

BÁO CÁO HỘI THẢO
MỘT SỐ VẤN ĐỀ TRONG XÁC SUẤT VÀ
THỐNG KÊ ỨNG DỤNG

BÀI TOÁN MÔ PHỎNG SINH MẪU BẰNG
PHẦN MỀM R

ThS. Nguyễn Thị Hằng

HÀ NỘI, NGÀY 15/04/2021

LỜI MỞ ĐẦU

- R là một ngôn ngữ lập trình và môi trường phần mềm dành cho tính toán và đồ họa thống kê.
- Nền tảng đồ họa của R, có thể tạo ra những đồ thị chất lượng cao cùng các biểu tượng toán học. R cũng có định dạng văn bản riêng tương tự như LaTeX, dùng để cung cấp tài liệu hướng dẫn toàn diện, có trực tuyến ở các định dạng khác nhau
- Trong khuôn khổ đề tài nhỏ này, xin được trình bày bài toán sinh mẫu trên phần mềm thống kê R với các phân phối quen thuộc trên không gian 1 chiều và 2 chiều.

CHƯƠNG 1. CƠ SỞ LÝ THUYẾT

1. Giới thiệu về R

2. Ước lượng Bayes

3. Sinh mẫu bằng phần mềm R

CHƯƠNG 2. BÀI TOÁN MÔ PHỎNG SINH MẪU BẰNG PHẦN MỀM R

1. Sinh mẫu ngẫu nhiên mô phỏng phân bố nhị thức

Sinh mẫu ngẫu nhiên mô phỏng phân bố nhị thức với $n = 100$ và $p = 0.5$: $\pi(k) = C_n^k 0.5^k 0.5^{100-k}$, sử dụng ma trận đối xứng $A_{100 \times 100}$ có các

phần tử như sau:
$$a_{ij} = \begin{cases} \frac{1}{2} & \text{nếu } j = i \pm 1 \\ \frac{1}{2} & \text{nếu } j = i = 0 \text{ hoặc } j = i = 100 \\ 0 & \text{còn lại} \end{cases}$$

Lời giải.

Giả sử ta cần sinh một mẫu ngẫu nhiên gồm 10.000 giá trị để mô phỏng π . Chọn ngẫu nhiên giá trị ban đầu, giả sử là $x_k = i$ ($i \in \{0,1,2, \dots, 100\}$ và k bất kỳ). Giải thuật như sau:

- Sinh trạng thái mới j : Sinh giá trị ngẫu nhiên u từ phân bố đều $U(0,1)$:
 - Cho i ($i \neq 0,100$): Nếu $u \geq \frac{1}{2}$ thì nhận trạng thái mới $j = i + 1$, ngược lại đặt $j = i - 1$
 - Cho $i = 0$: Nếu $u \geq \frac{1}{2}$ thì nhận trạng thái mới $j = 1$, ngược lại đặt $j = 0$
 - Cho $i = 100$: Nếu $u < \frac{1}{2}$ thì nhận trạng thái mới $j = 99$, ngược lại đặt $j = 100$

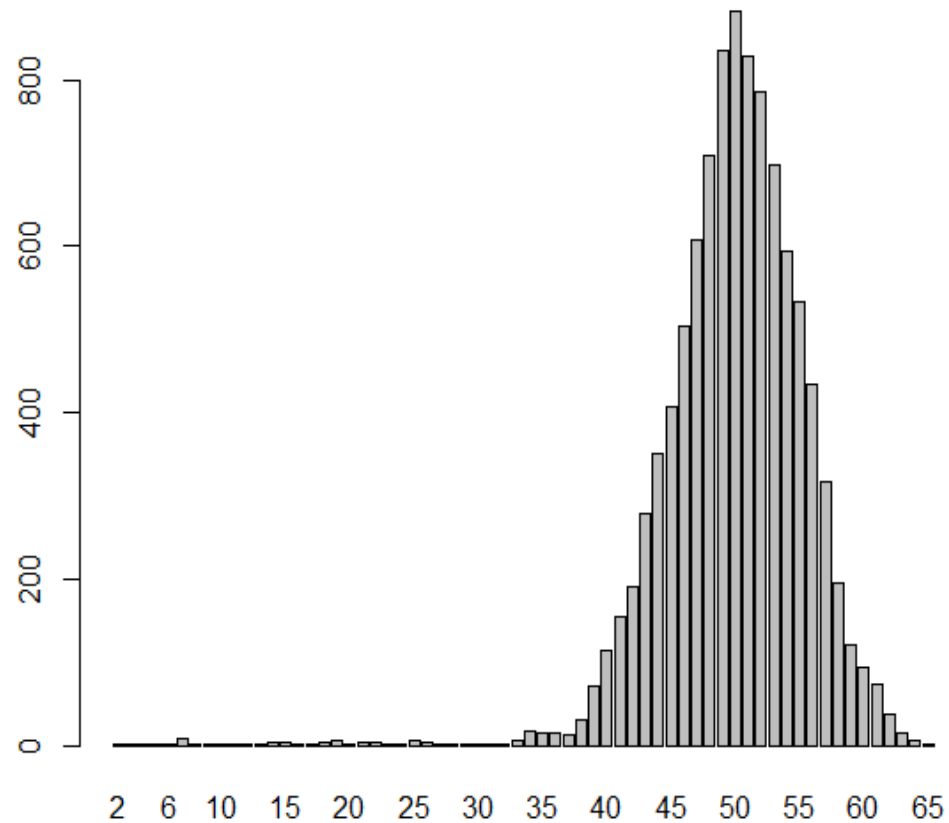
- Tính tỉ số:

