TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI

———————o0o———————

ĐỒ ÁN TỐT NGHIỆP

PHƯƠNG PHÁP SUY LUẬN BAYES VÀ ỨNG DỤNG CHO MÔ HÌNH HỒI QUY TUYẾN TÍNH

Bùi Thị Quyên

quyen.bt163420@sis.hust.edu.vn

Chuyên ngành: Toán ứng dụng

Giảng viên hướng dẫn:	TS. Đỗ Văn Cường
Bộ môn:	Toán ứng dụng
Viện:	Toán ứng dụng và Tin học

HÀ NỘI, 2022

Lời cảm ơn

Em xin chân thành cảm ơn thầy Đỗ Văn Cường - Giảng viên bộ môn Toán ứng dụng, Viện Toán ứng dụng và Tin học đã tận tình hướng dẫn, góp ý trong quá trình thực hiện đồ án. Do thời gian cũng như kiến thức còn có hạn, đồ án có thể có những thiếu sót, rất mong nhận được những ý kiến đóng góp của thầy cô và các bạn đọc.

Hà Nội, tháng ... năm 2022

Tác giả đồ án

Bùi Thị Quyên

Mục lục

Danh mục từ viết tắt và kí hiệu

$\hat{\theta}_{MLE}$	Ước lượng hợp lý cực đại (Maximum likelyhood estimation)
$\hat{\theta}_{MAP}$	Ước nghiệm hậu nghiệm lớn nhất (Maximum a posterior)
$\hat{\theta}_{MSE}$	Ước lượng trung bình bình phương sai số nhỏ nhất (Minimum of mean square errors).
$i.i.d$	Độc lập cùng phân phối (Independently identical distributed)
v^2	$\frac{1}{n}\sum(x_i-\mu)^2$
s^2	$\frac{1}{n}\sum(x_i-\bar{x})^2$
S^{*2}	$\frac{1}{n-1}\sum(x_i-\bar{x})^2$
$\underline{x}$	$(x_1, x_2, ..., x_n)$

Danh sách hình vẽ

Lời mở đầu

Tổng quan vấn đề

Trong phân tích dữ liệu Frequentist và Bayesian là hai trường phái có nhiều đối lập, tuy nhiên cả hai trường phái vẫn song song tồn tại và có những ưu thế cũng như những hạn chế nhất định so với trường phái còn lại. Nguồn gốc quan trọng nhất của sự khác biệt Bayesian và Frequentist là cách hiểu về xác suất. Theo định nghĩa cổ điển về xác suất , cho một phép thử có n kết cục đồng khả năng, trong đó có m kết cục thuận lợi cho A, khi đó xác suất xảy ra A bằng m/n, tưc là bằng số kết cục thuận lợi cho A chia tổng số kết cục có thể. Tuy nhiên, cách tính xác suất này rất hạn chế, chỉ cho các loại phép thử hữu hạn kết cục đồng khả năng. Để khắc phục hạn chế về việc chỉ áp dụng cho các phép thử có kết cục hữu hạn, người ta đã đưa ra định nghĩa hình học của xác suất. Tuy nhiên điều kiện đồng khả năng không phải lúc nào cũng được đảm bảo. Có một cách khác để xác định xác suất của một sự kiện. Giả sử tiến hành một loạt n_1 phép thử cùng loại, nếu sự kiện A nào đó xuất hiện trong m_1 phép thử, thì ta gọi m_1/n_1 là tần suất xuất hiện A trong loạt phép thử đã cho. Tương tự với loại phép thử thứ hai, thứ ba, ..., ta có các tần suất tương ứng $m_2/n_2, m_3/n_3$,... Trên cơ sở quan sát lâu dài các thí nghiệm khác nhau, người ta nhận thấy tần suất xuất hiện các sự kiện có tính ổn định, thay đổi rất ít trong các loạt phép thử khác nhau và sao động xung quanh một hằng số xác định. Sự khác biệt đó càng ít khi số phép thử tăng lên nhiều lần. Đặc tính ổn định của tần suất khi số phép thử tăng lên khá lớn cho phép ta định nghĩa xác suất của sự kiện là trị số ổn định đó của tần suất xuất hiện sự kiện. Người ta lấy tần suất khi số phép thử đủ lớn làm xác suất xuất hiện sự kiện. Trường phái Frequentist chính là hiểu xác suất theo nghĩa tần suất như vậy. Ví dụ, xác suất tung đồng xu mặt sấp bằng 0,5 có nghĩa là ta tung đồng xu nhiều lần và đếm số lần xuất hiện mặt sấp, càng tung nhiều lần thì tần suất xuất hiện nó đâu đó xấp xỉ con số 0,5. Tuy vậy, trường phái Bayesian cho rằng, con số 0.5 chỉ là một sự đo lường không chắc chắn. Tức là, với Bayesian, xác suất là một ý kiến chủ quan, phụ thuộc vào niềm tin của người phân tích, còn với Frequentist, xác suất là một con số khách quan, hoàn toàn độc lập với niềm tin của người phân tích.

Sự khác biệt về cách hiểu xác suất ở trên dẫn tới những sự khác biệt quan trọng trong việc mô hình dữ liệu. Gia sử chúng ta cần ước lượng xác suất mặt sấp của một đồng xu. Với Frequentist, vì xác suất này là khách quan, nó là một tham số cố định. Trong khi đó, với Bayesian, vì xác suất là chủ quan, nó biến động tuỳ theo người phân tích, nên nó sẽ là một biến ngẫu nhiên. Trước khi thu thập dữ liệu, bạn cũng đã có một niềm tin nhất định, thể hiện qua một phân phối tiên nghiệm. Hệ quả là, trường phái Frequentist xây dựng mô hình tìm giá trị của tham số hợp lý nhất với những gì quan sát được, còn trường phái Bayesian xây dựng mô hình để cập nhật niềm tin về biến ngẫu nhiên. Các phương pháp Frequentist thường yêu cầu số lượng quan sát lớn, còn phương pháp của trường phái Bayesian có thể áp dụng với tất cả số lượng quan sát. Đây là một ưu điểm cực lớn của trường phái Bayesian.

Tóm lại, các mô hình Bayesian đã tăng tính linh hoạt và tăng sự lựa chọn cho việc phân tích dữ liệu hơn rất nhiều cho khoa học dữ liệu nói chung, còn trường phái Frequentist thể hiện điểm mạnh về tính hiệu quả trong việc tính toán và tính giải thích. Ngày nay, sự phát triển và nhu cầu cao của việc phân tích dữ liệu làm cho sự phân chia trường phái ít quan trọng hơn; thay vào đó, người phân tích dữ liệu cần sử dụng linh hoạt cả 2 trường phái để phân tích dữ liệu một cách hiệu quả nhất.

Mục tiêu của đồ án

Phát triển bài toán thống kê theo phương pháp suy luận Bayes.

Kết quả chính

- Công thức cụ thể và chứng minh chi tiết.
- Quy trình và lập trình trên R.
- Áp dụng phương pháp Bayes cho số liệu thực tế.

Cấu trúc của đồ án

Nội dung đồ án gồm 4 phần chính:

- **Chương 1: Cơ sở lý thuyết.** Trình bày về định lý Bayes, một số phân phối xác suất, nêu tóm tắt về một số phương pháp ước lượng tham số, trình bày phương pháp Monte Carlo Markov Chain và ví dụ.
- **Chương 2: Ước lượng tham số.** Trình bày ước lượng tham số (Ước lượng hợp lý cực đại và ước lượng Bayes) cho một vài phân phối xác suất, trình bày suy luận Bayes cho mô hình hồi quy tuyến tính, nêu khái niệm, và trình bày các phương pháp ước lượng tham số mô hình, mô phỏng các phương pháp ước lượng tham số.
- **Chương 3: Ứng dụng.** Trình bày một số kết quả về dữ liệu thật sau khi đã áp dụng phương pháp Bayesian sử dụng Monte Carlo Markov Chain.
- **Chương 4: Kết luận.** Rút ra một số kết luận về phương pháp ước lượng hợp lý cực đại và phương pháp Bayes.

Chương 1

Cơ sở lý thuyết

Trong chương này trình bày về định lý Bayes, một số phân phối xác suất, nêu tóm tắt về một số phương pháp ước lượng tham số.

1.1 Định lý Bayes

1.1.1 Trường hợp rời rạc

Giả sử ta có một nhóm đầy đủ $A_1, A_2, ..., A_n$ và có thêm sự kiện H nào đó. Ta muốn xác định xác suất $P(A_i|H), i = \overline{1,n}$ khi đó công thức Bayes đưa ra cách tính như sau:

$$P(A_i|H) = \frac{P(A_i)P(H|A_i)}{P(H)}$$

Hay

$$P(A_i|H) = \frac{P(A_i)P(H|A_i)}{\sum_{i=1}^{n} P(A_i)P(H|A_i)}$$

Các xác suất $P(A_i)$ đã được xác định từ trước, gọi là xác suất tiên nghiệm. Các xác suất $P(A_i|H)$ được xác định sau khi đã có kết quả thí nghiệm thể hiện qua sự xuất hiện của sự kiện H và được gọi là xác suất hậu nghiệm.

1.1.2 Trường hợp liên tục

Trong trường hợp biến ngẫu nhiên liên tục, ta có định lý sau đây.

Định lý 1.1.1 (Định lý Bayes) *Đặt $\mathcal{P} = \{P_\theta | \theta \in \Theta\}$ là một mô hình xác suất với tham số θ thuộc Θ và quan sát D trong không gian mẫu $\mathcal{X}$. Kí hiệu $\pi(\theta)$ là phân phối tiên nghiệm của tham số θ, $f(D|\theta)$ là hàm mật độ xác suất của mô hình, $p(\theta|D)$ là phân phối hậu nghiệm của tham số θ được cho bởi quan sát D. Định lý Bayes phát biểu rằng:*

$$p(\theta|D) = \frac{f(D|\theta) \times \pi(\theta)}{\int_\Theta f(D|\theta) \times \pi(\theta) d\theta},$$

trong đó, $f(D|\theta)$ được gọi là hàm hợp lý (Likelihood), $\int_\Theta f(D|\theta) \times \pi(\theta) d\theta$ được gọi là giá trị hợp lý biên. Nó là một hằng số xác định nhưng không dễ tính toán. Do đó, ta thường viết:

$$p(\theta|D) \propto f(D|\theta) \times \pi(\theta).$$

Nghĩa là, phân phối hậu nghiệm của tham số θ tỷ lệ với tích của hàm hợp lý và phân phối tiên nghiệm.

1.2 Một số hàm phân phối xác suất

1.2.1 Phân phối Khi bình phương

Định nghĩa 1.2.1 (Phân phối Khi-bình phương) *Biến ngẫu nhiên X được gọi là tuân theo luật phân phối Khi - bình phương n bậc tự do nếu hàm mật độ xác suất của nó có dạng:*

$$f_X(x) = K.x^{n/2-1} e^{-x/2}, x > 0, n \in N^*$$

với $K = \frac{1}{2^{n/2}\gamma(n/2)}$

Kí hiệu $X \sim \chi^2_n$.

Kỳ vọng, phương sai và mốt

Định lý 1.2.2 *Nếu biến ngẫu nhiên $X \sim \chi_n^2$ thì ta có các kết quả sau:*

1. *Giá trị kỳ vọng: E(X) = n*

2. *Phương sai: Var(X) = 2n*

3. *Mốt: mode(X) = n − 2*

1.2.2 Phân phối Gamma

Ta đưa ra định nghĩa cho phân phối gamma với tham số α và β như sau.

Định nghĩa 1.2.3 (Phân phối gamma) *Ta nói biến ngẫu nhiên $X \in \mathbb{R}^+$ tuân theo phân phối gamma nếu hàm mật độ của nó có dạng*

$$f_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$$

trong đó α là tham số **hình dạng**, *β là một hằng số dương và được gọi là* **cường độ** *của phân phối Gamma.*
Khi đó ta kí hiệu $X \sim Gamma(\alpha, \beta)$

Đồ thị của hàm mật độ

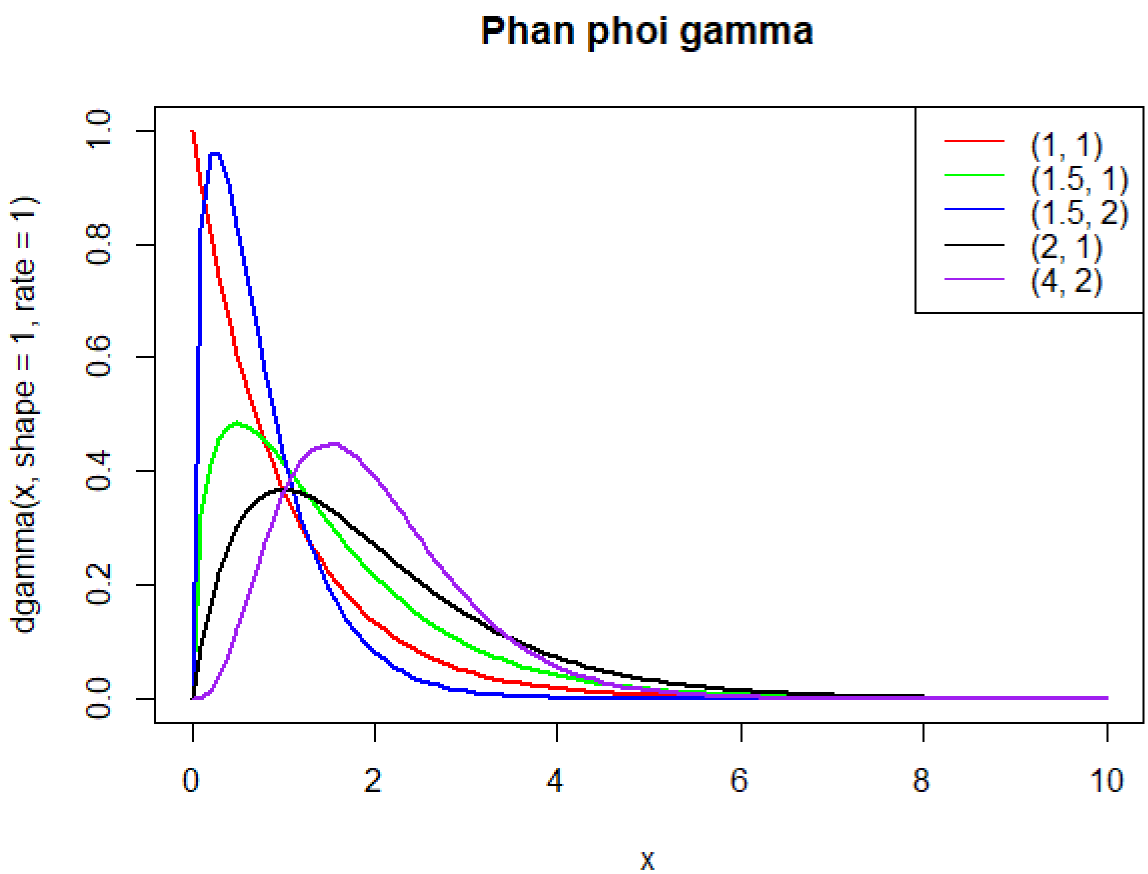
Sau đây là đồ thị của hàm mật độ phân phối gamma với một vài giá trị hình dạng và giá trị cường độ khác nhau.

Kỳ vọng, phương sai và mốt

Định lý 1.2.4 *Nếu biến ngẫu nhiên $X \sim Gamma(\alpha, \beta)$ thì ta có các kết quả sau:*

1. *Giá trị kỳ vọng: E(X) = $\frac{\alpha}{\beta}$*

2. *Phương sai: Var(X) = $\frac{\alpha}{\beta^2}$*



Hình 1.1: Đồ thị của hàm mật độ phân phối Gamma

3. *Mốt:* $Mode(X) = \frac{\alpha-1}{\beta}$

Định lý 1.2.5 *Ta có một số kết quả sau.*

1. *Nếu biến ngẫu nhiên* $X \sim Gamma(\alpha, \beta)$ *thì* $kX \sim Gamma(\alpha, \frac{\beta}{k})$.

2. $\chi^2_n \equiv Gamma(n/2, 1/2)$.

1.2.3 Phân phối Gamma nghịch đảo

Ta đưa ra định nghĩa cho phân phối gamma nghịch đảo với tham số α và β như sau.

Định nghĩa 1.2.6 (Phân phối gamma nghịch đảo) *Ta nói biến ngẫu nhiên $X \in \mathbb{R}^+$ tuân theo phân phối gamma nghịch đảo nếu hàm mật độ của nó có dạng*

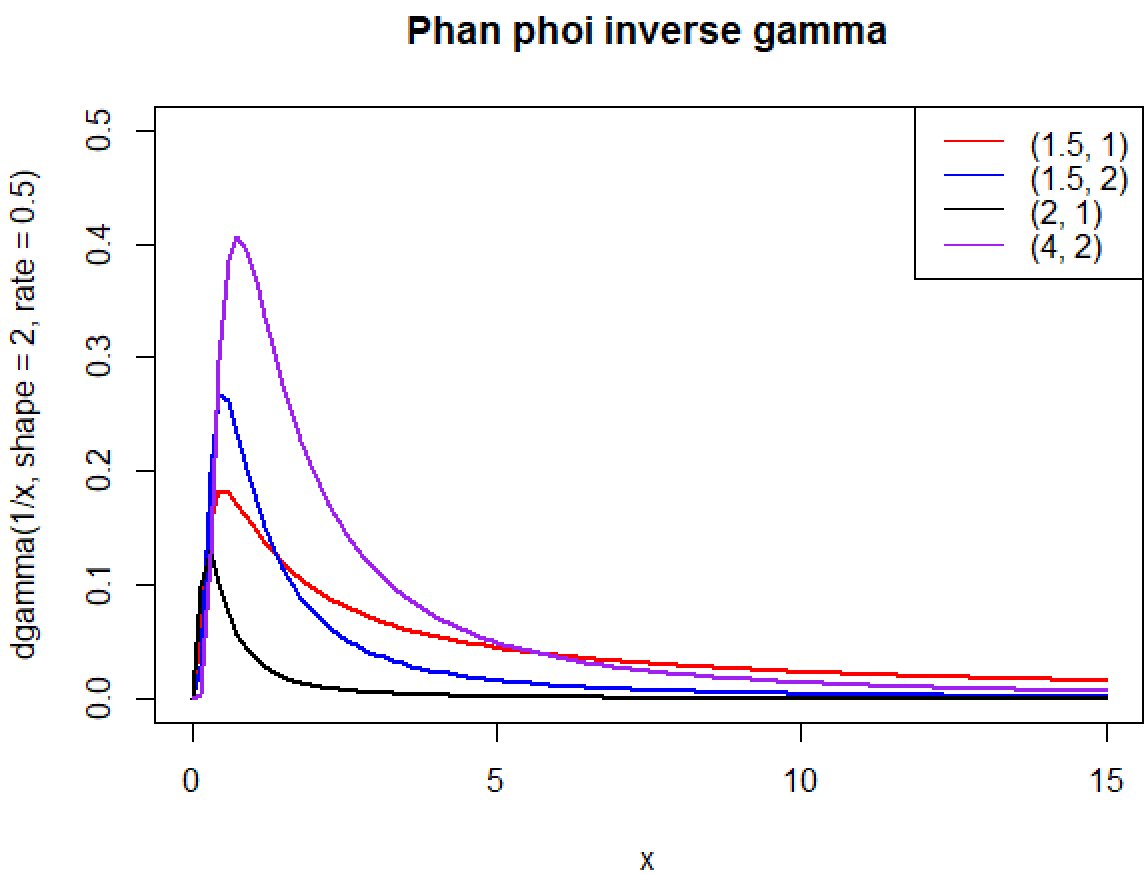
$$f_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} e^{-x/\beta}$$

trong đó α là một hằng số dương, được gọi là tham số **hình dạng** *và β là hằng số dương, được gọi là tham số* **thang đo**.

Khi đó ta kí hiệu $X \sim IGamma(\alpha, \beta)$

Đồ thị của hàm mật độ

Sau đây là đồ thị của hàm mật độ phân phối gamma nghịch đảo với một vài giá trị các tham số khác nhau.



Hình 1.2: Đồ thị của hàm mật độ phân phối Gamma nghịch đảo

Kỳ vọng, phương sai và mốt

Định lý 1.2.7 *Nếu biến ngẫu nhiên* $X \sim IGamma(\alpha, \beta)$ *thì ta có các kết quả sau:*

1. *Giá trị kỳ vọng:* $E(X) = \frac{\beta}{\alpha-1}$, *với* $\alpha > 1$.
2. *Phương sai:* $Var(X) = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)}$ *với* $\alpha > 2$.
3. *Mốt:* $Mode(X) = \frac{\beta}{\alpha+1}$.

Ta có một số tính chất của phân phối Gamma nghịch đảo được phát biểu trong định lý sau đây.

Định lý 1.2.8

1. *Nếu biến ngẫu nhiên* $X \sim IGamma(\alpha, \beta)$ *thì* $kX \sim IGamma(\alpha, k\beta)$.

2. *Nếu biến ngẫu nhiên* $X \sim Gamma(\alpha, \beta)$ *thì* $1/X \sim IGamma(\alpha, \beta)$.

1.2.4 Phân phối chuẩn

Định nghĩa 1.2.9 (Phân phối chuẩn) *Biến ngẫu nhiên* $X \in R$ *tuân theo phân phối chuẩn nếu hàm mật độ của nó có dạng:*

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{1}{2\sigma^2}(x-\mu)^2\}$$

với trung bình $\mu \in R$ *, phương sai* $\sigma^2 \in R^+$

Kí hiệu: $X \sim N(\mu, \sigma^2)$

Định lý 1.2.10 *Nếu các biến ngẫu nhiên* $X_1, ..., X_n \overset{\text{i.i.d}}{\sim} N(0,1)$ *thì* $X_1^2 + X_2^2 + ... + X_n^2 \sim \chi_n^2$.

1.3 Một số phương pháp ước lượng tham số

1.3.1 Phương pháp ước lượng hợp lý cực đại

Giả sử tham số cần ước lượng là θ. Nguyên lý của phương pháp này là ước lượng tham số θ dựa trên hàm của quan sát $\underline{x} = (x_1, x_2, ..., x_n)$ sao cho xác suất thu được các quan sát đó là lớn nhất. Giả sử ta có một biến ngẫu nhiên gốc X tuân theo luật phân phối bất kỳ nào đó với hàm mật độ xác suất là $f_{X|\theta}(x)$, khi đó hàm hợp lý, ký hiệu là $L(\theta)$ được xác định là:

$$L(\theta) = \prod_{i=1}^{n} f(x_i|\theta)$$

Ước lượng hợp lí cực đại chính là việc đi tìm tham số θ sao cho hàm hợp lý là lớn nhất.

1.3.2 Phương pháp ước lượng mô men

Trước khi đi vào phương pháp ta có một số định nghĩa sau:

- $E(X^k)$ là mô men lý thuyết bậc k của phân phối, với $k = 1, 2, ...$
- $E[(X-\mu)^k]$ là mô men lý thuyết bậc k của phân phối về trung bình, với $k = 1, 2, ...$
- $\bar{x^k} = \frac{1}{n}\sum_{i=1}^{n} x_i^k$ là mô men mẫu bậc k, với $k = 1, 2, ...$
- $s^k = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^k$ là mô men mẫu bậc k về trung bình, với $k = 1, 2, ...$

Phương pháp mô men liên quan đến việc cân bằng giữa các mô men mẫu với các mô men lý thuyết tương ứng.

Ý tưởng đằng sau phương pháp mô men là đặt mô men mẫu bậc 1 bằng với mô men lý thuyết bậc 1

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = E(X).$$

Đặt mô men mẫu bậc 2 bằng với mô men lý thuyết bậc 2

$$\bar{x^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2 = E(X^2).$$

Tiếp tục đặt mô men mẫu bậc k bằng với mô men lý thuyết bậc k cho đến khi k bằng số tham số. Giải quyết k phương trình trên ta thu được ước lượng cho các tham số. Các giá trị kết quả này được gọi là phương pháp ước lượng mô men.

1.3.3 Phương pháp ước lượng Bayes

Dưới đây là hai phương pháp ước lượng Bayes mà ta sẽ đề cập đến.

Phương pháp cực đại hậu nghiệm - Maximum A Posterior (MAP)

Chúng ta đã biết rằng với phương pháp suy luận Bayes thì phân phối hậu nghiệm tỷ lệ với tiên nghiệm và hàm hợp lý. Đối với phương pháp MAP thì việc tìm tham số ước lượng θ là việc tìm giá trị của θ để hàm phân phối hậu nghiệm đạt giá trị lớn nhất.
Khi ta biết hàm mật độ phân phối hậu nghiệm thì ước lượng MAP chính là mốt của hàm phân phối hậu nghiệm.

Phương pháp trung bình bình phương sai số nhỏ nhất - Mean Square Error (MSE)

Hàm bình phương mất mát (Quadratic loss function)

Giả sử biến ngẫu nhiên X có hàm phân phối $F_X(x)$ nào đó, a là tham số cần ước lượng. Khi đó, ước lượng trung bình bình phương sai số nhỏ nhất nghĩa là đi tìm a sao cho:

$$f(a) = E(a - X)^2 \rightarrow min$$

Ta có,

$$f^{'}(a) = 2a - 2E(x) = 0$$
$$\Longrightarrow a = E(X)$$

Do đó, ước lượng Bayes bằng phương pháp ước lượng trung bình bình phương sai số nhỏ nhất là:

$$a = E(X)$$

Phân phối tiên nghiệm

Để có thể sử dụng định lý Bayes ta cần có hàm phân phối tiên nghiệm $\pi(\theta)$ được coi như là "niềm tin" của chúng ta về giá trị khả thi của tham số trước khi có dữ liệu. Nghĩa là những dữ liệu quan sát được không có bất kỳ ảnh hưởng nào tới việc lựa chọn tiên nghiệm.

Sau đây là một số phân phối tiên nghiệm cho tham số θ.

Tiên nghiệm liên hợp tự nhiên (conjugate prior)

Định nghĩa 1.3.1 (Tiên nghiệm liên hợp tự nhiên) *Nếu phân phối tiên nghiệm và phân phối hậu nghiệm có cùng dạng hàm phân phối thì ta nói tiên nghiệm là tiên nghiệm liên hợp tự nhiên.*

Sử dụng tiên nghiệm liên hợp tự nhiên làm ta dễ dàng xác định được phân phối hậu nghiệm, từ đó dễ dàng trong việc ước lượng tham số mô hình.

