Mô Hình Hồi Quy Bayesian

Khương Quỳnh Long

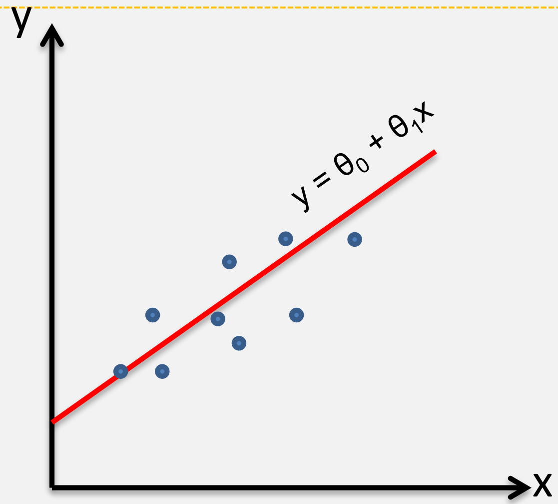
Hà Nội, 08/2019

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

Nội dung

- So sánh sự khác biệt giữa “Classical model” và Bayesian model
- Markov Chain Monte Carlo
- Hàm loss và decision making
- Một số phầm mềm sử dụng trong Bayesian

“Classical model”



Maximum Likelihood Estimation

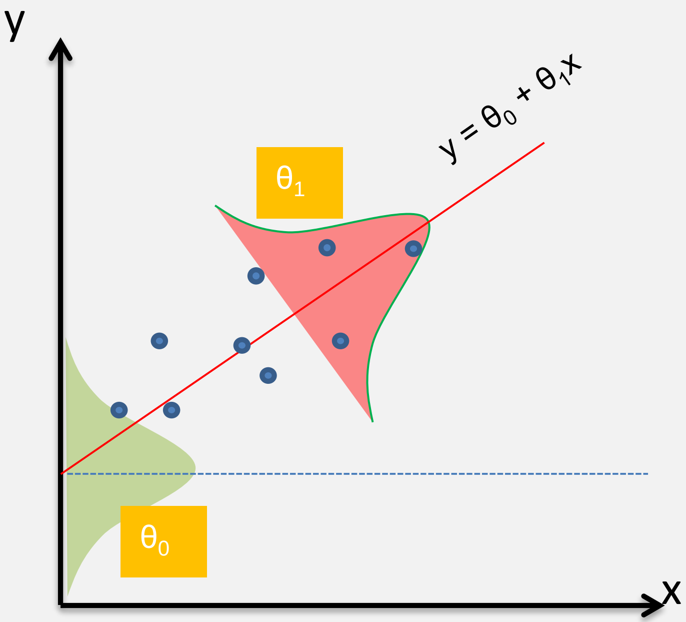
$$y = \theta_0 + \theta_1 x + \varepsilon$$

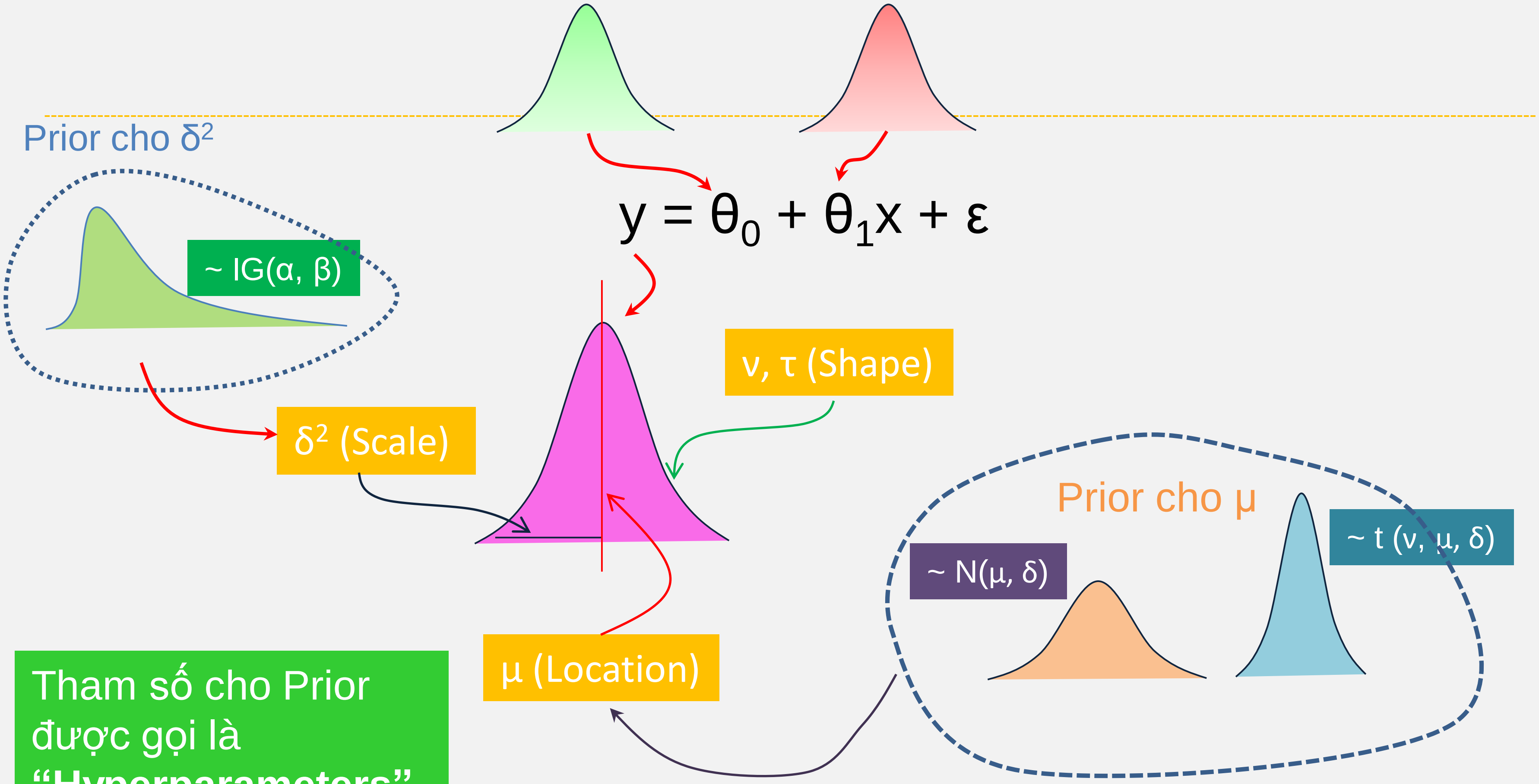
Fixed

~ N(0, δ^2)

- Tham số mô hình là cố định (point estimate)
- Dựa hoàn toàn vào likelihood của dữ liệu