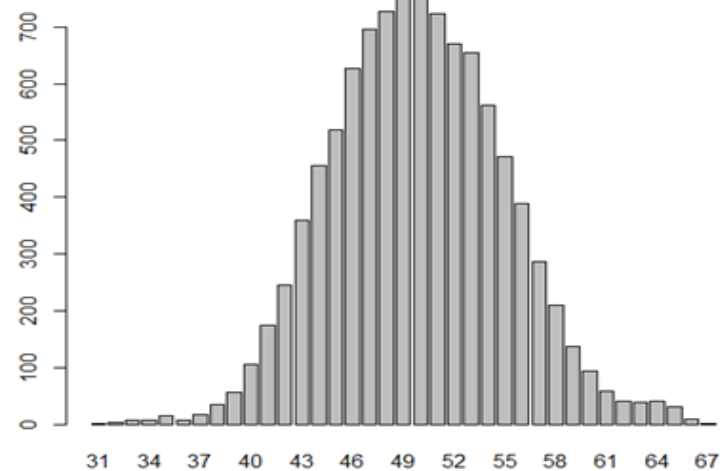
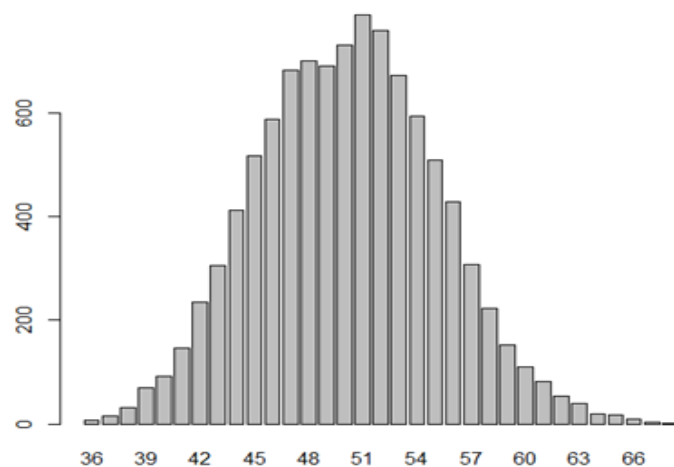
$$\frac{\pi(j)}{\pi(i)} = \frac{C_{100}^j 0.5^{100}}{C_{100}^i 0.5^{100}} = \frac{100!}{j!(100-j)!} \cdot \frac{i!(100-i)!}{100!} = \frac{i!(100-i)!}{j!(100-j)!}$$

$$\text{Đặt } r = \begin{cases} 1 & \text{nếu } i = j \\ \frac{100-i}{i+1} & \text{nếu } j = i + 1 \\ \frac{i}{100-i+1} & \text{nếu } j = i - 1 \end{cases}$$

- So sánh:
 - Nếu $r \geq 1$ đặt $x_{k+1} = j$, ngược lại sinh u_1 từ $U(0,1)$:
 - Nếu $u < r$ đặt $x_{k+1} = j$, ngược lại đặt $x_{k+1} = x_k$
- Thay $k = k + 1$, quay lại bước đầu.