Tiên nghiệm phẳng

Định nghĩa 1.3.2 (Lượng thông tin Fisher) *Giả sử X là biến ngẫu nhiên có hàm mật độ xác suất là $f(x, \theta)$, lượng thông tin Fisher về θ của biến ngẫu nhiên X, kí hiệu là $I(\theta)$, là đại lượng:*

$$I(\theta) = -E\Big(\frac{\partial^2}{\partial^2\theta} \log f(x, \theta)\Big)$$

Khi có N tham số, ta có ma trận thông tin Fisher:

$$[I(\theta)]_{ij} = -E\Big(\frac{\partial^2}{\partial\theta_i\partial\theta_j} \log f(x, \theta)\Big)$$

Định nghĩa 1.3.3 (Jeffreys'prior) *Trong thống kê Bayes, Jeffreys'prior là một phân phối tiên nghiệm không thông tin cho vecto tham số θ, hàm mật độ phân phối của nó tỉ lệ với căn bậc hai định thức của ma trận thông tin Fisher:*

$$\pi(\theta) \propto \sqrt{det[I(\theta)]}$$

Tiên nghiệm đều

Nếu chúng ta biết được khoảng giới hạn giá trị của tham số θ cần ước lượng và chúng ta không muốn đặt "niềm tin" cá nhân vào suy luận. Trong trường hợp này việc lựa chọn phân phối tiên nghiệm đều là lựa chọn hợp lý, để tất các giá trị khả năng của θ có cùng trọng số.

1.4 Tính toán Bayes

Trong suy luận Bayesian, khi có tiên nghiệm liên hợp, ta chỉ cần quan tâm đến dạng của phân phối hậu nghiệm mà không cần đến hệ số K. Tuy nhiên, không phải lúc nào ta cũng có tiên nghiệm liên hợp tự nhiên. Lúc này, ta cần phải biết chính xác hàm phân phối hậu nghiệm, nhưng hệ số K tính toán đôi khi rất khó khăn. Do vậy, ta sẽ phải có một phương pháp nào đó để ước lượng được tham số mà không cần tìm ra giá trị K cụ thể. Markov Chain Monte Carlo là phương pháp có thể giải quyết khó khăn này.
Trong chương này, ta sẽ tìm hiểu về phương pháp Markov chain Monte Carlo.

1.4.1 Phương pháp Monte Carlo

Vấn đề cơ bản trong thống kê là tính toán một tích phân có dạng:

$$\mathcal{I} = E[h(X)] = \int_{\chi} h(x)p(x)dx$$

trong đó X là một vecto ngẫu nhiên nhiều chiều, h là một hàm của X trên $\mathcal{X}$, $p(x)$ là hàm mật độ của véc tơ ngẫu nhiên X.
Trong hầu hết các trường hợp chúng ta không thể tính toán tích phân $\mathcal{I}$ mà chúng ta phải dùng máy tính.
Bài toán: Giả sử X là biến ngẫu nhiên một chiều, ta cần ước lượng tích phân

$$\mathcal{I} = E[h(X)] = \int_a^b h(X)p(x)dx$$

Phương pháp tất định(Deterministic numerical methods)
Ý tưởng phương pháp:

1. Chia miền thành các khoảng nhỏ $a = x_0 < x_1 < ... < x_n = b$

2. Ước lượng $\mathcal{I}$ bởi $\hat{\mathcal{I}} = \sum_{i=1}^{n} h(\xi_i)p(\xi_i)(x_i - x_{i-1})$ với $\xi_i \in [x_{i-1}, x_i)$. $\hat{\mathcal{I}}$ sẽ rất gần với $\mathcal{I}$ khi mà miền $[x_{i-1}, x_i)$ nhỏ đi.

Nhược điểm:

1. Không hiệu quả nếu X nhiều chiều.

2. Miền xác định phải hữu hạn: hữu hạn trên đoạn a, b

Phương pháp Monte Carlo

Giả sử rằng, ta có thể dùng máy tính để sinh ra tập mẫu gồm n phần tử sao cho tập biến ngẫu nhiên $X_i \overset{\text{i.i.d}}{\sim} p(x), i = 1, .., n$

Ta có:

$$\hat{\mathcal{I}}_n := \frac{1}{n} \sum_{i=1}^{n} h(X_i)$$

Theo luật số lớn, $\hat{\mathcal{I}}_n \xrightarrow{a.s} \mathcal{I}, Var(\hat{\mathcal{I}}_n) = Var(h(X_1))/n \to 0$ vì $n \to \infty$ bất kể số chiều của tích phân.

Ta thấy rằng, phương pháp Monte carlo vẫn làm việc tốt khi mà số chiều dù có nhiều.

Phương pháp số chia miền từ a đến b thành rất nhiều miền, và chúng ta quan tâm cả đến những vùng mà nó không đóng góp gì vào tích phân, trong khi phương pháp Monte Carlo chỉ lấy mẫu từ vùng có xác suất cao, chúng ta chỉ tập trung tài nguyên tính toán vào vùng mà chúng ta cần phải tập trung.

Nhưng làm cách nào để chúng ta dạy được máy tính sinh ra được các biến ngẫu nhiên theo phân phối p mà chúng ta quan tâm.

Sinh ra biến ngẫu nhiên

Ta có thể dùng máy tính để sinh ra biến ngẫu nhiên, nhưng chỉ sinh ra duy nhất từ phân phối đều $\mathcal{U}(0, 1)$ mà thôi.

Thuật toán như sau:

1. Thiết kế thuật toán máy tính để sinh ra dãy số $u_1, ..., u_n$ trong đoạn [0,1]

2. Thực hiện một loạt các phương pháp kiểm định mà ta chấp nhận giả thuyết dãy số sinh ra từ biến ngẫu nhiên tuân theo phân phối đều $\mathcal{U}(0,1)$.

3. Nếu ta không thể bác bỏ giả thuyết dãy số này không ổn khi n đủ lớn thì tập số sinh ra được xem như độc lập cùng phân phối từ $\mathcal{U}$(0,1).

Định nghĩa 1.4.1 (Thuật toán random number generator) *Bộ sinh số ngẫu nhiên là một thuật toán mà bắt đầu từ giá trị ban đầu u_0 và sử dụng quy tắc chuyển đổi D, tạo ra một chuỗi $u_i = D(u_{i-1})$ với giá trị trong khoảng [0,1]. Tập mẫu ngẫu nhiên sinh ra $u_1, ..., u_n$ độc lập cùng phân phối* $\mathcal{U}(0,1)$

Random seed (hạt giống ngẫu nhiên) cung cấp một giá trị ban đầu u_0 cho trình tạo ngẫu nhiên.

Máy tính ngày nay đều được trang bị bộ sinh số ngẫu nhiên như vậy.

Bài toán đặt ra: Giả sử rằng, ta có thể sinh ra biến ngẫu nhiên $U \sim \mathcal{U}(0,1)$, vậy làm thế nào để sinh ra một biến ngẫu nhiên $X \sim p(x)$ cho bất kì hàm mật độ xác suất $p(x)$ nào.

Phương pháp 1: Phương pháp ngược.

Mục đích là sinh ra một biến ngẫu nhiên X với mật độ xác suất $p(x)$ và hàm phân phối tích lũy $F(x)$.

Giả sử hàm nghịch đảo F^{-1} tồn tại và dễ dàng tính toán được.

1. Sinh ra $U \sim \mathcal{U}(0,1)$

2. Đặt $X = F^{-1}(U)$

Khi đó, X được tạo ra là một biến ngẫu nhiên có hàm phân phối tích luỹ là $F(x)$. Bởi vì:

$F_X(x) = P(X \leq x) = P(F^{-1}(U) \leq x) = P(U \leq F(x)) = F(x)$

Phương pháp 2: Phương pháp chuyển

Trong một số tình huống, biến ngẫu nhiên mà chúng ta quan tâm nó là hàm của các biến ngẫu nhiên khác mà giả sử ta có thể sinh ra nó được bằng máy

tính.
Ví dụ, nếu $Z \sim N(0,1), Y \sim \chi_v^2$, Y và Z độc lập, thì

$$T = \frac{Z}{\sqrt{Y/v}}$$

là biến ngẫu nhiên tuân theo luật phân phối Student t. Do đó, nếu ta biết cách để sinh ra mẫu ngẫu nhiên từ N(0,1) và χ_v^2 thì ta có thể sinh ra được T.
Tuy nhiên phương thức nghịch đảo và phương thức chuyển cũng có nhiều khó khăn:

1. Khó khăn trong việc tính nghịch đảo hàm phân phối tích luỹ F^{-1}. Ví dụ, nghịch đảo của hàm tích luỹ chuẩn.

2. Trong nhiều tình huống, ta không thể biểu diễn phân phối bằng hàm của các phân phối đơn giản khác được.

3. Trong một số trường hợp, mật độ $p(x)$ tỉ lệ với hàm số $k(x)$, $p(x) =$ k(x)/C với hằng số chuẩn hoá $C = \int k(x)dx$ chưa biết.

Từ đó, ta có ý tưởng về phương pháp chấp nhận - từ chối (Accept-Reject methods).
Phương pháp 3: Phương pháp chấp nhận - từ chối (Accept-Reject methods)
Sử dụng mật độ đơn giản hơn là g, dễ dàng lấy mẫu từ đó, g là ột hàm mật độ công cụ và để hỗ trợ sinh ra biến ngẫu nhiên từ p.
Mục đích: Sinh ra biến ngẫu nhiên từ phân phối $p(x) =$ k(x)/C, trong đó $C = \int k(x)dx$ là hằng số có thể chưa biết. Ta viết $p(x) \propto k(x)$.
Thuật toán: Tìm một hàm mật độ g dễ dàng sinh ra được biến ngẫu nhiên và hằng số $M > 0$ sao cho:

$$k(x) \leq Mg(x)$$

với $\forall x$ trong miền của p

1. Sinh ra biến ngẫu nhiên$X \sim g$ và $U \sim \mathcal{U}(0,1)$

2. Ta chấp nhận biến ngẫu nhiên Y = X nếu $U \leq \frac{k(X)}{Mg(x)}$, trường hợp khác thì quay lại (1).

Xác suất chấp nhận Y=X là

$$P(U \leq \frac{k(X)}{Mg(x)}) = \frac{C}{M}$$

Vì vậy, M càng nhỏ thì thuật toán càng hiệu quả.
Ưu điểm của phương pháp chấp nhận - từ chối:

1. Cung cấp một phương pháp sinh để mô phỏng từ mật độ p bất kì với một hằng số chuẩn hoá.

2. Các mẫu được tạo ra là độc lập với nhau.

Nhược điểm của phương pháp là có thể khó để chọn ra g và M(nếu M lớn thì khó sinh ra biến ngẫn nhiên mong muốn).
Ngoài ra, có một số phương pháp thay thế để lấy mẫu từ $p(x)$ như:

1. Phương pháp Markov chain Monte Carlo

2. Gibbs sampling

3. Phương pháp nghịch đảo, biến đổi, chấp nhận - từ chối là các phương pháp dùng để sinh ra biến ngẫu nhiên từ phân phối $p(x)$.

Tổng kết lại, phương pháp Monte Carlo để giải quyết 2 vấn đề chính là

1. Sinh ra tập mẫu từ phân phối xác suất của hàm mật độ xác suất $p(x)$.

2. Ước lượng tích phân

$$I = \int h(x)p(x)dx = E_{X \sim p(x)(h(X))}$$

1.4.2 Phương pháp Markov chain Monte Carlo (MCMC)

Ta muốn sinh ra dãy các quan sát từ hàm mật độ $p(x) \propto k(x)$. Như đã trình bày ở trên, ta có thể dùng phương pháp Accept - Reject để sinh ra các mẫu ngẫu nhiên độc lập từ hàm $p(x)$ đơn giản (1 chiều, 2 chiều, 3 chiều). Nhưng có một khó khăn là không dễ dàng để xác định được hàm mật độ đề xuất $g(x)$ và M thỏa mãn điều kiện $k(x) \leq Mg(x)$. Dù thỏa mãn điều kiện đó rồi thì M phải nhỏ, vì nếu M mà lớn thì thuật toán sẽ không hiệu quả.
Có một thuật toán khác được phát triển dựa trên phương pháp Accept - Reject là Markov Chain Monte Carlo. Các biến ngẫu nhiên được sinh ra không phải là độc lập như phương pháp Accept - Reject nữa mà là phụ thuộc, ta sinh ra một dãy các biến ngẫu nhiên theo một xích Markov, và xích đó hội tụ về phân phối $p(x)$ mà chúng ta đang quan tâm. Lý thuyết xích Markov là nền tảng của MCMC.

Xích Markov

Định nghĩa 1.4.2 (Xích Markov). *Gọi $\{X_n, n \geq 0\}$ là một dãy các biến ngẫu nhiên nhận các giá trị trong không gian trạng thái $\mathcal{X}$, $\{X_n, n \geq 0\}$ được gọi là xích Markov nếu thỏa mãn:*

$$P(X_{n+1} \in A | X_n = x_n, X_{n-1} = x_{n-1}, ..., X_0 = x_0) = P(X_{n+1} \in A | X_n = x_n)$$

với mọi n, $A \subset \mathcal{X}$.

Định nghĩa 1.4.3 (Mật độ chuyển). *Hàm mật độ có điều kiện $k(y|x)$ cho bởi*

$$P(X_{n+1} \in A | X_n = x) = \int_A k(y|x) dx$$

được gọi là mật độ chuyển của xích Markov.
Nếu không gian trạng thái $\mathcal{X}$ rời rạc thì $k(x|y)$ được gọi là ma trận chuyển.

Định nghĩa 1.4.4 (Phân phối ổn định). *p được gọi là phân phối ổn định của xích Markov với hàm mật độ chuyển k nếu*

$$p(x) = \int p(y)k(x|y)dy$$

Điều đó có nghĩa là, hệ thống vẫn ở trạng thái cũ sau mỗi bước chuyển.

Định lý 1.4.5 (Tính ergodic). *Xích Markov $\{X_n, n \geq 0\}$ được gọi là ergodic với phân phối ổn định p nếu*

$$\frac{1}{n}\sum_{i=1}^{n} h(X_i) \xrightarrow{a.s} \int h(x)p(x)dx = E_{X\sim p(x)}(h(X))$$

với bất kì hàm chặn h nào đó.

Markov Chain Monte Carlo

Trong lí thuyết MCMC, p(x) đã biết, ta muốn tạo ra một xích Markov ergodic mà nhận p(x) là phân phối ổn định.

Trong ứng dụng của chúng ta, p(x) đã biết và ta muốn ước lượng:

$$\mathcal{I} = \int h(x)p(x)dx$$

Nếu ta có thể tạo ra một xích Markov ergodic $\{X_i, i \geq 0\}$ với phân phối ổn định p, thì:

$$\frac{1}{n}\sum_{i=1}^{n} h(X_i) \xrightarrow{a.s} \int h(x)p(x)dx = \mathcal{I}$$

Thuật toán Random walk Metropolis-Hastings (trường hợp một chiều)(RWMH)

Mục đích: Đặt $p(x) \propto k(x)$ là hàm mật độ. Ta muốn thiết kế một xích Markov $X_n, n \geq 0$ sao cho nó nhận phân phối p là phân phối ổn định.

Thuật toán 1. Bắt đầu từ một giá trị X_0. Tại các bước $n = 0, 1, ...$

1. Tạo một đề xuất $\epsilon \sim N(0, \sigma^2)$, và đặt $Y = X_n + \epsilon$.

2. Tính toán xác suất chấp nhận.
   $\alpha(Y, X_n) = min\left(\frac{p(Y)}{p(X_n)}, 1\right) = min\left(\frac{k(Y)}{k(X_n)}, 1\right)$

3. Sinh ra một biến ngẫu nhiên $U \sim U(0, 1)$.
   Nếu $U \leq \alpha(Y, X_n)$ thì đặt $X_{n+1} = Y$.
   Ngược lại, đặt $X_{n+1} = X_n$.

Khi mà chúng ta thực hiện thuật toán trên thì ta sẽ sinh ra được một dãy $\{X_n, n \geq 0\}$ là một xích Markov ergodic nhận $p(x)$ là phân phối ổn định.
Giả sử rằng ta đã có xích Markov $X_n, n = 1, ..., N$ được tạo bởi thuật toán RWMH từ phân phối hậu nghiệm $p(\theta|\underline{x})$, vậy kỳ vọng $E(\theta|\underline{x}) = \int \theta p(\theta|\underline{x}) d\theta$ được ước lượng bởi:

$$\hat{\theta} = \frac{1}{N} \sum_{n=1}^{N} X_n$$

Phương sai hậu nghiệm

$$Var(\theta|\underline{x}) = \int (\theta - E(\theta|\underline{x}))^2 p(\theta|\underline{x}) d\theta$$

được ước lượng bởi:

$$\frac{1}{N} \sum_{n=1}^{N} (X_n - \hat{\theta})^2$$

Nhận xét 1.4.6

1. *Độ hiệu quả của thuật toán RWMH phụ thuộc vào giá trị của σ^2.*

2. *Trong cả hai trường hợp, σ^2 quá lớn hoặc quá nhỏ sẽ dẫn đến kết quả ước lượng MCMC sẽ có phương sai rất lớn.*

Thuật toán Random walk Metropolis-Hastings (Trường hợp nhiều chiều)

Mục đích: Đặt $p(x) \propto k(x)$ là hàm mật độ. Ta muốn thiết kế một xích Markov $\{X_n, n \geq 0\}$ sao cho nó nhận phân phối p là phân phối ổn định.

Thuật toán 2. Bắt đầu từ một vecto X_0. Tại các bước $n = 0, 1, ...$

1. Tạo một đề xuất $\epsilon \sim N(0, \sum)$ là một phân phối chuẩn nhiều chiều với véc tơ kỳ vọng băng 0 và ma trận hiệp phương sai $\sum$, và đặt $Y = X_n + \epsilon$.

2. Tính toán xác suất chấp nhận.
$\alpha(Y, X_n) = min\left(\frac{p(Y)}{p(X_n)}, 1\right) = min\left(\frac{k(Y)}{k(X_n)}, 1\right)$

3. Sinh ra một biến ngẫu nhiên $U \sim U(0, 1)$.
Nếu $U \leq \alpha(Y, X_n)$ thì đặt $X_{n+1} = Y$.
Ngược lại, đặt $X_{n+1} = X_n$.

Thuật toán Independent Metropolis-Hastings (IMH)

Để sinh ra X_{n+1}, ta làm như sau:

1. Sinh ra $Y \sim g$.

2. Tính toán xác suất chấp nhận.
$\alpha(Y, X_n) = min\left(\frac{k(Y)g(X_n)}{k(X_n)g(Y)}, 1\right)$

3. Sinh ra một biến ngẫu nhiên $U \sim U(0, 1)$.
Nếu $U \leq \alpha(Y, X_n)$ thì đặt $X_{n+1} = Y$.
Ngược lại, đặt $X_{n+1} = X_n$.

Thuật toán General Metropolis-Hastings

Thuật toán RWMH và IMJH là trường hợp đặc biệt của thuật toán General Metropolis-Hastings.

Gọi $q(y|x)$ là một phân phối đề xuất.

Để sinh ra X_{n+1}, ta làm như sau:

1. Sinh ra một đề xuất $Y \sim q(\cdot|X_n)$.

2. Tính toán xác suất chấp nhận.
   $\alpha(Y, X_n) = min\left(\frac{k(Y)q(X_n|Y)}{k(X_n)q(Y|X_n)}, 1\right)$

3. Sinh ra một biến ngẫu nhiên $U \sim U(0, 1)$.
   Nếu $U \leq \alpha(Y, X_n)$ thì đặt $X_{n+1} = Y$.
   Ngược lại, đặt $X_{n+1} = X_n$.

Chú ý

1. Điểm xuất phát X_0 về mặt lí thuyết không quan trọng, nhưng trong thực hành thì nếu xuất phát điểm tốt thì xích Markov sẽ hội tụ nhanh hơn.

2. Xích Markov có thể mất một vài vòng lặp trước khi nó hội tụ về phân phối dừng. Khoảng vòng lặp trước khi nó hội tụ về phân phối dừng gọi là burn -in.

3. Hiệu quả của thuật toán MH phụ thuộc vào mật độ đề xuất q.

Thuật toán Gibbs sampling

Phân phối mà chúng ta quan tâm $p(x) = p(x_1, x_2)$, trong đó, $x = (x_1, x_2)$, x_1, x_2 một chiều hoặc nhiều chiều.

Mục đích là sinh ra mẫu từ hàm mật độ có điều kiện $p(x_1|x_2)$ và $p(x_2|x_1)$.

Thuật toán. Tại bước thứ n có $x^{(n)} = (x_1^{(n)}, x_2^{(n)})$

1. Sinh ra $x_1^{(n+1)} \sim p(x_1|x_2^{(n)})$

2. Sinh ra $x_2^{(n+1)} \sim p(x_2|x_1^{(n+1)})$

Dãy $\{x^{(n)}, n \geq 0\}$ được tạo bởi thuật toán là xích Markov ergodic nhận $p(x_1, x_2)$ là phân phối ổn định.

Thuật toán Gibbs sampling (trường hợp tổng quát)

Đặt $p(x) = p(x_1, x_2, ..., x_p)$, trong đó, $x = (x_1, x_2, ...x_p)$, x_i một chiều hoặc nhiều chiều.

Mục đích là sinh ra mẫu từ hàm mật độ có điều kiện $p(x_i|x_1, x_2, ..., x_{i-1}, x_{i+1}, ..., x_p)$ với $i = 1, ..., p$.

Thuật toán. Tại bước thứ n có $x^{(n)} = (x_1^{(n)}, ..., x_p^{(n)})$

1. Sinh ra $x_1^{(n+1)} \sim p(x_1|x_2^{(n)}, ..., x_p^{(n)})$
2. Sinh ra $x_2^{(n+1)} \sim p(x_2|x_1^{(n+1)}, x_3^{(n)}, ..., x_p^{(n)})$

   ...
3. Sinh ra $x_p^{(n+1)} \sim p(x_p|x_1^{(n+1)}, ..., x_{p-1}^{(n+1)})$

Dãy $\{x^{(n)}, n \geq 0\}$ được tạo bởi thuật toán là xích Markov ergodic nhận $p(x_1, ..., x_p)$ là phân phối ổn định.

Ví dụ thuật toán RWMH

Giả sử ta có n =50 quan sát từ phân phối chuẩn có trung bình μ và phương sai bằng 1, cho $\bar{x} = 5$.

Ta đã biết hàm hợp lý có dạng:

$$f(\underline{x}|\mu) \propto \exp\{-\frac{n}{2}(\mu - \bar{x})^2\}$$

Phân phối tiên nghiệm phẳng:

$$\pi(\mu) \propto 1$$

Phân phối hậu nghiệm

$$f(\underline{x}|\mu) \propto \exp\{-\frac{n}{2}(\mu - \bar{x})^2\} \propto \exp\{n\bar{x}\mu - \frac{n}{2}\mu^2\}$$

Ta nhận thấy rằng

$$\mu|\underline{x} \sim N(\bar{x}, \frac{1}{n})$$

Nên $\hat{\mu}_{MSE} = \bar{x} = \hat{\mu}_{MLE}$

Giả sử như chúng ta không hề biết dạng hàm mật độ của phân phối hậu nghiệm, ta sẽ dùng thuật toán RWMH để sinh ra mẫu từ phân phối hậu nghiệm.

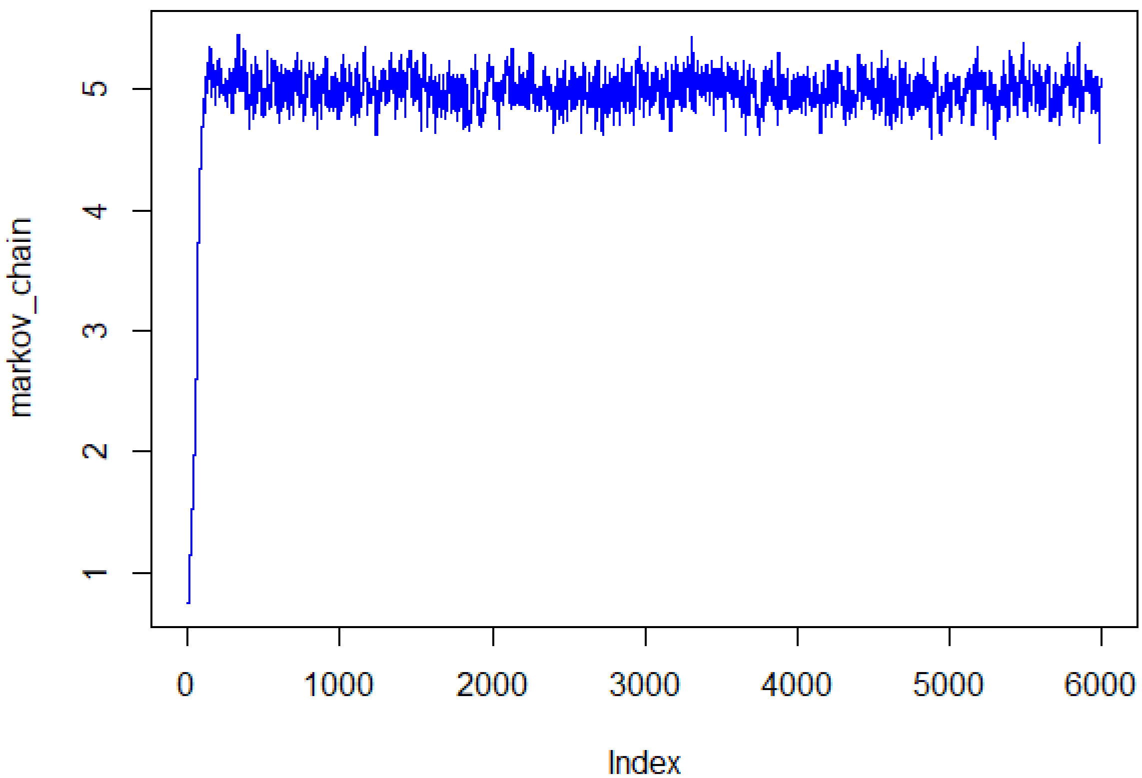
(i) Thuật toán RWMH với $\sigma^2 = 0.1$, với tổng số vòng lặp là 6000, trong đó có 1000 vòng là burn-in.

