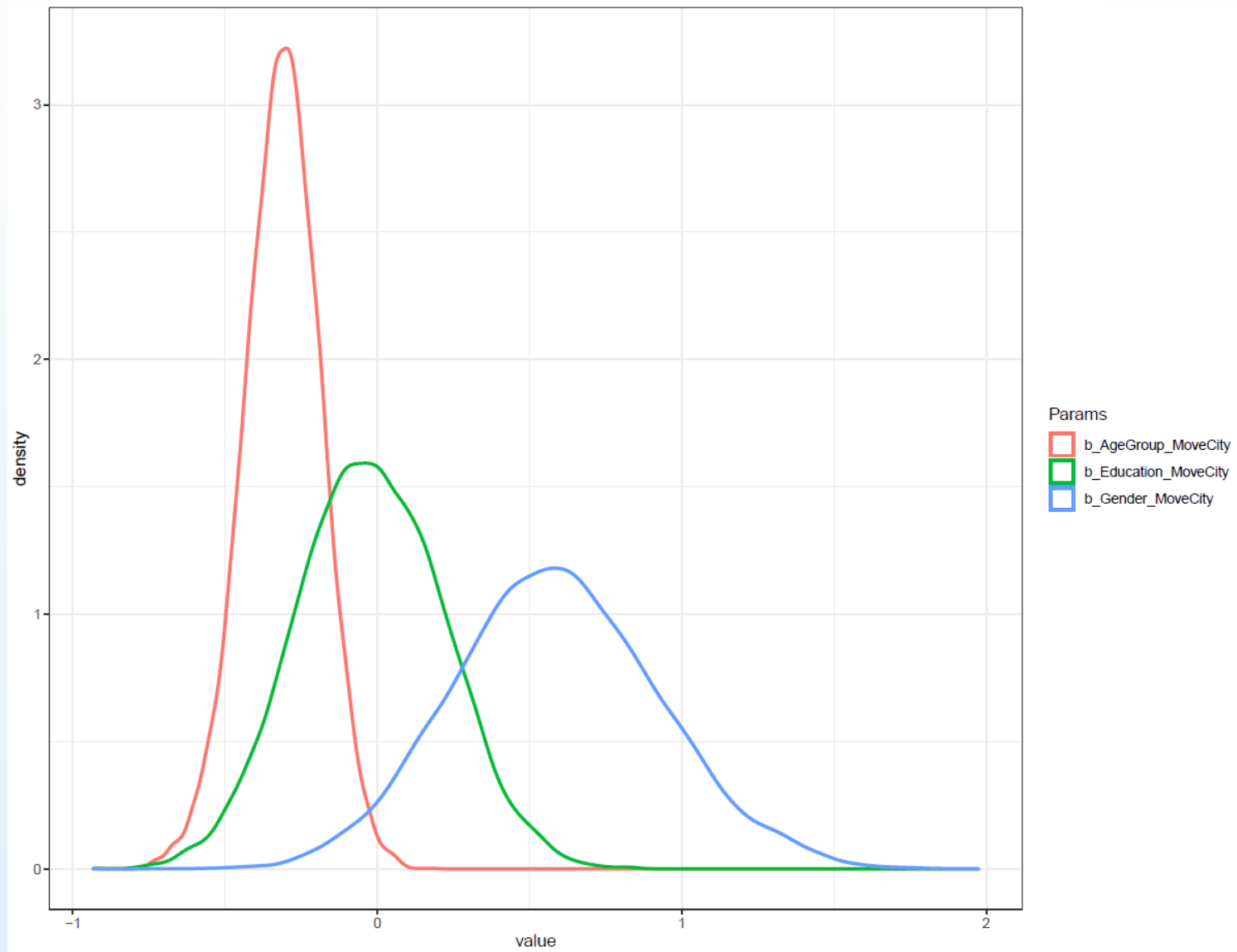
1

2

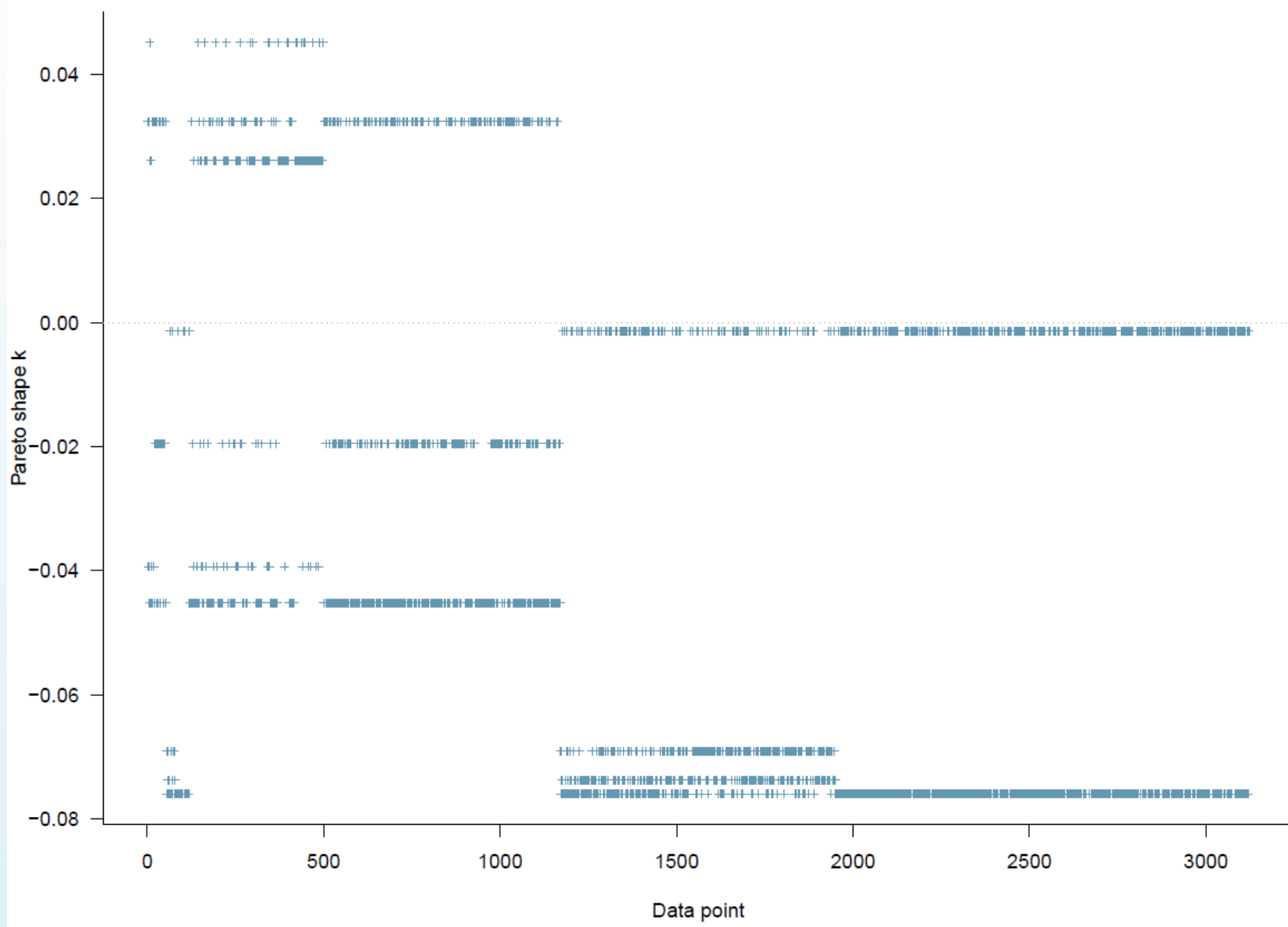
3

4

Biểu đồ vết (trace plot)



Phân phối xác suất hậu nghiệm (Posterior density plot)



Kiểm định PSIS-LOO





# **\*\*\* Hỏi và Đáp \*\*\***





# Các bước phân tích Bayes



## Các bài toán

- Gánh nặng tài chính trong chăm sóc sức khỏe
  - <https://www.mdpi.com/2306-5729/4/2/57>
  - <https://github.com/sshpa/bayesvl>
- Nhận thức về biến đổi khí hậu và sự ủng hộ chính sách bảo vệ biển
  - <https://www.mdpi.com/1660-4601/18/7/3681/htm>

# Các bước thực hiện phân tích Bayes

Bước 1: Xây dựng mô hình

Bước 2: Xác định xác suất tiên nghiệm

Bước 3: Thử nghiệm mô hình

Bước 4: Đánh giá kết quả

Bước 5: So sánh



# Thực hành phần mềm bayesvl



## Bước 1: Xây dựng mô hình/cây quan hệ

1. Tạo dựng các biến
2. Tạo dựng mối liên hệ giữa các biến
3. Tạo ngôn ngữ Stan
4. Thực hiện mô phỏng/phân tích



## Bước 2: Đánh giá mô hình

| Đồ họa                                         | Hàm                            |
|------------------------------------------------|--------------------------------|
| Bước mô phỏng Markov<br>(Traceplot)            | <code>bvl_plotTrace()</code>   |
| Đồ thị Gelman<br>(Gelman plot)                 | <code>bvl_plotGelmans()</code> |
| Đồ thị tự tương quan<br>(Autocorrelation plot) | <code>bvl_plotAcfs()</code>    |

Ngoài ra, hàm `bvl_stanLoo()` cũng có thể dùng để đánh giá độ phù hợp của mô hình thống kê.