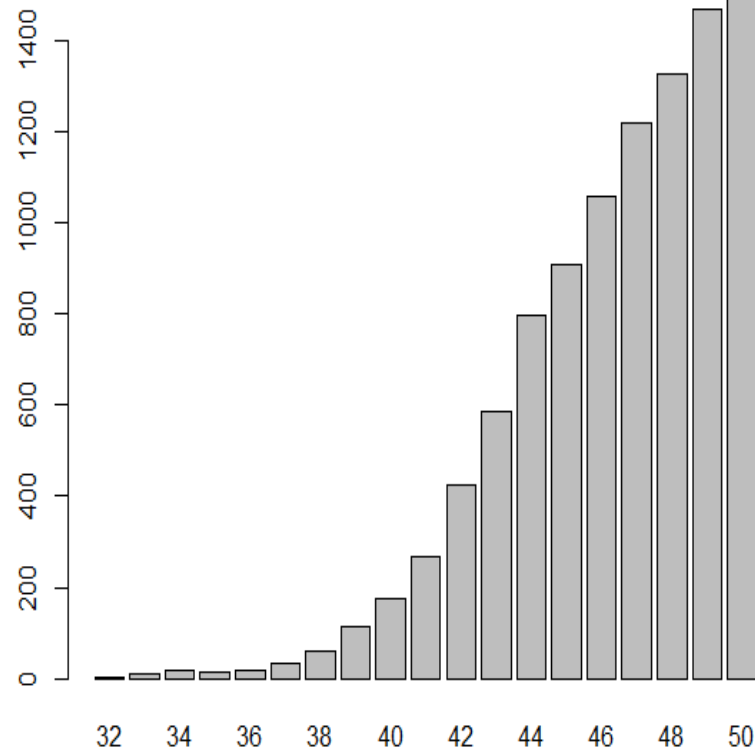
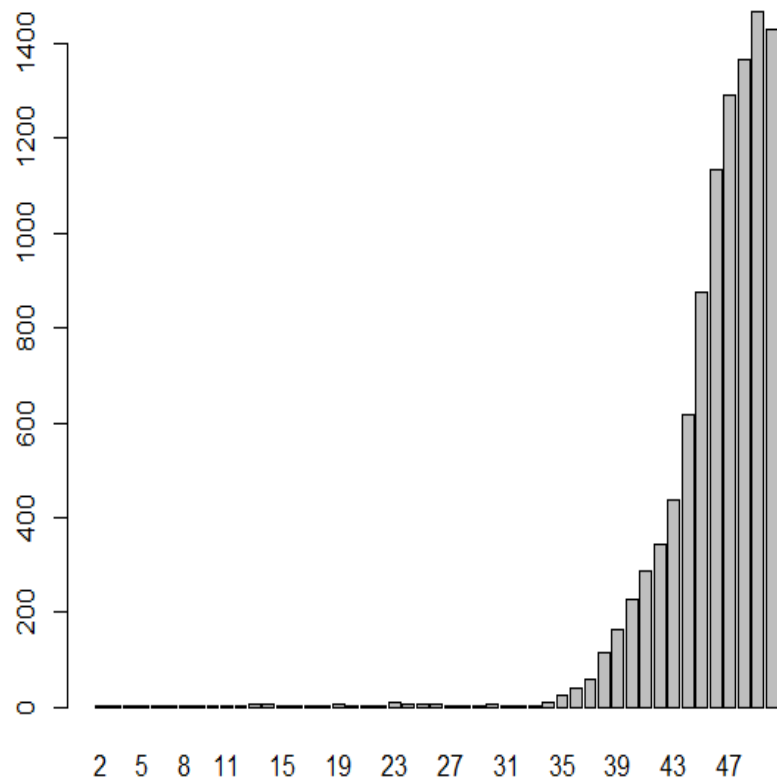
Sử dụng R để lập trình, với thông tin ban đầu nhập vào là cỡ mẫu ngẫu nhiên cần sinh ($n = 10.000$), giả sử giá trị xuất phát ban đầu của chuỗi là 2 (được chọn ngẫu nhiên), ta được biểu đồ tần suất các giá trị trạng thái xuất hiện như sau:





- ◆ Thay đổi các giá trị xuất phát ban đầu khác nhau, ta đều nhận được biểu đồ có hình dạng tương tự như trên, và tần số tập trung chủ yếu vào khoảng giá trị trạng thái từ 30 đến 65, có nghĩa là cách chọn giá trị ban đầu không ảnh hưởng đến kết quả của thuật toán:

- ◆ Giảm tham số n của phân phối nhị thức xuống dưới 100, ta thấy rằng biểu đồ tần suất không thể hiện đầy đủ sự phân bố mật độ các giá trị như trên nữa:



2. Sinh mẫu ngẫu nhiên từ phân bố mũ

Sinh mẫu ngẫu nhiên từ phân bố mục tiêu là phân bố mũ:

$$\pi(x) \sim f(x, \lambda) = \begin{cases} \lambda e^{-\lambda x} & \text{nếu } x \geq 0 \\ 0 & \text{nếu } x < 0 \end{cases} \quad (\lambda > 0) \text{ và phân bố đều nghi}$$

$$q_x(y) \sim U(-1, 1).$$

Lời giải.