Bayesian model





Tham số cho Prior được gọi là **“Hyperparameters”**

Markov Chain Monte Carlo

Likelihood

Độ khả dĩ của dữ liệu: Xác suất thu thập được dữ liệu dưới điều kiện giả thuyết H đúng (Đơn giản là dữ liệu thu thập)

Prior

Xác suất tiền định: Xác suất giả thuyết H mà chúng ta tin là xảy ra (đúng) trước khi thu thập dữ liệu

$$P(H|D) = \frac{P(D|H) * P(H)}{P(D)}$$

Posterior

Xác xuất hậu định: Cần tìm Xác suất giả thuyết H đúng cho bởi dữ liệu thu thập

Marginal likelihood (normalizing constant)

Hằng số: Xác suất của dữ liệu (tổng tất cả likelihood*prior của tất cả các giả thuyết có khả năng xảy ra)

MCMC

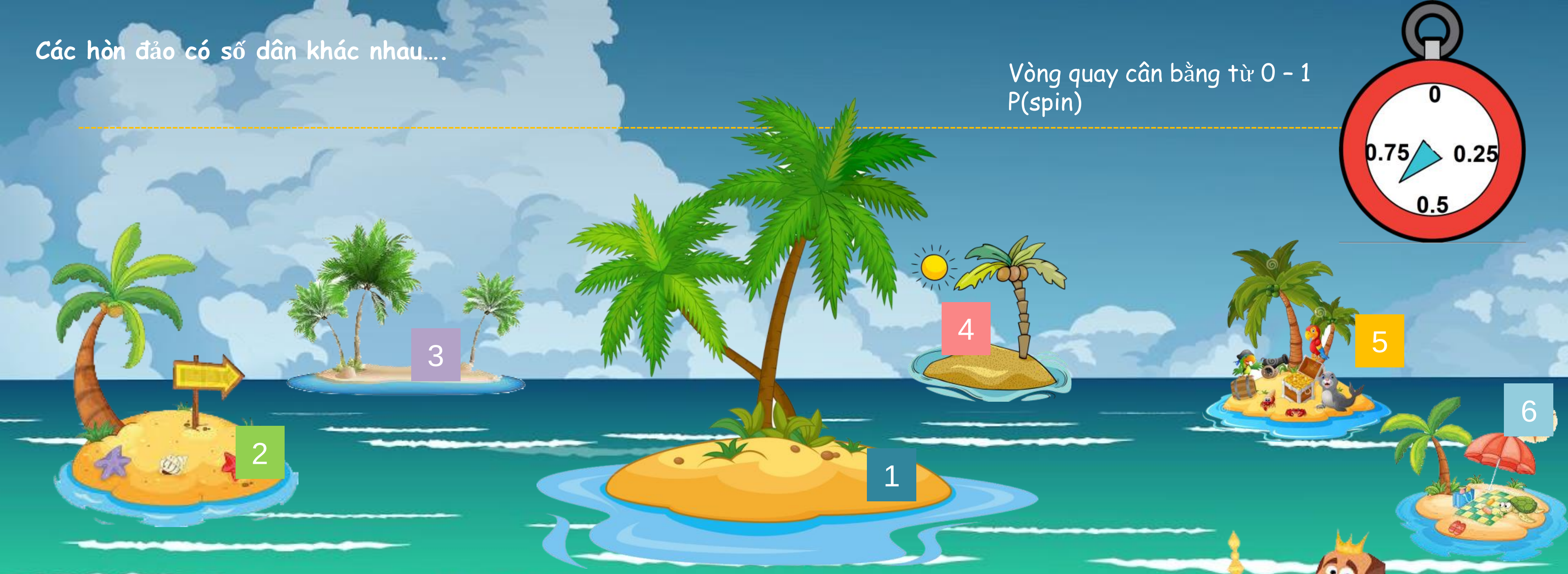
- Thuật toán dạng mô phỏng

- Có nhiều phương pháp sử dụng trong Bayesian: Metropolis–Hastings, Gibbs, Hamiltonian Monte Carlo (HMC), No-U Turn Sampler (NUTS)…

- Metropolis–Hastings

Các hòn đảo có số dân khác nhau….

Vòng quay cân bằng từ 0 – 1
P(spin)



Nếu hòn đảo định tới có số dân lớn hơn hòn đảo hiện tại → đi

Nếu nhỏ hơn. Tính P(move) = Dân số (proposed)/Dân số (current)

Nếu P(move) > P(spin) → đi
Nếu không, ở lại thêm 1 ngày….

Số ngày ở lại mỗi hòn đảo ~ số dân của hòn đảo đó !!!

