

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

Câu 1. (2.5 điểm)

- a) Dùng bước ngẫu nhiên Metropolis, thiết kế thuật toán lấy mẫu cho
 - 1) Phân phối Poisson có kì vọng $\lambda = 4$ với độ dời $\varepsilon \sim \mathcal{U}\{-1, 1\}$.
 - 2) Phân phối chuẩn tắc với độ dời $\varepsilon \sim \mathcal{U}(-1, 1)$.
- b) So sánh các thuật toán trên với các thuật toán lấy mẫu độc lập đã học về thời gian chạy và sai số trong ước lượng kì vọng của phân phối.

Câu 2. (1.5 điểm) Dùng phương pháp Gibbs lấy mẫu cho (X, Y) có hàm mật độ

$$f(x, y) = 2e^{-(x+2y)}, 0 < x, y < \infty$$

để ước lượng xác suất $P(X < Y)$ và kì vọng $E(XY)$.

Câu 3. (2.5 điểm) Có 2 hộp bi: hộp I gồm 7 bi đỏ và 3 bi đen, hộp II gồm 2 bi đỏ và 8 bi đen. Việc bốc bi được tiến hành theo các bước sau:

- *Bước 1:* Chọn hộp bi I hoặc II với xác suất chọn hộp I là p ($0 \leq p \leq 1$).
- *Bước 2:* Bốc ngẫu nhiên một viên bi trong hộp đã chọn.

Dùng suy diễn Bayes, “xác định” p trong các trường hợp:

- a) Bốc 5 lần có hoàn lại thì thấy có 4 lần được bi đỏ và 1 lần được bi đen.
- b) Tương tự (a) nhưng biết thêm thông tin $p \leq 0.25$.
- c) Bốc 3 lần không hoàn lại thì thấy lần lượt được bi đỏ, đen, đỏ.
- d) Bốc 3 lần không hoàn lại thì thấy lần lượt được bi đỏ, đỏ, đen.

Câu 4. (3.5 điểm) Tìm hiểu bộ dữ liệu Auto MPG tại <https://archive.ics.uci.edu/ml/datasets/Auto+MPG>. Download tập tin dữ liệu auto-mpg.data (<https://archive.ics.uci.edu/ml/machine-learning-databases/auto-mpg/auto-mpg.data>) và bỏ cột car name. Dùng suy diễn Bayes

- a) Xác định sự phụ thuộc của mpg vào các biến còn lại.
- b) Xác định sự phụ thuộc của origin vào các biến còn lại.
- c) Xác định sự phụ thuộc của cylinders vào các biến còn lại.

Lưu ý:

- Trình bày bài làm (lời giải, công thức Toán, mã Python, kết quả, ...) trong tập tin notebook.
- Cần trình bày, cài đặt, kiểm tra và đánh giá các thuật toán được yêu cầu.
- Câu 1 chỉ được phép dùng $\mathcal{U}(0, 1)$. Dùng PyMC để thực hiện các Câu 3, 4.

--- HẾT ---