

Bayesian statistics

# Report

## Data analysis & Results

---

Khương Quỳnh Long

Hà Nội, 08/2019

<https://gitlab.com/LongKhuong/adhere-bayesian-statistics>

# Nội dung thảo luận

---

1

- Thông tin cần cung cấp khi phân tích Bayesian

2

- Xây dựng checklist cho phần “data analysis và results”

- 
- ▶ Thống kê Bayes hiện tại chưa được dùng như một quy trình chuẩn trong nhiều lĩnh vực
  - ▶ Chưa có chuẩn về khai báo thông tin trong phân tích và kết quả
  - ▶ Hiện tại, đa số phải dựa vào kiến thức nền của người đọc để viết cho phù hợp

# Thông tin cần cung cấp<sup>(1)</sup>

---

- ▶ **Essential points:** Cần có trong tất cả phân tích theo Bayesian
- ▶ **Optional points:** Không nhất thiết phải có trong tất cả các trường hợp, nhưng có thể cân nhắc trong tình huống cụ thể
- ▶ **Helpful points:** Có thể làm tăng tầm ảnh hưởng của kết quả

# Essential points

## 1. Thúc đẩy việc sử dụng phân tích Bayesian (non-NHST):

- Một số reviewers và người đọc chưa quen với Bayesian  
→ thường phải có câu dẫn và lý do sử dụng Bayesian
- Nêu các ưu điểm của Bayesian so với NHST

### Ưu điểm “chung”:

- ✓ *“The inferences from a Bayesian analysis are richer and more informative than NHST because the posterior distribution reveals joint probabilities of combinations of parameter values”<sup>(1)</sup>*
- ✓ *“There is no reliance on sampling distributions and  $p$  values to interpret the parameter estimates”*
- ✓ ...

1. John Kruschke (2014). Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan 2nd Edition. Elsevier Inc

# Essential points

## Ưu điểm “cụ thể” cho từng tình huống:

- ✓ *“Bayesian models are designed to be appropriate to the data structure, without having to make approximation assumptions typical in NHST (e.g., homogeneity of variances across groups, and normally distributed noise), and that.....” (1)*
  - ✓ *“The power of Bayesian methods in estimating complex models for complex data analysis is indisputable. Besides the capability for implementing estimation procedures, which generally cannot be done in MLE, the ease, flexibility, and computation time were also very acceptable” (2)*
  - ✓ ...
1. John Kruschke (2014). Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan 2nd Edition. Elsevier Inc
  2. Zhang, Z., Hamagami, F., Lijuan Wang, L., Nesselroade, J. R., & Grimm, K. J. (2007). Bayesian analysis of longitudinal data using growth curve models. International Journal of Behavioral Development, 31(4), 374–383.

# Essential points

---

Ưu điểm “cụ thể” cho từng tình huống:

- ✓ *“Bayesian methods provide a clear way to incorporate prior information that both increases the statistical power of analysis and formulizes the accumulation of scientific findings, which is especially helpful in the context of small data set”*

# Essential points

---

## 2. Mô tả rõ ràng về cấu trúc dữ liệu, mô hình sử dụng và các tham số mô hình:

- ✓ Biến tiên lượng, biến kết cuộc...
- ✓ Mô tả mô hình sử dụng (e.g., logistic, linear, multilevel...) và lý do mô hình đó phù hợp
- ✓ Tham số của mô hình...
- ✓ Các bước xây dựng mô hình...



# Essential points

---

## 3. Mô tả rõ về Prior sử dụng:

- ✓ Prior là “nghệ thuật” của Bayesian
- ✓ Cam kết có kế hoạch về Prior trước khi quan sát dữ liệu
- ✓ Nêu rõ các Prior sử dụng và lý do sử dụng (ngay cả đối với non-informative)
- ✓ Cách xây dựng Prior (đối với informative prior)
- ✓ Nếu có nhiều Prior, có thể thực hiện sensitivity analysis

# Essential points

---

## 4. Cung cấp thông tin chuỗi MCMC:

- ✓ MCMC chứa đựng toàn bộ thông tin của mô hình
- ✓ Thuật toán MCMC gì?: MCMC, Gibbs, HMC...
- ✓ Cung cấp thông tin: số vòng lặp, burn-in, thinning, ESS..
- ✓ Thông tin về trạng thái hội tụ:

# Essential points

---

## 5. Diễn giải phân phối hậu nghiệm:

- ✓ Posterior central tendency (mean, median hay mode)
- ✓ CI hay HDI
- ✓ Kiểm định giả thuyết
  - CompVal: ngưỡng so sánh → lý do chọn ngưỡng đó
  - ROPE: cung cấp giới hạn của ROPE
  - Bayes Factor: cung cấp tiêu chuẩn phân loại

# Optional points

---

- ▶ Robustness of posterior for different priors:

Trong tình huống sử dụng nhiều prior, kết luận về posterior là thống nhất → tăng tính tin cậy

- ▶ Posterior predictive check: kiểm tra performance của mô hình, có thể sử dụng cross-validation (LOO-CV, 10-fold-CV...)

# Helpful points

---

- ▶ Cung cấp data thô
- Nhà nghiên cứu khác có thể xây dựng mô hình khác từ data đó
- Kết hợp nhiều data có cấu trúc tương tự thành data lớn hơn
- ▶ Cung cấp chuỗi MCMC của posterior:
- Vì MCMC chứa rất nhiều thông tin, một báo cáo khó có thể tóm tắt hết nội dung của nó → cung cấp MCMC, nhà nghiên cứu khác có thể tự khai thác thông tin
- Sử dụng thông tin của chuỗi MCMC để làm informative prior cho nghiên cứu tiếp theo

**Thank you !**