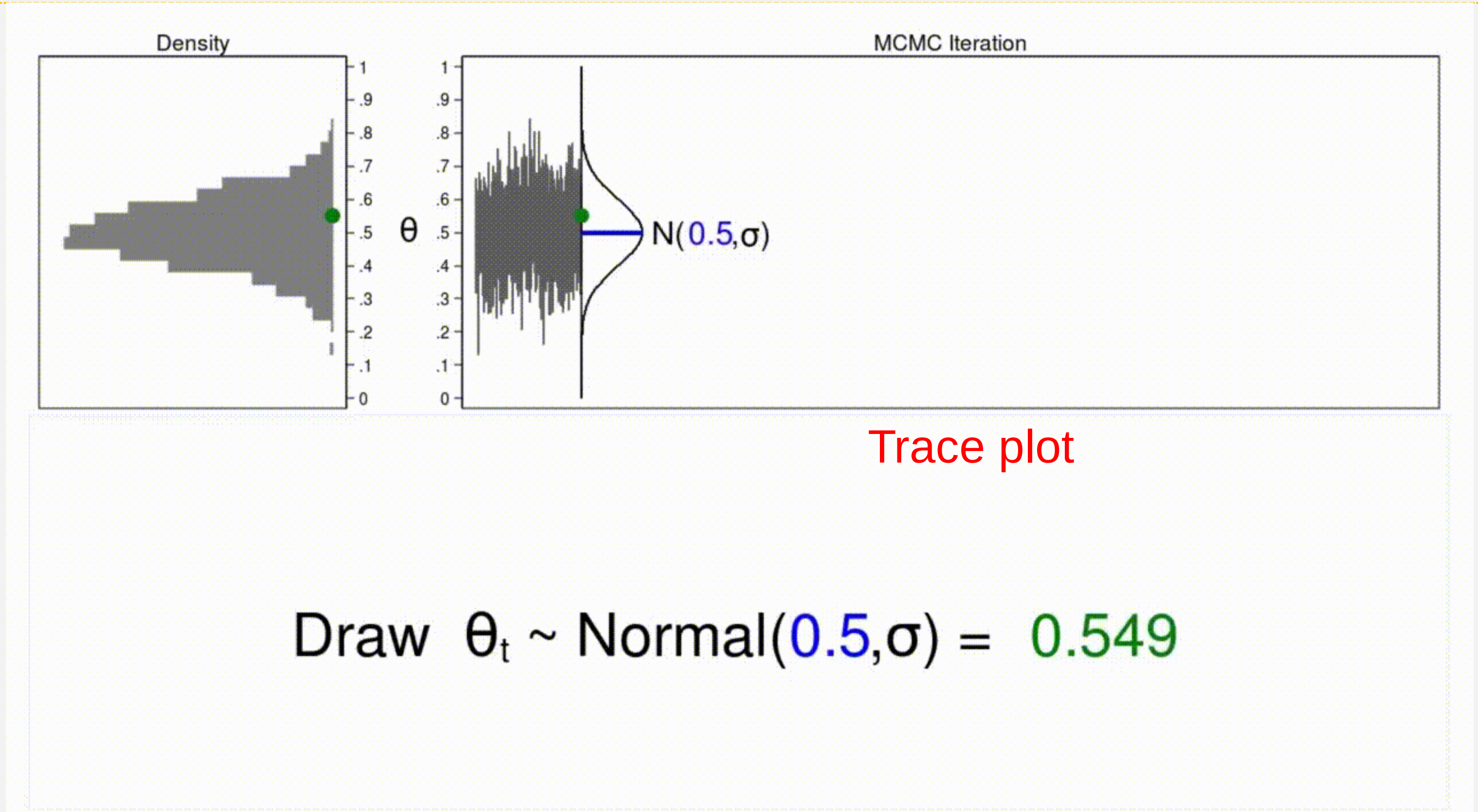
Metropolis-Hastings

- Mô phỏng Monte Carlo
- Markov process
- Metropolis-Hastings algorithms

“Random walk”

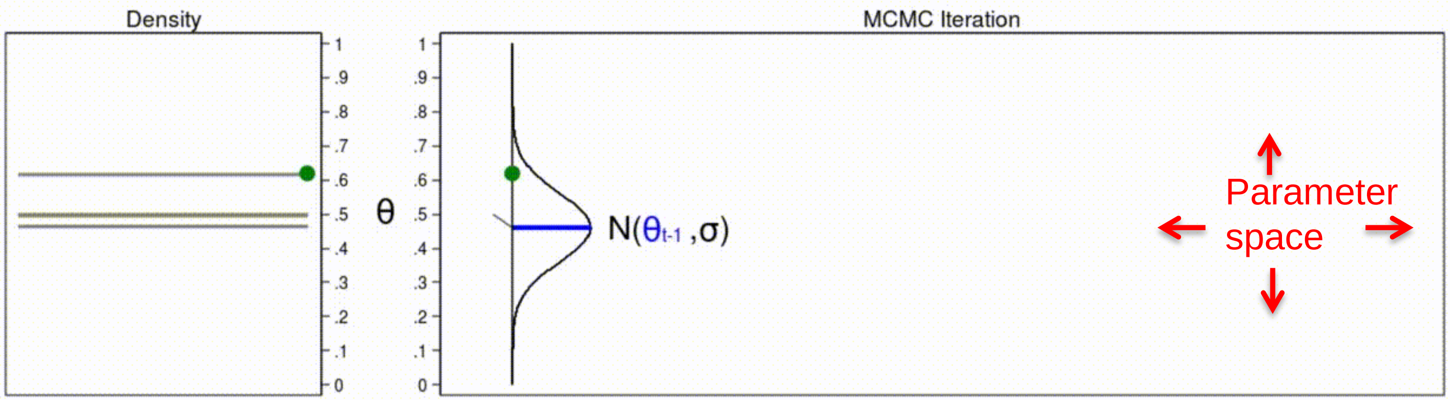
$$P(\theta|Data) = \frac{P(Data|\theta) * P(\theta)}{P(Data)}$$

Monte Carlo



Source: https://www.youtube.com/watch?v=OTO1DygELpY&index=2&list=PLN5IskQdgXWktwVOxs3vAVkI4jpMX3pIv

Markov process



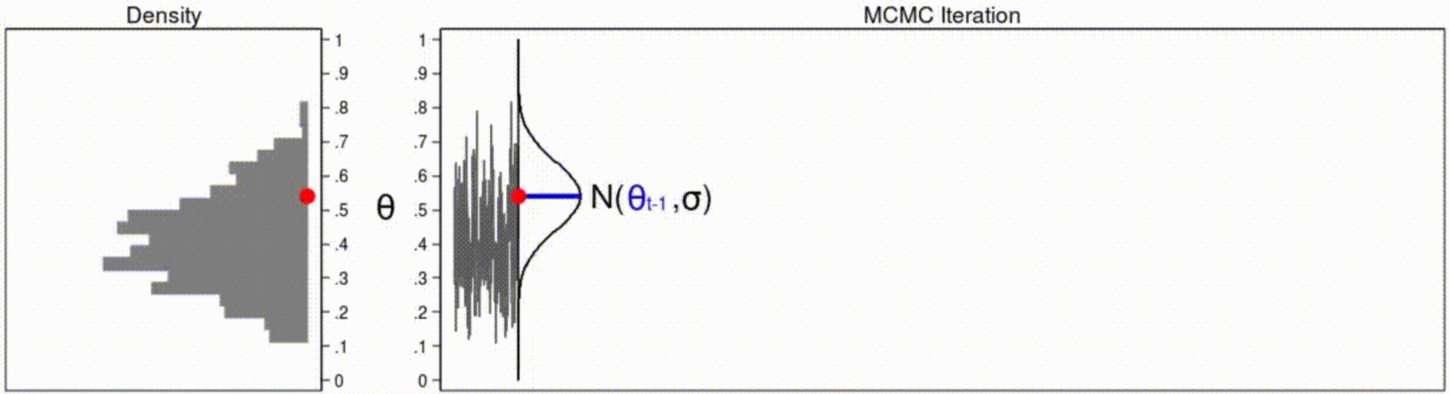
Random walk

Draw $\theta_t \sim \text{Normal}(\theta_{t-1}, \sigma)$

$\text{Normal}(0.460, \sigma) = 0.621$

Giá trị trung bình của phân phối ở vòng lặp sau chính bằng giá trị θ vừa lấy từ vòng lặp trước

Metropolis-Hastings algorithms



Step 1: $r(\theta_{new}, \theta_{t-1}) = \dfrac{\text{Posterior}(\theta_{new})}{\text{Posterior}(\theta_{t-1})} = \dfrac{\text{Beta}(1,1,0.767) \times \text{Binomial}(10,4,0.767)}{\text{Beta}(1,1,0.540) \times \text{Binomial}(10,4,0.540)} = 0.069$