# BƯỚC 3: ĐỌC KẾT QUẢ

- Khi đọc kết quả cần chú ý tới 3 yếu tố:
  - `n_eff`: mẫu độc lập (effective sample size)
  - `Rhat`: chỉ số Gelman
  - Mean và standard deviation: phân phối của mô phỏng hậu nghiệm

| Đồ họa                                                  | Hàm                              |
|---------------------------------------------------------|----------------------------------|
| Phân phối mật độ<br>(Density plot)                      | <code>bvl_plotDensity()</code>   |
| Phân phối mật độ theo khoảng<br>(Interval plot)         | <code>bvl_plotIntervals()</code> |
| Phân phối tần suất<br>(Param plot)                      | <code>bvl_plotParams()</code>    |
| Phân phối mật độ trên mặt phẳng 2D<br>(2D Density plot) | <code>bvl_plotDensity2d()</code> |



## Bước 4: So sánh mô hình

- Để so sánh mô hình hàm `bvl_stanWAIC()` và `bvl_compareWAIC()` có thể để so sánh mức độ phù hợp giữa 2 mô hình.
- Ngoài ra, **bayesvl** đã được điều chỉnh để tương thích với các mẫu code so sánh mức độ phù hợp giữa các mô hình của phần mềm **Loo**



# Các bộ dữ liệu có thể sử dụng

- Kinh tế y tế và hệ thống y tế: <https://www.mdpi.com/2306-5729/4/2/57> / <https://github.com/sshpa/bayesvl>
- Mối liên kết giữa người thành thị và đa dạng sinh học: <https://direct.mit.edu/dint/article/3/4/578/107428/Multifaceted-Interactions-between-Urban-Humans-and>
- Hành vi của người chơi trong thế giới ảo và nhận thức về môi trường: <https://direct.mit.edu/dint/article/3/4/606/107672/A-Multinational-Data-Set-of-Game-Players-Behaviors?searchresult=1>
- Hành vi sử dụng thông tin và lựa chọn nhà cung cấp dịch vụ của bệnh nhân: <https://www.sciencedirect.com/science/article/pii/S2352340916302803?via%3Dihub>
- Kế hoạch chủ quan và nhận thức về khả năng thành công của nhà khởi nghiệp: <https://www.sciencedirect.com/science/article/pii/S2352340916000408>
- Năng lực kỹ thuật số của thanh niên: <https://www.mdpi.com/2306-5729/4/2/69>
- Nhận thức, thái độ và kỹ năng của học sinh về công dân toàn cầu: <https://www.sciencedirect.com/science/article/pii/S2352340921004467#bib0002>
- Trải nghiệm học online của học sinh: <https://www.sciencedirect.com/science/article/pii/S2352340922008678>
- Nhận thức của công chúng về biến đổi khí hậu và giá trị của hệ sinh thái biển và ven biển: <https://www.sciencedirect.com/science/article/pii/S2352340923000422>



- Tâm lý sợ hãi của nhà đầu tư chứng khoán cá nhân trong khủng hoảng: <https://osf.io/54z9y/>
- Hành vi lái xe an toàn: <https://www.sciencedirect.com/science/article/pii/S2352340923004559>
- Ý định và hành vi sử dụng công nghệ của sinh viên điều dưỡng: <https://www.sciencedirect.com/science/article/pii/S235234092100514X>
- Nhận thức về sự khan hiếm nước và tái sử dụng nước uống: <https://www.sciencedirect.com/science/article/pii/S2352340920301839>