Sau khi chạy thuật toán, ta có được xích Markov sau:

```
  [1] 0.7431357 0.7431357 0.7431357 0.7515420 0.8735228 0.8735228 1.0037737
  [8] 1.1131508 1.1303779 1.1475318 1.1475318 1.1475318 1.1471260 1.1471260
 [15] 1.1482375 1.3296228 1.3296228 1.3296228 1.3296228 1.3703439 1.3703439
 [22] 1.3703439 1.3703439 1.4284756 1.4284756 1.5308666 1.5308666 1.5308666
 [29] 1.5863045 1.5863045 1.5863045 1.7055393 1.7055393 1.7556613 1.7556613
 [36] 1.8899032 1.8899032 1.9000145 1.9000145 1.9724297 2.0109692 2.0577195
 [43] 2.1396426 2.1636674 2.2643418 2.2453329 2.3718553 2.3718553 2.4627156
 [50] 2.5935183 2.5984139 2.5984139 2.5984131 2.6911327 2.6983515 2.6983515
 [57] 2.9026407 3.0010028 3.1579485 3.2376134 3.2906966 3.4534544 3.6290326
 [64] 3.6290326 3.7301834 3.7301834 3.7301834 3.7518962 3.8947967 3.8947967
 [71] 3.8947967 3.8922988 4.0561245 4.1363677 4.1800719 4.1800719 4.1800719
 [78] 4.2702831 4.3477010 4.3477010 4.4694518 4.4995493 4.5885349 4.5309970
 [85] 4.5508293 4.5711059 4.4714960 4.4969805 4.5707622 4.6784208 4.6784208
 [92] 4.6940349 4.6940349 4.6940349 4.6940349 4.7954395 4.8161723 4.8161723
 [99] 4.7937379 4.7135365 4.8418173 4.8325915 4.8325915 4.9194393 4.8774745
[106] 4.8702813 4.8702813 4.8352092 4.8026001 4.8345693 4.8817055 5.0635438
[113] 4.9367306 5.0100156 5.1094158 5.0651635 5.0683450 4.9887071 4.9667991
[120] 4.9794015 5.0438538 5.0776702 5.1206057 5.0063674 5.0832914 5.0832914
[127] 5.0832914 5.1290661 5.2233674 5.1968268 5.0941496 5.0941496 5.0941496
[134] 5.2573353 5.2573353 5.1904943 5.1613511 5.1613511 5.3130489 5.3586539
[141] 5.2722211 5.2465627 5.2585994 5.2585994 5.2213525 5.0661737 5.0321713
[148] 5.0061037 5.0061037 5.0234605 4.9470526 5.2143924 5.3058989 5.3058989
[155] 5.3451116 5.3451116 5.3280426 5.1102689 5.2093275 5.2105907 5.2105907
[162] 5.1167246 5.1167246 5.1167246 5.0983776 5.1110454 5.1511386 5.0256971
[169] 5.1567540 5.1673647 5.1557543 5.1557543 5.0000472 5.0496365 4.8810741
[176] 4.8968511 4.9892020 5.0698158 5.0663291 4.9190008 4.8949315 5.1261584
[183] 5.1261584 4.9714786 5.0722004 5.1100916 5.0881913 5.0881913 5.1468152
[190] 5.1865850 5.1865850 5.2082384 5.2082384 5.2465980 5.0958241 5.1367118
[197] 5.0380147 5.1159938 5.0679559 5.1577797 5.2159849 5.2583400 5.2435534
[204] 5.2435534 5.2263431 5.1421825 5.0872261 4.9355871 5.0115483 5.0910678
```

Hình 1.3: Xích Markov sinh ra từ phân phối hậu nghiệm

Như ta thấy, xích Markov hội tụ khá nhanh, chỉ sau khoảng 100 giá trị đầu. Sau đây là hình ảnh hội tụ của xích.



Hình 1.4: Sự hội tụ của xích với $\sigma^2 = 0.1$

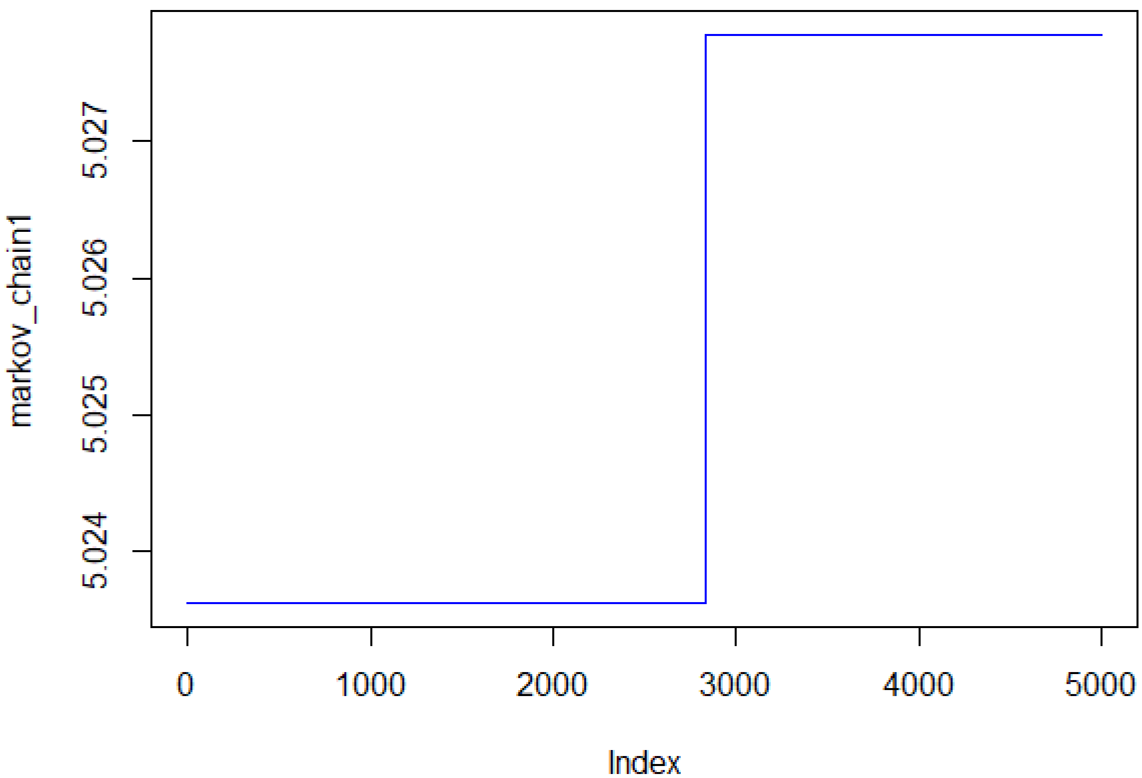
Sau khi có xích Markov, ta tính trung bình của hậu nghiệm:

$$\mu = 4.991845$$

Phương sai hậu nghiệm: 0.1322137.

(ii) Thực hiện thuật toán giống như phần (i) 10 lần và tính độ lệch chuẩn của các ước lượng. Độ lệch chuẩn này chính là sai số chuẩn Monte carlo : 0.005026685.

(iii) Thực hiện thuật toán với $\sigma^2 = 10000$

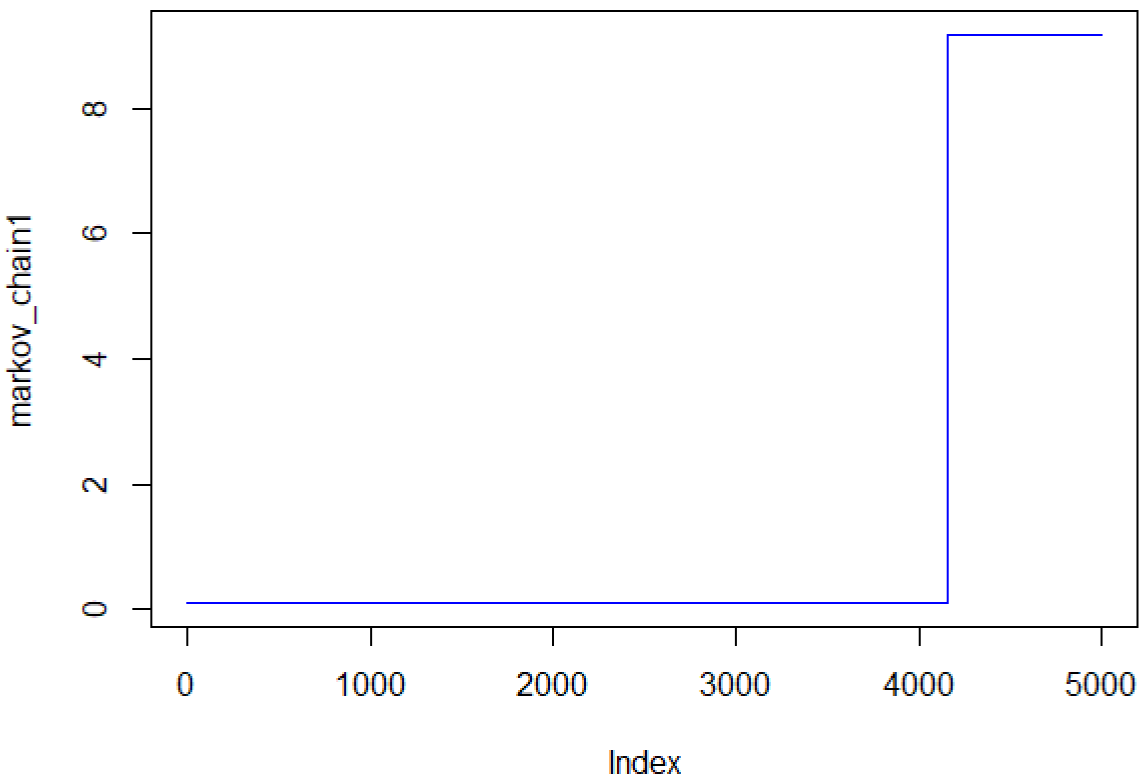


Hình 1.5: RWMH với $\sigma^2 = 10000$

Ta có thể thấy rằng, với $\sigma^2 = 10000$ quá lớn, xích không hội tụ.

Sai số chuẩn Monte carlo: 1.180966.

(iv) Thực hiện thuật toán với $\sigma^2 = 0.0001$



Hình 1.6: RWMH với $\sigma^2 = 0.0001$

Ta có thể thấy rằng, với $\sigma^2 = 0.0001$ quá nhỏ, xích không hội tụ.

Sai số chuẩn Monte carlo: 1.55626.

Kết luận: việc lựa chọn σ^2 ảnh hưởng trực tiếp đến sự hội tụ của xích Markov.

Chương 2

Ước lượng tham số

Trong chương này sẽ trình bày phương pháp ước lượng hợp lý cực đại và phương pháp ước lượng Bayes cho một vài mô hình xác suất.

2.1 Phân phối chuẩn

Định nghĩa 2.1.1 (Hàm hợp lý - Likelihood) *Giả sử các biến ngẫu nhiên* $X_1, X_2, .., X_n \overset{\text{i.i.d}}{\sim} N(\mu, \sigma^2)$*. Khi đó, hàm mật độ xác suất đồng thời của các biến ngẫu nhiên* $(X_1, X_2, .., X_n)$ *có dạng:*

$$
\begin{aligned}
f_{X_1,..,X_n}(x_1, ..., x_n|\mu, \sigma^2) &= \prod_{i=1}^{n} f_{X_i}(x_i) \\
&= (2\pi\sigma^2)^{-n/2} exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}
\end{aligned}
$$

Vì hàm số này phụ thuộc vào $(x_1, x_2, ..., x_n), \mu$ *và* σ^2 *nên ta coi nó là hàm của* μ, σ^2 *nếu biết các quan sát* $(x_1, ..., x_n)$*. Ta gọi đó là hàm hợp lý của tham số* μ, σ^2*. Đặt* $\underline{x} = (x_1, ..., x_n)$*. Để đơn giản, ta sẽ kí hiệu hàm mật độ xác suất đồng thời là* $f(\underline{x}|\mu, \sigma^2)$*. Hàm hợp lý của tham số* μ, σ^2 *là:*

$$L(\mu, \sigma^2) = f(\underline{x}|\mu, \sigma^2) = (2\pi\sigma^2)^{-n/2} exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}$$

2.1.1 Trường hợp 1: σ^2 đã biết

Khi đó, hàm hợp lý có dạng:

$$L(\mu) = (2\pi\sigma^2)^{-n/2} exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}$$

Sau khi đã có hàm hợp lý, ta sẽ ước lượng tham số μ theo hai phương pháp sau đây.

a. Phương pháp ước lượng hợp lý cực đại

Mục tiêu của phương pháp ước lượng hợp lý cực đại là đi tìm tham số μ để hàm hợp lý đạt giá trị lớn nhất.

Định lý 2.1.2 *Đặt $\underline{x} = (x_1, ..., x_n)$ là các quan sát độc lập cùng phân phối chuẩn $N(\mu, \sigma^2)$. Gọi $\hat{\mu}_{MLE}$ là ước lượng hợp lý cực đại của μ. Khi đó,*

$$\hat{\mu}_{MLE} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Chứng minh. Ta có hàm hợp lý:

$$L(\mu) = (2\pi\sigma^2)^{-n/2} exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}$$

Ta tìm giá trị cực đại của hàm hợp lý trên cũng tương đương với việc tìm giá trị cực đại hàm Logarit của hàm hợp lý. Do đó, ta có:

$$l(\mu) = logL(\mu) = -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 + const$$

. Ta tiến hành đạo hàm bậc một hàm $l(\mu)$ để tìm điểm dừng.

$$\frac{\partial}{\partial\mu} l(\mu) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) = \frac{n}{\sigma^2}(\bar{x}-\mu) = 0$$

Suy ra, $\mu = \bar{x}$.

Ta có đạo hàm bậc hai của hàm $l(\mu)$

$$\frac{\partial^2}{\partial^2\mu} l(\mu) = -\frac{n}{\sigma^2} < 0$$

Do đó, $\mu = \bar{x}$ là giá trị cực đại của hàm $l(\mu)$.

Vậy, ước lượng hợp lý cực đại của tham số μ là

$$\hat{\mu}_{MLE} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Định lý đã được chứng minh. $\square$

Tiếp theo, ta xét đến một số tính chất của ước lượng hợp lý cực đại.

Tính chất của ước lượng hợp lý cực đại:

i) $\hat{\mu}_{MLE}$ là ước lượng không chệch của μ.

ii) $\hat{\mu}_{MLE}$ là ước lượng vững của μ.

Chứng minh. i) $\hat{\mu}_{MLE}$ là ước lượng không chệch của μ.

Ta có:

$$\hat{\mu}_{MLE} = \frac{s_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n}$$

Suy ra giá trị kỳ vọng của $\hat{\mu}_{MLE}$ là:

$$E(\hat{\mu}_{MLE}) = \frac{1}{n}E(\sum_{i=1}^{n} X_i) = \frac{1}{n}.n\mu = \mu.$$

Vậy, $\hat{\mu}_{MLE}$ là ước lượng không chệch của μ.

ii) $\hat{\mu}_{MLE}$ là ước lượng vững của μ.

Ta có, phương sai của $\hat{\mu}_{MLE}$ là

$$Var(\hat{\mu}_{MLE}) = Var(\frac{\sum_{i=1}^{n} X_i}{n}) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{1}{n^2}.n\sigma^2 = \frac{\sigma^2}{n}$$

Do đó, trung bình bình phương sai số:

$$MSE(\hat{\mu}_{MLE}, \mu) = \frac{\sigma^2}{n} \xrightarrow[n \to \infty]{} 0.$$

Vậy, $\hat{\mu}_{MLE}$ là ước lượng vững của μ.

Vậy, các tính chất của ước lượng hợp lý cực đại đã được chứng minh. $\square$

b. Phương pháp Bayes

Tiếp theo, ta xét đến ước lượng bayes của tham số μ.

Ta có hàm hợp lý:

$$L(\mu) = (2\pi\sigma^2)^{-n/2} exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}$$

Để tìm được phân phối hậu nghiệm, ta cần có hàm phân phối tiên nghiệm $\pi(\mu)$. Khi đó, phân phối hậu nghiệm tỷ lệ với tích của hàm hợp lý và phân phối tiên nghiệm:

$$p(\mu|\underline{x}) \propto f(\underline{x}|\mu) \times \pi(\mu)$$

Ta sẽ xét một số phân phối tiên nghiệm sau đây.

i) Tiên nghiệm liên hợp tự nhiên

Định lý sau đây chỉ ra rằng phân phối chuẩn là một tiên nghiệm liên hợp tự nhiên của phân phối chuẩn; ước lượng Bayes bằng phương pháp trung bình bình phương sai số nhỏ nhất.

Định lý 2.1.3 *Tiên nghiệm liên hợp tự nhiên cho phân tích Bayes của phân phối chuẩn là phân phối chuẩn với trung bình a_0, phương sai b_0^2 và phân phối hậu nghiệm là phân phối chuẩn với trung bình a_1, phương sai b_1^2. Tham số cập nhật*

$$\begin{cases} \frac{1}{b1^2} = \frac{1}{\sigma^2/n} + \frac{1}{b_0^2} \\ a_1 = \frac{b_0^2}{\sigma^2/n+b_0^2}\bar{x} + \frac{\sigma^2/n}{\sigma^2/n+b_0^2}a_0 \end{cases}$$

Khi đó, ước lượng Bayes bằng phương pháp trung bình bình phương sai số nhỏ nhất:

$$\hat{\mu}_{MSE} = a_1$$

Chứng minh. Ta có hàm hợp lý:

$$f(\underline{x}|\mu) \propto \exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\}$$

Mà

$$\begin{aligned}\sum_{i=1}^{n}(x_i-\mu)^2 &= \sum_{i=1}^{n}(x_i^2-2\mu x_i+\mu^2)\\ &= n(\mu-\bar{x})^2+n\bar{x^2}-\bar{x}^2\\ &= n(\mu-\bar{x})^2+ns^2\end{aligned}$$

Do đó,

$$f(\underline{x}|\mu) \propto \exp\{-\frac{1}{2(\sigma/\sqrt{n})^2}(\mu-\bar{x})^2\}$$

Để tạo ra được một tiên nghiệm liên hợp tự nhiên, ta bắt chước dạng hàm của hàm hợp lý, phân phối tiên nghiệm liên hợp tự nhiên sẽ có dạng:

$$\pi(\mu) \propto \exp\{-\frac{1}{2b_0^2}(\mu-a_0)^2\}$$

Tiên nghiệm liên hợp này chính là phân phối chuẩn với trung bình a_0 và phương sai b_0^2.

$$\mu \sim N(a_0, b_0^2)$$

Theo định lý Bayes, phân phối hậu nghiệm tỷ lệ với tích của hàm hợp lý và phân phối tiên nghiệm:

$$f(\underline{x}|\mu)\pi(\mu) \propto \exp\{-\frac{1}{2(\sigma/\sqrt{n})^2}(\mu-\bar{x})^2-\frac{1}{2b_0^2}(\mu-a_0)^2\}$$

Sau một vài bước biến đổi, ta được phân phối hậu nghiệm có dạng sau:

$$f(\underline{x}|\mu)\pi(\mu) \propto \exp\{-\frac{1}{2b_1^2}(\mu-a_1)^2\}$$

Trong đó:

$$\begin{cases} \frac{1}{b1^2} = \frac{1}{\sigma^2/n} + \frac{1}{b_0^2} \\ a_1 = \frac{b_0^2}{\sigma^2/n+b_0^2}\bar{x} + \frac{\sigma^2/n}{\sigma^2/n+b_0^2}a_0 \end{cases}$$

Suy ra, $\mu|\underline{x} \sim N(a_1, b_1^2)$.

Vậy, ước lượng Bayes bằng phương pháp ước lượng trung bình bình phương sai số nhỏ nhất là:

$$\hat{\mu}_{MSE} = a_1$$

Định lý đã được chứng minh. □

Tiếp theo ta có nhận xét về tính chất tổ hợp lồi của ước lượng hợp lý cực đại và kỳ vọng của phân phối tiên nghiệm.

Nhận xét 2.1.4 *(i) Nhắc lại:*

$$E(\mu) = a_0$$

$$\hat{\mu}_{MLE} = \bar{x}$$

Ta nhận thấy rằng, $\hat{\mu}_{MSE}$ là tổ hợp lồi của ước lượng hợp lí cực đại và kỳ vọng của phân phối tiên nghiệm.

$$\hat{\mu}_{MSE} = k\hat{\mu}_{MLE} + (1-k)E(\mu)$$

với $k = \frac{b_0^2}{\sigma^2/n+b_0^2}$.

(ii) Khi số lượng quan sát càng lớn thì $\hat{\mu}_{MSE} \approx \hat{\mu}_{MLE}$, tức là lúc này phân phối tiên nghiệm không còn quan trọng nữa.

Trích xuất tiên nghiệm cho phân phối tiên nghiệm chuẩn

Đặt $N(a_0, b_0^2)$ là tiên nghiệm liên hợp cho tham số μ. Đặt $g_{\mu,1}$ là dự đoán cho giá trị μ, $g_{\mu,2}$ là dự đoán cho độ lệch. Ta có giá trị kỳ vọng và phương sai của biến ngẫu nhiên X tuân theo luật phân phối $N(a_0, b_0^2)$ là:

$$E(X) = a_0,$$
$$Var(X) = b_0^2.$$

Ta sẽ đặt giá trị kỳ vọng và độ lệch chuẩn bằng "niềm tin" của ta:

$$a_0 = g_{\mu,1}$$
$$b_0 = g_{\mu,2}$$

Mô phỏng

Để nghiên cứu về tiên nghiệm liên hợp tự nhiên, ta so sánh giữa phương pháp ước lượng Bayes và phương pháp ước lượng hợp lý cực đại dựa trên dữ liệu mô phỏng. Toàn bộ quá trình mô phỏng sẽ được thực hiện trong R.

Giá trị đầu vào:

$$\mu = 5$$
$$\sigma = 2$$

Sinh ngẫu nhiên n quan sát từ phân phối chuẩn có kỳ vọng và phương sai là các giá trị đầu vào. Ta xây dựng 27 kịch bản khác nhau theo kích thước mẫu và dự đoán của chuyên gia. Dưới đây là bảng kết quả mô phỏng.

Sample-size	Prior guess		Bayesian estimates
n	$g_{\mu,1}$	$g_{\mu,2}$	$\hat{\mu}$
10	4.0	3.6	4.331135
		2.4	4.319189
		1.2	4.267148
	4.9999	0.59997	4.354112
		1.19994	4.369380
		1.79991	4.440762
	6.0	5.4	4.363800
		3.6	4.391015
		1.8	4.523624
	MLE		**4.341355**

Sample-size	Prior guess		Bayesian estimates
n	$g_{\mu,1}$	$g_{\mu,2}$	$\hat{\mu}$
100	4.0	3.6	5.425562
		2.4	5.420100
		1.2	5.391314
	4.9999	0.59997	5.429114
		1.19994	5.428059
		1.79991	5.422450
	6.0	5.4	5.430743
		3.6	5.431716
		1.8	5.436914
	MLE		**5.429962**

Sample-size	Prior guess		Bayesian estimates
n	$g_{\mu,1}$	$g_{\mu,2}$	$\hat{\mu}$
1000	4.0	3.6	5.079542
		2.4	5.079126
		1.2	5.076884
	4.9999	0.59997	5.079859
		1.19994	5.079840
		1.79991	5.079733
	6.0	5.4	5.080001
		3.6	5.080159
		1.8	5.081010
	MLE		**5.079875**

Từ các kết quả trên, ta rút ra được nhận xét sau

- Đối với phương pháp ước lượng Bayes thì phân phối tiên nghiệm ảnh hưởng nhiều đến kết quả của phân phối hậu nghiệm, nhưng khi mà số lượng quan sát càng nhiều thì sự ảnh hưởng này có xu hướng giảm, ước lượng hợp lý cực đại và ước lượng Bayes xấp xỉ nhau.

- Đối với phương pháp ước lượng hợp lý cực đại thì ước lượng của tham số hoàn toàn phụ thuộc vào dữ liệu quan sát được, số lượng quan sát càng lớn thì ước lượng càng chính xác.

ii) Tiên nghiệm phẳng

Định lý sau đây chỉ ra phân phối tiên nghiệm phẳng, phân phối hậu nghiệm và ước lượng Bayes cho tham số μ.

Định lý 2.1.5 *Đặt $\underline{x} = (x_1, ..., x_n)$ là các quan sát độc lập cùng phân phối chuẩn $N(\mu, \sigma^2)$. Theo như hàm bình phương mất mát và phân phối tiên nghiệm phẳng $\pi(\mu) \propto 1$, ta thu được các kết quả sau:*

(i) Phân phối hậu nghiệm $p(\mu|\underline{x}) \sim N(\bar{x}, \frac{\sigma^2}{n})$

(ii) Ước lượng Bayes $\hat{\lambda}_{MSE} = \bar{x}$

Chứng minh.

Để tìm được phân phối tiên nghiệm, ta cần tìm lượng thông tin Fisher.

$$I(\mu) = -E[\frac{\partial^2}{\partial\mu^2} l(\mu)]$$

Ta có hàm logarit của hàm hợp lý như sau:

$$l(\mu) = -\frac{1}{2\sigma^2/n}(\mu - \bar{x})^2 + const$$

$$=> \frac{\partial}{\partial\mu} l(\mu) = -\frac{1}{\sigma^2/n}(\mu - \bar{x})$$

$$=> \frac{\partial^2}{\partial^2\mu} l(\mu) = -\frac{n}{\sigma^2}$$

Do đó, ta có lượng thông tin Fisher

$$I(\mu) = \frac{n}{\sigma^2}$$

Theo Jeffrey's rule

$$\pi \propto \sqrt{I(\mu)}$$

Suy ra, phân phối tiên nghiệm

$$\pi(\mu) \propto 1$$

Ta đã có hàm hợp lý

$$f(\underline{x}|\mu) \propto \exp\{-\frac{1}{2(\sigma/\sqrt{n})^2}(\mu - \bar{x})^2\}$$

Theo định lý Bayes, phân phối hậu nghiệm tỉ lệ với tích của phân phối tiên nghiệm và hàm hợp lý

$$f(\underline{x}|\mu)\pi(\mu) \propto \exp\{-\frac{1}{2(\sigma/\sqrt{n})^2}(\mu - \bar{x})^2\}$$

Do đó, phân phối hậu nghiệm:

$$p(\mu|\underline{x}) \propto \exp\{-\frac{1}{2(\sigma/\sqrt{n})^2}(\mu - \bar{x})^2\}$$

Suy ra

$$\mu|\underline{x} \sim N(\bar{x}, \frac{\sigma^2}{n})$$

Vậy, theo bình phương hàm mất mát, ước lượng Bayes cho tham số μ bằng phương pháp trung bình bình phương sai số nhỏ nhất là $\hat{\mu}_{MSE} = \bar{x}$.