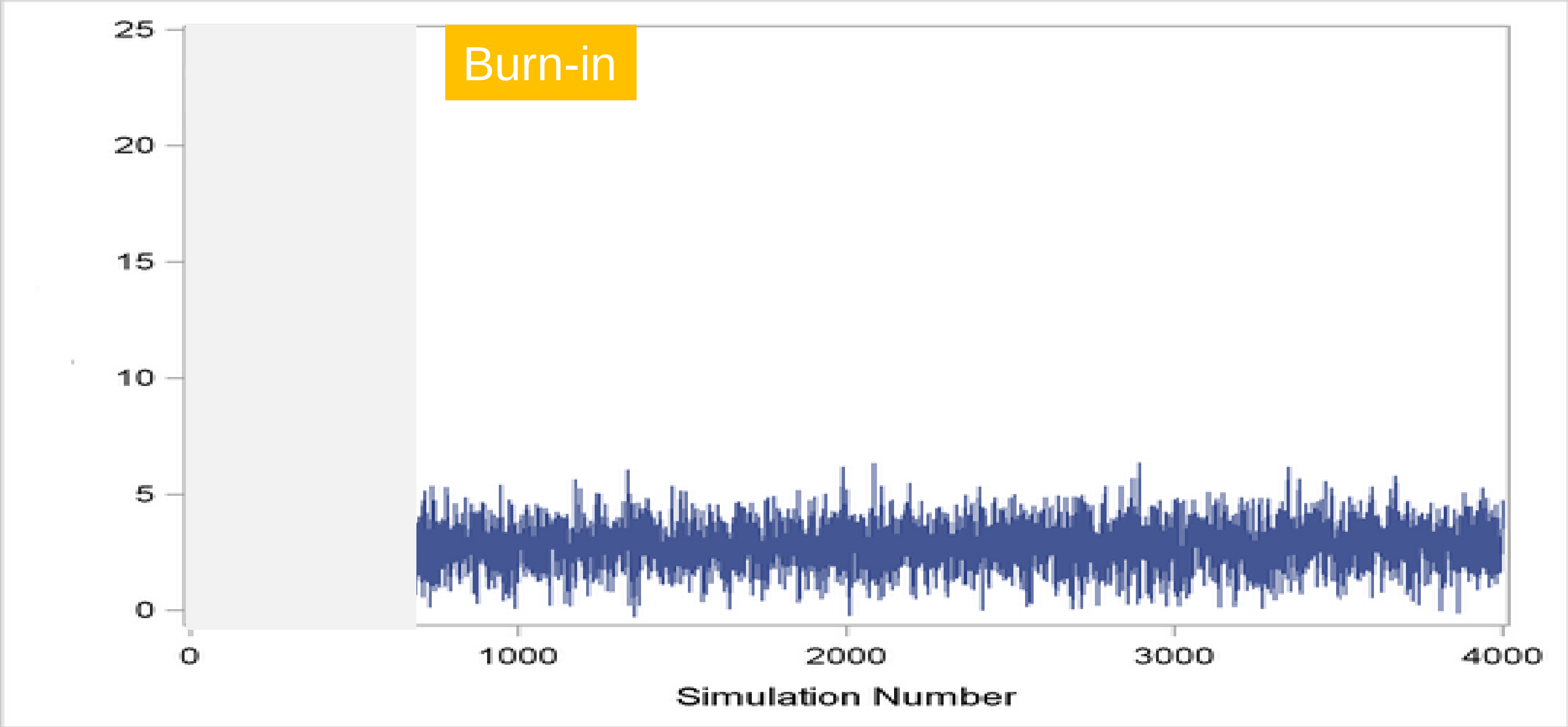
Step 2: Acceptance probability $\alpha(\theta_{new}, \theta_{t-1}) = \min\{r(\theta_{new}, \theta_{t-1}), 1\} = \min\{0.069, 1\} = 0.069$

Step 3: Draw $u \sim \text{Uniform}(0,1) = 0.233$

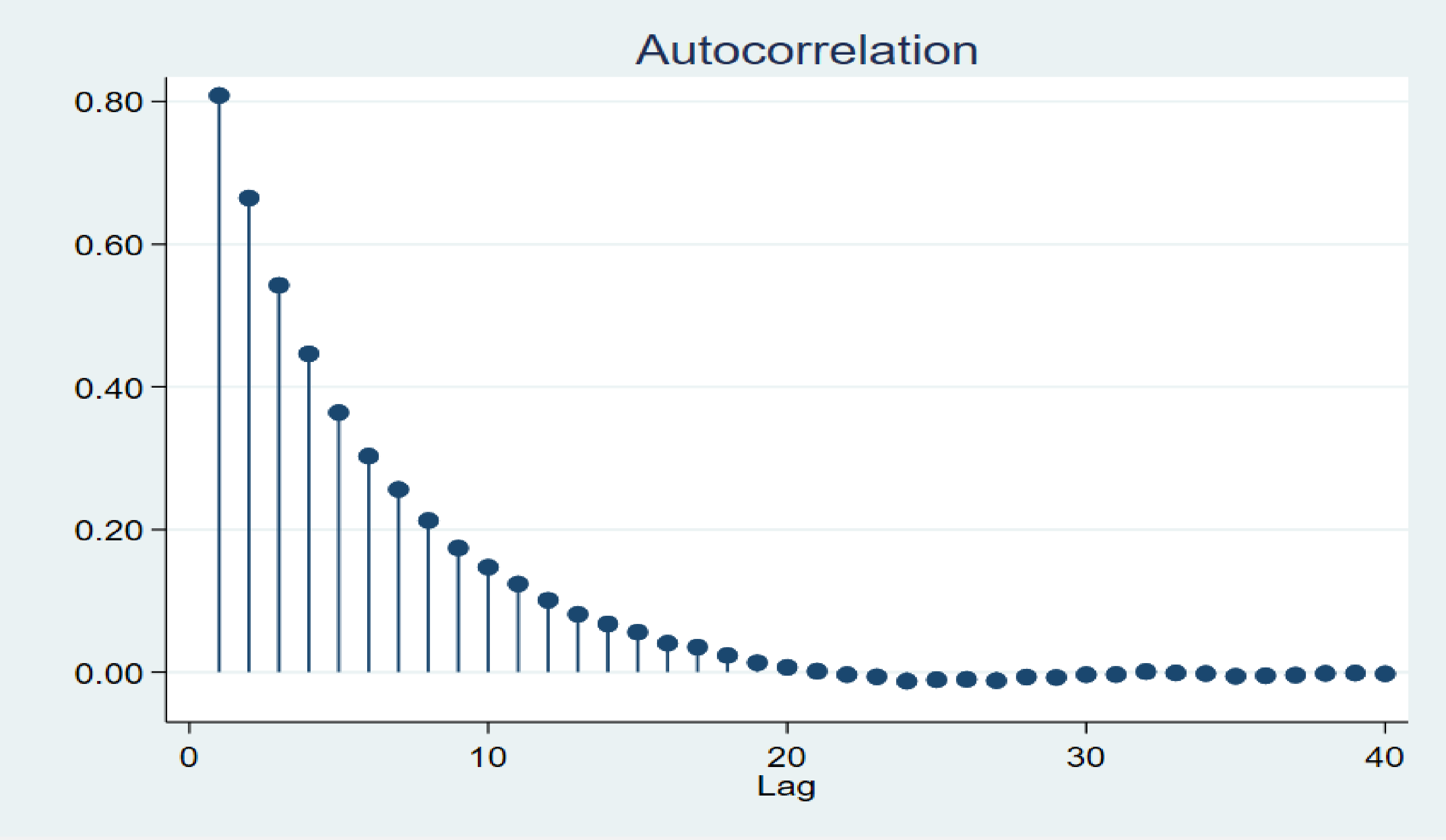
Step 4: If $u < \alpha(\theta_{new}, \theta_{t-1})$ $\rightarrow$ If $0.233 < 0.069$ Then $\theta_t = \theta_{new} = 0.767$