Đây là trường hợp phân bố mục tiêu cần mô phỏng là một phân bố liên tục, vì thế phân bố đề nghị cũng phải liên tục, như vậy không gian trạng thái là liên tục.. Giả sử ta cần sinh mẫu ngẫu nhiên gồm 20.000 giá trị. Chọn ngẫu nhiên một giá trị ban đầu $x^{(t)} = x_0 \geq 0$. Giải thuật như sau:

- Sinh giá trị trạng thái mới y từ phân bố đề nghị $U(-1,1)$: giá trị y phải thỏa mãn $(y - x^{(t)}) \sim U(-1,1)$, như vậy $y \sim U(x^{(t)} - 1, x^{(t)} + 1)$.
- Tính $r = \frac{f(y,\lambda)}{f(x^{(t)},\lambda)} = \frac{\lambda e^{-\lambda y}}{\lambda e^{-\lambda x^{(t)}}} = e^{\lambda(x^{(t)} - y)}$ với $y \geq 0$

- So sánh:

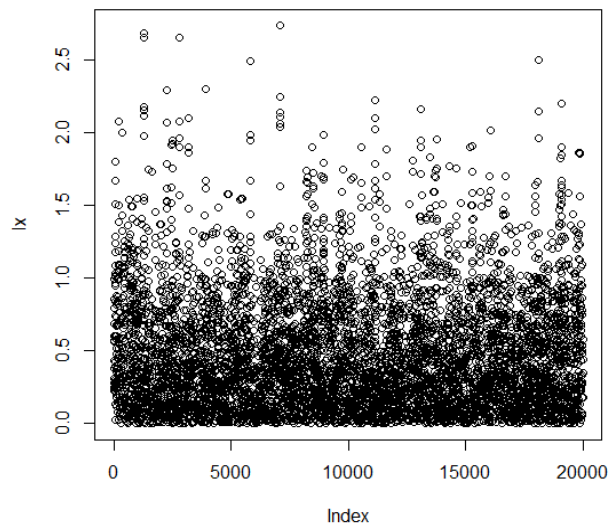
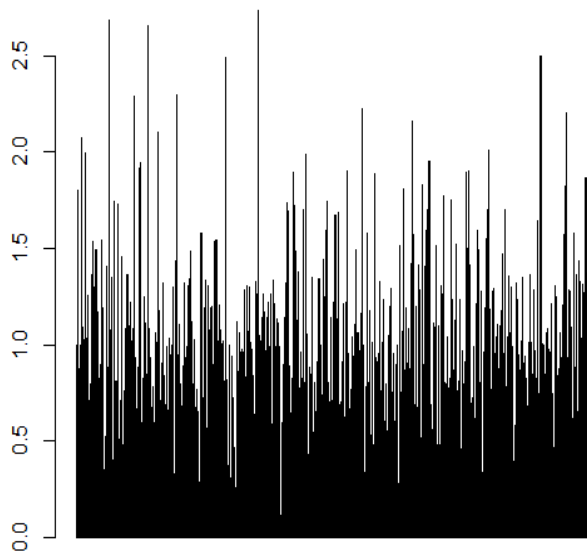
Nếu $r \geq 1$, chấp nhận giá trị mới $x^{(t+1)} = y$,
ngược lại sinh u từ phân bố đều $U(0,1)$.