Định lý đã được chứng minh. □

Nhận xét 2.1.6 *Trong trường hợp tiên nghiệm phẳng thì* $\hat{\mu}_{MSE} = \hat{\mu}_{MLE}$.

iii) Tiên nghiệm đều

Hàm hợp lý có dạng

$$f(\underline{x}|\mu) \propto exp\left\{-\frac{1}{2(\sigma/\sqrt{n})^2}(\mu - \bar{x})^2\right\}$$

Phân phối tiên nghiệm là phân phối đều trên đoạn $[a, b]$

$$\pi(\mu) \propto \mathbb{I}_{[a,b]}(\mu)$$

$$\text{với} \quad \mathbb{I}_{[a,b]}(\lambda) = \begin{cases} 1, & \mu \in [a,b] \\ 0, & \mu \notin [a,b] \end{cases}$$

Theo định lý Bayes, phân phối hậu nghiệm tỷ lệ với tích của hàm hợp lý và phân phối tiên nghiệm:

$$f(\underline{x}|\mu)\pi(\mu) \propto \exp\left\{-\frac{1}{2(\sigma/\sqrt{n})^2}(\mu - \bar{x})^2\right\}.\mathbb{I}_{[a,b]}(\mu)$$

Ta nhận thấy rằng phân phối hậu nghiệm không thuộc một trong những phân phối đã biết, do đó ta không thể dễ dàng tìm được ước lượng theo phương pháp Bayes như đã đề cập ở trên. Vì vậy, lúc này ta cần phải biết chính xác hàm phân phối hậu nghiệm, tức là đi tìm hệ số K trong hàm hậu nghiệm sau:

$$\begin{aligned} p(\mu|\underline{x}) &= Kf(\underline{x}|\mu) \times \pi(\mu) \\ &= K\exp\left\{-\frac{1}{2(\sigma/\sqrt{n})^2}(\mu - \bar{x})^2\right\}.\mathbb{I}_{[a,b]}(\mu) \end{aligned}$$

Suy ra

$$\begin{aligned} K^{-1} &= \int_a^b \exp\left\{-\frac{1}{2(\sigma/\sqrt{n})^2}(\mu - \bar{x})^2\right\} \\ &= \Phi(\frac{b - \bar{x}}{\sigma}\sqrt{n}) - \Phi(\frac{a - \bar{x}}{\sigma}\sqrt{n}) \end{aligned}$$

Vậy, ta có hàm phân phối hậu nghiệm là

$$p(\mu|\underline{x}) = K.\exp\left\{-\frac{1}{2\sigma^2/n}(\mu - \bar{x})^2\right\}.\mathbb{I}_{[a,b]}(\mu)$$

trong đó,

$$K^{-1} = \Phi(\frac{b - \bar{x}}{\sigma}\sqrt{n}) - \Phi(\frac{a - \bar{x}}{\sigma}\sqrt{n})$$

Sau khi ta đã có chính xác hàm phân phối hậu nghiệm, ta tìm ước lượng sao cho hàm hậu nghiệm đạt giá trị lớn nhất.

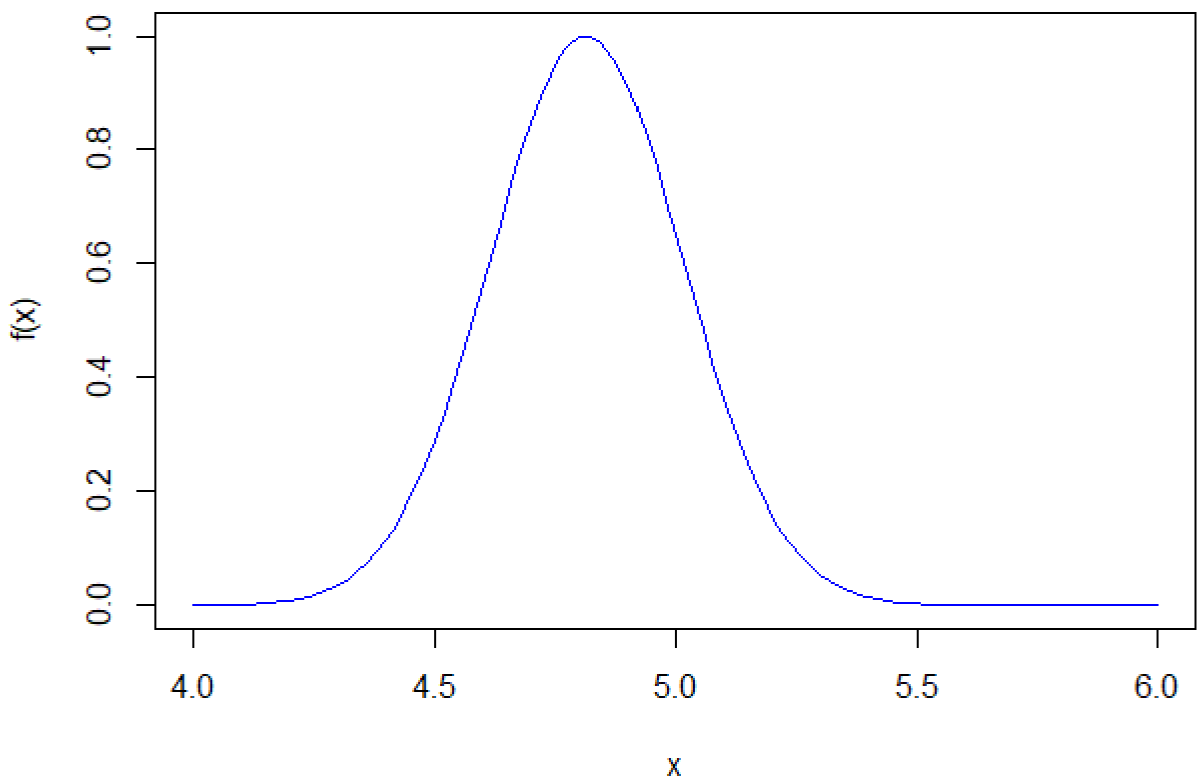
Mô phỏng

Giá trị đầu vào là

$$\mu = 5$$

$$\sigma = 2$$

Sinh ra n = 10 quan sát từ phân phối chuẩn có kỳ vọng và phương sai như trên. Khi lấy a = 4, b = 6, ta được phân phối hậu nghiệm sau



Hình 2.1: Phân phối hậu nghiệm

2.1.2 Trường hợp 2: μ đã biết

Khi đó, hàm hợp lý có dạng:

$$L(\sigma^2) = (2\pi\sigma^2)^{-n/2} exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 \right\}$$

Sau khi đã có hàm hợp lý, ta sẽ ước lượng tham số σ theo hai phương pháp sau đây.

a. Ước lượng hợp lý cực đại

Mục tiêu của phương pháp ước lượng hợp lý cực đại là đi tìm tham số σ^2 để hàm hợp lý đạt giá trị lớn nhất.

Định lý 2.1.7 *Đặt* $\underline{x} = (x_1, ..., x_n)$ *là các quan sát độc lập cùng phân phối chuẩn* $N(\mu, \sigma^2)$. *Gọi* $\hat{\sigma^2}_{MLE}$ *là ước lượng hợp lý cực đại của* σ^2. *Khi đó,*

$$\hat{\sigma^2}_{MLE} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2$$

Chứng minh. Ta có hàm hợp lý:

$$L(\sigma^2) = (2\pi\sigma^2)^{-n/2} exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 \right\}$$

Ta tìm giá trị cực đại của hàm hợp lý trên cũng tương đương với việc tìm giá trị cực đại hàm Logarit của hàm hợp lý. Do đó, ta có:

$$l(\sigma^2) = logL(\sigma^2) = -\frac{n}{2} log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 + const$$

Ta tiến hành đạo hàm bậc một hàm $l(\sigma^2)$ để tìm điểm dừng.

$$\frac{\partial}{\partial \sigma^2} l(\sigma^2) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{n} (x_i - \mu)^2 = -\frac{1}{2\sigma^2} (n\sigma^2 - \sum_{i=1}^{n} (x_i - \mu)^2) = 0$$

Suy ra,

$$\hat{\sigma^2}_{MLE} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2$$

Dễ thấy đạo hàm bậc hai

$$\frac{\partial^2}{\partial^2\sigma^2}l(\sigma^2) < 0$$

Do đó, $\hat{\sigma^2}_{MLE} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$ là giá trị cực đại của hàm $l(\sigma^2)$.

Vậy, ước lượng hợp lý cực đại của tham số σ^2 là

$$\hat{\sigma^2}_{MLE} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$$

Định lý đã được chứng minh. □

Tiếp theo, ta xét đến một số tính chất của ước lượng hợp lý cực đại.

Tính chất của ước lượng hợp lý cực đại:

i) Ước lượng hợp lý cực đại là ước lượng không chệch của σ^2.

ii) Ước lượng hợp lý cực đại là ước lượng vững của σ^2.

Trước khi chứng minh các tính chất trên, ta có định lý sau đây.

Định lý 2.1.8 (Luật số lớn) *Nếu các biến ngẫu nhiên $X_1, ..., X_n$ độc lập cùng phân phối với biến ngẫu nhiên X có kỳ vọng $\mu = E(X)$ và phương sai $\sigma^2 = Var(X)$ thì $\frac{X_1+...+X_n}{n} \xrightarrow{h.c.c} \mu$.*

Tiếp theo, ta chứng minh các tính chất của ước lượng hợp lí cực đại.

Chứng minh. i) $\hat{\sigma^2}_{MLE}$ là ước lượng không chệch của σ^2.

Ta có,

$$E(\hat{\sigma^2}_{MLE}) = E(\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2) = \frac{1}{n}.n\sigma^2 = \sigma^2.$$

Do đó, ước lượng hợp lý cực đại là ước lượng không chệch.

ii) $\hat{\sigma^2}_{MLE}$ là ước lượng vững của σ^2.

Ta có, phương sai của $\hat{\sigma^2}_{MLE}$ là

$$Var(\hat{\sigma^2}_{MLE}) = \frac{1}{n^2}Var\sum_{i=1}^{n}(X_i - \mu)^2$$

Do đó, trung bình bình phương sai số:

$$MSE(\hat{\mu}_{MLE}, \mu) = \frac{\sigma^2}{n} \xrightarrow[n \to \infty]{} 0.$$

Vì $(X_1 - \mu)^2, ..., (X_n - \mu)^2$ độc lập cùng phân phối với

$$\begin{cases} E(X-\mu)^2 = \sigma^2 \\ Var(X-\mu)^2 = E(X-\mu)^4 - (E(X-\mu)^2)^2 = \sigma_4 - \sigma_2^2 \end{cases}$$

Suy ra, $\frac{(X_1-\mu)^2 + ... + (X_n-\mu)^2}{n} \xrightarrow{h.c.c} \sigma^2$.

Do đó, $\hat{\sigma^2}_{MLE}$ là ước lượng vững của σ^2.

Vậy, các tính chất của ước lượng hợp lý cực đại đã được chứng minh. □

b. Phương pháp Bayes

Tiếp theo, ta xét đến ước lượng bayes của tham số σ^2.

Ta có hàm hợp lý:

$$L(\sigma^2) = f(\underline{x}|\sigma^2) = (2\pi\sigma^2)^{-n/2} exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 \right\}$$

Để dễ dàng cho việc tính toán, ta tham số hóa lại. Đặt $\frac{1}{\sigma^2} = \tau$.

Khi đó, hàm hợp lý trở thành

$$f(\underline{x}|\mu, \tau) = (2\pi)^{-n/2} \tau^{n/2} exp\left\{ -\frac{\tau}{2} \sum_{i=1}^{n} (x_i - \mu)^2 \right\}$$

Để tìm được phân phối hậu nghiệm, ta cần có hàm phân phối tiên nghiệm $\pi(\tau)$. Khi đó, phân phối hậu nghiệm tỷ lệ với tích của hàm hợp lý và phân phối tiên nghiệm:

$$p(\tau|\underline{x}) \propto f(\underline{x}|\tau) \times \pi(\tau)$$

Định lý sau đây chỉ ra phân phối gamma là một tiên nghiệm liên hợp tự nhiên của phân phối gamma; chỉ ra ước lượng Bayes bằng phương pháp trung bình bình phương sai số nhỏ nhất.

Định lý 2.1.9 *Tiên nghiệm liên hợp tự nhiên cho phân tích Bayes của phân phối gamma là phân phối gamma $\tau \sim Gamma(a_0, b_0)$ và phân phối hậu nghiệm là phân phối gamma $\tau|\underline{x} \sim Gamma(a_1, b_1)$. Tham số cập nhật*

$$\begin{cases} a_1 = a_0 + \frac{n}{2} \\ b_1 = b_0 + \frac{1}{2}nv^2 \end{cases}$$

trong đó, $v^2 = \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n}$*. Khi đó, ước lượng Bayes bằng phương pháp trung bình bình phương sai số nhỏ nhất:*

$$\hat{\tau}_{MSE} = \frac{a_1}{b_1}$$

Chứng minh. Ta có hàm hợp lý có dạng:

$$f(\underline{x}|\mu,\tau) = (2\pi)^{-n/2}\tau^{n/2}exp\left\{-\frac{\tau}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}$$

Do đó,

$$f(\underline{x}|\tau) \propto \tau^{n/2}exp\left\{-\frac{\tau}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}$$

Để tạo ra được một tiên nghiệm liên hợp tự nhiên, ta bắt chước dạng hàm của hàm hợp lý, phân phối tiên nghiệm liên hợp tự nhiên sẽ có dạng:

$$\pi(\tau) \propto \tau^{a_0-1}\exp\{-\tau b_0\}$$

Tiên nghiệm liên hợp này chính là phân phối gamma

$$\tau \sim Gamma(a_0, b_0)$$

Theo định lý Bayes, phân phối hậu nghiệm tỷ lệ với tích của hàm hợp lý và phân phối tiên nghiệm:

$$p(\tau|\underline{x}) \propto \tau^{a_0+n/2-1}exp\left\{-\tau(b_0+\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2)\right\}$$

Suy ra, $\tau|\underline{x} \sim Gamma(a_1, b_1)$.

$$\begin{cases} a_1 = a_0 + \frac{n}{2} \\ b_1 = b_0 + \frac{1}{2}nv^2 \end{cases}$$

trong đó, $v^2 = \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n}$.

Vậy, ước lượng Bayes bằng phương pháp trung bình bình phương nhỏ nhất là:

$$\hat{\tau}_{MSE} = \frac{a_1}{b_1}$$

Định lý đã được chứng minh. $\square$

Tiếp theo ta có nhận xét về tính chất tổ hợp lồi của ước lượng hợp lý cực đại và kỳ vọng của phân phối tiên nghiệm.

Nhận xét 2.1.10 *Nhắc lại:*

$$\hat{\sigma^2}_{MLE} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2 = v^2$$

$$=> \hat{\tau}_{MLE} = \frac{1}{\hat{\sigma^2}_{MLE}} = \frac{1}{v^2}$$

Ta có thể thấy rằng $\hat{\tau}_{MSE}$ có thể biểu diễn bằng tổ hợp lồi của $\hat{\tau}_{MLE}$ và giá trị kì vọng của phân phối tiên nghiệm là

$$\hat{\tau}_{MSE} = k\hat{\tau}_{MLE} + (1-k)E(\tau)$$

với $k = \frac{nv^2/2}{b_0+nv^2/2}$.

2.1.3 Trường hợp 3: μ, σ^2 chưa biết

Hàm hợp lý có dạng:

$$L(\mu, \sigma^2) = (2\pi\sigma^2)^{-n/2} exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right\}$$

Sau khi đã có hàm hợp lý, ta sẽ ước lượng tham số theo hai phương pháp sau đây.

a. Phương pháp ước lượng hợp lý cực đại

Mục tiêu của phương pháp ước lượng hợp lý cực đại là đi tìm tham số μ, σ^2 để hàm hợp lý đạt giá trị lớn nhất.

Định lý 2.1.11 *Đặt $\underline{x} = (x_1, ..., x_n)$ là các quan sát độc lập cùng phân phối chuẩn $N(\mu, \sigma^2)$. Gọi $\hat{\mu}_{MLE}$ là ước lượng hợp lý cực đại của μ, $\hat{\sigma^2}_{MLE}$ là ước lượng hợp lý cực đại của σ^2. Khi đó,*

$$\hat{\mu}_{MLE} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\hat{\sigma^2}_{MLE} = s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Chứng minh. Ta có hàm hợp lý:

$$L(\mu, \sigma^2) = (2\pi\sigma^2)^{-n/2} exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right\}$$

Ta tìm giá trị cực đại của hàm hợp lý trên cũng tương đương với việc tìm giá trị cực đại hàm Logarit của hàm hợp lý. Do đó, ta có:

$$l(\mu, \sigma^2) = \log L(\mu, \sigma^2) = -\frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2 + const$$

Ta tiến hành đạo hàm riêng bậc một hàm $l(\mu, \sigma^2)$ theo μ , σ^2 để tìm điểm dừng.

$$\frac{\partial}{\partial \mu} l(\mu, \sigma^2) = \frac{1}{\sigma^2}(\sum_{i=1}^{n} x_i - n\mu) = 0$$

$$\frac{\partial}{\partial \sigma^2} l(\mu, \sigma^2) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - \bar{x})^2 = 0$$

Suy ra,

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$$
$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Ta có đạo hàm bậc hai của hàm $l(\mu, \sigma^2)$ theo μ , σ^2

$$\frac{\partial^2}{\partial^2\mu} l(\mu, \sigma^2) = -\frac{n}{\sigma^2} < 0$$
$$\frac{\partial^2}{\partial^2\sigma^2} l(\mu, \sigma^2) < 0$$

Vậy, ước lượng hợp lý cực đại của tham số μ, σ^2

$$\begin{cases} \hat{\mu}_{MLE} = \bar{x} \\ \hat{\sigma^2}_{MLE} = s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \end{cases}$$

Định lý đã được chứng minh. □

b. Phương pháp Bayes

Tiếp theo, ta xét đến ước lượng bayes.

Hàm hợp lý có dạng:

$$L(\mu, \sigma^2) = (2\pi\sigma^2)^{-n/2} exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right\}$$

Để dễ dàng cho việc tính toán, ta tham số hóa lại. Đặt $\frac{1}{\sigma^2} = \tau$.

Khi đó, hàm hợp lý trở thành

$$L(\mu, \tau) = f(\underline{x}|\mu, \tau) \propto \tau^{n/2} exp\left\{-\frac{\tau}{2}\sum_{i=1}^{n}(x_i - \mu)^2\right\}$$

Để tìm được phân phối hậu nghiệm, ta cần có hàm phân phối tiên nghiệm $\pi(\mu, \tau)$. Khi đó, phân phối hậu nghiệm tỷ lệ với tích của hàm hợp lý và phân phối tiên nghiệm:

$$p(\mu, \tau|\underline{x}) \propto f(\underline{x}|\mu, \tau) \times \pi(\mu, \tau)$$

Ta sẽ xét một số phân phối tiên nghiệm sau đây.

i) Tiên nghiệm liên hợp tự nhiên

Trước tiên, ta định nghĩa về phân phối hai chiều chuẩn gamma.

Định nghĩa 2.1.12 (Phân phối 2 chiều N-G). *Ta nói biên ngẫu nhiên 2 chiều $(X, Y) \in \mathbb{R} \times \mathbb{R}^+$ tuân theo phân phối chuẩn gamma nếu ta có hàm mật độ đồng thời*

$$f_{X,Y}(x, y) = \frac{c^a}{\Gamma(a)} \sqrt{\frac{b}{2\pi}} y^{a-\frac{1}{2}} exp\{-\frac{y}{2}[b(x-m)^2 + 2c]\}$$

kí hiệu $(X, Y) \sim N - G(a, b, c, m)$ với $a, b, c > 0; m \in \mathbb{R}$

Định lý về phân phối biên của Y và phân phối có điều kiện của X.

Định lý 2.1.13 (Phân phối biên của Y và phân phối có điều kiện của X). *Giả sử $(X, Y) \sim N - G(a, b, c, m)$ với $a, b, c > 0; m \in \mathbb{R}$ .Khi đó, ta có:*

1. $Y \sim Gamma(a, c)$

2. $X|Y = y \sim N(m, by)$

Chứng minh.

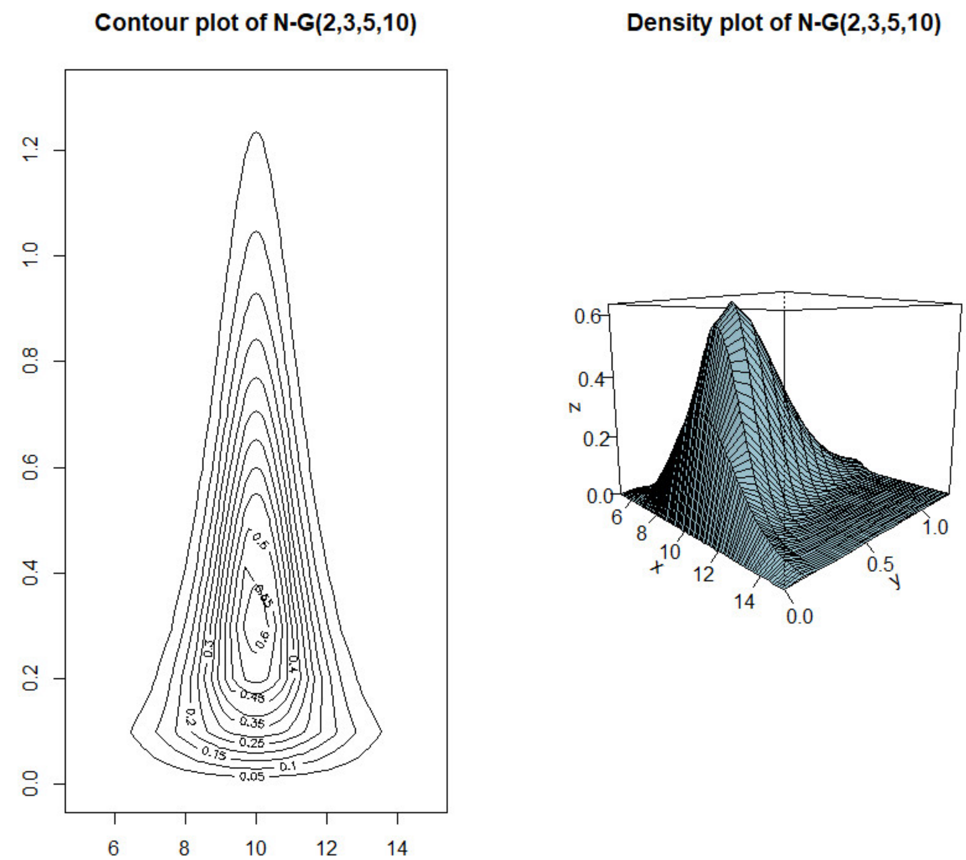
1. $f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(x, y)dx = \dfrac{c^a}{\Gamma(a)} y^{a-1} e^{-cy}$
   Do đó: $Y \sim Gamma(a, c)$

2. $f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \sqrt{\frac{b}{2\pi} y}.exp\{-\frac{yb}{2}(x-m)^2\}$
   Do đó: $X|Y = y \sim N(m, by)$

Định lý đã được chứng minh. □

Đồ thị của hàm mật độ

Sau đây là đồ thị của hàm mật độ phân phối hai chiều N-G.



Định lý 2.1.14 (Kì vọng, phương sai của X,Y).

1. *Kì vọng, phương sai của* Y

   $E(Y) = \frac{a}{c}$

   $Var(Y) = \frac{a}{c^2}$

2. *Kì vọng, phương sai của* X

   $E(X) = m$

   $Var(X) = \frac{c}{b(a-1)}$

Chứng minh.

1. Vì $Y \sim Gamma(a, c)$ nên $E(Y) = a/c$ và $Var(Y) = \frac{a}{c^2}$

2. Vì $X|Y = y \sim N(m, by)$
   nên $E(X) = E(E(X|Y)) = \int_0^{+\infty} m f_Y(y)dy = E(m) = m$
   Vì $Y \sim Gamma(a, c)$ nên $\frac{1}{Y} \sim IGamma(a, c)$. Do đó, $E(\frac{1}{Y}) = \frac{c}{a-1}$
   Ta có: $E(X^2|Y) = Var(X|Y) + E(X|Y)^2 = \frac{1}{by} + m^2$
   $=> E(X^2) = E(E(X^2|Y)) = E(\frac{1}{by} + m^2) = \frac{1}{b}E(\frac{1}{y}) + m^2 = m^2 + \frac{c}{b(a-1)}$
   $=> Var(X) = E(X^2) - (EX)^2 = \frac{c}{b(a-1)}$

Định lý đã được chứng minh. □

Định lý sau đây chỉ ra phân phối hai chiều chuẩn gamma là một tiên nghiệm liên hợp tự nhiên của phân phối hai chiều chuẩn gamma; chỉ ra ước lượng Bayes bằng phương pháp trung bình bình phương sai số nhỏ nhất.

Định lý 2.1.15 *Tiên nghiệm liên hợp tự nhiên cho phân tích Bayes của phân phối hai chiều chuẩn gamma là phân phối hai chiều chuẩn gamma* $(\mu, \tau) \sim N - G(a_0, b_0, c_0, m_0)$ *và phân phối hậu nghiệm là phân phối hai chiều chuẩn gamma* $(\mu, \tau)|\underline{x} \sim N - G(a_1, b_1, c_1, m_1)$. *Tham số cập nhật*

$$\begin{cases} a_1 = a_0 + \frac{n}{2} \\ b_1 = b_0 + n \\ c_1 = c_0 + \frac{n}{2}\left[s^2 + \frac{b_0}{b_0+n}(\bar{x} - m_0)^2\right] \\ m_1 = \frac{n}{b_0+n}\bar{x} + \frac{b_0}{b_0+n}m_0 \end{cases}$$

Khi đó, $\mu|\tau, \underline{x} \sim N(m_1, b_1\tau)$ *và* $\tau|\underline{x} \sim Gamma(a_1, c_1)$.