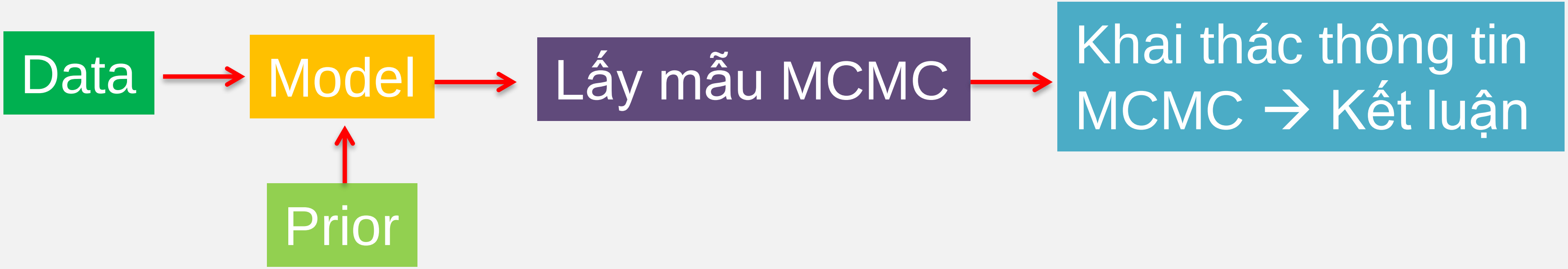
Otherwise $\theta_t = \theta_{t-1} = 0.540$

Burn-in (warm up)



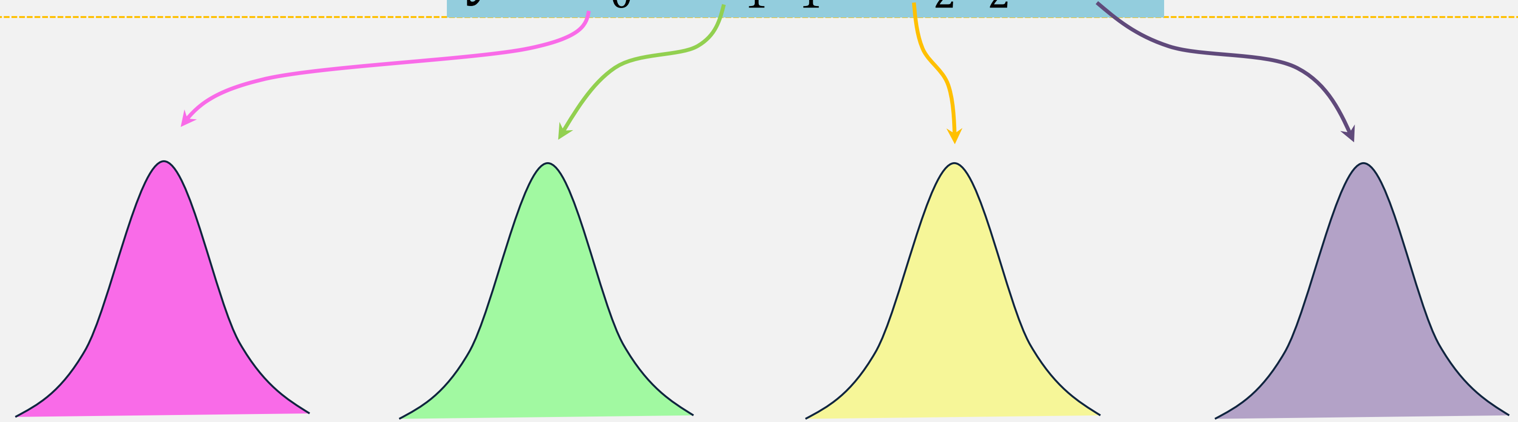
Thinning




Data
Model
Prior
Lấy mẫu MCMC
Khai thác thông tin MCMC → Kết luận

Khai thác thông tin từ phân phối hậu nghiệm

$$y = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \varepsilon$$



- Khai thác thông tin gì từ phân phối này?
- Point estimates (central tendency)
- 95%, 99% Credible interval hoặc Highest Density Interval (HDI)

Point estimate

- Mean hay median hay mode?

- Hàm loss
  - L_0: 0/1 loss
  - L_1: linear loss
  - L_2: squared loss

- Chọn chỉ số sao cho loss nhỏ nhất

L_0: 0/1 loss

$$L_{0,}\, i(g) = \begin{cases} 0 \text{ nếu } g = xi \\ 1 \text{ nếu } g \neq xi \end{cases}$$

```
(47, 33, 35, 32, 19, 33, 34, 36, 47, 32, 35, 41, 32, 29, 35,
 25, 32, 36, 20, 47, 37, 32, 35, 25, 37, 40, 36, 38, 40, 35, 49,
 23, 33, 35, 38, 28, 36, 4, 28, 45, 37, 39, 34, 41, 28, 33, 27,
 26, 30, 34, 23)
# mean(x) = 33.45098
# median(x) = 34
# mode(x) = 35
```

➔ L_0 nhỏ nhất khi g là **mode** của x

L_1: linear loss

$$L_1(g) = \sum |xi - g|$$

```
(47, 33, 35, 32, 19, 33, 34, 36, 47, 32, 35, 41, 32, 29, 35,
 25, 32, 36, 20, 47, 37, 32, 35, 25, 37, 40, 36, 38, 40, 35, 49,
 23, 33, 35, 38, 28, 36, 4, 28, 45, 37, 39, 34, 41, 28, 33, 27,
 26, 30, 34, 23)
# mean(x) = 33.45098
# median(x) = 34
# mode(x) = 35
```

➔ L_1 nhỏ nhất khi g là **median** của x

```r
x <- c(47, 33, 35, 32, 19, 33, 34, 36, 47, 32, 35, 41, 32, 29, 35,
       25, 32, 36, 20, 47, 37, 32, 35, 25, 37, 40, 36, 38, 40, 35, 49,
       23, 33, 35, 38, 28, 36, 4, 28, 45, 37, 39, 34, 41, 28, 33, 27,
       26, 30, 34, 23)
# mean(x) = 33.45098
# median(x) = 34

j1 = rep(NA, length(x))
j2 = rep(NA, length(x))
for (i in 1:length(x)) {
    j1[i] = abs(x[i] - mean(x))
    j2[i] = abs(x[i] - median(x))
    loss_mean = sum(j1)
    loss_median = sum(j2)
}
loss_mean
# 284.7451
loss_median
# 282
```