Nếu $u < r$, chấp nhận giá trị mới $x^{(t+1)} = y$,
ngược lại giữ nguyên giá trị cũ $x^{(t+1)} = x^{(t)}$

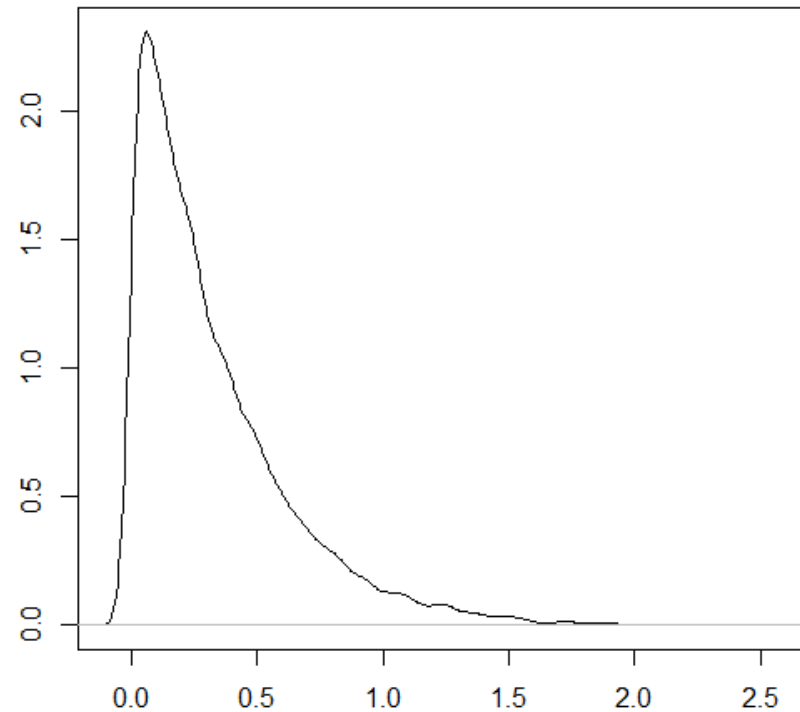
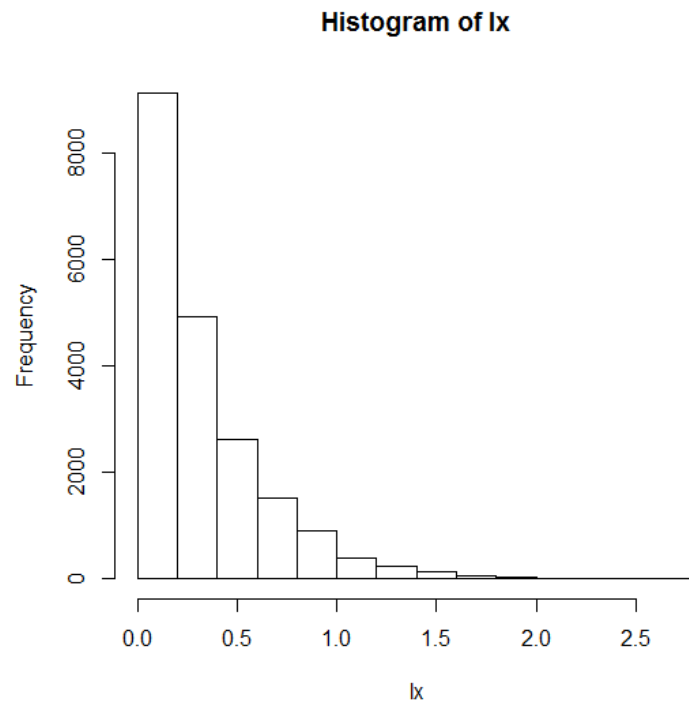
- Thay $t = t + 1$ quay lại bước đầu.

Sử dụng R để lập trình, với cỡ mẫu ngẫu nhiên sinh ra là 20.000 giá trị, chọn ngẫu nhiên giá trị khởi đầu là $x_0 = 0.25$, phân bố $\pi(x)$ có tham số $\lambda = 3$, ta được kết quả sau:

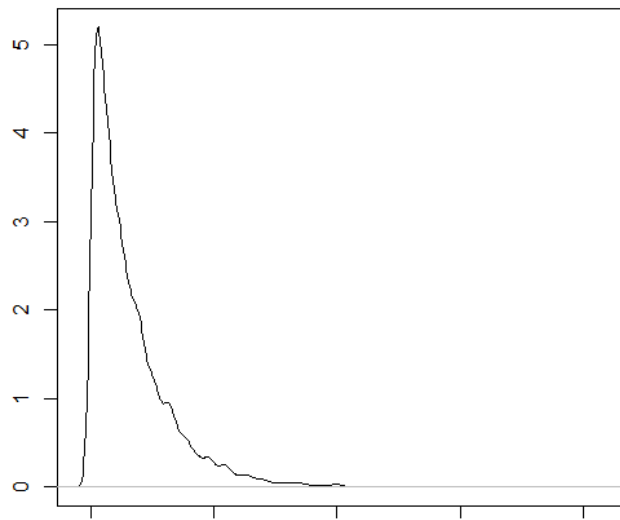
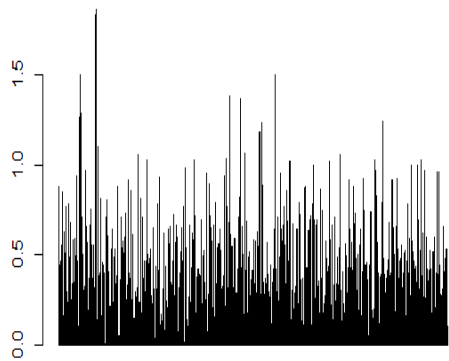
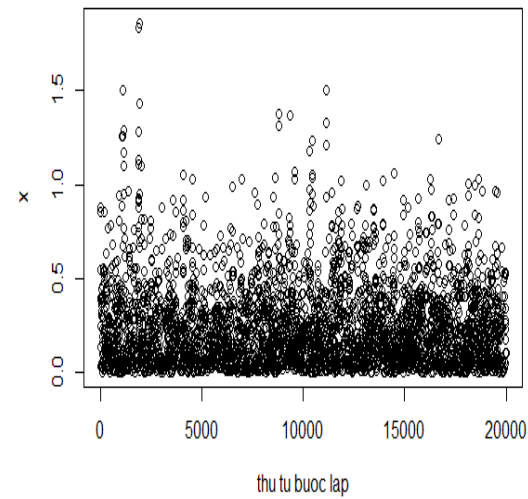
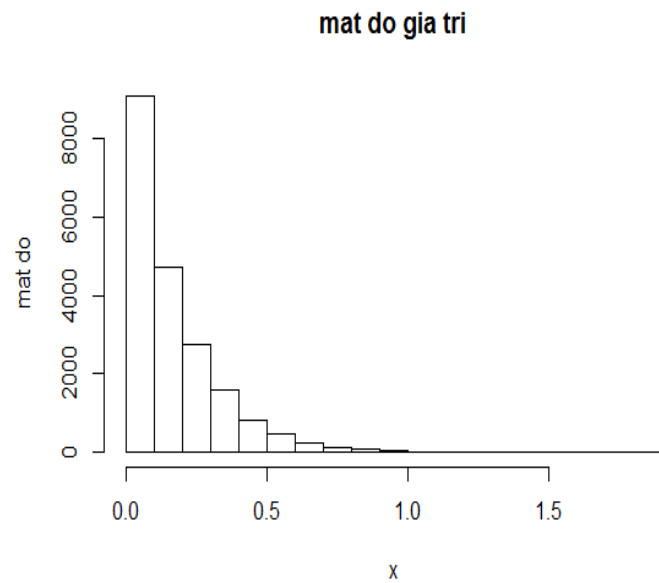
- Biểu diễn giá trị hàm tại các số được sinh ra chia theo số bước lặp:



- Các số sinh ra có giá trị hàm tập trung từ 0 đến 2, trong đó tần số lớn nhất là trong khoảng (0,0.2):



- ◆ Thay đổi giá trị của tham số λ lên gấp đôi: $\lambda = 6$, các thông số khác vẫn giữ nguyên, ta có kết quả sau:



3. Sinh mẫu ngẫu nhiên từ phân bố chuẩn

Sinh mẫu ngẫu nhiên từ phân bố mục tiêu $\pi(x) \sim N(0,1) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})$ với phân bố đề nghị $q_x(y) \sim U(0,1)$.

Lời giải.

Đây cũng là trường hợp mô phỏng phân bố $\pi(x)$ liên tục, với phân bố đề nghị đối xứng $U(-1,1)$. Đương nhiên với giải thuật Metropolis ta vẫn cần điều kiện phân bố $q(x,y)$ đối xứng, trong trường hợp này là phân bố đều trên $(0,1)$.