Do đó, ước lượng Bayes bằng phương pháp trung bình bình phương sai số nhỏ nhất:

$$\begin{cases} \hat{\mu}_{MSE} = m_1 \\ \hat{\tau}_{MSE} = \frac{a_1}{c_1} \end{cases}$$

Chứng minh. Hàm hợp lý có dạng:

$$f(\underline{x}|\mu, \tau) = (2\pi)^{-n/2}\tau^{n/2} exp\left\{-\frac{\tau}{2}\sum_{i=1}^{n}(x_i - \mu)^2\right\}$$

Ta không cần quan tâm đến hằng số trong hàm hợp lý, do đó,

$$f(\underline{x}|\mu,\tau) \propto \tau^{n/2} exp\left\{-\frac{\tau}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}$$

Ta có biến đổi sau

$$\begin{aligned}\sum_{i=1}^{n}(x_i-\mu)^2 &= \sum_{i=1}^{n}(x_i^2-2\mu x_i+\mu^2)\\ &= n(\mu-\bar{x})^2+n\bar{x^2}-\bar{x}^2\\ &= n(\mu-\bar{x})^2+ns^2\end{aligned}$$

Vậy,

$$f(\underline{x}|\mu,\tau) \propto \tau^{n/2} exp\left\{-\frac{\tau}{2}\left[n(\mu-\bar{x})^2+ns^2\right]\right\}$$

với $\mu \in R$, $\tau \in R_*^+$.

Để tạo ra được một tiên nghiệm liên hợp tự nhiên, ta bắt chước dạng hàm của hàm hợp lý, phân phối tiên nghiệm liên hợp tự nhiên sẽ có dạng:

$$\pi(\mu,\tau) \propto \tau^{a_0-\frac{1}{2}} exp\{-\frac{\tau}{2}\left[b_0(\mu-m_0)^2+2c_0\right]\}$$

Tiên nghiệm liên hợp này chính là phân phối hai chiều chuẩn gamma

$$\pi(\mu,\tau) \sim N-G(a_0,b_0,c_0,m_0)$$

Theo định lý Bayes, phân phối hậu nghiệm tỷ lệ với tích của hàm hợp lý và phân phối tiên nghiệm:

$$f(\underline{x}|\mu,\tau)\pi(\mu,\tau) \propto \tau^{a_0+\frac{n}{2}-\frac{1}{2}} exp\{-\frac{\tau}{2}\left[n(\mu-\bar{x})^2+b_0(\mu-m_0)^2+2c_0+ns^2\right]\}$$

Ta có:

$n(\mu-\bar{x})^2+b_0(\mu-m_0)^2 = (n+b_0)\left[\mu-(\frac{n}{b_0+n}\bar{x}+\frac{b_0}{b_0+n}m_0)^2\right]+\frac{b_0 n}{b_0+n}(\bar{x}-m_0)^2$

Suy ra,

$$p(\mu,\tau|\underline{x}) \propto \tau^{a_0+n/2-1/2}$$

$$\times \exp\left\{-\frac{\tau}{2}\left\{(n+b_0)\big[\mu-(\frac{n}{b_0+n}\bar{x}+\frac{b_0}{b_0+n}m_0)^2\big]+\frac{b_0 n}{b_0+n}(\bar{x}-m_0)^2+2c_0+ns^2\right\}\right\}$$

Ta dễ dàng thấy được, phân phối hậu nghiệm $(\mu,\tau|\underline{x}) \sim N-G(a_1,b_1,c_1,m_1)$ với

$$\begin{cases} a_1 = a_0 + \frac{n}{2} \\ b_1 = b_0 + n \\ c_1 = c_0 + \frac{n}{2}\big[s^2 + \frac{b_0}{b_0+n}(\bar{x}-m_0)^2\big] \\ m_1 = \frac{n}{b_0+n}\bar{x} + \frac{b_0}{b_0+n}m_0 \end{cases}$$

Vậy, ước lượng Bayes bằng phương pháp trung bình bình phương nhỏ nhất là:

$$\begin{cases} \hat{\mu}_{MSE} = \frac{n}{b_0+n}\bar{x} + \frac{b_0}{b_0+n}m_0 \\ \hat{\tau}_{MSE} = \frac{a_0+n/2}{c_0+\frac{n}{2}\left[s^2+\frac{b_0}{b_0+n}(\bar{x}-m_0)^2\right]} \end{cases}$$

Định lý đã được chứng minh. $\square$

Mô phỏng

Để nghiên cứu về tiên nghiệm liên hợp tự nhiên, ta so sánh giữa phương pháp ước lượng Bayes và phương pháp ước lượng hợp lý cực đại dựa trên dữ liệu mô phỏng. Toàn bộ quá trình mô phỏng sẽ được thực hiện trong R.

Giá trị đầu vào:

$$\mu = 5$$
$$\tau = 0.25$$

Sinh ngẫu nhiên n quan sát từ phân phối chuẩn có kỳ vọng và phương sai là các giá trị đầu vào. Ta xây dựng tất cả kịch bản khác nhau theo kích thước mẫu và dự đoán của chuyên gia. Dưới đây là bảng kết quả mô phỏng một số kịch bản khi kích thước mẫu nhỏ.

Sample-size	Prior guess				Bayesian estimates	
n	$g_{\tau,1}$	$g_{\tau,2}$	$g_{\mu,1}$	$g_{\mu,2}$	$\hat{\tau}$	$\hat{\mu}$
10	0,1	0,09	2	1,8	0.1367411	2.998999
			4,999	4,9991	0,2548048	4,699806
			6	5,4	0,2419066	4,832526
		0,06	2	1,2	0,1349408	3,257419
			4,999	2,9994	0,1951376	4,677623
			6	3,6	0,1907512	4,770088
		0,03	2	0,6	0,1081469	2,64696
			4,999	1,4997	0,1306453	4,745611
			6	1,8	0,1283843	4,970857
	0,2499	0,2249	2	1,8	0,2047143	3,588942
			4,999	4,9991	0,367118	4,657426
			6	5,4	0,3551478	4,714642
		0,14994	2	1,2	0,2322025	3,82806
			4,999	2,9994	0,3361668	4,646317
			6	3,6	0,330582	4,685236
		0,07497	2	0,6	0,2798427	4,786503
			4,999	1,4997	0,2848829	4,683493
			6	1,8	0,2798427	4,786503

Sample-size	Prior guess				Bayesian estimates	
n	$g_{\tau,1}$	$g_{\tau,2}$	$g_{\mu,1}$	$g_{\mu,2}$	$\hat{\tau}$	$\hat{\mu}$
10	0,3	0,27	2	1,8	0,2213739	3,70084
			4,999	4,9991	0,3861108	4,651952
			6	5,4	0,37492	4,700018
		0,18	2	1,2	0,2561575	3,925406
			4,999	2,9994	0,3656162	4,642448
			6	3,6	0,3600647	4,675052
		0,09	2	0,6	0,242836	3,299531
			4,999	1,4997	0,3280514	4,674634
			6	1,8	0,3223734	4,761792
		MLE			**0.4175408**	**4.621806**

ii) Tiên nghiệm phẳng

Định lý sau đây chỉ ra phân phối tiên nghiệm phẳng, phân phối hậu nghiệm và ước lượng Bayes.

Định lý 2.1.16 *Đặt $\underline{x} = (x_1, ..., x_n)$ là các quan sát độc lập cùng phân phối chuẩn $N(\mu, \sigma^2)$. Theo như hàm bình phương mất mát và phân phối tiên nghiệm phẳng $\pi(\mu, \tau) \propto \tau^{3/2}$, ta thu được các kết quả sau:*

(i) Phân phối hậu nghiệm $p(\mu, \tau|\underline{x}) \sim N - G(\frac{n}{2} + 2, n, \frac{ns^2}{2}, \bar{x})$

(ii) Ước lượng Bayes

$$\begin{cases} \hat{\mu}_{MSE} = \bar{x} \\ \hat{\tau}_{MSE} = \frac{n/2+2}{ns^2/2} = (1 + \frac{4}{n}).\frac{1}{s^2} \end{cases}$$

Chứng minh. Để tìm được phân phối tiên nghiệm, ta cần tìm ma trận thông tin Fisher.

$$\mathbf{I}(\mu, \sigma^{\mathbf{2}}) = \begin{pmatrix} -E(\frac{\partial^2}{\partial^2 \mu}) & -E(\frac{\partial^2}{\partial \mu \partial \sigma^2}) \\ -E(\frac{\partial^2}{\partial \sigma^2 \partial \mu}) & -E(\frac{\partial^2}{\partial^2 (\sigma^2)}) \end{pmatrix}$$

Ta có hàm logarit của hàm hợp lý như sau:

$$l(\mu, \sigma^2) = logL(\mu, \sigma^2) = -\frac{n}{2}log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$$

Suy ra

$$\frac{\partial}{\partial \mu}l(\mu, \sigma^2) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu) = \frac{1}{\sigma^2}(n\bar{x} - n\mu)$$

$$\frac{\partial}{\partial \sigma^2}l(\mu, \sigma^2) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - \mu)^2$$

$$\frac{\partial^2}{\partial^2 \mu} = \frac{-n}{\sigma^2}$$

$$\frac{\partial^2}{\partial^2 (\sigma^2)} = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{i=1}^{n}(x_i - \mu)^2$$

$$\frac{\partial^2}{\partial \mu \partial \sigma^2} = \frac{\partial^2}{\partial \sigma^2 \partial \mu} = -\frac{n}{\sigma^4}(\bar{x} - \mu)$$

Suy ra, ta có ma trận thông tin Fisher:

$$\mathbf{I}(\mu,\sigma^{\mathbf{2}}) = \begin{pmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{pmatrix}$$

Do đó,

$$\det I(\mu,\sigma^2) = \frac{n^2}{2\sigma^6}$$

Suy ra,

$$\det I(\mu,\tau) = \frac{n^2\tau^3}{2}$$

Jeffrey'rule:

$$\pi(\mu,\tau) \propto \sqrt{det(I)}$$

$$=> \pi(\mu,\tau) \propto \tau^{3/2}$$

Ta đã có hàm hợp lý

$$f(\underline{x}|\mu,\tau) \propto \tau^{\frac{n}{2}} exp\{-\frac{\tau}{2}[n(\mu-\bar{x})^2 + ns^2]\}$$

Hàm phân phối tiên nghiệm phẳng

$$\pi(\mu,\tau) \propto \tau^{3/2}$$

Theo định lý Bayes, phân phối hậu nghiệm tỉ lệ với tích của phân phối tiên nghiệm và hàm hợp lý

$$f(\underline{x}|\mu,\tau)\pi(\mu,\tau) \propto \tau^{n/2+2-\frac{1}{2}} exp\left\{-\frac{\tau}{2}[n(\mu-\bar{x})^2 + 2\frac{ns^2}{2}]\right\}$$

Vậy,

$$(\mu,\tau)|\underline{x} \sim N-G(\frac{n}{2}+2, n, \frac{ns^2}{2}, \bar{x})$$

Vậy, theo bình phương hàm mất mát, ước lượng Bayes bằng phương pháp trung bình bình phương sai số nhỏ nhất là

$$\begin{cases} \hat{\mu}_{MSE} = \bar{x} = \hat{\mu}_{MLE} \\ \hat{\tau}_{MSE} = \frac{n/2+2}{ns^2/2} = (1+\frac{4}{n}).\frac{1}{s^2} \approx \hat{\tau}_{MLE} \end{cases}$$

Định lý đã được chứng minh. $\square$

Nhận xét 2.1.17 *Trong trường hợp tiên nghiệm phẳng thì*

$$\begin{cases} \hat{\mu}_{MSE} = \hat{\mu}_{MLE} \\ \hat{\tau}_{MSE} \approx \hat{\tau}_{MLE} \end{cases}$$

2.2 Mô hình hồi quy tuyến tính

Định nghĩa 2.2.1 *Đặt $X_1, X_2, ..., X_k, Y$ là các biến ngẫu nhiên, véc tơ $\boldsymbol{\beta}$ và số thực σ^2. Mô hình hồi quy tuyến tính đa biến với các tham số $(\boldsymbol{\beta}, \sigma)$ có dạng:*

$$Y = \boldsymbol{X}\boldsymbol{\beta} + \epsilon \tag{2.1}$$

Trong đó, $\epsilon \sim N(0, \sigma^2)$, $\boldsymbol{X} = (X_1, ..., X_k)$, $\boldsymbol{\beta} = (\beta_1, ..., \beta_k)^T$.
Kí hiệu $(\boldsymbol{X}, Y) \sim MLR(\boldsymbol{\beta}, \sigma^2)$.

Theo như mô hình, mỗi biến phụ thuộc Y_i, trên điều kiện có bộ dữ liệu quan sát $\mathbf{X}_{i\cdot} = (X_{i1}, ..., X_{ik})$ và các tham số $\boldsymbol{\beta}, \sigma^2$, tuân theo luật phân phối chuẩn:

$$Y_i|\mathbf{X}_{i\cdot}, \boldsymbol{\beta}, \sigma^2 \sim N(\mathbf{X}_{i\cdot}\boldsymbol{\beta}, \sigma^2)$$

Do đó, hàm mật độ xác suất của $Y_i|\mathbf{X}_{i\cdot}, \boldsymbol{\beta}, \sigma^2$ có dạng:

$$f_{Y_i|\mathbf{X}_{i\cdot}, \boldsymbol{\beta}, \sigma^2}(y) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left\{-\frac{1}{2\sigma^2}(y_i - \mathbf{x}_{i\cdot}\boldsymbol{\beta})^2\right\}$$

Để thuận tiện, ta sẽ kí hiệu $f_{Y_i|\mathbf{X}_{i\cdot}, \boldsymbol{\beta}, \sigma^2}(y)$ thành $f(y_i|\mathbf{x}_{i\cdot}, \boldsymbol{\beta}, \sigma^2)$.
Ta tiến hành n quan sát độc lập, vì mỗi biến phụ thuộc Y_i độc lập nên hàm mật độ xác suất đồng thời có dạng:

$$f(\mathbf{y}|\mathbf{x}, \boldsymbol{\beta}, \sigma^2) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mathbf{x}_{i\cdot}\boldsymbol{\beta})^2\right\}$$

Trong đó

$$\mathbf{x}_{i\cdot} = (x_{i1}, x_{i2}, ..., x_{ik}), \text{với } i = \overline{1, n}$$

$$\mathbf{x} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1k} \\ x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix}$$

Vì hàm mật độ đồng thời chỉ phụ thuộc vào $\boldsymbol{\beta}, \sigma^2$ nên ta coi nó là hàm của $\boldsymbol{\beta}, \sigma^2$. Do đó, ta được hàm hợp lý:

$$L(\boldsymbol{\beta}, \sigma^2) = \Big(\frac{1}{\sqrt{2\pi\sigma^2}}\Big)^n exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mathbf{x}_{i\cdot}\boldsymbol{\beta})^2\right\}$$

2.2.1 Phương pháp ước lượng hợp lý cực đại

Mục tiêu của phương pháp ước lượng hợp lý cực đại là đi tìm tham số $\boldsymbol{\beta}, \sigma^2$ để hàm hợp lý đạt giá trị lớn nhất.

Định lý 2.2.2 *Giả sử ta có n quan sát từ mô hình $MLR(\boldsymbol{\beta}, \sigma^2)$. Gọi $\hat{\boldsymbol{\beta}}_{MLE}, \hat{\sigma^2}_{MLE}$ lần lượt là ước lượng hợp lý cực đại của $\boldsymbol{\beta}, \sigma^2$. Khi đó,*

$$\hat{\boldsymbol{\beta}}_{MLE} = (\mathbf{x}^{\mathbf{T}}\mathbf{x})^{-1}\mathbf{x}^{\mathbf{T}}\mathbf{y}$$
$$\hat{\sigma^2}_{MLE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \mathbf{x}_{i\cdot}\hat{\boldsymbol{\beta}}_{\boldsymbol{MLE}})^2$$

Chứng minh. Ta có hàm hợp lý có dạng:

$$L(\boldsymbol{\beta}, \sigma^2) = \Big(\frac{1}{\sqrt{2\pi\sigma^2}}\Big)^n exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mathbf{x}_{i\cdot}\boldsymbol{\beta})^2\right\}$$

Ta tìm giá trị cực đại của hàm hợp lý trên cũng tương đương với việc tìm giá trị cực đại hàm Logarit của hàm hợp lý. Do đó, ta có:

$$l(\boldsymbol{\beta}, \sigma^2) = \log L(\boldsymbol{\beta}, \sigma^2) = -\frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mathbf{x}_{i\cdot}\boldsymbol{\beta})^2$$

Để lấy giá trị cực đại hàm hợp lý, ta suy ra:

$$h(\boldsymbol{\beta}) = \sum_{i=1}^{n} (y_i - \mathbf{x}_{i\cdot}\boldsymbol{\beta})^2 \to min$$

Ta có:

$$\begin{aligned}
&h(\boldsymbol{\beta}) = \sum_{i=1}^{n} (y_i - \mathbf{x}_{i\cdot}\boldsymbol{\beta})^2 \\
=> &\frac{\partial}{\partial \beta_1} h(\boldsymbol{\beta}) = -2\sum_{i=1}^{n} (y_i - \mathbf{x}_{i\cdot}\boldsymbol{\beta})x_{i1} = -2\sum_{i=1}^{n} (y_i x_{i1} - \mathbf{x}_{i\cdot}\boldsymbol{\beta}x_{i1}) \\
&\frac{\partial}{\partial \beta_2} h(\boldsymbol{\beta}) = -2\sum_{i=1}^{n} (y_i - \mathbf{x}_{i\cdot}\boldsymbol{\beta})x_{i2} = -2\sum_{i=1}^{n} (y_i x_{i2} - \mathbf{x}_{i\cdot}\boldsymbol{\beta}x_{i2}) \\
&\dots \\
&\frac{\partial}{\partial \beta_k} h(\boldsymbol{\beta}) = -2\sum_{i=1}^{n} (y_i - \mathbf{x}_{i\cdot}\boldsymbol{\beta})x_{ik} = -2\sum_{i=1}^{n} (y_i x_{ik} - \mathbf{x}_{i\cdot}\boldsymbol{\beta}x_{ik})
\end{aligned}$$

Nên ta có hệ phương trình:

$$\begin{cases}
\sum_{i=1}^{n} (y_i x_{i1} - \mathbf{x}_{i\cdot}\boldsymbol{\beta}x_{i1}) = 0 \\
\sum_{i=1}^{n} (y_i x_{i2} - \mathbf{x}_{i\cdot}\boldsymbol{\beta}x_{i2}) = 0 \\
\dots \\
\sum_{i=1}^{n} (y_i x_{ik} - \mathbf{x}_{i\cdot}\boldsymbol{\beta}x_{ik}) = 0
\end{cases}$$

Suy ra:

$$\mathbf{x^T x}\boldsymbol{\beta} = \mathbf{x^T y}$$

Do đó,

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{MLE}} = (\mathbf{x^T x})^{\mathbf{-1}}\mathbf{x^T y}$$

Ta có:

$$l(\boldsymbol{\beta}, \sigma^2) = \log L(\boldsymbol{\beta}, \sigma^2) = -\frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n} (y_i - \mathbf{x}_{i\cdot}\boldsymbol{\beta})^2$$

nên

$$\frac{\partial}{\partial \sigma^2} l(\boldsymbol{\beta}, \sigma^2) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{n} (y_i - \mathbf{x}_{i\cdot}\boldsymbol{\beta})^2 = 0$$

Do đó,

$$\hat{\sigma^2}_{MLE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \mathbf{x}_{i\cdot}\hat{\boldsymbol{\beta}}_{\boldsymbol{MLE}})^2$$

Định lý đã được chứng minh. $\square$

2.2.2 Phương pháp Bayes

Tiếp theo, ta xét đến ước lượng bayes.

a) Tiên nghiệm liên hợp tự nhiên

Ta sẽ xét hai trường hợp sau đây.

Trường hợp 1: σ^2 đã biết, cần ước lượng $\boldsymbol{\beta}$

Định lý 2.2.3 *Tiên nghiệm liên hợp tự nhiên cho $\boldsymbol{\beta}|\sigma^2$ tuân theo luật phân phối chuẩn nhiều chiều $\boldsymbol{\beta}|\sigma^2 \sim MN(\boldsymbol{\mu}_0, \Sigma_0)$ và hậu nghiệm tuân theo luật phân phối chuẩn nhiều chiều $\boldsymbol{\beta}|\sigma^2, \mathbf{y}, \mathbf{x} \sim MN(\boldsymbol{\mu}_1, \Sigma_1)$ với các tham số cập nhật:*

$$\Sigma_1^{-1} = \Sigma_0^{-1} + \Sigma_y^{-1}$$
$$\boldsymbol{\mu}_1 = \Sigma_1(\Sigma_0^{-1}\boldsymbol{\mu}_0 + \Sigma_y^{-1}\boldsymbol{\mu}_y)$$

trong đó,

$$\boldsymbol{\mu}_y = (\mathbf{x}^{\mathbf{T}}\mathbf{x})^{-1}\mathbf{x}^{\mathbf{T}}\mathbf{y}$$
$$\Sigma_y = \sigma^2(\mathbf{x}^{\mathbf{T}}\mathbf{x})^{-1}$$

Khi đó,

$$\hat{\boldsymbol{\beta}}_{MSE} = \boldsymbol{\mu}_{\mathbf{1}}$$

Chứng minh. Ta có hàm mật độ xác suất đồng thời có dạng:

$$
\begin{aligned}
f(\mathbf{y}|\mathbf{x},\boldsymbol{\beta},\sigma^2) &= \Big(\frac{1}{\sqrt{2\pi\sigma^2}}\Big)^n exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mathbf{x}_{i\cdot}\boldsymbol{\beta})^2\right\} \\
&\propto (\sigma^2)^{-n/2} exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mathbf{x}_{i\cdot}\boldsymbol{\beta})^2\right\} \\
&\propto \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{x}\boldsymbol{\beta})^T(\boldsymbol{y}-\boldsymbol{x}\boldsymbol{\beta})\right\} \\
&\propto \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{\beta}^{\boldsymbol{T}}\boldsymbol{x}^{\boldsymbol{T}}\boldsymbol{x}\boldsymbol{\beta} - 2\boldsymbol{\beta}^{\boldsymbol{T}}\boldsymbol{x}^{\boldsymbol{T}}\boldsymbol{y})\right\} \\
&\propto \exp\left\{-\frac{1}{2}(\boldsymbol{\beta}^T\Sigma_y^{-1}\boldsymbol{\beta} - 2\boldsymbol{\beta}^T\Sigma_y^{-1}\boldsymbol{\mu}_y)\right\}
\end{aligned}
$$

với,

$$
\begin{aligned}
\boldsymbol{\mu}_y &= (\mathbf{x}^{\mathbf{T}}\mathbf{x})^{-1}\boldsymbol{x}^{\boldsymbol{T}}\mathbf{y} \\
\Sigma_y &= \sigma^2(\mathbf{x}^{\mathbf{T}}\mathbf{x})^{-1}
\end{aligned}
$$

Phân phối tiên nghiệm liên hợp tự nhiên sẽ có dạng:

$$
\pi(\boldsymbol{\beta}|\sigma^2) \propto \exp\left\{-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\mu}_0)^T\Sigma_0^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu}_0)\right\}
$$

Ta biến đổi

$$
\begin{aligned}
&(\boldsymbol{\beta}-\boldsymbol{\mu}_0)^T\Sigma_0^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu}_0) \\
=\ &\boldsymbol{\beta}^T\Sigma_0^{-1}\boldsymbol{\beta} - \boldsymbol{\mu}_0^T\Sigma_0^{-1}\boldsymbol{\beta} - \boldsymbol{\beta}^T\Sigma_0^{-1}\boldsymbol{\mu}_0 + \boldsymbol{\mu}_0^T\Sigma_0^{-1}\boldsymbol{\mu}_0 \\
=\ &\boldsymbol{\beta}^T\Sigma_0^{-1}\boldsymbol{\beta} - 2\boldsymbol{\beta}^T\Sigma_0^{-1}\boldsymbol{\mu}_0
\end{aligned}
$$

Do đó, phân phối hậu nghiệm:

$$\begin{aligned}
p(\boldsymbol{\beta}|\sigma^2, \mathbf{y}, \mathbf{x}) &\propto f(\mathbf{y}|\mathbf{x}, \boldsymbol{\beta}, \sigma^2) \times \pi(\boldsymbol{\beta}|\sigma^2) \\
&\propto \exp\left\{-\frac{1}{2}\left[\boldsymbol{\beta}^T\left(\Sigma_0^{-1} + \Sigma_y^{-1}\right)\boldsymbol{\beta} - 2\boldsymbol{\beta}^T\left(\Sigma_0^{-1}\boldsymbol{\mu}_0 + \Sigma_y^{-1}\boldsymbol{\mu}_y\right)\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[\boldsymbol{\beta}^T\Sigma_1^{-1}\boldsymbol{\beta} - 2\boldsymbol{\beta}^T\Sigma_1^{-1}\left(\Sigma_1\Sigma_0^{-1}\boldsymbol{\mu}_0 + \Sigma_1\Sigma_y^{-1}\boldsymbol{\mu}_y\right)\right]\right\} \\
&= \exp\left\{-\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\mu}_1)^T\Sigma_1^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu}_1)\right\}
\end{aligned}$$

với,

$$\begin{aligned}
\Sigma_1^{-1} &= \Sigma_0^{-1} + \Sigma_y^{-1} \\
\boldsymbol{\mu}_1 &= \Sigma_1(\Sigma_0^{-1}\boldsymbol{\mu}_0 + \Sigma_y^{-1}\boldsymbol{\mu}_y)
\end{aligned}$$

Do đó, $\boldsymbol{\beta}|\sigma^2, \mathbf{y}, \mathbf{x} \sim MN(\boldsymbol{\mu}_1, \Sigma_1)$.