L_2: Squared loss

$$L_2(g) = \sum(xi - g)2$$

```
(47, 33, 35, 32, 19, 33, 34, 36, 47, 32, 35, 41, 32, 29, 35,
 25, 32, 36, 20, 47, 37, 32, 35, 25, 37, 40, 36, 38, 40, 35, 49,
 23, 33, 35, 38, 28, 36, 4, 28, 45, 37, 39, 34, 41, 28, 33, 27,
 26, 30, 34, 23)
# mean(x) = 33.45098
# median(x) = 34
# mode(x) = 35
```

→ L_2 nhỏ nhất khi g là **mean** của x

```r
x <- c(47, 33, 35, 32, 19, 33, 34, 36, 47, 32, 35, 41, 32, 29, 35,
       25, 32, 36, 20, 47, 37, 32, 35, 25, 37, 40, 36, 38, 40, 35, 49,
       23, 33, 35, 38, 28, 36, 4, 28, 45, 37, 39, 34, 41, 28, 33, 27,
       26, 30, 34, 23)
# mean(x) = 33.45098
# median(x) = 34

j1 = rep(NA, length(x))
j2 = rep(NA, length(x))
for (i in 1:length(x)) {
    j1[i] = (x[i] - mean(x))^2
    j2[i] = (x[i] - median(x))^2
    loss_mean = sum(j1)
    loss_median = sum(j2)
}
loss_mean
# 3124.627
loss_median
# 3140
```

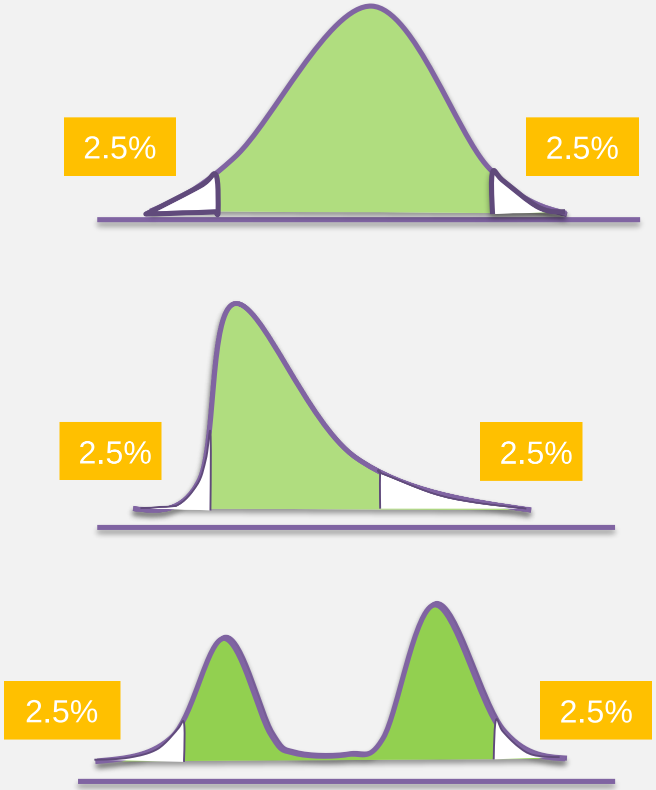
Interval

Equal-tailed Credible Interval (CI)

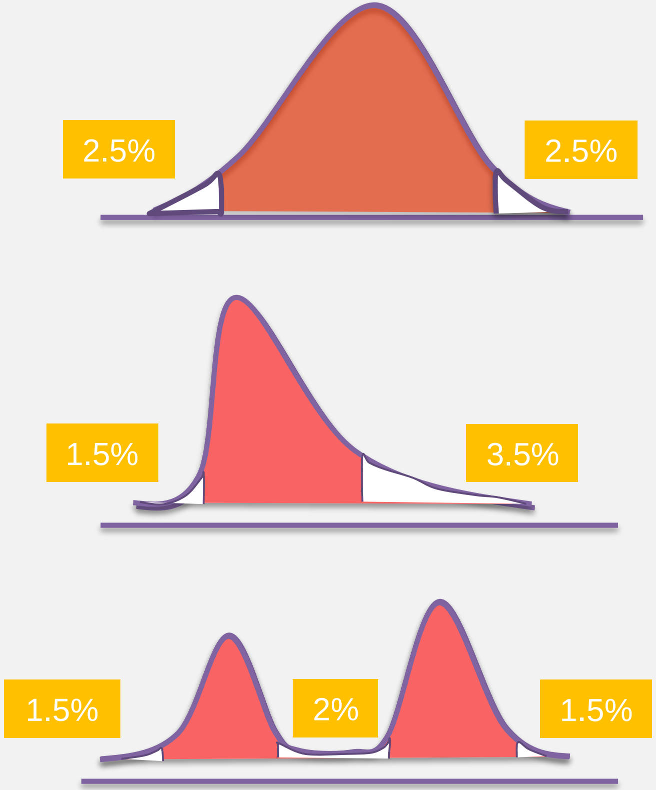
Hay

Highest Density Interval (HDI)