Giả sử ta cần sinh mẫu ngẫu nhiên gồm 1000 giá trị. Chọn ngẫu nhiên một giá trị ban đầu $x^{(t)} = x_0$. Giải thuật như sau:

- Sinh trạng thái mới y từ phân bố $U(0,1)$: giá trị y phải thỏa mãn $(y -$

- Tính :
$$r = \frac{\pi(y)}{\pi(x^{(t)})} = \frac{\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right)}{\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^{(t)2}}{2}\right)} = \exp\left(\frac{x^{(t)2} - y^2}{2}\right)$$

- So sánh:

Nếu $r \geq 1$ thì chấp nhận trạng thái mới: $x^{(t+1)} = y$,

ngược lại, sinh u từ $U(0,1)$

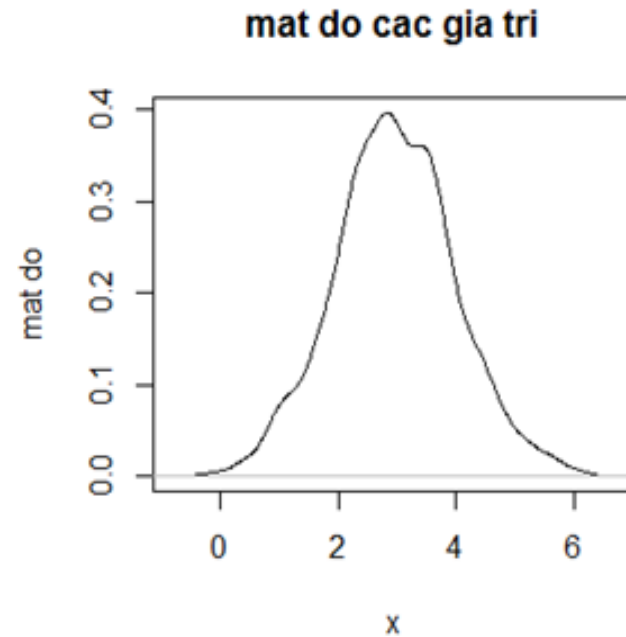
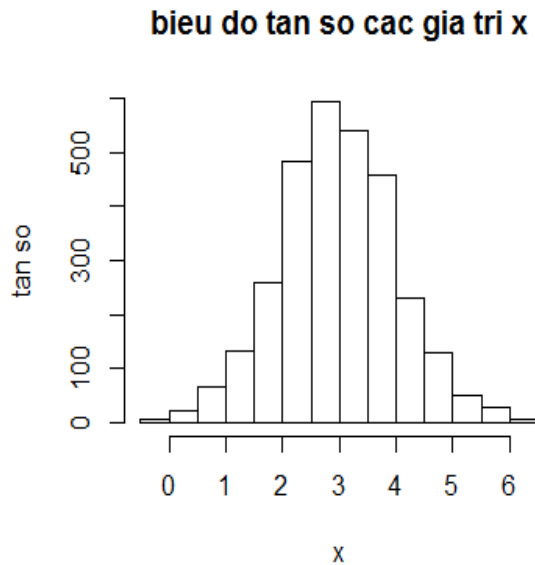
Nếu $u < r$ thì chấp nhận trạng thái mới: $x^{(t+1)} = y$,

ngược lại giữ nguyên trạng thái trước: $x^{(t+1)} = x^{(t)}$

- Thay $t = t + 1$, quay trở lại bước đầu.

Sử dụng R để lập trình, với giả sử lấy mẫu ngẫu nhiên gồm 1000 giá trị, chọn ngẫu nhiên giá trị ban đầu, giả sử là $x_0 = 0.5$, ta được kết quả mô phỏng như sau:

- ◆ Vẫn với giải thuật trên, giả thiết phân bố đề nghị là phân bố đều $U(0,1)$, điểm ban đầu xuất phát là $x_0 = 0.5$, lấy mẫu ngẫu nhiên 4000 giá trị, nhưng xét với phân bố mục tiêu $\pi(x) \sim N(3,1)$, chuỗi mô phỏng lấy từ giá trị ở bước lặp thứ 1000. Ta có kết quả như sau:



VIỆN HÀN LÂM KHOA HỌC & CÔNG NGHỆ VIỆT NAM
VIỆN TOÁN HỌC

HỘI THẢO KHOA HỌC
MỘT SỐ VẤN ĐỀ TRONG XÁC SUẤT VÀ THỐNG KÊ
ỨNG DỤNG

Hà Nội, ngày 15 – 04 – 2021

CHƯƠNG TRÌNH
&
TÓM TẮT BÁO CÁO

HÀ NỘI, 2021

Bài toán mô phỏng sinh mẫu bằng phần mềm R

Nguyễn Thị Hằng

Dại học Mở - Địa chất

R là một phần mềm sử dụng cho phân tích thống kê và vẽ biểu đồ. Thật ra, về bản chất, R là ngôn ngữ máy tính đa năng, có thể sử dụng cho nhiều mục tiêu khác nhau, từ tính toán đơn giản, toán học giải trí (recreational mathematics), tính toán ma trận (matrix), đến các phân tích thống kê phức tạp. Vì là một ngôn ngữ, cho nên người ta có thể sử dụng R để phát triển thành các phần mềm chuyên môn cho một vấn đề tính toán cá biệt.