Vậy,

$$\hat{\boldsymbol{\beta}}_{MSE} = \boldsymbol{\mu_1}$$

Định lí đã được chứng minh. □

Trường hợp 2: $\boldsymbol{\beta}$ đã biết, cần ước lượng σ^2

Định lý 2.2.4 *Tiên nghiệm liên hợp tự nhiên cho* $\sigma^2|\boldsymbol{\beta}$ *tuân theo luật phân phối gamma nghịch đảo* $\sigma^2|\boldsymbol{\beta} \sim IGamma(a_0, b_0)$ *và hậu nghiệm tuân theo luật phân phối gamma nghịch đảo* $\sigma^2|\boldsymbol{\beta}, \mathbf{y}, \mathbf{x} \sim IGamma(a_1, b_1)$ *với các tham số cập nhật:*

$$\begin{aligned}
a_1 &= a_0 + \frac{n}{2} \\
b_1 &= b_0 + \frac{1}{2}(\mathbf{y} - \mathbf{x}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{x}\boldsymbol{\beta})
\end{aligned}$$

Khi đó, ước lượng Bayes bằng phương pháp trung bình bình phương sai số nhỏ nhất là

$$\hat{\sigma^2}_{MSE} = \frac{b_1}{a_1 - 1}$$

Chứng minh. Ta có hàm mật độ xác suất đồng thời có dạng:

$$\begin{aligned}
f(\mathbf{y}|\mathbf{x}, \boldsymbol{\beta}, \sigma^2) &= \Big(\frac{1}{\sqrt{2\pi\sigma^2}}\Big)^n exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mathbf{x}_{i\cdot}\boldsymbol{\beta})^2\right\} \\
&\propto (\sigma^2)^{-n/2} exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mathbf{x}_{i\cdot}\boldsymbol{\beta})^2\right\} \\
&\propto \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{x}\boldsymbol{\beta})^T(y - \mathbf{x}\boldsymbol{\beta})\right\}
\end{aligned}$$

Phân phối tiên nghiệm liên hợp tự nhiên sẽ có dạng:

$$\pi(\sigma^2|\boldsymbol{\beta}) \propto (\sigma^2)^{-a_0-1} \exp\left\{-\frac{b_0}{\sigma^2}\right\}$$

Do đó, phân phối hậu nghiệm:

$$\begin{aligned}
p(\sigma^2|\boldsymbol{\beta}, \mathbf{y}, \mathbf{x}) &\propto f(\mathbf{y}|\mathbf{x}, \boldsymbol{\beta}, \sigma^2) \times \pi(\sigma^2|\boldsymbol{\beta}) \\
&\propto (\sigma^2)^{-a_0-\frac{n}{2}-1} \exp\left\{-\frac{1}{\sigma^2}\left[b_0 + \frac{1}{2}(\mathbf{y} - \mathbf{x}\boldsymbol{\beta})^T(y - \mathbf{x}\boldsymbol{\beta})\right]\right\}
\end{aligned}$$

Do đó, $\sigma^2|\boldsymbol{\beta}, \mathbf{y}, \mathbf{x} \sim IGamma(a_1, b_1)$

trong đó,

$$\begin{aligned}
a_1 &= a_0 + \frac{n}{2} \\
b_1 &= b_0 + \frac{1}{2}(\mathbf{y} - \mathbf{x}\boldsymbol{\beta})^T(y - \mathbf{x}\boldsymbol{\beta})
\end{aligned}$$

Khi đó, ước lượng Bayes bằng phương pháp trung bình bình phương sai số nhỏ nhất là

$$\hat{\sigma^2}_{MSE} = \frac{b_1}{a_1 - 1}$$

Định lí đã được chứng minh. □

Mô phỏng

Toàn bộ quá trình mô phỏng sẽ được thực hiện trong R.

Giá trị đầu vào:

$$\boldsymbol{\beta} = (1, 2, 3, 4, 5)$$
$$\sigma = 0.5$$

Đầu tiên, sinh ngẫu nhiên n giá trị $\boldsymbol{x} = (x_1, ..., x_4)$ từ phân phối chuẩn nhiều chiều có kỳ vọng là véc tơ $(1, 2, 3, 4)$ và ma trận hiệp phương sai là ma trận đơn vị cấp 4. Với mỗi $\boldsymbol{x} = (x_1, ..., x_4)$, sinh ngẫu nhiên ra y tương ứng từ phân phối chuẩn có kỳ vọng là $(1, x) * \boldsymbol{\beta}$ và phương sai $\sigma = 0.5$ như đã cho.

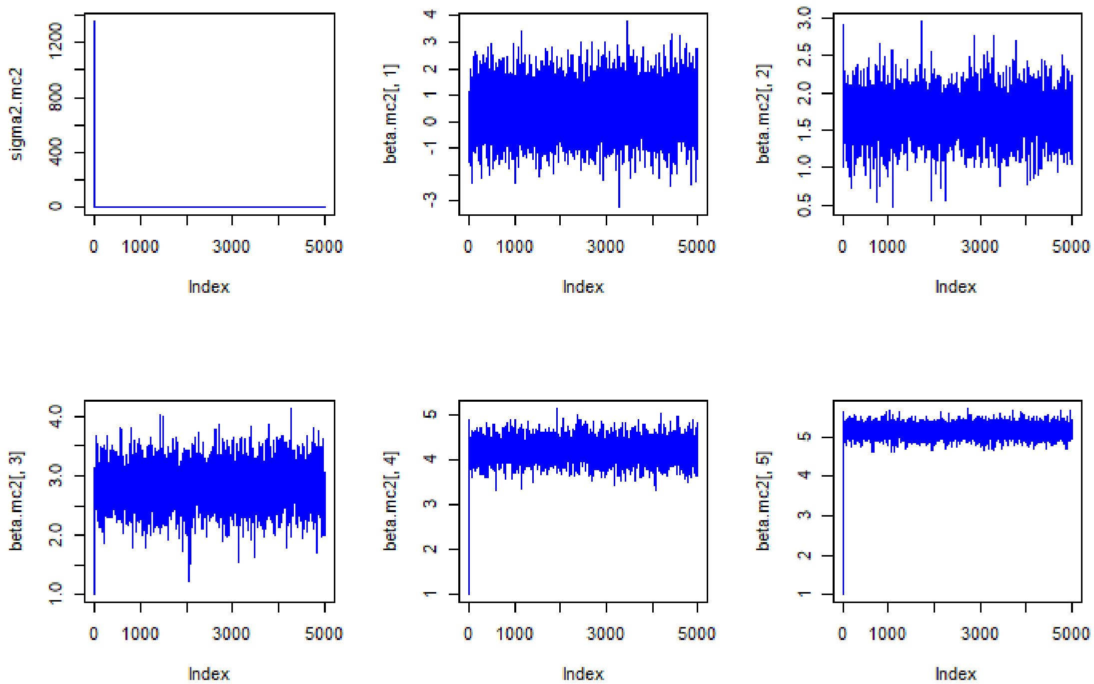
Sử dụng thuật toán Gibbs sampling để sinh ra xích Markov cần tìm.

Dưới đây là kết quả mô phỏng.

Khi n = 10

	β_1	β_2	β_3	β_4	β_5
MLE	-3.163099	2.135291	3.471609	4.662085	5.316256
MSE	0.5051601	1.8692737	2.9120023	4.2416253	5.0283090

Sau đây là hình ảnh hội tụ của xích.

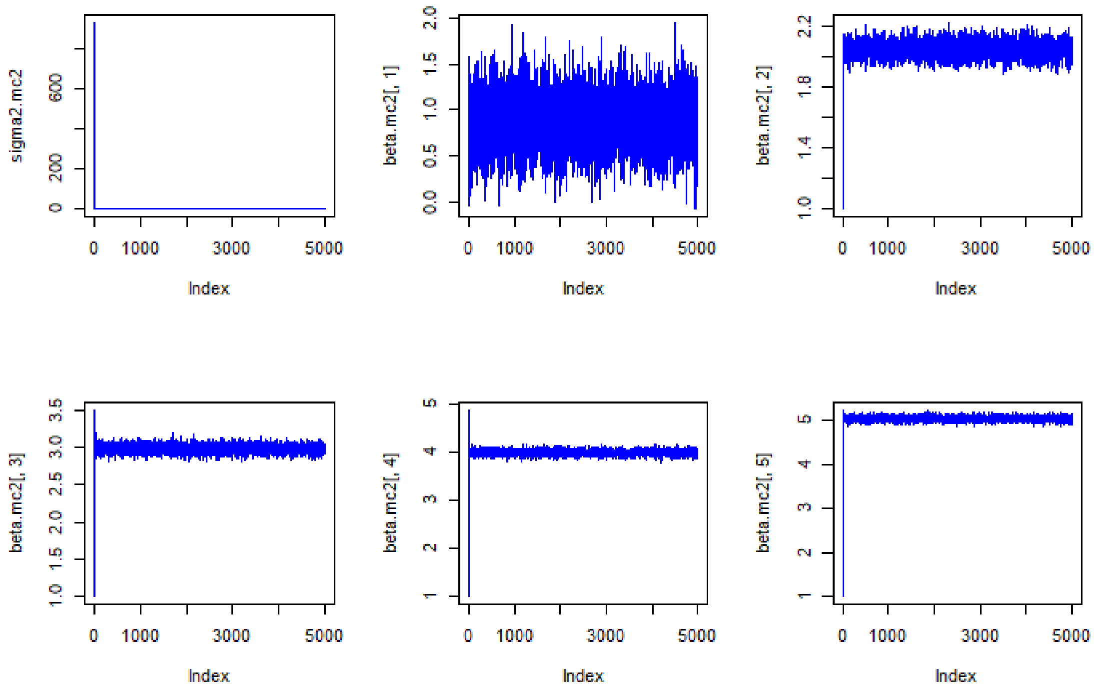


Hình 2.2: Sự hội tụ của xích Markov n = 10

Khi n = 100

	β_1	β_2	β_3	β_4	β_5
MLE	0.9021695	2.0444764	2.9739440	3.9906523	5.0381899
MSE	0.8860276	2.0445855	2.9751749	3.9915452	5.0394853

Sau đây là hình ảnh hội tụ của xích.

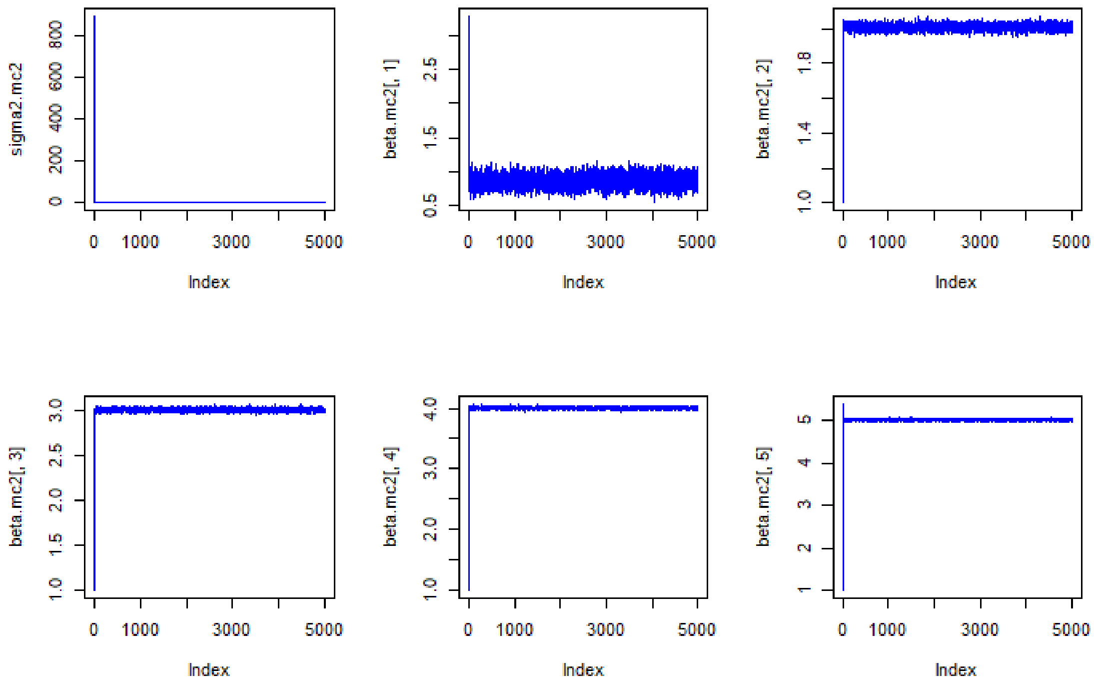


Hình 2.3: Sự hội tụ của xích Markov n =100

Khi n = 1000

	β_1	β_2	β_3	β_4	β_5
MLE	0.8622019	2.0100663	3.0051328	4.0034197	5.0276165
MSE	0.8601638	2.0097422	3.0046074	4.0030341	5.0272572

Sau đây là hình ảnh hội tụ của xích.



Hình 2.4: Sự hội tụ của xích Markov n = 1000

b) Tiên nghiệm phẳng

Định lý sau đây chỉ ra phân phối tiên nghiệm phẳng, phân phối hậu nghiệm cho phân tích Bayes của mô hình $MLR(\boldsymbol{\beta}, \sigma^2)$.

Định lý 2.2.5 *Giả sử ta có n quan sát từ mô hình $MLR(\boldsymbol{\beta}, \sigma^2)$. Theo như hàm bình phương mất mát và phân phối tiên nghiệm phẳng $\pi(\alpha, \beta) \propto \frac{1}{\sigma^2}$, ta thu được các kết quả sau:*

(i)

$$\boldsymbol{\beta}|\mathbf{y}, \mathbf{x}, \sigma^2 \sim MN(\boldsymbol{\mu}, \Sigma)$$

trong đó,

$$\boldsymbol{\mu} = (\mathbf{x^T x})^{-1}\mathbf{x^T y}$$
$$\Sigma = \sigma^2(\mathbf{x^T x})^{-1}$$

(ii)

$$\sigma^2|\mathbf{y},\mathbf{x},\boldsymbol{\beta} \sim IGamma\Big(\frac{n}{2},\frac{1}{2}\sum_{i=1}^{n}(y_i-\mathbf{x}_{i\cdot}\boldsymbol{\beta})^2\Big)$$

Mô phỏng

Toàn bộ quá trình mô phỏng sẽ được thực hiện trong R.

Giá trị đầu vào:

$$\boldsymbol{\beta} = (1,2,3,4,5)$$
$$\sigma = 0.5$$

Đầu tiên, sinh ngẫu nhiên n giá trị $\boldsymbol{x} = (x_1,...,x_4)$ từ phân phối chuẩn nhiều chiều có kỳ vọng là véc tơ $(1,2,3,4)$ và ma trận hiệp phương sai là ma trận đơn vị cấp 4. Với mỗi $\boldsymbol{x} = (x_1,...,x_4)$, sinh ngẫu nhiên ra y tương ứng từ phân phối chuẩn có kỳ vọng là $(1,x) * \boldsymbol{\beta}$ và phương sai $\sigma = 0.5$ như đã cho.

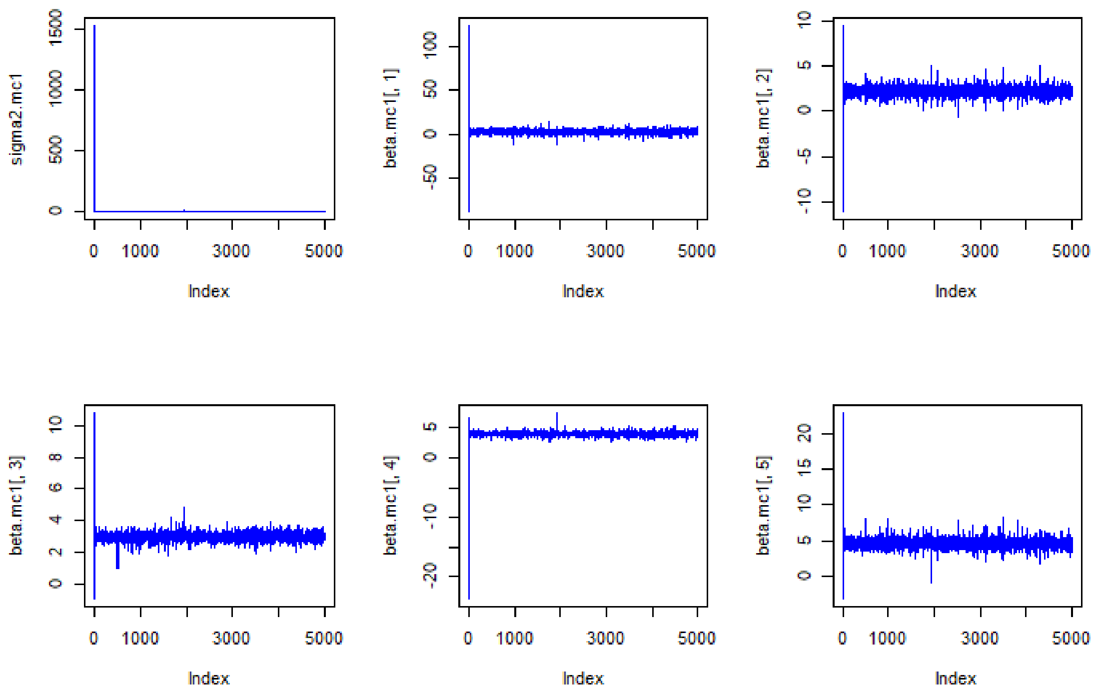
Sử dụng thuật toán Gibbs sampling để sinh ra xích Markov cần tìm.

Dưới đây là kết quả mô phỏng.

Khi n = 10

	β_1	β_2	β_3	β_4	β_5
MLE	2.421866	2.240562	2.981204	4.019570	4.593324
MSE	2.382155	2.227155	2.980494	4.008524	4.612936

Sau đây là hình ảnh hội tụ của xích.

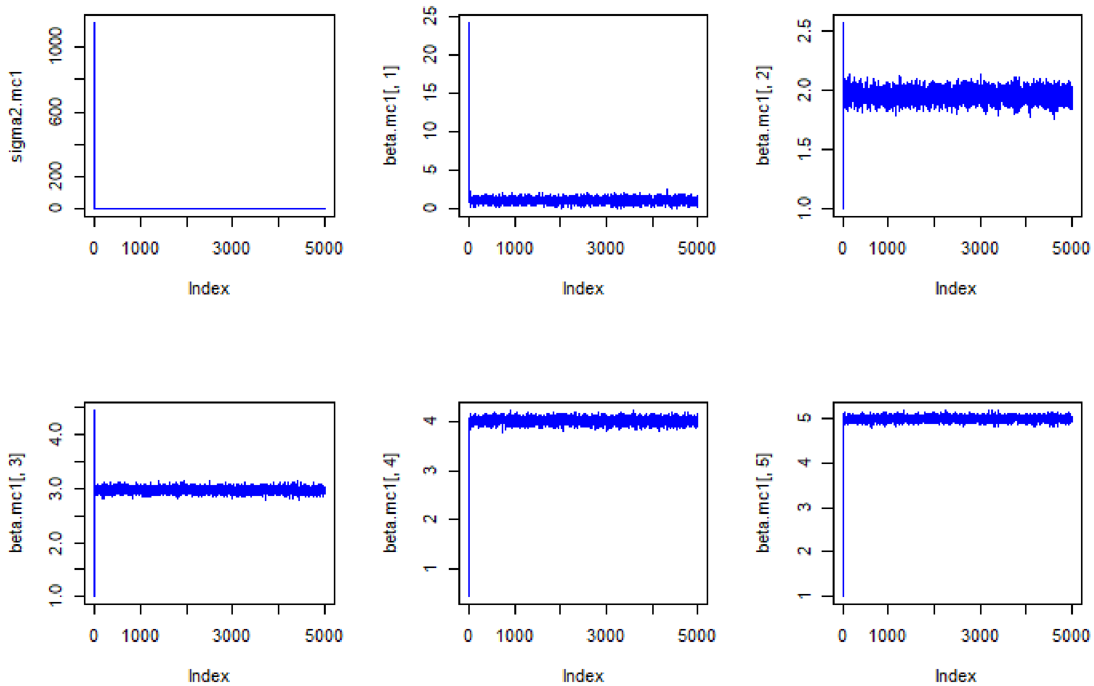


Hình 2.5: Sự hội tụ của xích Markov n =10

Khi n = 100

	β_1	β_2	β_3	β_4	β_5
MLE	1.021015	1.956848	2.970668	4.018061	4.998938
MSE	1.023323	1.956955	2.970311	4.017452	4.997596

Sau đây là hình ảnh hội tụ của xích.

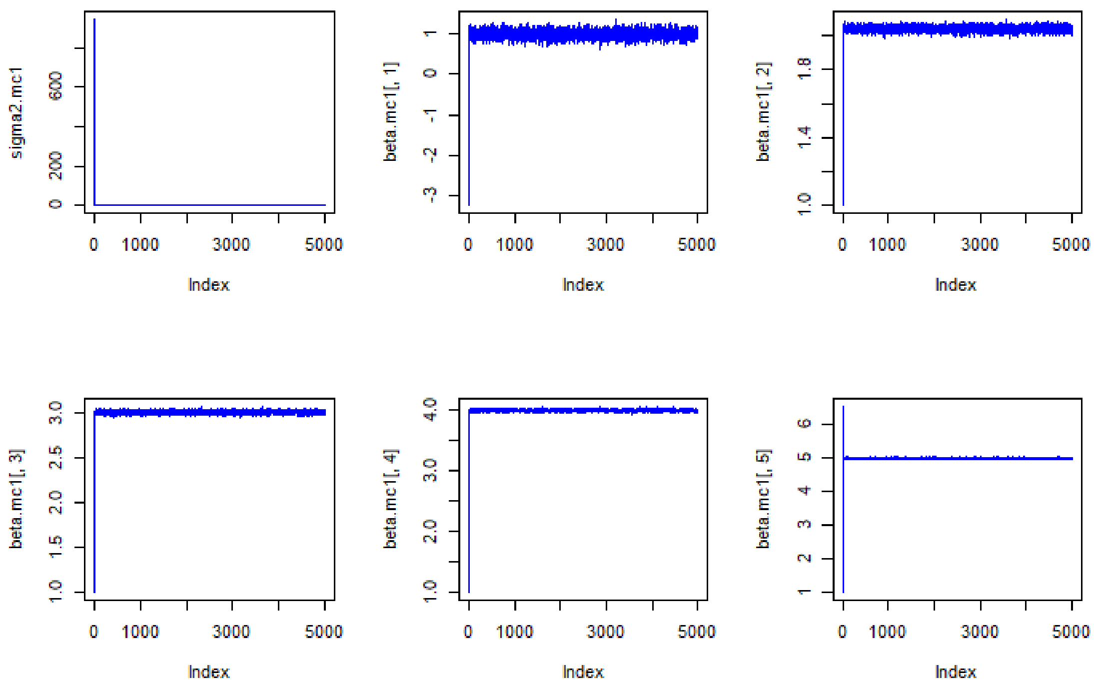


Hình 2.6: Sự hội tụ của xích Markov n = 100

Khi n = 1000

	β_1	β_2	β_3	β_4	β_5
MLE	0.9831367	2.0412784	3.0041335	4.0007088	4.9892627
MSE	0.9831123	2.0408774	3.0035420	4.0000083	4.9887625

Sau đây là hình ảnh hội tụ của xích.



Hình 2.7: Sự hội tụ của xích Markov n = 1000

Chương 3

Ứng dụng

Trong chương này, ta sẽ trình bày kết quả ứng dụng phương pháp Bayes trên bộ dữ liệu thực tế.

3.1 Giới thiệu bộ dữ liệu

Bộ dữ liệu "DirectMarketing.csv" nằm trong quyển sách "Business Analytics for Managers" của tác giả Wolfgang Jank. Bộ dữ liệu bao gồm thông tin của 1000 khách hàng trong cơ sở dữ liệu khách hàng của công ty Direct Marketing(DM).
Mô tả thông tin trường dữ liệu quan trọng

- **Salary:** Tiền lương
- **Children:** Số con
- **Gender_b:** Giới tính, 1 là giới tính nữ, 0 là giới tính nam.
- **Catalogs:** Số sản phẩm mà khách hàng đã mua.
- **AmountSpent:** Tổng số tiền mà khách hàng mua sản phẩm của DM trong một năm.

	AmountSpent	Salary	Children	Gender_b	Catalogs
1	755	47500	0	1	6
2	1318	63600	0	0	6
3	296	13500	0	1	18
4	2436	85600	1	0	18
5	1304	68400	0	1	12
6	495	30400	0	0	6
7	782	48100	0	1	12
8	1155	68400	0	0	18
9	158	51900	3	1	6
10	3034	80700	0	0	18
11	927	43700	1	0	12
12	2065	111800	3	0	18
13	704	44100	1	1	24
14	2136	111400	0	0	12
15	5564	110000	0	1	24
16	2766	83100	1	1	12

Hình 3.1: Minh họa một số quan sát của bộ dữ liệu

3.2 Bài toán

Yêu cầu đặt ra

DM muốn xây dựng một mô hình cho biến AmountSpent, đó là tổng số tiền mà mỗi khách hàng mua sản phẩm của DM trong 1 năm.

Mô hình bài toán

Xét mô hình hồi quy tuyến tính với biến phụ thuộc AmountSpent và các biến độc lập Salary, Children, Gender_b, Catalogs. Khi đó, ta có mô hình như sau:

$$AmountSpent = \beta_0 + \beta_1 Salary + \beta_2 Children + \beta_3 Gender\_b + \beta_4 Catalogs + \epsilon$$

trong đó, $\epsilon \sim N(0, \sigma^2)$.