## References

- [1] Vuong QH. (2022). *The kingfisher story collection*. <https://www.amazon.com/dp/B0BFV9L58W>
- [2] Vuong QH. (2018). The (ir)rational consideration of the cost of science in transition economies. *Nature Human Behaviour*, 2(1), 5. <https://www.nature.com/articles/s41562-017-0281-4>
- [3] Vuong QH. (2019). Breaking barriers in publishing demands a proactive attitude. *Nature Human Behaviour*, 3, 1034. <https://www.nature.com/articles/s41562-019-0667-6>
- [4] Vuong QH. (2017). Open data, open review and open dialogue in making social sciences plausible. *Scientific Data Updates*. <https://blogs.nature.com/scientificdata/2017/12/12/authors-corner-open-data-open-review-and-open-dialogue-in-making-social-sciences-plausible/>
- [5] Vuong QH. (2020). Reform retractions to make them more transparent. *Nature*, 582, 149. <https://www.nature.com/articles/d41586-020-01694-x>
- [6] Vuong QH, Nguyen MH, La VP. (2022). *The mindsp sponge and BMF analytics for innovative thinking in social sciences and humanities*. De Gruyter. [https://www.google.com.vn/books/edition/The\\_mindsp sponge\\_and\\_BMF\\_analytics\\_for\\_inn/EGeEEAAQBAJ](https://www.google.com.vn/books/edition/The_mindsp sponge_and_BMF_analytics_for_inn/EGeEEAAQBAJ)
- [7] Nguyen MH, La VP, Le TT, Vuong QH. (2022). Introduction to Bayesian Mindsp sponge Framework analytics: An innovative method for social and psychological research. *MethodsX*, 9, 101808. <https://linkinghub.elsevier.com/retrieve/pii/S2215016122001881>
- [8] Vuong QH., et al. (2020). Bayesian analysis for social data: A step-by-step protocol and interpretation. *MethodsX*, 7, 100924. <https://www.sciencedirect.com/science/article/pii/S2215016120301448>
- [9] Vuong QH. (2023). *Mindsp sponge Theory*. De Gruyter. [https://www.google.com.vn/books/edition/Mindsp sponge\\_Theory/OSiGEAAQBAJ](https://www.google.com.vn/books/edition/Mindsp sponge_Theory/OSiGEAAQBAJ)
- [10] Nguyen MH, Jones TE. (2022). Predictors of support for biodiversity loss countermeasure and bushmeat consumption among Vietnamese urban residents. *Conservation Science and Practice*, 4(12), e12822. <https://conbio.onlinelibrary.wiley.com/doi/10.1111/csp2.12822>
- [11] Nguyen MH, Jones TE. (2022). Building eco-surplus culture among urban residents as a novel strategy to improve finance for conservation in protected areas. *Humanities and Social Sciences Communications*, 9, 426. <https://www.nature.com/articles/s41599-022-01441-9>
- [12] Kantabutra S, Ketprapakorn N. (2021). Toward an Organizational Theory of Resilience: An Interim Struggle. *Sustainability*, 13, 13137. <https://www.mdpi.com/2071-1050/13/23/13137>
- [13] Ho MT, et al. (2022). Rethinking technological acceptance in the age of emotional AI: Surveying Gen Z (Zoomer) attitudes toward non-conscious data collection. *Technology in Society*, 70, 102011. <https://www.sciencedirect.com/science/article/abs/pii/S0160791X2200152X>
- [14] Mantello P, et al. (2023). Machines that feel: behavioral determinants of attitude towards affect recognition technology—upgrading technology acceptance theory with the mindsp sponge model. *Humanities and Social Sciences Communications*, 10, 430. <https://www.nature.com/articles/s41599-023-01837-1>
- [15] Vuong QH, et al. (2018). Cultural additivity: behavioural insights from the interaction of Confucianism, Buddhism and Taoism in folktales. *Palgrave Communications*, 4, 143. <https://www.nature.com/articles/s41599-018-0189-2>
- [16] Nguyen MH, Le TT, Vuong QH. (2023). Ecomindsp sponge: A Novel Perspective on Human Psychology and Behavior in the Ecosystem. *Urban Science*, 7(1), 31. <https://www.mdpi.com/2413-8851/7/1/31>
- [17] Jin R, Wang X. (2022). “Somewhere I belong?” A study on transnational identity shifts caused by “double stigmatization” among Chinese international student returnees during COVID-19 through the lens of mindsp sponge mechanism. *Frontiers in Psychology*, 13, 1018843. <https://www.frontiersin.org/articles/10.3389/fpsyg.2022.1018843/full>
- [18] Årleskog C, Vackerberg N, Andersson AC. (2021). Balancing power in co-production: introducing a reflection model. *Humanities and Social Sciences Communications*, 8, 108. <https://www.nature.com/articles/s41599-021-00790-1>
- [19] Yanagisawa H, Kawamata O, Ueda K. (2019). Modeling Emotions Associated With Novelty at Variable Uncertainty Levels: A Bayesian Approach. *Frontiers in Computational Neuroscience*, 13, 2. <https://www.frontiersin.org/articles/10.3389/fncom.2019.00002/full>
- [20] Gill J. (2015). *Bayesian Methods: A Social and Behavioral Sciences Approach*. Chapman & Hall. <https://www.routledge.com/Bayesian-Methods-A-Social-and-Behavioral-Sciences-Approach-Third-Edition/Gill/p/book/9781439862483>
- [21] McElreath R. (2016). *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. Chapman and Hall/CRC. <https://www.taylorfrancis.com/books/mono/10.1201/9781315372495/statistical-rethinking-richard-mcelreath>
- [22] La VP, Vuong QH. (2019). bayesvl: Visually Learning the Graphical Structure of Bayesian Networks and Performing MCMC with 'Stan'. *The Comprehensive R Archive Network (CRAN)*. <https://cran.r-project.org/web/packages/bayesvl/index.html>