R có chứa nhiều loại kỹ thuật thống kê (mô hình hóa tuyến tính và phi tuyến, kiểm thử thống kê cổ điển, phân tích chuỗi thời gian, phân loại, phân nhóm...) và đồ họa. Được thiết kế xoay quanh một ngôn ngữ máy thực thụ, R cho phép người dùng thêm các tính năng bổ sung bằng cách định nghĩa các hàm mới. Nhiều hệ thống trong R được viết bằng chính ngôn ngữ của nó, giúp cho người dùng dễ sử dụng các giải thuật.

Trong khuôn khổ báo cáo, tôi xin phép không đi sâu vào cách thiết lập chương trình để chạy trên R mà chỉ đề cập đến R như một công cụ giúp giải quyết bài toán sinh mẫu mô phỏng phân phối mục tiêu cho trước.

Hàm phân phối của biến ngẫu nhiên giá trị lớn nhất của quá trình ngẫu nhiên và Ứng dụng trong thống kê kiểm định

Phạm Việt Hùng

Phạm Việt Hùng

Trong các bài toán thống kê kiểm định, để chấp nhận hoặc bác bỏ giả thiết đưa ra, ta thường so sánh giá trị thống kê với một giá trị ngưỡng. Trong báo cáo này, chúng tôi đề cập đến một lớp các bài toán kiểm định mà giá trị ngưỡng được xác định từ hàm phân phối của biến ngẫu nhiên giá trị lớn nhất của một quá trình ngẫu nhiên có phân phối chuẩn. Một số ứng dụng trong thực tế sẽ được trình bày chi tiết.

1. **Nguyễn Thị Lan Anh**
Trưởng Đại học phòng cháy chữa cháy
nguyenlananh090289@gmail.com
2. **Nguyễn Tuấn Anh**
Trưởng Đại học phòng cháy chữa cháy
ntuananhpc@gmail.com
3. **Tô Thị Vân Anh**
Đại học Thương mại
Vananhdhtm@gmail.com
4. **Nguyễn Bảo Châu**
Đại học Hà Nội
chau_eii@yahoo.fr
5. **Đỗ Văn Cường**
Đại học Xây dựng
kuongdv@gmail.com
6. **La Thùy Dung**
Đại học Khoa học Tự nhiên Hà Nội
ltdung153@gmail.com
7. **Nguyễn Trung Dũng**
Trưởng Đại học Sư phạm Hà Nội 2
nguyentrongdung@hpu2.edu.vn
8. **Đỗ Minh Đức**
Đại học Khoa học Tự nhiên Hà Nội
dangthihuy_T57@hus.edu.vn
9. **Võ Thị Trúc Giang**
Trưởng Đại học Tiền Giang
vothitrucgiang@tgu.edu.vn
10. **Hoàng Mạnh Hà**
Đại học Thủ Dầu Một
hahm@tdmu.edu.vn
11. **Hoàng Thị Thu Hà**
Đại học Thương mại
hoangha.math@tmu.edu.vn
12. **Trần Thị Thu Hà**
Trưởng Đại học Y Dược Thái Bình
tranhaytb@gmail.com
13. **Nguyễn Văn Hạnh**
Học viện Nông nghiệp Việt Nam
nvhanh@vnua.edu.vn
14. **Nguyễn Thị Hằng**
Đại học Mở Địa chất
hangntmdc@gmail.com
15. **Nguyễn Thu Hằng**
Đại học Dược Hà Nội
hangnt@hup.edu.vn
16. **Vũ Thị Thuý Hằng**
Đại học Thương mại
hangtmdt@tmu.edu.vn
17. **Mai Thị Hoa**
Công ty Itsol
maihoa.x.92@gmail.com
18. **Phan Thanh Hồng**
Đại học Thăng Long
hongpt@thanglong.edu.vn
19. **Nguyễn Thế Hùng**
Đại học Bách Khoa Hà Nội
hung96ad@gmail.com
20. **Phạm Việt Hùng**
Viện Toán học
pvhung@math.ac.vn
21. **Vũ Thị Hương**
Đại học Giao thông Vận tải
vthuong@utc.edu.vn
22. **Đỗ Quang Khải**
Đại học Bách Khoa Hà Nội
doquangkhai1995@gmail.com