Sử dụng thuật toán Gibbs sampling để ước lượng các hệ số hồi quy. Chúng ta xét phân phối tiên nghiệm phẳng

$$\pi(\boldsymbol{\beta}, \sigma^2) \propto \frac{1}{\sigma^2}$$

khi đó, phân phối hậu nghiệm có dạng

$$p(\boldsymbol{\beta}, \sigma^2|\mathbf{x}, \mathbf{y}) \propto (\sigma^2)^{-\frac{n}{2}-1} exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mathbf{x}_{i\cdot}\boldsymbol{\beta})^2\}$$

Phân phối hậu nghiệm $p(\boldsymbol{\beta}, \sigma^2|\mathbf{x}, \mathbf{y})$ là phân phối mà chúng ta quan tâm. Mục đích là sinh ra mẫu từ hàm mật độ có điều kiện $p(\boldsymbol{\beta}|\mathbf{y}, \mathbf{x}, \sigma^2)$ và $p(\sigma^2|\mathbf{y}, \mathbf{x}, \boldsymbol{\beta})$. Theo Định lý 2.2.5, ta có

(i)

$$\boldsymbol{\beta}|\mathbf{y}, \mathbf{x}, \sigma^2 \sim MN(\boldsymbol{\mu}, \Sigma)$$

trong đó,

$$\boldsymbol{\mu} = (\mathbf{x}^{\mathbf{T}}\mathbf{x})^{-1}\mathbf{x}^{\mathbf{T}}\mathbf{y}$$
$$\Sigma = \sigma^2(\mathbf{x}^{\mathbf{T}}\mathbf{x})^{-1}$$

(ii)

$$\sigma^2|\mathbf{y}, \mathbf{x}, \boldsymbol{\beta} \sim IGamma\Big(\frac{n}{2}, \frac{1}{2}\sum_{i=1}^{n}(y_i - \mathbf{x}_{i\cdot}\boldsymbol{\beta})^2\Big)$$

3.3 Kết quả thu được

```
beta[,1]
  [1] -473.1089 -419.2100 -434.5161 -456.1344 -384.0797 -492.7782 -579.3081
  [8] -461.7709 -540.7859 -585.1846 -462.6970 -519.9764 -504.6584 -440.5052
 [15] -623.7878 -517.0539 -470.2654 -485.7148 -395.4530 -524.9818 -387.3075
 [22] -461.7140 -575.9148 -505.0966 -551.0670 -543.6311 -474.3705 -502.8164
 [29] -501.9367 -513.2390 -452.7103 -423.4167 -476.1085 -514.7655 -423.8208
 [36] -516.8104 -533.5285 -457.9558 -325.8256 -523.2219 -492.6488 -484.1347
 [43] -432.2540 -425.3832 -517.4614 -504.7911 -409.6126 -539.7967 -544.6885
 [50] -617.4809 -537.0367 -277.8780 -397.4961 -422.1085 -527.5614 -550.8304
 [57] -499.8784 -502.1376 -455.8218 -392.5222 -536.5860 -514.8592 -511.5226
 [64] -415.4781 -577.1499 -464.6988 -468.6120 -451.0480 -486.8512 -619.0638
 [71] -495.6588 -417.4421 -414.4094 -524.2239 -461.5857 -455.3429 -452.2878
 [78] -513.5209 -440.2598 -555.1709 -545.0123 -386.8021 -499.7455 -449.2460
 [85] -457.3596 -472.3365 -371.1624 -484.9676 -457.5355 -590.4634 -509.1653
 [92] -417.1842 -573.0156 -531.0092 -420.6209 -423.9549 -471.7335 -563.0778
 [99] -469.7868 -390.0071 -524.7148 -421.9426 -494.9324 -572.1816 -509.1476
[106] -522.1829 -476.6295 -530.3579 -529.6181 -438.0353 -506.4979 -408.4866
[113] -543.9036 -576.9635 -426.8483 -423.7249 -415.4520 -380.3939 -534.1588
[120] -489.3352 -571.5095 -425.6060 -558.0203 -524.3061 -451.1613 -518.9897
[127] -482.6105 -491.7184 -540.3017 -450.2377 -482.2772 -430.6314 -524.6786
[134] -458.3355 -497.1238 -440.3129 -433.4522 -475.8268 -475.5104 -430.9216
[141] -543.5962 -502.5860 -478.2667 -435.1798 -549.4879 -521.5564 -378.4251
[148] -423.3666 -583.8095 -474.4273 -578.9945 -426.6648 -442.5230 -400.5955
```

Hình 3.2: Xích Markov của β_0

```
beta[,2]
  [1] 0.02058651 0.02005938 0.02052174 0.02013556 0.02020095 0.02150797 0.02161270
  [8] 0.02061973 0.02105945 0.02152117 0.01972971 0.02158967 0.02002603 0.02081972
 [15] 0.02145440 0.02024682 0.02026037 0.02027560 0.02041581 0.02123147 0.01979381
 [22] 0.02017754 0.02090537 0.02017409 0.02160709 0.02124463 0.01901954 0.02030768
 [29] 0.02025318 0.02181509 0.01968820 0.02060168 0.02097271 0.02107158 0.01985452
 [36] 0.01977945 0.02166996 0.01944797 0.01914554 0.02059119 0.02095536 0.02045835
 [43] 0.02039454 0.02019261 0.01985341 0.02041429 0.01962089 0.02115566 0.02151585
 [50] 0.02209746 0.02001845 0.02026581 0.01949524 0.02103661 0.02072366 0.02213369
 [57] 0.02130008 0.02047960 0.02067016 0.02007574 0.02104345 0.02146809 0.01996247
 [64] 0.02055032 0.02051101 0.02016307 0.02068335 0.02020822 0.02023653 0.02157485
 [71] 0.02010462 0.01977783 0.01989627 0.02045012 0.02008206 0.02084174 0.02063630
 [78] 0.02172181 0.02003570 0.02219888 0.02117548 0.02001823 0.02058716 0.01917354
 [85] 0.01981930 0.02103427 0.02138993 0.02044191 0.02049245 0.02177347 0.02107668
 [92] 0.02024331 0.02129421 0.02124484 0.02038416 0.01903286 0.02056087 0.02132560
 [99] 0.02004554 0.01950088 0.01988254 0.02046850 0.02064874 0.02047584 0.02080556
[106] 0.02069413 0.02029137 0.02089416 0.02163543 0.02028254 0.02160348 0.02029571
[113] 0.02141378 0.02090703 0.02016803 0.02087483 0.02009431 0.01899691 0.02053084
[120] 0.02011804 0.02141272 0.02030718 0.02119370 0.02044709 0.02010954 0.02047819
[127] 0.02031778 0.02043484 0.02135242 0.02024380 0.02099289 0.02043935 0.02034483
[134] 0.02147355 0.02137247 0.01978445 0.02154388 0.01990945 0.02071110 0.01979139
```

Hình 3.3: Xích Markov của β_1

```
beta[,3]
  [1] -201.0164 -172.7775 -199.5629 -195.6809 -231.7340 -196.2444 -185.0664
  [8] -227.7718 -195.3089 -213.4939 -179.1748 -200.6490 -197.0613 -189.0935
 [15] -156.7993 -164.2784 -191.8452 -194.2013 -219.5434 -194.2883 -208.0307
 [22] -214.5623 -178.6177 -193.7819 -222.5988 -203.2332 -189.8617 -220.1126
 [29] -212.1041 -204.3021 -185.3559 -211.7602 -182.6134 -197.5090 -220.7667
 [36] -199.0693 -208.4001 -192.0472 -196.2691 -197.5906 -181.5310 -197.5954
 [43] -203.4347 -235.3955 -202.4715 -199.2720 -207.6721 -202.0135 -219.1926
 [50] -197.0087 -174.6544 -210.4375 -204.4357 -208.5682 -176.0350 -198.0644
 [57] -182.5819 -211.5991 -197.5113 -214.9220 -187.5528 -217.8014 -202.1823
 [64] -220.5060 -180.4413 -152.4530 -221.5902 -223.9637 -190.3314 -175.7851
 [71] -228.0442 -179.6218 -204.0914 -197.3768 -203.0174 -181.0114 -208.7152
 [78] -170.1220 -231.7642 -220.2715 -209.8370 -198.5293 -170.5028 -191.4481
 [85] -216.7720 -194.7465 -229.7213 -203.3745 -215.9720 -214.2488 -211.5680
 [92] -215.1994 -191.0788 -228.7989 -208.4109 -181.0433 -205.4118 -187.6167
 [99] -221.2121 -216.6189 -187.0175 -204.9029 -227.8494 -195.9093 -202.9543
[106] -198.2035 -196.8239 -175.0920 -198.6492 -214.9996 -195.5838 -208.3136
[113] -193.6986 -173.8390 -224.8160 -203.0756 -201.5700 -225.4119 -183.0482
[120] -171.3162 -192.4645 -218.9923 -188.6253 -170.1691 -197.9878 -184.2108
[127] -209.0909 -165.8513 -195.5327 -185.3979 -215.0044 -213.6080 -187.6628
[134] -204.6311 -222.2210 -221.8694 -199.3980 -198.3580 -198.1276 -225.4343
```

Hình 3.4: Xích Markov của β_2

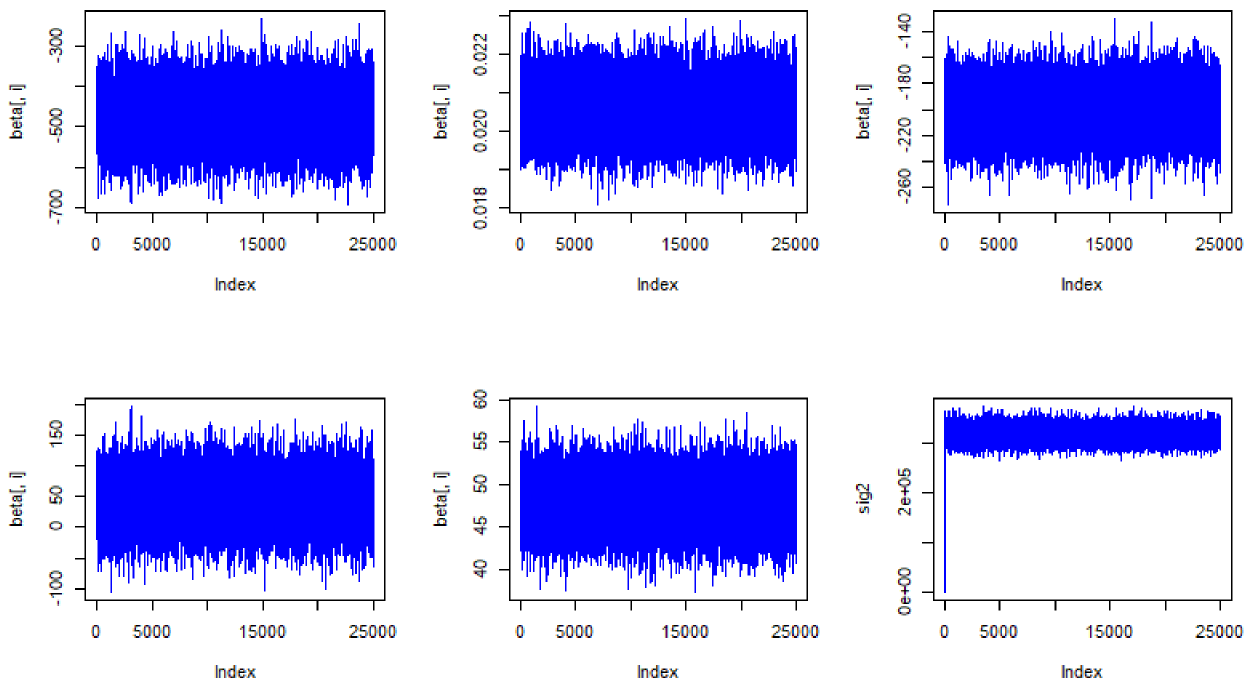
```
beta[,4]
  [1]  41.78279489  24.54008619  18.91104183  72.14204738  38.68058383  72.52700423
  [7]  10.18043962  75.69737833  66.93652440  55.90665765  33.43912724  54.63152464
 [13]  55.27022585  68.10223025 116.15789032 -11.42537000  43.87724023  24.14492726
 [19]  69.51119636  42.85354913  48.52071839  21.38730819  30.79683069  84.02608169
 [25]  28.22639147  40.60011524  37.51878986  31.54862041  84.08214333  46.41738415
 [31]  48.73422177   8.02189827 105.71289502  29.23908289  90.02500519  74.57416807
 [37]  48.83334162  95.66892259  -6.65240047 127.31679410  13.90236660 100.08732511
 [43]  42.58371387  46.92864981  65.38945793  55.93773154 -43.67395722  52.27019993
 [49]  34.10709441 101.20481587  42.52764817  17.10396742  31.26105133  15.72267611
 [55]  17.85315927  39.80032527  19.51058663  74.39304326  27.48404928  28.63519266
 [61]  70.09276422 127.10421319  73.75339157  35.62953170  30.47627968  36.32869768
 [67]  43.88870495  17.00369769  38.81933186 118.16451449  74.31972153 -38.28161319
 [73]   8.14245507  98.14597350  93.62879475 -17.76387861  94.02436179  48.70643562
 [79]  41.12944542   9.03178933  53.67194566  12.95266375  81.42945994   0.50760202
 [85]  86.69988833  64.76319243  -4.37208689  61.39396866  14.49362548 108.97341064
 [91]  67.80186033 -12.90742762  41.49953615  41.07507111  63.11640109  43.67051736
 [97]  16.60534641 122.94994568 132.50698290  29.47931698  53.60720538  62.44174074
[103]  87.27212612  84.72224730  60.21440891  36.09897097  42.40467396  31.87663171
```

Hình 3.5: Xích Markov của β_3

```
beta[,5]
  [1] 47.77724 45.90640 46.32287 46.65504 46.47060 43.44687 51.66642 46.79076
  [9] 48.56549 52.36643 49.45949 46.27051 51.15978 43.71158 47.96597 50.67940
 [17] 49.19799 50.04740 45.42642 48.10808 46.45601 49.39531 52.96148 48.03411
 [25] 50.53461 49.87750 52.03851 53.50842 48.67400 48.61339 49.26948 46.47864
 [33] 43.72985 48.69101 47.24556 51.68410 48.38106 47.57355 43.60634 46.72150
 [41] 47.89238 46.78614 45.84456 48.37230 52.83926 49.73761 48.95403 49.67812
 [49] 48.68610 50.49724 53.38284 40.73527 48.59750 44.29982 50.75577 46.07387
 [57] 45.67824 48.87537 46.32643 46.77068 49.57986 45.44896 53.00001 46.70693
 [65] 52.11955 46.46843 49.62494 47.68815 48.05076 49.55005 51.49710 49.72387
 [73] 46.63945 49.84377 48.50388 46.71435 46.87205 44.15689 47.82327 48.86061
 [81] 50.39834 45.69173 47.66286 53.59456 48.31707 45.83535 43.42000 49.58126
 [89] 46.31397 48.89877 48.48976 47.53146 50.86259 47.60529 46.56888 49.96134
 [97] 48.04375 48.50191 48.62529 44.96815 52.35494 45.69539 49.40736 51.86739
[105] 47.47007 51.74482 50.36430 48.93826 45.62560 48.79885 44.85629 45.78288
[113] 50.55470 53.89531 46.33490 44.01733 46.33278 49.03817 50.43318 47.80324
[121] 51.21459 46.98665 49.35945 51.57247 46.77766 49.81956 50.41319 45.60071
[129] 49.73361 48.32656 48.10692 44.41131 50.66221 44.94945 46.90838 49.31699
[137] 45.54965 48.51591 48.58544 48.80964 51.03239 51.32065 48.34703 49.26941
[145] 48.99153 48.95083 40.90314 46.56999 48.91420 49.30315 56.01646 48.86347
[153] 46.93101 48.75864 47.46255 44.33640 46.68999 47.80813 45.79785 50.30348
[161] 52.22010 43.59168 47.71352 49.81597 48.20085 47.12577 43.61743 52.15923
[169] 49.19005 51.62330 51.23485 47.16055 45.26652 48.61268 46.18120 45.88999
[177] 50.18797 53.75458 45.78005 49.80975 47.60632 48.85216 51.72255 45.40144
```

Hình 3.6: Xích Markov của β_4

Sau đây là một số hình ảnh về sự hội tụ của các chuỗi Markov mà thuật toán đã sinh ra.



Hình 3.7: Sự hội tụ của các xích Markov

Kết quả đạt được

	β_0	β_1	β_2	β_3	β_4
β_{MLE}	-472.98901851	0.02058483	-200.99042723	41.74898100	47.77752725
β_{MSE}	-473.64312804	0.02058663	-200.89638925	41.90915189	47.79649979

Suy ra ta có được phương trình sau đây

$$AmountSpent = -473.64312804 + 0.02058663 \times Salary$$

$$- 200.89638925 \times Children + 41.90915189 \times Gender\_b + 47.79649979 \times Catalogs$$

Ý nghĩa hệ số hồi quy

1. Khi các yếu tố khác không thay đổi, nếu tiền lương của khách hàng tăng thêm 1 đơn vị thì sô tiền mà khách hàng mua sản phẩm của DM trong

một năm tăng thêm 0.02058663 đơn vị.

2. Khi các yếu tố khác không thay đổi, nếu số con của khách hàng tăng thêm 1 thì sô tiền mà khách hàng mua sản phẩm của DM trong một năm giảm đi 200.89638925 đơn vị.

3. Khi các yếu tố khác không thay đổi, nếu Catalogs của khách hàng tăng thêm 1 đơn vị thì sô tiền mà khách hàng mua sản phẩm của DM trong một năm tăng thêm 47.79649979 đơn vị.

4. Khi các yếu tố khác không thay đổi, nếu khách hàng là nữ giới, thì sô tiền mà khách hàng mua sản phẩm của DM trong một năm lớn hơn số tiền mà khách hàng nam giới mua sản phẩm của DM trong một năm là 41.90915189 đơn vị.

Chương 4

Kết luận

Trường phái Cổ điển cho rằng tham số là cố định, chúng ta không biết giá trị cố định thực sự là gì không có nghĩa là chúng ta có thể coi một giá trị cố định như thể nó là ngẫu nhiên. Trường phái Bayes lại có quan điểm ngược lại, tức là giá trị tham số cố định chưa biết nhưng coi nó là một biến ngẫu nhiên. Các nhà thống kê Bayes đưa ra phân phối tiên nghiệm là "niềm tin" của bản thân, mang tính chủ quan và sau đó cập nhật phân phối này theo cách mà nó vẫn hoàn toàn phù hợp với dữ liệu quan sát được (sử dụng Quy tắc Bayes).

Vậy tại sao những người theo trường phái Bayes lại khăng khăng về điểm này, và tại sao họ không chấp nhận rằng "cố định" có nghĩa là "cố định" chứ không phải "ngẫu nhiên"? Đối với một số người theo trường phái Bayes, đó là vì họ giải thích xác suất theo "mức độ tin tưởng" chủ quan của họ - đặc biệt là mức độ tin tưởng ban đầu của họ về giá trị của một tham số cố định không xác định được biểu thị bằng phân phối tiên nghiệm.

Vậy tại sao các nhà thống kê Cổ điển lại khăng khăng rằng những lợi ích của phương pháp Bayes phải bị từ chối và không được chấp nhận một cách đơn giản? Không có câu trả lời duy nhất hoặc đơn giản cho câu hỏi này, nhưng một yêu cầu thiết yếu của phương pháp Bayes là cần phải xác định phân phối tiên nghiệm cho tham số chưa biết trước khi phân tích bất kỳ dữ liệu nào. Vì mục đích xác định phân phối tiên nghiệm này mà phán đoán chủ quan được áp dụng. Tuy nhiên, trường phái Cổ điển chỉ ra rằng chủ nghĩa chủ quan này không phù hợp với "phương pháp khoa học" vốn phải khách quan nhất có thể, và đặc biệt không được phụ thuộc vào người thực hiện thí nghiệm hay người phân tích kết quả.

Tài liệu tham khảo

Tiếng Việt

[1] Trần Minh Ngọc, Đỗ Văn Cường, tài liệu *Khóa bồi dưỡng giảng viên năm 2021 "Một số chủ đề hiện đại trong thống kê ứng dụng "*, Viện Nghiên cứu cao cấp về Toán (VIASM).

[2] Tống Đình Quỳ (2007), *Xác suất thống kê*, NXB Bách Khoa Hà Nội.

[3] Lê Xuân Lý, Slide bài giảng Phân tích số liệu.

[4] Nguyễn Văn Tuấn (2014), Phân tích số liệu với R, NXB tổng hợp TP HCM.

Tiếng Anh

[5] William M. Bolstad. Introduction to Bayesian statistics, 2nd Edition. Wiley, 2007.

[6] https://statswithr.github.io/book/introduction-to-bayesian-regression.html

Phụ lục

Mã nguồn mô phỏng ước lượng Bayes và ước lượng hợp lý cực đại cho tham số μ của phân phối chuẩn khi σ^2 đã biết.

```
# Dau vao
mu<-5
sigma<-2
# Mo phong
n<-10
x<-rnorm(n, mean = mu, sd=sigma)
# MLE
s<-sum(x)
x.bar<-s/n
mu.MLE<-x.bar
mu.MLE
############################
# 1. Tien nghiem lien hop
# Prior information
#Scenario 1.1.1: underguest, low confidence (du doan duoi, tu tin th???p)
g1<-4
rho<-0.9
g2<-g1*rho
# Prior elicitation trich xuat gia tri sieu tham so
a0<-g1
b0<-g2
# Prior distribution
```

```
curve(dnorm(x, mean = a0, sd=b0), col="red", xlim = c(-1, 11), ylim=c(0,
   2.5))
# Posterior distribution
k<-b0^2/(b0^2+sigma^2/n)
a1<-k*x.bar+(1-k)*a0
b1<-sigma*sqrt(k/n)
curve(dnorm(x, mean = a1, sd=b1), xlim = c(-1, 11), col="blue", add = T)
# Estimate
mu.MSE<-a1
c(mu.MLE, mu.MSE)

#Scenario 1.1.2: underguest, medium confidence (du doan duoi, tu tin trung
   binh)
g1<-4
rho<-0.6
g2<-g1*rho
# Prior elicitation trich xuat gia tri sieu tham so
a0<-g1
b0<-g2
# Prior distribution
curve(dnorm(x, mean = a0, sd=b0), col="red", xlim = c(-1, 11), ylim=c(0,
   2.5))
# Posterior distribution
k<-b0^2/(b0^2+sigma^2/n)
a1<-k*x.bar+(1-k)*a0
b1<-sigma*sqrt(k/n)
curve(dnorm(x, mean = a1, sd=b1), xlim = c(-1, 11), col="blue", add = T)
# Estimate
mu.MSE<-a1
c(mu.MLE, mu.MSE)

#Scenario 1.1.3: underguest, high confidence (du doan duoi, tu tin cao)
```

```
g1<-4
rho<-0.3
g2<-g1*rho
# Prior elicitation trich xuat gia tri sieu tham so
a0<-g1
b0<-g2
# Prior distribution
curve(dnorm(x, mean = a0, sd=b0), col="red", xlim = c(-1, 11), ylim=c(0,
   2.5))
# Posterior distribution
k<-b0^2/(b0^2+sigma^2/n)
a1<-k*x.bar+(1-k)*a0
b1<-sigma*sqrt(k/n)
curve(dnorm(x, mean = a1, sd=b1), xlim = c(-1, 11), col="blue", add = T)
# Estimate
mu.MSE<-a1
c(mu.MLE, mu.MSE)

#Scenario 1.2.1: precise guest, low confidence (du doan chinh xac, tu tin
   th???p)
g1<-5
rho<-0.9
g2<-g1*rho
# Prior elicitation trich xuat gia tri sieu tham so
a0<-g1
b0<-g2
# Prior distribution
curve(dnorm(x, mean = a0, sd=b0), col="red", xlim = c(-1, 11), ylim=c(0,
   2.5))
# Posterior distribution
k<-b0^2/(b0^2+sigma^2/n)
a1<-k*x.bar+(1-k)*a0
```

```
b1<-sigma*sqrt(k/n)
curve(dnorm(x, mean = a1, sd=b1), xlim = c(-1, 11), col="blue", add = T)
# Estimate
mu.MSE<-a1
c(mu.MLE, mu.MSE)
#Scenario 1.2.2: precise guest, medium confidence (du doan duoi, tu tin
   trung bnh)
g1<-5
rho<-0.6
g2<-g1*rho
# Prior elicitation trich xuat gia tri sieu tham so
a0<-g1
b0<-g2
# Prior distribution
curve(dnorm(x, mean = a0, sd=b0), col="red", xlim = c(-1, 11), ylim=c(0,
   2.5))
# Posterior distribution
k<-b0^2/(b0^2+sigma^2/n)
a1<-k*x.bar+(1-k)*a0
b1<-sigma*sqrt(k/n)
curve(dnorm(x, mean = a1, sd=b1), xlim = c(-1, 11), col="blue", add = T)
# Estimate
mu.MSE<-a1
c(mu.MLE, mu.MSE)
#Scenario 1.2.3: precise guest, high confidence (du doan duoi, tu tin cao)

g1<-5
rho<-0.3
g2<-g1*rho
# Prior elicitation trich xuat gia tri sieu tham so
a0<-g1
b0<-g2
```

```
# Prior distribution
curve(dnorm(x, mean = a0, sd=b0), col="red", xlim = c(-1, 11), ylim=c(0,
   2.5))
# Posterior distribution
k<-b0^2/(b0^2+sigma^2/n)
a1<-k*x.bar+(1-k)*a0
b1<-sigma*sqrt(k/n)
curve(dnorm(x, mean = a1, sd=b1), xlim = c(-1, 11), col="blue", add = T)
# Estimate
mu.MSE<-a1
c(mu.MLE, mu.MSE)
#Scenario 1.3.1: overguest, low confidence (du doan trn, tu tin th???p)

g1<-6
rho<-0.9
g2<-g1*rho
# Prior elicitation trich xuat gia tri sieu tham so
a0<-g1
b0<-g2
# Prior distribution
curve(dnorm(x, mean = a0, sd=b0), col="red", xlim = c(-1, 11), ylim=c(0,
   2.5))
# Posterior distribution
k<-b0^2/(b0^2+sigma^2/n)
a1<-k*x.bar+(1-k)*a0
b1<-sigma*sqrt(k/n)
curve(dnorm(x, mean = a1, sd=b1), xlim = c(-1, 11), col="blue", add = T)
# Estimate
mu.MSE<-a1
c(mu.MLE, mu.MSE)

#Scenario 1.3.2: overguest, medium confidence (du doan trn, tu tin trung
```

```
   bnh)

g1<-6
rho<-0.6
g2<-g1*rho
# Prior elicitation trich xuat gia tri sieu tham so
a0<-g1
b0<-g2
# Prior distribution
curve(dnorm(x, mean = a0, sd=b0), col="red", xlim = c(-1, 11), ylim=c(0,
   2.5))
# Posterior distribution
k<-b0^2/(b0^2+sigma^2/n)
a1<-k*x.bar+(1-k)*a0
b1<-sigma*sqrt(k/n)
curve(dnorm(x, mean = a1, sd=b1), xlim = c(-1, 11), col="blue", add = T)
# Estimate
mu.MSE<-a1
c(mu.MLE, mu.MSE)
#Scenario 1.3.3: overguest, high confidence (du doan trn, tu tin cao)

g1<-6
rho<-0.3
g2<-g1*rho
# Prior elicitation trich xuat gia tri sieu tham so
a0<-g1
b0<-g2
# Prior distribution
curve(dnorm(x, mean = a0, sd=b0), col="red", xlim = c(-1, 11), ylim=c(0,
   2.5))
# Posterior distribution
k<-b0^2/(b0^2+sigma^2/n)
```

```
a1<-k*x.bar+(1-k)*a0
b1<-sigma*sqrt(k/n)
curve(dnorm(x, mean = a1, sd=b1), xlim = c(-1, 11), col="blue", add = T)
# Estimate
mu.MSE<-a1
c(mu.MLE, mu.MSE)
############################
# 2. Tien nghiem deu
# Uniform prior [4, 6]
a<-4
b<-6
s<-sum(x)
x.bar<-s/n
ub<-pnorm(q=(b-x.bar)*sqrt(n)/sigma)
lb<-pnorm(q=(a-x.bar)*sqrt(n)/sigma)
K<-1/(ub-lb)
K
f<-function(x){exp(log(K)-n*(x-x.bar)^2/(2*sigma^2))}
curve(f(x), xlim = c(4, 8))
```