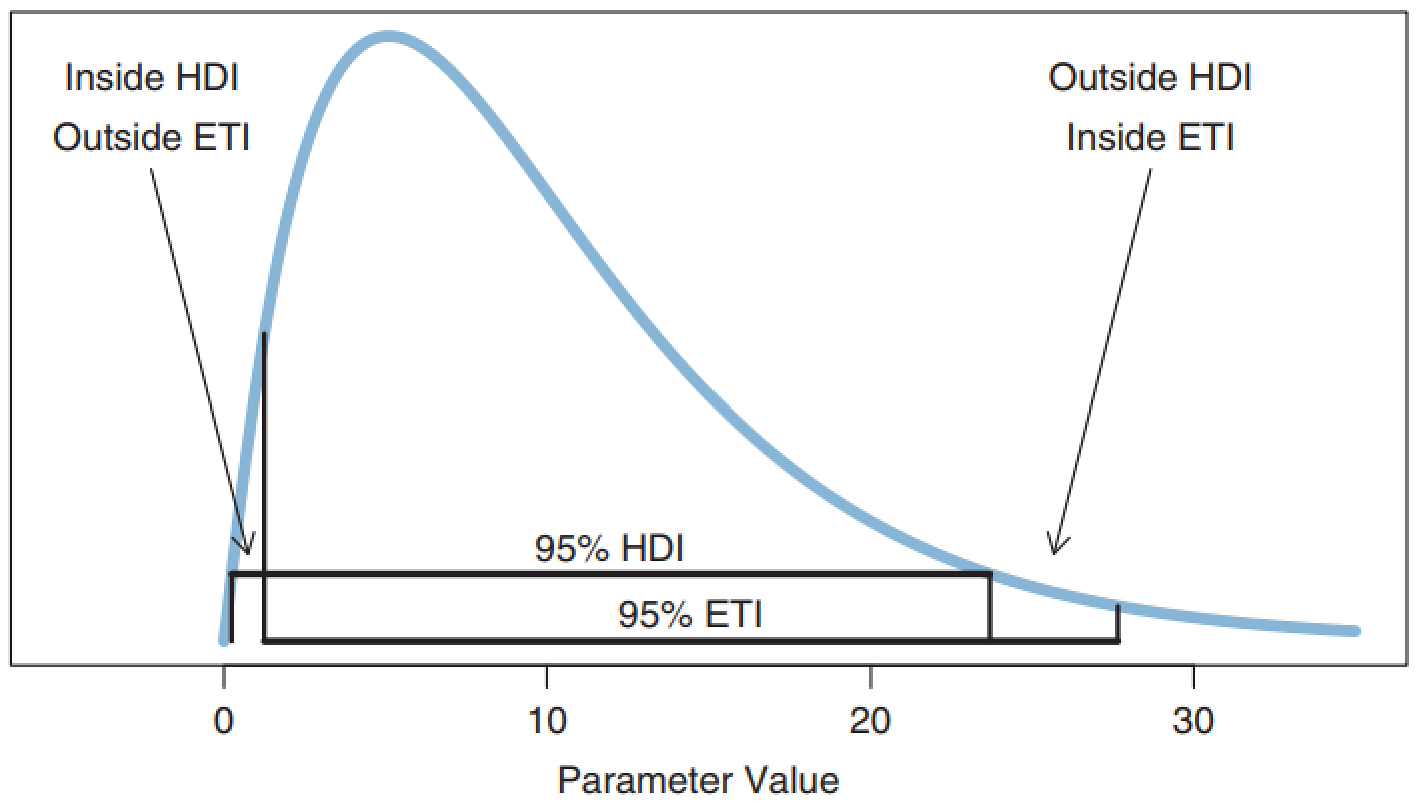
Equal-tailed Credible Interval



Highest Density Interval



ETI vs. HDI

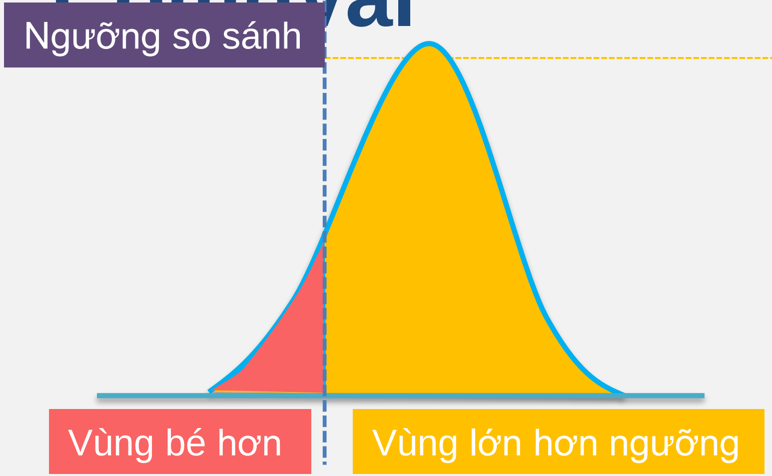

Inside HDI
Outside ETI
Outside HDI
Inside ETI
95% HDI
95% ETI
0
10
20
30
Parameter Value

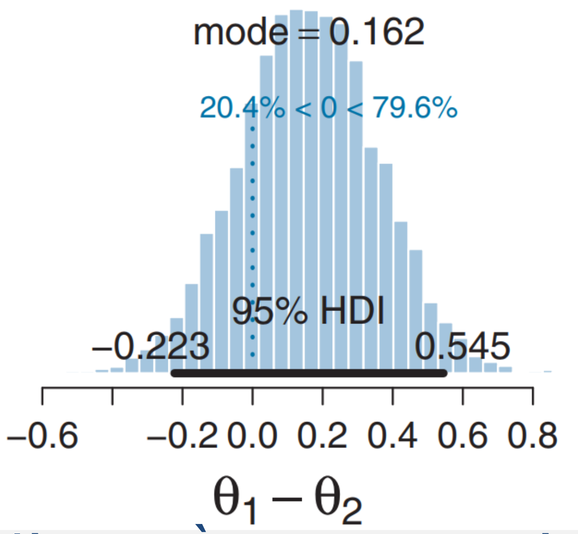
Kiểm định giả thuyết

Kiểm định giả thuyết

- Bộ 3:
  - Comp value (ngưỡng so sánh)
  - ROPE (Region of Practical Equivalent)
  - Bayes Factor

