

BÀI TẬP 4  
(Phương pháp MCMC và Thống kê Bayes tính toán)  
THỐNG KÊ MÁY TÍNH VÀ ỨNG DỤNG (CLC)

1. Phân phối nửa chuẩn (half-normal distribution) có hàm mật độ

$$f(x) = \frac{2}{\sqrt{2\pi}} e^{-x^2/2} \mathbb{I}_{[0,\infty)}(x).$$

Lấy mẫu cho phân phối nửa chuẩn bằng các phương pháp sau:

- (a) Metropolis-Hasting với phân phối đề cử là phân phối  $\text{Exp}(1)$  với hàm mật độ

$$g(x'|x) = e^{-x'} \mathbb{I}_{[0,\infty)}(x').$$

- (b) Bước ngẫu nhiên Metropolis với độ dời  $\epsilon \sim \mathcal{U}(-2, 2)$ .

- (c) Bước ngẫu nhiên Metropolis với độ dời  $\epsilon \sim \mathcal{U}(-0.2, 0.2)$ .

Đánh giá 3 phương pháp.

2. Dùng phương pháp Gibbs lấy mẫu cho phân phối 2 chiều có hàm mật độ

$$f(x, y) \propto e^{-\frac{1}{2}x(1+y^2)} \quad (0 < x, -\infty < y < \infty).$$

Đánh giá kết quả.

3. Một hộp 20 viên bi chỉ gồm bi đỏ và bi đen. Giả sử tỉ lệ bi đỏ không nhỏ hơn 40% và không quá 75%. Lấy ngẫu nhiên 5 viên bi không hoàn lại từ hộp thì được 4 bi đỏ. Dùng suy diễn Bayes, “xác định” tỉ lệ bi đỏ trong hộp.
4. Từ Iris Dataset (xem Ví dụ trong Slide), dùng suy diễn Bayes
- (a) Xác định sự phụ thuộc của petal-length vào sepal-length, sepal-width, petal-width và class.
- (b) Xác định sự phụ thuộc của class vào petal-length, sepal-length, sepal-width và petal-width.

*Lưu ý:*

- Các thuật toán cần được trình bày bằng mã giả và cài đặt bằng Python.
- Câu 1 chỉ được phép dùng  $\mathcal{U}(0, 1)$ .
- Dùng PyMC3 để phân tích các Câu 3, 4.

— HẾT —