Mã nguồn mô phỏng ước lượng Bayes và ước lượng hợp lý cực đại của phân phối chuẩn khi μ, σ^2 chưa biết.

```
# Dau vao
mu<-5
tau <- 0.25
# Mo phong
########################################################
#Sample size n=10
n<-10
x<-rnorm(n, mean = mu, sd=1/sqrt(tau))
# MLE
x.bar<-sum(x)/n
```

```
x1 <- matrix(x,nrow = 1)
x1<- (x1 - x.bar)^2
S2 <- 1/n * sum(x1)
############################
# 1. Tien nghiem lien hop
# Prior information
#Scenario 1.1.1: underguest, low confidence (du doan duoi, tu tin th???p)
g1<-0.3
rho<-0.9
g2<-g1*rho
# Prior elicitation trich xuat gia tri sieu tham so cho tau
a0<-(g1/g2)^2
c0<-g1/g2^2
# Prior elicitation trich xuat gia tri sieu tham so cho Mu
g3<-2
g4<-g3*rho
m0<-g3
b0<- c0/(g4^2*(a0-1))
k<- b0/(b0+n)
a1<- a0+n/2
c1 <- c0+n/2*(S2+k*(x.bar-m0)^2)
m1<-(1-k)*x.bar + k*m0
b1 <- b0 +n
# Estimate
mu_MSE <- m1
mu_MLE <-x.bar
tau_MSE <- a1/c1
tau_MLE <- 1/S2
c(mu_MSE,mu_MLE)
c(tau_MSE,tau_MLE)

#Scenario 1.1.1: underguest, low confidence (du doan duoi, tu tin trung
```

```
   binh)
g1<-0.3
rho<-0.9
g2<-g1*rho
# Prior elicitation trich xuat gia tri sieu tham so cho tau
a0<-(g1/g2)^2
c0<-g1/g2^2
# Prior elicitation trich xuat gia tri sieu tham so cho Mu
g3<-4.9999
g4<-g3*rho
m0<-g3
b0<- c0/(g4^2*(a0-1))
k<- b0/(b0+n)
a1<- a0+n/2
c1 <- c0+n/2*(S2+k*(x.bar-m0)^2)
m1<-(1-k)*x.bar + k*m0
b1 <- b0 +n
# Estimate
mu_MSE <- m1
mu_MLE <-x.bar
tau_MSE <- a1/c1
tau_MLE <- 1/S2
c(mu_MSE,mu_MLE)
c(tau_MSE,tau_MLE)


#Scenario 1.1.1: underguest, low confidence (du doan duoi, tu tin trung
   binh)
g1<-0.3
rho<-0.9
g2<-g1*rho
# Prior elicitation trich xuat gia tri sieu tham so cho tau
```

```
a0<-(g1/g2)^2
c0<-g1/g2^2
# Prior elicitation trich xuat gia tri sieu tham so cho Mu
g3<-6
g4<-g3*rho
m0<-g3
b0<- c0/(g4^2*(a0-1))
k<- b0/(b0+n)
a1<- a0+n/2
c1 <- c0+n/2*(S2+k*(x.bar-m0)^2)
m1<-(1-k)*x.bar + k*m0
b1 <- b0 +n
# Estimate
mu_MSE <- m1
mu_MLE <-x.bar
tau_MSE <- a1/c1
tau_MLE <- 1/S2
c(mu_MSE,mu_MLE)
c(tau_MSE,tau_MLE)

#Scenario 1.1.1: underguest, medium confidence (du doan duoi, tu tin trung
   binh)
g1<-0.3
rho<-0.6
g2<-g1*rho
# Prior elicitation trich xuat gia tri sieu tham so cho tau
a0<-(g1/g2)^2
c0<-g1/g2^2
# Prior elicitation trich xuat gia tri sieu tham so cho Mu
g3<-2
g4<-g3*rho
m0<-g3
```

```
b0<- c0/(g4^2*(a0-1))
k<- b0/(b0+n)
a1<- a0+n/2
c1 <- c0+n/2*(S2+k*(x.bar-m0)^2)
m1<-(1-k)*x.bar + k*m0
b1 <- b0 +n
# Estimate
mu_MSE <- m1
mu_MLE <-x.bar
tau_MSE <- a1/c1
tau_MLE <- 1/S2
c(mu_MSE,mu_MLE)
c(tau_MSE,tau_MLE)

#Scenario 1.1.1: underguest, medium confidence (du doan duoi, tu tin trung
   binh)
g1<-0.3
rho<-0.6
g2<-g1*rho
# Prior elicitation trich xuat gia tri sieu tham so cho tau
a0<-(g1/g2)^2
c0<-g1/g2^2
# Prior elicitation trich xuat gia tri sieu tham so cho Mu
g3<-4.999
g4<-g3*rho
m0<-g3
b0<- c0/(g4^2*(a0-1))
k<- b0/(b0+n)
a1<- a0+n/2
c1 <- c0+n/2*(S2+k*(x.bar-m0)^2)
m1<-(1-k)*x.bar + k*m0
b1 <- b0 +n
```

```
# Estimate
mu_MSE <- m1
mu_MLE <-x.bar
tau_MSE <- a1/c1
tau_MLE <- 1/S2
c(mu_MSE,mu_MLE)
c(tau_MSE,tau_MLE)

#Scenario 1.1.1: underguest, medium confidence (du doan duoi, tu tin trung
   binh)
g1<-0.3
rho<-0.6
g2<-g1*rho
# Prior elicitation trich xuat gia tri sieu tham so cho tau
a0<-(g1/g2)^2
c0<-g1/g2^2
# Prior elicitation trich xuat gia tri sieu tham so cho Mu
g3<-6
g4<-g3*rho
m0<-g3
b0<- c0/(g4^2*(a0-1))
k<- b0/(b0+n)
a1<- a0+n/2
c1 <- c0+n/2*(S2+k*(x.bar-m0)^2)
m1<-(1-k)*x.bar + k*m0
b1 <- b0 +n
# Estimate
mu_MSE <- m1
mu_MLE <-x.bar
tau_MSE <- a1/c1
tau_MLE <- 1/S2
c(mu_MSE,mu_MLE)
```

```
c(tau_MSE,tau_MLE)

#Scenario 1.1.1: underguest, medium confidence (du doan duoi, tu tin trung
   binh)
g1<-0.3
rho<-0.3
g2<-g1*rho
# Prior elicitation trich xuat gia tri sieu tham so cho tau
a0<-(g1/g2)^2
c0<-g1/g2^2
# Prior elicitation trich xuat gia tri sieu tham so cho Mu
g3<-6
g4<-g3*rho
m0<-g3
b0<- c0/(g4^2*(a0-1))
k<- b0/(b0+n)
a1<- a0+n/2
c1 <- c0+n/2*(S2+k*(x.bar-m0)^2)
m1<-(1-k)*x.bar + k*m0
b1 <- b0 +n
# Estimate
mu_MSE <- m1
mu_MLE <-x.bar
tau_MSE <- a1/c1
tau_MLE <- 1/S2
c(mu_MSE,mu_MLE)
c(tau_MSE,tau_MLE)
############################
#Scenario 1.1.2: underguest, low confidence (du doan duoi, tu tin thap)
g1<-0.1
rho<-0.9
g2<-g1*rho
```

```
# Prior elicitation trich xuat gia tri sieu tham so cho tau
a0<-(g1/g2)^2
c0<-g1/g2^2
# Prior elicitation trich xuat gia tri sieu tham so cho Mu
g3<-2
g4<-g3*rho
m0<-g3
b0<- c0/(g4^2*(a0-1))
k<- b0/(b0+n)
a1<- a0+n/2
c1 <- c0+n/2*(S2+k*(x.bar-m0)^2)
m1<-(1-k)*x.bar + k*m0
b1 <- b0 +n
# Estimate
mu_MSE <- m1
mu_MLE <-x.bar
tau_MSE <- a1/c1
c(mu_MSE,mu_MLE)
#Scenario 1.1.3: underguest, high confidence (du doan duoi, tu tin cao)
g1<-0.1
rho<-0.3
g2<-g1*rho
# Prior elicitation trich xuat gia tri sieu tham so cho tau
a0<-(g1/g2)^2
c0<-g1/g2^2
# Prior elicitation trich xuat gia tri sieu tham so cho Mu
g3<-2
g4<-g3*rho
m0<-g3
b0<- c0/(g4^2*(a0-1))
k<- b0/(b0+n)
a1<- a0+n/2
```

```
c1 <- c0+n/2*(S2+k*(x.bar-m0)^2)
m1<-(1-k)*x.bar + k*m0
b1 <- b0 +n
# Estimate
mu_MSE <- m1
mu_MLE <-x.bar
tau_MSE <- a1/c1
c(mu_MSE,mu_MLE)
#Scenario 1.2.1: precise guest, low confidence (du doan chinh xac, tu tin
   thap)
g1<-0.25
rho<-0.9
g2<-g1*rho
# Prior elicitation trich xuat gia tri sieu tham so cho tau
a0<-(g1/g2)^2
c0<-g1/g2^2
# Prior elicitation trich xuat gia tri sieu tham so cho Mu
g3<-2
g4<-g3*rho
m0<-g3
b0<- c0/(g4^2*(a0-1))
k<- b0/(b0+n)
a1<- a0+n/2
c1 <- c0+n/2*(S2+k*(x.bar-m0)^2)
m1<-(1-k)*x.bar + k*m0
b1 <- b0 +n
# Estimate
mu_MSE <- m1
mu_MLE <-x.bar
tau_MSE <- a1/c1
tau_MLE <- 1/S2
c(mu_MSE,mu_MLE)
```

```
c(tau_MSE,tau_MLE)
#Scenario 1.2.2: precise guest, medium confidence (du doan chinh xac, tu
   tin trung binh)
g1<-0.25
rho<-0.6
g2<-g1*rho
# Prior elicitation trich xuat gia tri sieu tham so cho tau
a0<-(g1/g2)^2
c0<-g1/g2^2
# Prior elicitation trich xuat gia tri sieu tham so cho Mu
g3<-2
g4<-g3*rho
m0<-g3
b0<- c0/(g4^2*(a0-1))
k<- b0/(b0+n)
a1<- a0+n/2
c1 <- c0+n/2*(S2+k*(x.bar-m0)^2)
m1<-(1-k)*x.bar + k*m0
b1 <- b0 +n
# Estimate
mu_MSE <- m1
mu_MLE <-x.bar
tau_MSE <- a1/c1
tau_MLE <- 1/S2
c(mu_MSE,mu_MLE)
c(tau_MSE,tau_MLE)
#Scenario 1.2.3: precise guest, high confidence (du doan chinh xac, tu tin
   cao)
g1<-0.25
rho<-0.3
g2<-g1*rho
# Prior elicitation trich xuat gia tri sieu tham so cho tau
```

```
a0<-(g1/g2)^2
c0<-g1/g2^2
# Prior elicitation trich xuat gia tri sieu tham so cho Mu
g3<-2
g4<-g3*rho
m0<-g3
b0<- c0/(g4^2*(a0-1))
k<- b0/(b0+n)
a1<- a0+n/2
c1 <- c0+n/2*(S2+k*(x.bar-m0)^2)
m1<-(1-k)*x.bar + k*m0
b1 <- b0 +n
# Estimate
mu_MSE <- m1
mu_MLE <-x.bar
tau_MSE <- a1/c1
tau_MLE <- 1/S2
c(mu_MSE,mu_MLE)
c(tau_MSE,tau_MLE)
```

Mô phỏng ước lượng Bayes và ước lượng hợp lý cực đại cho mô hình hồi quy tuyến tính đa biến.

```
# Simulation study for multiple linear regression (MLR)
# Input
beta<- c(1, 2, 3, 4, 5)# he so hoi quy thuc te
sigma<-0.5#do lech chuan cua sai so
# Generate (X, Y)
n<- 1000# co mau
k<-4# so bien doc lap
D<- diag(k)# ma tran don vi cap k
x<-matrix(rep(0, n*k), ncol = k) #ma tran n dong k cot
y<- rep(0, n)#khoi tao vector y
```

```
for (i in 1:n) {
  x[i,]<- mvrnorm(1, mu = c(1, 2, 3, 4), Sigma = D)# sinh bnn phan phoi
      chuan nhieu chieu
  mu<-c(1, x[i,])%*%beta# trung binh cua y|x
  y[i]<- rnorm(1, mean = mu, sd = sigma)
}
head(x)
head(y)
d<-data.frame(y, x)
pairs(d)
cor(d)
# lm() function
reg<- lm(d$y~d$X1+d$X2+d$X3+d$X4)
summary(reg)
###########################
# MLE
x0<-rep(1, n)
x<-cbind(x0, x)
k<-ncol(x)
A<-solve(t(x)%*%x)
beta.MLE<- A%*%t(x)%*%y
sigma2.MLE<-t(y-x%*%beta.MLE)%*%(y-x%*%beta.MLE)/(n-1)
sigma.MLE<- sqrt(sigma2.MLE)
beta.MLE
sigma.MLE
##############################
# MSE with flat prior
N<- 5000
sigma2.mc1<- rep(0, N)
beta.mc1<- matrix(rep(0, N*k), ncol = k)
sigma2.mc1[1]<- 1
beta.mc1[1, ]<- rep(1, k)
```

```
for (i in 1:(N-1)) {
sigma2.mc1[i+1]<-1/rgamma(1, shape = n/2, rate =
   t(y-x%*%beta.mc1[i,])%*%(y-x%*%beta.mc1[i,])/2)
require(MASS)
beta.mc1[i+1,]<-mvrnorm(mu=beta.MLE, Sigma =
   sigma2.mc1[i+1]*solve(t(x)%*%x))
}
head(sigma2.mc1)
tail(sigma2.mc1)
head(beta.mc1)
tail(beta.mc1)
par(mfrow=c(2, 3))
plot(sigma2.mc1, type = "s", col = "blue")
plot(beta.mc1[, 1], type = "s", col = "blue")
plot(beta.mc1[, 2], type = "s", col = "blue")
plot(beta.mc1[, 3], type = "s", col = "blue")
plot(beta.mc1[, 4], type = "s", col = "blue")
plot(beta.mc1[, 5], type = "s", col = "blue")
posterior_mean_beta1 <- mean(beta.mc1[, 1])
posterior_mean_beta2 <- mean(beta.mc1[, 2])
posterior_mean_beta3 <- mean(beta.mc1[, 3])
posterior_mean_beta4 <- mean(beta.mc1[, 4])
posterior_mean_beta5 <- mean(beta.mc1[, 5])
beta.mc.FLAT<-c(posterior_mean_beta1, posterior_mean_beta2,
           posterior_mean_beta3, posterior_mean_beta4, posterior_mean_beta5)
############################################
# Conjugate prior
N<- 5000
sigma2.mc2<- rep(0, N)
beta.mc2<- matrix(rep(0, N*k), ncol = k)
sigma2.mc2[1]<- 1
beta.mc2[1, ]<- rep(1, k)
```

```
a0<-2
b0<-1
MU0<- c(0.5, 1.5, 2.5, 3.5, 4.5)
SIGMA0<-diag(k)
for (i in 1:(N-1)) {
  a1<-a0+n/2
  b1<-b0+t(y-x%*%beta.mc2[i,])%*%(y-x%*%beta.mc2[i,])/2
  sigma2.mc2[i+1]<-1/rgamma(1, shape = a1, rate = b1)
  SIGMA.y<-sigma2.mc2[i+1]*A
  MU.y<-A%*%t(x)%*%y
  SIGMA1<-solve(solve(SIGMA0)+solve(SIGMA.y))
  MU1<-SIGMA1%*%solve(SIGMA0)%*%MU0+SIGMA1%*%solve(SIGMA.y)%*%MU.y
  require(MASS)
  beta.mc2[i+1,]<-mvrnorm(mu=MU1, Sigma = SIGMA1)
}
head(sigma2.mc2)
tail(sigma2.mc2)
head(beta.mc2)
tail(beta.mc2)
par(mfrow=c(2, 3))
plot(sigma2.mc2, type = "s", col = "blue")
plot(beta.mc2[, 1], type = "s", col = "blue")
plot(beta.mc2[, 2], type = "s", col = "blue")
plot(beta.mc2[, 3], type = "s", col = "blue")
plot(beta.mc2[, 4], type = "s", col = "blue")
plot(beta.mc2[, 5], type = "s", col = "blue")

posterior_mean_beta1 <- mean(beta.mc2[, 1])
posterior_mean_beta2 <- mean(beta.mc2[, 2])
posterior_mean_beta3 <- mean(beta.mc2[, 3])
posterior_mean_beta4 <- mean(beta.mc2[, 4])
posterior_mean_beta5 <- mean(beta.mc2[, 5])
```

```
beta.MSE<-c(posterior_mean_beta1, posterior_mean_beta2,
              posterior_mean_beta3, posterior_mean_beta4,
                 posterior_mean_beta5)
```

Mã nguồn ứng dụng trên bộ dữ liệu thực tế

```
#################################### Q3 ##################################
setwd("E:/DoAnTotNghiep_Quyen")
data <- read.csv("DirectMarketing.csv", header = TRUE)
y1 <- data["AmountSpent"]
data[, c('AmountSpent', 'Salary', 'Children','Gender_b','Catalogs')]
y = matrix(unlist(y1), ncol = 1, byrow = TRUE)
n <- lengths(y1)
a <- matrix(rep(1,n),ncol = 1)
b <- matrix(unlist(data["Salary"]), ncol = 1, byrow = TRUE)
c <- matrix(unlist(data["Children"]), ncol = 1, byrow = TRUE)
d <- matrix(unlist(data["Gender_b"]), ncol = 1, byrow = TRUE)
e <- matrix(unlist(data["Catalogs"]), ncol = 1, byrow = TRUE)
X <- cbind(a,b,c,d,e)
X
dim(X)
dim <- dim(X)[2] # s??? h??? s??? = 5
A <- t(X)%*%X
Ainv<-solve(A)
Ainv
B <- t(X)%*%y
AmB <- Ainv%*%B
AmB
beta.MLE<-AmB[,1]
beta.MLE
sigma2.MLE<-t(y-X%*%beta.MLE)%*%(y-X%*%beta.MLE)/(n-dim)
sigma2.MLE
```

```r
reg<-lm(y~b+c+d+e)
summary(reg)
N <- 25000
beta <- matrix(rep(0,N*dim),ncol = dim)
sig2 <- matrix(rep(0,N),ncol = 1)
sig2[1] <- 1
library(MASS)
beta[1,] <- mvrnorm(mu=AmB,Sigma = sig2[1]*Ainv)
for(i in 1:(N-1)){
  gamma_shape <- n/2
  a <- matrix(beta[i,],ncol=dim,byrow = TRUE)
  xb <- X%*%t(a)
  a2 <- y-xb
  gamma_rate <- t(a2)%*%(a2/2)
  auxilary <- rgamma(n = 1, shape = gamma_shape, rate = gamma_rate[1])
  sig2[i+1] <- 1/auxilary
  beta[i+1,] <- mvrnorm(mu=AmB, Sigma=sig2[i+1]*Ainv)
}
par(mfrow=c(2,3))
for(i in 1:dim){
  plot(beta[,i],type = "s", col = "blue")
}
plot(sig2,type = "s", col = "blue")

nburn <- 5000
beta <- beta[(nburn+1):N,]
sig2 <- sig2[(nburn+1):N,]
posterior_mean_beta1 <- mean(beta[, 1])
posterior_mean_beta1
posterior_mean_beta2 <- mean(beta[, 2])
posterior_mean_beta2
posterior_mean_beta3 <- mean(beta[, 3])
```

```
posterior_mean_beta3
posterior_mean_beta4 <- mean(beta[, 4])
posterior_mean_beta4
posterior_mean_beta5 <- mean(beta[, 5])
posterior_mean_beta5
beta.MSE<-c(posterior_mean_beta1, posterior_mean_beta2,
           posterior_mean_beta3, posterior_mean_beta4, posterior_mean_beta5)
beta.MSE
beta.MLE
posterior_mean_sigma2 <- mean(sig2)
sigma2.MSE<-posterior_mean_sigma2
sigma2.MSE
posterior_std <- sqrt(mean(beta^2)-posterior_mean_beta1^2)
posterior_std
```

Mã nguồn ví dụ về thuật toán RWMH

```
# Bayesian computation
# Question 1
# part(i)
N.burnin <- 1000# Number of burnin
N.Iter <- N.burnin+5000# Number of iterations
n <- 50 #sample size
x.bar <- 5 #sample mean
# kernel of posterior distribution
k <- function(mu){1/(1+mu^2)*exp(n*x.bar*mu-n/2*mu^2)}
sigma = 10^(-1)# variance of epsilon
markov_chain<-rep(0, N.Iter) #create a vector to store the chain
mu0<-rnorm(1, mean = 0, sd = 1) # starting value
markov_chain[1] <-mu0 # first state of markov chain
i<-1
while(i< N.Iter){
  epsilon <- rnorm(1, mean = 0, sd = sigma)
```

```
  Y <- markov_chain[i] + epsilon
  alpha <- min(k(Y)/k(markov_chain[i]),1)
  u <- runif(1)
  if(u < alpha){markov_chain[i+1] <-Y}
  else{markov_chain[i+1] <- markov_chain[i]}
  i<-i+1
}
plot(markov_chain,type = "s", col = "blue")
markov_chain
markov_chain1 = markov_chain[(N.burnin+1):N.Iter]
hist(markov_chain1, probability = T, xlim = c(4, 6), breaks = 1000)
lines(density(markov_chain), col="red", lwd=3)
posterior_mean_estimate <-mean(markov_chain1)
posterior_mean_estimate
posterior_variance_estimate <- mean((markov_chain -
   posterior_mean_estimate)^2)
posterior_variance_estimate
########################################
# Simulation study for normal distribution
# True parameter
mu<-5
# sample size
n <- 100
# Observations
x<-rnorm(n, mean=mu, sd=1)
s<-sum(x)
x.bar <- s/n
###############
N.burnin <- 1000# Number of iterations
N.Iter <- N.burnin+5000# Number of iterations
# kernel of posterior distribution
k <- function(mu){1/(1+mu^2)*exp(n*x.bar*mu-n/2*mu^2)}
```

```
sigma = 10^(-1)# variance of epsilon
markov_chain<-rep(0, N.Iter) #create a vector to store the chain
mu0<-rnorm(1, mean = 0, sd = 1) # starting value
markov_chain[1] <-mu0 # first state of markov chain
i<-1
while(i< N.Iter){
  epsilon <- rnorm(1, mean = 0, sd = sigma)
  Y <- markov_chain[i] + epsilon
  alpha <- min(k(Y)/k(markov_chain[i]),1)
  u <- runif(1)
  if(u < alpha){markov_chain[i+1] <-Y}
  else{markov_chain[i+1] <- markov_chain[i]}
  i<-i+1
}
plot(markov_chain,type = "s", col = "blue")
markov_chain
markov_chain1 = markov_chain[(N.burnin+1):N.Iter]
hist(markov_chain1, probability = T, xlim = c(4, 6), breaks = 1000)
lines(density(markov_chain), col="red", lwd=3)
posterior_mean_estimate <-mean(markov_chain1)
posterior_mean_estimate
posterior_variance_estimate <- mean((markov_chain -
   posterior_mean_estimate)^2)
posterior_variance_estimate
#############################
# Part (ii)
# Function with input being sigma
mu.MCMC<-function(sigma1)
{
  # number of iterations
  N <- 20000
  # kernel of posterior distribution
```

```
  k <- function(mu){1/(1+mu^2)*exp(n*x.bar*mu-n/2*mu^2)}
  # create a vector to store the chain
  mc<-rep(0, N)
  # gia tri ngau nhien ban dau theo N(0,1)
  mu0<-rnorm(1, mean = 0, sd = 1)
  mc[1] <-mu0
  i<-1
  while(i< N){
    e <- rnorm(1, mean = 0, sd = sigma1)
    Y <- mc[i] + e
    f<-(1+Y^2)/(1+mc[i]^2)*exp(n*x.bar*(Y-mc[i])-n/2*(Y^2-mc[i]^2))
    alpha <- min(f,1)
    u <- runif(1)
    if(u < alpha){mc[i+1] <-Y}
    else{mc[i+1] <- mc[i]}
    i<-i+1
  }
  mu.MCMC<-mean(mc)#posterior_mean_estimate
  mu.MCMC
}
####################
# sigma=0.1
mu.hat<-rep(0, 10)
for (i in 1:10) {
  mu.hat[i]<-mu.MCMC(0.1)
}
MC.se1<-sd(mu.hat)
MC.se1
##############################
# part (iii)
# sigma=0.0001: markov chain is divergent
mu.hat<-rep(0, 10)
```

```
for (i in 1:10) {
  mu.hat[i]<-mu.MCMC(0.0001)
}
MC.se2<-sd(mu.hat)
MC.se2
# part (iv)
# sigma=10000: markov chain is divergent
mu.hat<-rep(0, 10)
for (i in 1:10) {
  mu.hat[i]<-mu.MCMC(10000)
}
MC.se3<-sd(mu.hat)
MC.se3
# Conclusion:
c(MC.se1, MC.se2, MC.se3)
```