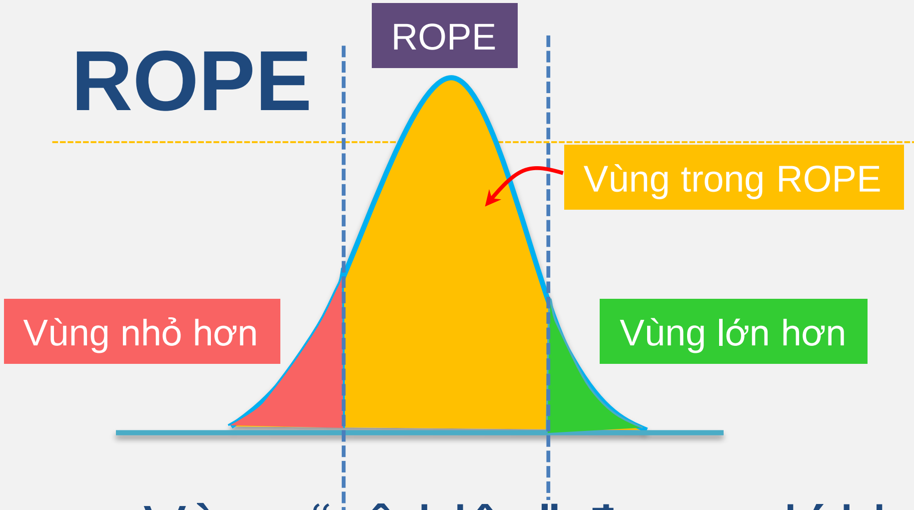
CompVal



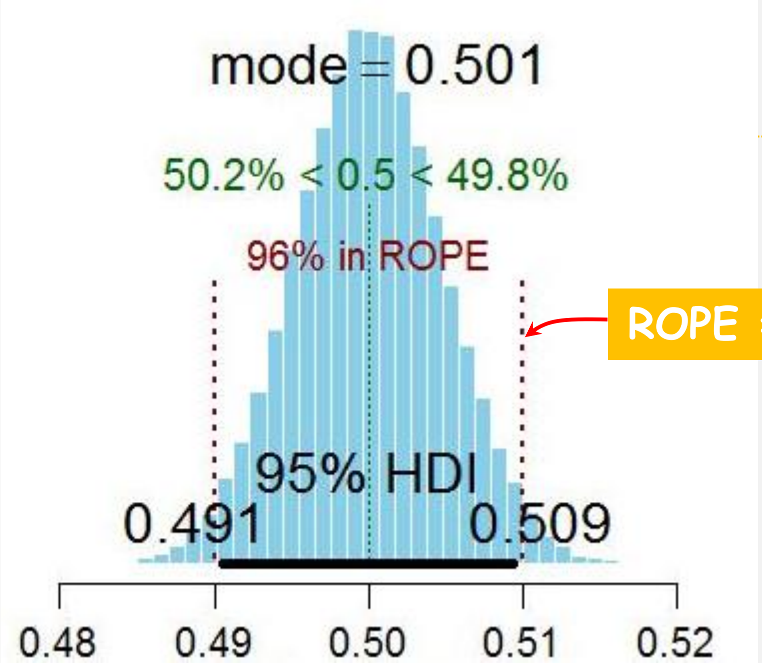


- CompVal chia phân phối hậu nghiệm là 2 phần: lớn hơn và bé hơn
- Ví dụ: phân phối hậu nghiệm của θ_1 - θ_2 có 95% HDI = -0.223 – 0.545)
- 79.6% phân phối hậu nghiệm nằm trên ngưỡng 0

ROPE

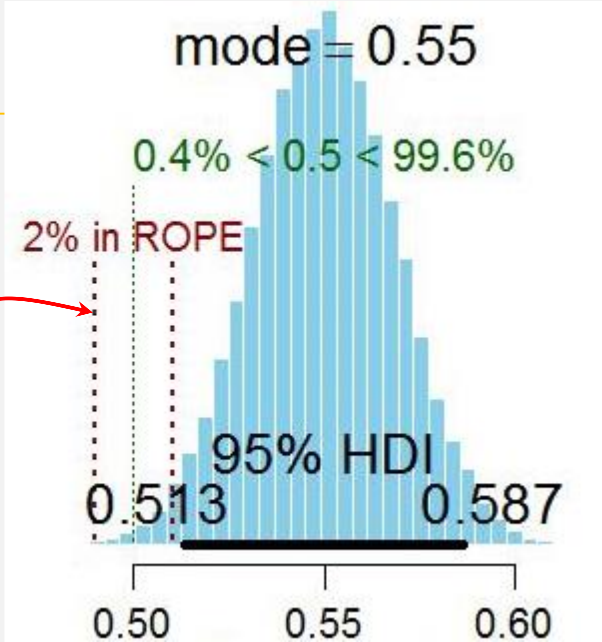


- Vùng "vô hiệu" được giới hạn bởi 2 ngưỡng trên-dưới, chia pp hậu nghiệm làm 3 vùng: vùng trong ROPE, vùng lớn hơn, vùng nhỏ hơn
- 95%HDI nằm hoàn toàn trong ROPE được coi như vô hiệu (không có khác biệt, không có hiệu quả…)
- 95%HDI nằm hoàn toàn ngoài ROPE → có hiệu ứng



$\theta_1 - \theta_2$

ROPE = 0.49 – 0.51



$\theta_3 - \theta_4$

- 95% HDI của θ_1-θ_2= 0.491 – 0.509 nằm hoàn toàn trong ROPE → sự khác biệt không có ý nghĩa trong thực hành
- 98% pp hậu nghiệm của θ_3-θ_4nằm hoàn toàn ngoài ROPE → θ_3 lớn hơn θ_4 có ý nghĩa trong thực hành

ROPE

- Tùy thuộc vào lĩnh vực, giả thuyết …
- ROPE lấy từ đâu?
- Bằng chứng từ y văn
- Ý kiến chuyên gia
- Kinh nghiệm thực hành
- …

Bayes Factor

- H_1 vs. H_2
- Giả thuyết nào có khả năng hơn?

Posterior odds = Bayes factor × Prior odds

$$\text{PO (H1:H2)} = \frac{P(H_1|data)}{P(H_2|data)}$$

$$= \frac{P(data|H_1) * P(H_1)/P(data)}{P(data|H_2) * P(H_2)/P(data)}$$

$$= \frac{P(data|H_1) * P(H_1)}{P(data|H_2) * P(H_2)}$$

Bayes factor

Prior odds

$$= \frac{P(data|H_1)}{P(data|H_2)} * \frac{P(H_1)}{P(H_2)}$$

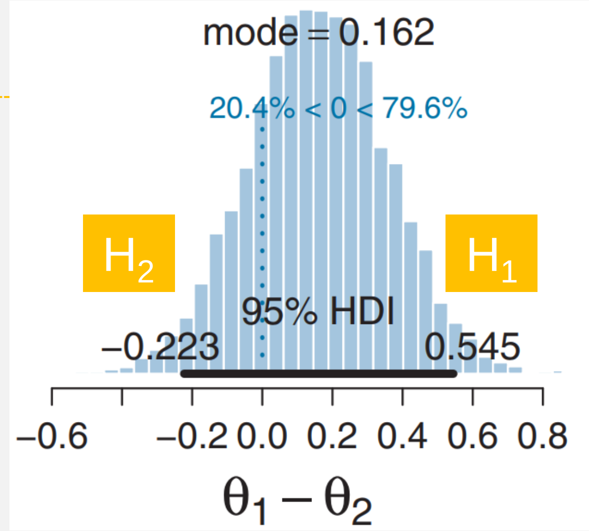
Bayes Factor

Kass and Raftery (1995)

$BF(H_1:H_2)$	$2\log[BF(H_1:H_2)]$	Bằng chứng chống lại H_2
1 - 3	0 - 2	Bare mention
3 - 20	2 - 6	Positive
20 - 150	6 - 10	Strong
> 150	> 10	Very strong

H_1: $\theta_1 - \theta_2 > 0$

H_2: $\theta_1 - \theta_2 < 0$



- Posterior odds = $P(H_1|data) / P(H_2|data)$ = 79.6/20.4 = 3.9
- Prior odds = $P(H_1) / P(H_2) \sim 1$
- Bayes Factor = Posterior odds / Prior odds = 3.9

➔ có bằng chứng cho thấy H_1 có khả năng xảy ra hơn

Một số phần mềm sử dụng trong Bayesian

MCMC sampler

- Dự án BUGS (*Bayesian inference Using Gibbs Sampling*) bắt đầu vào 1989 ở MRC Biostatistics Unit, Cambridge
- ✓ WinBUGS (1997)
- ✓ OpenBUGS
- ✓ MultiBUGS
- JAGS (*Just another Gibbs sampler*) - Martyn Plummer, 2007 – 2017
- Stan - Stan Development Team (*đặt theo tên của Stanislaw Ulam*), 2012 – 2017. Stan sử dụng HMC và NUTS
- Greta (Nick Golding, 2018) được đặt tên theo nhà toán học Grete Hermann (1901 – 1984). Greta sử dụng TensorFlow backend

Phần mềm thống kê

- R
- Python (PyStan, PyMC3)
- SAS
- AMOS
- Mplus
- Stata
- ….

Thank you !